Risk measures in finance, Backtesting, Sensitivity and Robustness

Manon Rivoire

INSTITUT POLYTECHNIQUE DE PARIS

NNT : 2024IPPAX042

Thèse de doctorat

Risk measures in finance, Backtesting, Sensitivity and Robustness

Thèse de doctorat de l'Institut Polytechnique de Paris
préparée à Ecole Polytechnique

École doctorale n°574 Ecole Doctorale de Mathématiques Hadamard (EDMH)
Spécialité de doctorat : Mathématiques Appliquées

Thèse présentée et soutenue à Palaiseau, le 05 Juillet 2024, par

RIVOIRE MANON

Composition du Jury :

Jean-François COEURJOLLY Professor, Université Grenoble Alpes (UGA)	Rapporteur
Alexandre BROUSTE Professor, Le Mans Université - Institut du Risque et de l'Assurance	Rapporteur
Charles-Albert LEHALLE Researcher, Imperial College	Examinateur
Zorana GRBAC Assistant Professor, Université Paris Cité	Examinatrice
Caroline HILLAIRET Professor, ENSAE-CREST	Présidente du jury
Nicole EL KAROUI Emeritus Professor, Sorbonne Université (Jussieu) - LPSM	Invitée
Emmanuel GOBET Professor, Ecole Polytechnique - CMAP	Directeur de thèse
Matthieu GARCIN Researcher, ESILV - DVRC	Invité

"Mathematics knows no races or geographic boundaries; for mathematics, the cultural world is one country."

(**David Hilbert**, International Congress of Mathematicians, 1900.)

Remerciements

Je souhaite remercier du fond du coeur toutes les personnes que j'ai eu la chance de croiser et qui m'ont offert l'opportunité de partager un bout de chemin plus ou moins long à leurs côtés. L'aboutissement de ma thèse est une victoire collective à laquelle toutes les personnes aux côtés desquelles j'ai évolué, ont contribué: mes directeurs de thèse, mes encadrants et professeurs, ma famille, mes amis, et toutes les personnes avec qui j'ai eu la chance d'échanger. En effet, sans eux, mon parcours n'aurait pu être le même. Ils m'ont donné les clés et la force nécessaires pour surmonter les obstacles me permettant ainsi d'atteindre mes objectifs. Je suis extrêmement reconnaissante de tout ce qu'ils ont fait pour moi et je leur adresse ma plus profonde gratitude.

J'exprime ma sincère gratitude à mes directeurs de thèse pour m'avoir offert la possibilité de réaliser cette thèse à leurs côtés. Leurs enseignements m'ont permis non seulement d'évoluer sur le plan académique mais aussi de grandir sur le plan personnel.

Je souhaite adresser mes sincères remerciements à mon directeur de thèse, **Emmanuel Gobet**, pour avoir été à l'initiative de ce projet de thèse et pour m'avoir offert la possibilité de réaliser mon doctorat sous sa direction. Je le remercie pour son engagement et son dévouement tout au long de mon parcours de recherche. Ses conseils éclairés et son exigence académique m'ont poussée à donner le meilleur de moi-même. Ce travail n'aurait pas pu voir le jour sans son soutien précieux.

Je remercie sincèrement ma directrice de thèse, **Laurence Carassus**, pour avoir contribué à la construction de ce projet de thèse et pour m'avoir fait confiance en m'offrant la possibilité de réaliser ma thèse sous sa direction.

Je remercie mes encadrants de thèse, **Matthieu Garcin et Stefano De-Marco**, pour avoir accepté de faire partie de mon projet de thèse et pour m'avoir ensuite suivie tout au long de ma thèse. Je les remercie notamment pour les échanges enrichissants qu'ils m'ont permis d'avoir avec eux.

Je tiens à exprimer ma plus profonde gratitude à la Professeure **Nicole El-Karoui**, pour son encadrement exceptionnel, ses conseils avisés et son soutien constant tout au long de ma thèse. Son expertise et sa passion pour les Mathématiques financières ont été une source d'inspiration inestimable. Ses retours constructifs ont grandement contribué à l'élaboration de ce travail. Je lui adresse mes sincères remerciements pour m'avoir guidée avec patience et bienveillance dans cette aventure académique. En tant que femme, elle incarne pour moi une source d'inspiration et de motivation, montrant par son exemple que la rigueur scientifique et la réussite professionnelle sont compatibles avec une grande humanité et un profond sens de l'éthique. Sa bienveillance, son écoute et son soutien ont été d'une aide sans précédent, me permettant de surmonter les défis auxquels j'ai pu être confrontée. Merci d'avoir été non seulement une encadrante exceptionnelle, mais aussi une mentor et un modèle à suivre.

Je souhaite adresser mes remerciements les plus sincères à mes rapporteurs, les Professeurs **Jean-François Coeurjolly et Alexandre Brouste** pour avoir rapporté ma thèse en détails. Je les remercie pour l'intérêt qu'ils ont porté à mes travaux, pour les commentaires intéressants ainsi que pour les suggestions d'améliorations qu'ils m'ont formulées.

Je voudrais également remercier les Professeurs qui ont accepté de participer à mon jury de thèse malgré leur emploi du temps très chargé: **Nicole El-Karoui, Caroline Hillairet, Zorana Grbac, Charles-Albert Lehalle, Matthieu Garcin**. Leur participation à mon jury est un véritable honneur.

Je souhaite remercier du fond du coeur la Professeure **Caroline Hillairet** qui m'a permise de décrocher ma thèse. Elle m'a suivie du début à la fin avec gentillesse, patience et bienveillance. Ses conseils et ses mots toujours justes et avisés à mon égard m'ont permis d'avancer en prenant les bonnes décisions. Son soutien sans commune mesure du début à la fin de ma thèse a été pour moi essentiel à ma réussite. Plus qu'une Professeure, Caroline représente pour moi un modèle de femme en sciences tant sur le plan des valeurs morales que sur le plan académique. Je la remercie pour tout ce qu'elle a fait pour moi.

J'adresse ma plus profonde gratitude à l'égard de tous les Professeurs que j'ai eu la chance de croiser durant mon parcours à l'Ecole Polytechnique et au sein du master Probabilités-Finance (ex. DEA El-Karoui) dans lequel j'ai suivi les cours nécessaires pour mon doctorat. Ils m'ont donné l'opportunité de faire un bout de chemin à leurs côtés. Sans eux, je n'aurai pas vécu la même aventure académique et je ne serai sans doute pas arrivée là où j'en suis aujourd'hui. J'adresse ainsi mes plus sincères remerciements à **Stéphanie Allassionnière, Erwan Le Pennec Charles-Albert Lehalle, Emilio Saïd, Basile Audoly, Aymeric Dieuleveut, Mathieu Rosenbaum, Matthieu Jonckheere, Gauthier Vermandel, Clément Rey, Vincent Bensaye, Julie Josse, Alexandre Gramfort, Olivier Le-Maître, Idriss Kharroubi, Lorenzo Zambotti, Thomas Duquesne, Jean-Philippe Bouchaud**. En effet, ils ont su m'épauler dans les moments difficiles de ma thèse, me soutenir et trouver les solutions les plus adaptées au bon déroulement de mon parcours académique afin que je réussisse ma thèse dans les meilleures conditions. Je les remercie pour leur présence, leur gentillesse, et leur soutien tout au long de ma thèse ainsi que pour le temps qu'ils m'ont accordé. Ils ont contribué à mon évolution tant sur le plan personnel qu'académique.

Tout d'abord, je souhaite adresser mes sincères remerciements au directeur du CMAP, le Professeur **Grégoire Allaire**, pour sa bienveillance, son dévouement, sa gentillesse et sa grande humanité. Il a su être présent dans les moments difficiles et m'aider à trouver des solutions pour que ma thèse se déroule dans les meilleures conditions possibles, je lui en suis très reconnaissante.

J'exprime ma profonde gratitude à l'égard du Professeur **Charles-Albert Lehalle et Emilio Said** pour m'avoir fait confiance pour collaborer avec eux sur leur projet de recherche. Cette collaboration a été pour moi très enrichissante, je les remercie pour tous les échanges que nous avons eus. Leur patience, leur gentillesse et leur pédagogie m'ont permis d'apprendre énormément à leurs côtés.

J'exprime ma sincère reconnaissance à l'égard de la Professeure **Stéphanie Allassonnière** pour sa gentillesse, sa patience, sa bienveillance, son écoute et son soutien sans commune mesure durant ma thèse.

J'adresse mes sincères remerciements au Maître de conférence **Clément Rey** pour sa gentillesse et sa bienveillance. J'ai eu la chance de croiser son chemin et je le remercie pour ses conseils tout au long de mon parcours ainsi que pour les échanges toujours intéressants qu'il m'a permis d'avoir avec lui. Son soutien a été d'une grande aide. Merci à lui.

Je souhaite exprimer mes plus sincères remerciements au Maître de conférence **Noufel Frikha** que j'ai eu la chance de rencontrer à la fin de mon doctorat. Je le remercie sincèrement pour ses conseils avisés, son temps et les échanges qu'il m'a permis d'avoir avec lui. Sa gentillesse et sa bienveillance m'ont profondément touchée et m'ont permis d'atteindre la soutenance dans les meilleures conditions. Merci à lui pour tout ce qu'il a fait pour moi.

Je remercie du fond du coeur Madame **Aldjia Mazzari**, pour son soutien, ses encouragements et ses conseils durant les periodes difficiles ainsi que pour m'avoir aidée à trouver des solutions à des moments clés de ma thèse.

Je remercie sincèrement le service informatique, Messieurs **Pierre Straebler et Sylvain Ferrand**, pour leur gentillesse, leur patience et leur dévouement. Ils ont fortement contribué au bon déroulement de ma thèse et je leur en suis extrêmement reconnaissante.

Je remercie également Monsieur **Josselin Massot** pour sa gentillesse, son enthousiasme et son soutien ainsi que pour toutes les magnifiques origamis qu'il m'a offert pour décorer mon bureau tout au long de ma thèse.

J'adresse mes sincères remerciements à tout le personnel du CMAP qui oeuvre au quotidien pour le bien-être des doctorants. Je remercie notamment Mesdames **Nassera Naar, Nicoletta Bourgeois, Leila Marzouk, Alexandra Noiret**, pour le soutien et l'aide inestimables qu'elles m'ont apporté tout au long de ma thèse. Leur efficacité, leur disponibilité et leur gentillesse ont grandement facilité mon parcours. Leurs conseils avisés et leur assistance précieuse ont joué un rôle crucial dans la réalisation de ce travail. Merci pour leur professionnalisme et pour avoir toujours été là pour répondre à mes questions et résoudre les problèmes administratifs avec rapidité et bienveillance.

Je remercie du fond du coeur l'Adjudant Chef de la 9ème Cie de l'Ecole Polytechnique **Mickaël Molle**, qui m'a permis de vivre ma thèse dans les meilleures conditions possible. En effet, les séances de sport qu'il m'a permis de réaliser à ses côtés dans la bienveillance, la bonne humeur et la gentillesse, ses conseils avisés, ainsi que les échanges que nous avons pu avoir ont été pour moi un immense soutien me permettant d'accéder à la soutenance dans les meilleures conditions. Ces entraînements ont été pour moi de vrais échappatoires qui ont fortement contribué au bon déroulement de ma thèse. Merci à lui pour tout.

Je remercie également toutes les personnes du CSX qui m'ont permis d'avoir une pratique sportive durant tout mon cursus à l'Ecole Polytechnique. Je remercie particulièrement Monsieur **Jauffrey Dussert** pour ses entraînements de natation, ses con-

seils, son écoute, sa gentillesse et son soutien.

J'adresse mes sincères remerciements à mon tuteur et mon encadrante de stage de fin de Master 2, **Alexandre Gramfort et Marine Le-Morvan**, pour m'avoir donné le goût de la recherche ainsi que pour leur soutien durant ma thèse.

J'adresse mes plus sincères remerciements aux membres de la **Chaire Stress Test** qui ont contribué au financement de mon doctorat. Je les remercie particulièrement pour toutes les connaissances que j'ai pu acquérir grâce à la chaire ainsi que pour les échanges très enrichissants que j'ai eu la chance d'avoir avec eux lors des séminaires de la chaire.

Je remercie du fond du coeur les doctorants et post-doctorants d'Emmanuel Gobet aux côtés de qui j'ai évolué durant toute ma thèse: **Clara Lage, Wanqing Wang, Elisa Ndiaye, Yushan Liu, Florian Bourgey, Michael Allouche, Maxime Grangereau, Linda Chamakh, Dorinel Bastide, David Métivier, Sharu Chardul, Jean Pachebat**. Ils m'ont permis d'entretenir des discussions académiques très intéressantes qui m'ont beaucoup appris. Grâce à eux, j'ai également eu un soutien très fort qui a joué un grand rôle dans la manière dont j'ai appréhendé la thèse et qui a été déterminant dans la réussite de mon parcours.

J'exprime ma profonde gratitude à l'égard de tous les doctorants et post-doctorants du CMAP sans qui ce voyage académique n'aurait pas été le même. En effet, les moments partagés ensemble resteront des souvenirs inoubliables. Je suis extrêmement reconnaissante de l'aide et du soutien qu'ils ont su m'apporter tout au long de cette aventure. C'est en grande partie grâce à eux que je suis parvenue au terme de mon parcours de doctorante. J'ai découvert des collègues merveilleux qui sont devenus de vrais amis: **Adam Perbost, Ali Aboud, Ali Baouan, Antoine Van Biesbroeck, Arianna Mingone, Armand Gissler, Arthur Loison, Amine Dhaou, Baptiste Goujaud, Baptiste Kerleger, Benjamin Riu, Christoph Schonle, Claire Ecotiere, Clara Lage, Clément Gauchy, Clément Mantoux, Constantin Philippenko, Corentin Houpert, Célia Escribe, David Métivier, Dorinel Bastide, Elisa Ndiaye, Manolis Sfendourakis, Florian Bourgey, Grégoire Pacreau, Grégoire Szymanski, Guillaume Chennetier, Ignacio Madrid, Jean Pachebat, Jessie Le Villain, Josué Tchouanti, Kang Liu, Leila Bassou, Loïc Balazi, Louis, Madeleine Kubasch, Mahmoud Hegazi, Margaux Zaffran, Margherita Castellano, Marius Chevallier, Matthias Rakotomalala, Maxence Noble-Bourillon, Mehdi Abou El Quassim, Michaël Allouche, Naoufal Acharki, Orso Forghieri, Oskar, Alexandre Perrin, Vanessa Dan, Pablo Jimenez, Pierre Clavier, Raphael Carpintero, Rémi, Renaud Gaucher, Solange Pruilh, Vincent Plassier, Wanqing Wang, Yushan Liu, Maxime Grangereau, Zoubair**. Ces personnes formidables que je n'oublierai jamais ont grandement contribué à la réussite de ma thèse.

Je remercie du fond du coeur **Antoine Van Biesbroeck** pour tout ce qu'il a fait pour moi tout au long de ma thèse. Son soutien inconditionnel, sa présence, et sa gentillesse m'ont permis d'évoluer, de m'améliorer et de surmonter les potentiels obstacles afin d'atteindre mes objectifs. Je le remercie également pour tous les moments partagés, ainsi que pour nos échanges. Ma thèse aurait été très différente sans sa présence, il a

fortement contribué à ma réussite.

Je remercie particulièrement mes amies **Clara Lage, Elisa Ndiaye, Jessie Le Villain, Leila Bassou, Madeleine Kubasch, Margaux Zaffran, Solange Pruilh, Wanqing Wang** pour leur soutien sans commune mesure, leur gentillesse et leurs encouragements ont été une source de motivation tout au long de ma thèse. Je les remercie également pour tous les bons moments passés à leurs côtés. Tout ce qu'elles ont fait pour moi durant mon parcours en thèse a fortement contribué à mon évolution, je leur en suis très reconnaissante.

Mes sincères remerciements vont également à mes amis **Ali Baouan, Ali Aboud, Arthur Loison, Benjamin Riu, Baptiste Goujaud, Baptiste Kerleger, Constantin Philippenko, Manolis Sfendourakis, Guillaume Chennetier, Kang Liu, Josué Tchouanti, Vincent Plassier et Loïc Balazi** pour leur soutien, leurs encouragements, leur présence et pour tous les moments partagés ensemble.

Je remercie aussi **Dorinel Bastide** pour sa profonde gentillesse, son soutien, sa bonne humeur constante ainsi que pour tous les échanges très enrichissants qu'il m'a permis d'avoir avec lui.

J'adresse mes sincère remerciements à **Florian Bourgey** pour sa présence tout au long de ma thèse. Son soutien tant sur le plan académique que moral, depuis Paris, Londres ou New-York m'ont grandement aidée durant ma thèse. Je le remercie également pour tous les échanges toujours très intéressants que j'ai eu la chance d'avoir avec lui. Il représente pour moi un modèle académique et une source de motivation pour atteindre mes objectifs. Merci à lui.

Je remercie sincèrement les membres de mon bureau avec qui j'ai partagé les moments difficiles de ma thèse comme les moments heureux: **Constantin Philippenko, Josué Tchouanti, Kang Liu, Ali Baouan, Loïc Balazi, Zoubair, Yushan Liu, Adam Perbost, Marius Chevallier**. Leur gentillesse, leur patience, leur écoute, m'ont permis de développer avec eux une complicité qui a rendu ma thèse très sympathique à leurs côtés. Je suis sincèrement reconnaissante de tout ce qu'ils m'ont apporté.

Parmi les doctorants du CMAP, plus que des collègues, j'ai découvert de vrais amis qui garderont à jamais une place particulière dans mon coeur.

J'adresse mes remerciements les plus sincères aux Professeurs ainsi qu'au personnel de l'ESILV qui m'ont encadrée durant mon parcours académique depuis ma première année à l'ESILV jusqu'à la fin de ma thèse: **Manon Lemoine, Pascal Clain, Marie-Noémie Thaï, Guillaume Guérard, Pierre Courbin, Kamel Hamdache, Sergio Focardi, Davide Mazza, Nicolas Travers, Aline Ellul, Laurence Carassus, Matthieu Garcin, Mohamad Ghassany, Thomas Cochet, Frédéric Fauberteau, Hélène Halconruy, Aurélien Vasseur, Walter Peretti, Basma Jaffal, Rémy Sart, Song He, Samir Yahiaoui, Swaminath Venkateswaran, Clément Duhart, Olivier Zanette, Cyril Grunspan, Vincent Lambert, Bastien Lejeune, Samira Essardy, Christina Keller, Sarah Massuelles**. Sans eux, mon parcours académique n'aurait pas été le même. En effet, ils ont joué un grand rôle dans mon évolution à la fois sur le plan personnel et académique. Grâce à eux, j'ai passé

de très belles années en école d'ingénieur, j'ai eu la chance d'avoir un accompagnement d'exception tant sur le plan personnel que sur le plan académique. Ils m'ont permis de réaliser ma dernière année d'école d'ingénieur en double diplôme avec l'Ecole Polytechnique et de continuer en thèse toujours à leurs côtés. Leur soutien, leur confiance en moi et leurs conseils avisés m'ont permis d'avancer et de m'améliorer en surmontant les obstacles nécessaires pour atteindre mes objectifs. Au-delà du plan académique, ils m'ont apporté un environnement propice à mon évolution. Je ne les remercierai jamais assez pour tout ce qu'ils ont fait pour moi.

En particulier, j'adresse mes remerciements les plus sincères à Madame **Manon Lemoine** sans qui toute cette aventure n'aurait pu être possible. Ses qualités humaines, sa sensibilité, et sa confiance en moi m'ont donné la chance de réaliser mon parcours académique à Paris et de grandir tant sur le plan personnel que scolaire. J'ai découvert une personne incroyable qui est devenue une véritable amie. Je la remercie pour son soutien, sa présence et tout ce qu'elle a fait pour moi depuis que nos chemins se sont croisés.

Je remercie du fond du coeur Monsieur **Pascal Clain**, sans qui mon évolution sur le plan personnel et sur le plan académique aurait été très différente. Je suis extrêmement reconnaissante de tout ce qu'il a fait pour moi depuis notre rencontre. Sa gentillesse, sa patience, sa confiance en moi et son soutien m'ont donné la force d'avancer, de surmonter les obstacles et de réaliser mes rêves. Je le remercie pour tous les moments partagés et les échanges que j'ai eu la chance d'avoir avec lui. Ses conseils avisés, son expérience et sa sagesse m'ont énormément appris. Il a été et restera pour moi une source d'inspiration et de motivation à tous points de vue.

J'adresse mes sincères remerciements au Professeur **Kamel Hamdache** pour les discussions qu'il m'a permis d'avoir avec lui. Son expérience, sa sagesse et ses conseils avisés m'ont beaucoup faite évoluer. Grâce à lui, j'ai trouvé la force de toujours me surpasser pour atteindre mes objectifs.

Je remercie sincèrement Monsieur **Pierre Courbin** dont le soutien sur le plan personnel comme scolaire a fait de mon parcours académique et de ma vie à Paris une aventure très enrichissante. Je le remercie pour son aide, sa gentillesse et sa patience ainsi que pour tous les moments passés ensemble à échanger sur des sujets enrichissants, à découvrir de nouvelles choses. Nos échanges ont fortement contribué à me faire grandir. Je le remercie pour tout ce qu'il a fait pour moi.

Je remercie Madame **Nathalie Bonatout** pour sa présence, sa gentillesse et son soutien durant ma thèse. Ses mots toujours réconfortants et son enthousiasme constant ont été pour moi des sources de motivation.

Je remercie ma Professeure de Mathématiques, **Marie-Noémie Thai**, pour son soutien, son accompagnement au quotidien et ses cours de Mathématiques qui m'ont permis de progresser, de m'élever et de me surpasser pour atteindre mes objectifs. Mon parcours académique n'aurait pas été le même sans sa présence. Je suis extrêmement reconnaissante de tout ce qu'elle a fait pour moi.

Je remercie sincèrement Monsieur **Guillaume Guérard** pour tout ce qu'il a fait pour moi depuis notre rencontre. En effet, il a fortement contribué à me donner le goût de la recherche. Son soutien, sa gentillesse, ses encouragements, et sa confiance en moi ont été d'une aide sans commune mesure. Il m'a permis de suivre mes rêves sans jamais baisser les bras. Il a fortement contribué à la réussite de mon parcours académique. Je le remercie également pour tous les échanges que nous avons eus et qui m'ont permis d'évoluer jusqu'à la fin de ma thèse.

J'adresse mes sincères remerciements à Madame **Aline Ellul**, qui a été d'un soutien indefectible depuis le début de mon parcours académique à Paris. Les échanges que nous avons eus ont été d'une grande aide.

Je souhaite remercier sincèrement Monsieur **Frédéric Fauberteau** pour les séances de course à pieds qui ont constitué un véritable échappatoire. Je le remercie également sa gentillesse, sa bienveillance ainsi que pour les échanges toujours très intéressants qu'il m'a permis d'avoir avec lui. Il a été un véritable soutien.

Je remercie du fond du coeur Monsieur **Benjamin Bobbia** pour sa gentillesse, son soutien sans commune mesure, et son aide précieuse. Sa présence a fortement contribué au bon déroulement de mon parcours en thèse, à ma motivation et à ma persévérance pour atteindre mes objectifs. Il m'a beaucoup faite évoluer tant sur le plan académique que personnel et je le remercie pour nos échanges ainsi que pour tous les moments partagés.

Je remercie sincèrement le Profeseur **Nicolas Travers** pour tous les échanges que j'ai eu l'occasion d'avoir avec lui. Sa gentillesse, ses conseils toujours avisés et son enthousiasme ont été pour moi un véritable soutien.

Je souhaite remercier le Professeur **Davide Mazza**, qui m'a suivie depuis l'école d'ingénieurs jusqu'à la fin de ma thèse, pour tout ce qu'il m'a apporté durant mon parcours académique.

Je souhaite exprimer ma plus profonde gratitude envers le Professeur **Sergio Focardi**. Plus qu'un professeur, il a été pour moi un modèle et une source d'inspiration inestimable durant ma thèse. Je le remercie du fond du coeur d'avoir toujours cru en moi. Son soutien, ses conseils et ses encouragements ont contribué à me donner la force de surmonter les obstacles pour suivre mes rêves et atteindre mes objectifs. Je suis très reconnaissante des échanges, toujours très intéressants, qu'il m'a permis d'avoir avec lui. Ses enseignements et le partage de ses connaissances m'ont énormément apporté sur le plan personnel comme académique. Je le remercie pour tout ce qu'il a fait pour moi.

J'adresse mes sincères remerciements à Mesdames et Messieurs **Bastien Lejeune, Samira Essardy et Christina Keller** pour leur gentillesse et leur soutien qui ont rendu mon parcours académique très agréable à leurs côtés.

Je remercie sincèrement les doctorants du DVRC: **Massinissa Ferhoune, Ihab Taleb, Walid Samah, Adrien Akar** sans qui cette aventure n'aurait pas été la même. Ils ont su m'accompagner et me soutenir tout au long de ma thèse et de mes périodes d'enseignement avec une gentillesse inégalable.

Durant mon doctorat, j'ai eu l'occasion d'enseigner en qualité de chargée de TDs à l'ESILV. Cette expérience a été très enrichissante, elle m'a fait grandir, et évoluer. Je remercie mes étudiants qui m'ont fait vivre une expérience inoubliable. Ils ont fortement contribué à mon évolution.

Je remercie du fond du coeur ma Professeure de Physique-Chimie, **Françoise Janvier**, au lycée Pierre-Béghin, qui a joué un rôle clé dans mon parcours académique. En effet, sans elle, mon parcours académique n'aurait pas été le même. Sa gentillesse, sa patience, son soutien, sa confiance en moi, et ses encouragements ont grandement contribué à mon évolution et à ce que je suis devenue aujourd'hui.

J'adresse mes sincères remerciements à la Proviseure de ma classe préparatoire, **Bétrice Duchemin**, qui m'a apporté un grand soutien dans les moments où j'en avais besoin pour que je puisse réaliser le parcours académique que je souhaitais.

Je remercie sincèrement mon Professeur de Mathématique de PCSI, au lycée Champollion, **Eric Labeye-Voisin**, qui a été un Professeur exceptionnel et qui m'a donné le goût de poursuivre mon cursus dans les Mathématiques. Sa pédagogie, ses qualités en enseignement et sa passion pour les Mathématiques qu'il savait partager avec ses élèves ont été pour moi une source d'inspiration qui a guidé mes études.

J'adresse mes sincères remerciements à mon Professeur d'anglais, **Jérôme Poirier**, pour sa gentillesse, ses conseils avisés et ses encouragements. Je le remercie pour tous les échanges enrichissants qu'il m'a permis d'avoir avec lui et qui ont fortement contribué à mon évolution.

J'adresse mes sincères remerciements à **Robin Escallier** qui a grandement contribué à mon évolution tant sur le plan personnel que sur le plan académique. Sa présence, sa bienveillance et sa gentillesse m'ont accompagnée tout le long de mon parcours. Il m'a permis de passer des moments inoubliables. Ses conseils toujours avisés m'ont aidée à faire les bons choix, à suivre mes rêves et à atteindre mes objectifs. Son soutien durant mon parcours académique a été sans commune mesure et a joué un rôle clé dans ce que je suis devenue aujourd'hui. Merci pour tout.

Je remercie sincèrement **Yohan Petetin** pour les moments partagés ensemble ainsi que pour son soutien pendant les moments difficiles de ma thèse.

J'adresse mes sincères remerciements à **Arnaud Dapogny** qui a fortement contribué à ma réussite aujourd'hui. Je le remercie de m'avoir donné la chance de croiser son chemin et partager des moments à ses côtés. Je le remercie particulièrement pour sa profonde gentillesse, son soutien et ses encouragements ainsi que pour tous les échanges et les bons moments partagés. Il a toujours su me motiver dans les moments difficiles. Sa vision des choses m'a grandement aidée à avancer avec enthousiasme et positivisme.

Je voudrais remercier du fond du coeur mes amis pour le soutien indéfectible qu'ils m'ont apporté tout au long de mon cursus. Leur présence, leurs encouragements et leurs mots bienveillants ont été essentiels pour moi durant cette période intense et exigeante. Leurs conseils, les moments de distraction bienvenus et leur compréhension ont été d'une aide précieuse. Leur amitié a rendu cette aventure beaucoup plus supportable et m'a donné la force de continuer, même dans les moments les plus difficiles. Merci d'avoir été là, de

m'avoir écoutée, motivée et soutenue. Je n'aurais pas pu y arriver sans eux.

J'adresse mes sincères remerciements à **Anne-Lise Maintenant et Keyvann Zertal**. Leur amitié, leur soutien et leur gentillesse ont fortement contribué à me motiver durant mon parcours académique. Les moments passés ensemble ainsi que nos échanges ont toujours été très sympathiques et m'ont donné la force d'atteindre mes objectifs.

Je remercie sincèrement **Caroline et Noémie Trenz** pour m'avoir accompagnée depuis mon plus jeune âge tout au long de mon parcours. Leur amitié m'est très chère. Leur soutien constant m'a permise d'avancer en croyant en mes rêves et d'atteindre mes objectifs.

Je remercie particulièrement mes amis de classe préparatoire au lycée Champollion : **Fanny Cecconello, Alice Protière, Loïc Bernard, Arthur Lacote, Dan Nguyen, Alexis Pernot, Mathias Laroche, Antoine Charlet, Charles Thoulon, Jean-Lou Quetin**, aux côtés de qui j'évolue depuis de nombreuses années et avec qui je passe des moments incroyables. Leur soutien et leur présence constituent pour moi un moteur au quotidien.

Je remercie mes amis de Master 2, **Clotilde Miura, Cédric Allain, Miguel Madrid-Vertez, Alexis Gerbeaux, Mithuran Gajendran** avec qui je partage ma vie au quotidien. En effet, ils me font évoluer, me conseillent, me soutiennent et me donnent le goût d'avancer sans jamais baisser les bras. Je remercie du fond du coeur Clotilde pour toutes les soirées passées ensemble, son soutien sans commune mesure, sa gentillesse, sa bienveillance et nos échanges qui me font toujours du bien. Je remercie Alexis pour sa profonde gentillesse, sa bienveillance et toutes les sessions de sport partagées toujours très sympathiques. Je remercie Cédric pour son soutien inconditionnel tout au long de ma thèse, les journées passées à la bibliothèque, les verres partagés, nos discussions ainsi que pour la préparation à ma soutenance. Je remercie Miguel pour sa gentillesse, ses encouragements, et son enthousiasme constant qui ont contribué à me motiver. Je remercie Mithuran pour son soutien sans commune mesure depuis la Suède. Nos échanges au cours de ses appels réguliers ont été pour moi une source de motivation et ont fortement contribué à la réussite de ma thèse. Je ne les remercierai jamais assez pour tout ce qu'ils font pour moi au quotidien.

Je remercie également mes amis d'école de l'ESILV: **Coralie Piriou, Madalina Nicolae, Manon Delrieux, Anthony Clark, Victor Dumas, Brice Parylusian, Grégor Jouet, Yliess Hati, Zacharie Guillaume, Loïc Hans, Thibault Herrera-Mione, Samy Injar, Pierre Heckly-Leydier, Mathys Ferriere, Victor Dumas, Marion Dussouillez, Marion Dalciet, Karim Kidiss**. J'ai pu partager avec eux les moments heureux comme les moments plus difficiles. Ils ont toujours su me divertir, me soutenir et m'offrir une oreille attentive lorsque j'en avais besoin. Je remercie particulièrement **Coralie Piriou**. Sa présence, sa générosité, sa gentillesse et son soutien sans commune mesure ont été des éléments clés de ma réussite. Elle m'a fortement aidée à surmonter les moments difficiles, à croire en mes rêves sans ne jamais baisser les bras. Nos échanges m'ont grandement faite évoluer. Je la remercie pour tout ce qu'elle a fait pour moi.

Je remercie sincèrement **Anthony Clark** pour tous les bons moments partagés ensemble depuis notre rencontre. Sa présence, nos échanges réguliers ainsi que ses mots toujours bienveillants ont constitué un véritable soutien pour moi. Sa mentalité de compétiteur et sa philosophie ont été une source d'inspiration et de motivation. Je le remercie d'avoir été un véritable modèle qui m'a permis d'avancer en visant les objectifs fixés et en gardant mes rêves en tête.

Je remercie les doctorants du DVIC **Madalina Nicolae, Brice Parylusian, Grégor Jouet, Yliess Hati, Zacharie Guillaume, Medhi Sam, Teddy Leclerc, Thomas Raynal, Valentin, Paul-Peter Arslan**, qui ont été de véritables piliers durant mon parcours académique. Leur gentillesse, leur bienveillance et leur support au quotidien m'ont fortement aidée à surmonter les obstacles pour atteindre mes objectifs sans ne jamais baisser les bras. Je les remercie pour tous les moments passés ensemble tout au long de ma scolarité.

J'adresse mes sincères remerciements au meilleur parrain **Victor Dumas**, qui est devenu un vrai ami. Il a rendu mon parcours académique très agréable. Je suis extrêmement reconnaissante de tout ce qu'il a fait pour moi depuis que je suis arrivée à Paris. Je le remercie pour tous les moments partagés et les échanges que j'ai eu la chance d'avoir avec lui. Sa gentillesse, sa patience, et sa bienveillance ont constitué un énorme soutien qui m'a permis d'arriver là où je suis aujourd'hui. Je ne le remercierai jamais assez pour son soutien inconditionnel. Je n'en serai jamais arrivée là sans lui.

Je remercie **Alexis Bogroff et Nicolas Pesci** qui ont été des chargés de TDs extraordinaires quand j'étais étudiantes et aux côtés de qui j'ai eu la chance de faire mes premiers pas dans l'enseignement. Leurs conseils, leur gentillesse et leur patience m'ont permis d'appréhender l'enseignement dans les meilleures conditions possibles avec plus de sérénité. Plus que des chargés de TDs ou des collègues, ils sont devenus de véritables amis. Je remercie Alexis pour toutes les sessions de sport partagées ainsi que pour nos échanges qui me faisaient beaucoup réfléchir et qui ont contribué à mon évolution et à ma réussite.

J'adresse mes sincères remerciements à l'équipe du Ministère de l'Intérieur au sein de laquelle j'ai réalisé mon stage de Master 1: **Philippe Bron, Fabien Antoine, Caroline Robillard, Cristian Perez, Victor Journé, Stanislas Ormieres**. Ce stage a été une vraie expérience pour moi au cours de laquelle j'ai découvert des personnes extraordinaires qui m'ont formée et qui m'ont donné les clés nécessaires pour que je puisse réaliser mes rêves et arriver là où je suis aujourd'hui. Leurs enseignements ont marqué mon parcours et m'ont permis d'évoluer à tous points de vue. Depuis mon stage, ils m'ont toujours accompagnée avec une gentillesse et une bienveillance sans commune mesure. Je les remercie pour tout ce qu'ils ont fait pour moi.

J'adresse ma profonde gratitude à mon médecin, le Docteur **Olivier Mitaut**, pour tout ce qu'il a fait pour moi depuis mon plus jeune âge. Ses compétences, son écoute, son dévouement, sa patience, sa gentillesse, ses qualités humaines et sa générosité ont été d'un grand réconfort durant les périodes difficiles. Ses conseils avisés m'ont guidée tout au long mon parcours académique. Il m'a permise de grandir et d'évoluer à ses côtés depuis mon plus jeune âge pour devenir ce que je suis aujourd'hui. Sa confiance en moi

et son soutien ont été pour moi une source de motivation et d'inspiration pour arriver à surmonter les obstacles et atteindre mes objectifs. Pour tout cela, je le remercie du fond du coeur.

Je remercie sincèrement mon osthéopathe, **Jean-Baptiste Audia** pour son accompagnement depuis mon plus jeune âge. Son soutien, sa gentillesse, son dévouement ont été d'une grande aide pour moi et ont fortement contribué à mon développement personnel et au bon déroulement de mon parcours académique. Je le remercie pour toutes les séances qu'il a su me trouver à la dernière minute, avec une gentillesse sans précédent. Sa confiance en moi et ses encouragement m'ont permise de rester motiver et de croire en mes rêves. Je le remercie sincèrement pour tout ce qu'il a fait pour moi.

Je tiens à exprimer ma profonde gratitude envers ma famille, dont le soutien indéfectible a été le pilier de cette aventure académique. Leur amour, leur patience et leur soutien ont été mes sources d'inspiration constantes tout au long de ce parcours, catalyseurs essentiels de mon succès.

À mes parents Marie-Agnès et Guy, sans qui je ne serai pas là où j'en suis aujourd'hui. Je suis infiniment reconnaissante pour leur encouragement inébranlable, leurs sacrifices et leur foi en moi. Leurs conseils avisés et leur soutien émotionnel m'ont permis de surmonter les défis et de persévérer dans la réalisation de cette thèse. Je suis extrêmement reconnaissante de tout ce qu'ils ont fait pour moi depuis ma naissance. En effet, depuis mon plus jeune âge, ils m'ont toujours soutenue et encouragée. Ils ont été mes premiers enseignants, mes premiers guides. Ils ont toujours été là pour moi dans les bons comme les mauvais moments. Leur amour inconditionnel et leur soutien indéfectible ont fait de moi la personne que je suis aujourd'hui. Mon admiration à leur égard pour les parents qu'ils sont depuis ma naissance a constitué la source de motivation principale durant mon parcours et ma vie en général. Je ne les remercierai jamais assez pour tout ce qu'ils ont fait et continue de faire pour moi.

À ma petite soeur Lucie, dont le soutien inconditionnel et les encouragements ont été une source de réconfort et de motivation inestimable. Ses mots d'encouragement, son soutien et sa présence m'ont donné la force nécessaire pour atteindre mes objectifs. En effet, elle est pour moi une source d'inspiration et de réconfort. Je suis extrêmement reconnaissante pour tout ce qu'elle a fait pour moi jusque-là. Ces quelques mots ne suffiront pas à exprimer toute la gratitude que je ressens à son égard pour son soutien inconditionnel, sa patience et sa compréhension. Elle a toujours su trouver les mots justes pour m'encourager et me guider, elle a toujours été là pour moi même dans les moments les plus difficiles. Son amour et sa générosité m'ont permise d'arriver là où je suis aujourd'hui. Merci également à Victoria.

À ma tante Joëlle, qui a été d'une aide sans commune mesure tout au long de mon parcours académique. Sa présence, son soutien et ses encouragements m'ont fourni la motivation dont j'avais besoin pour surmonter les éventuels obstacles et poursuivre mes rêves. Je suis extrêmement reconnaissante de tout ce qu'elle a fait pour moi. Nos échanges réguliers et son écoute attentive m'ont toujours permis de me sentir soutenue et comprise. Je la remercie pour tous les moments partagés, pour les gestes d'amour et pour le soutien qui ont toujours eu une grande importance pour moi et qui ont fait la

différence dans ma vie.

Je remercie également tous mes oncles et tantes, cousins et cousines. En particulier, je remercie **Jérémy, Amélie, Charlotte et Léo** pour tous les bons moments passés durant mon parcours académique à Paris.

Enfin j'adresse mes sincères remerciements à **Elisabeth, Bruno, Grégoire, Lysianne, Auriane et Valériane** qui font partie intégrante de ma famille et qui ont su m'accompagner depuis mon plus jeune âge pour me permettre de réaliser mes rêves. Je les remercie du plus profond de mon être pour tous les moments partagés, pour les échanges que nous avons eus et pour leur soutien depuis toujours.

Je souhaite remercier du fond du coeur **Pierre Jahjah-Oueis** pour tout ce qu'il a fait pour moi depuis notre rencontre. En effet, je le remercie sincèrement pour tout le soutien qu'il m'a apporté. Son amour, sa patience et sa gentillesse ont été des sources inestimables de force et de motivation qui m'ont permis de surmonter les moments de doute et de fatigue. Tout au long de mon parcours, ses encouragements, ses conseils avisés ainsi que sa confiance en moi m'ont donné la force nécessaire pour mener à bien ma thèse en gardant en tête mes objectifs. Je le remercie pour tous les moments qu'il m'a permis de passer à ses côtés, les discussions que nous avons eues, le partage de ses expériences, me donnant ainsi l'énergie de poursuivre mes rêves. Sa gentillesse, sa générosité et son attention m'ont fourni le réconfort nécessaire dont j'avais besoin dans les moments plus difficiles. Je le remercie également pour tous les moments de distraction et de complicité qu'il m'a offerts. Sans sa présence, mon parcours académique comme personnel n'aurait pu être le même. Je lui suis extrêmement reconnaissante et je ne le remercierai jamais assez pour tout ce qu'il a fait pour moi.

Abstract

In a financial context where the temporal and spatial effects of processes are decisive in describing many of the phenomena that occur, we are interested in studying the most common temporal and spatial effects necessary to properly describe these phenomena. We are wondering whether there exists characteristic properties related to these temporal and spatial effects that would have consequences on the processes. Addressing such questions is the purpose of Chapter 2. Indeed, in Chapter 2, we focus on two time transformations: the time-translation, also called time-origin change, and the time-scaling; and on the related properties called stationarity and self-similarity. The stationarity property is the invariance in time and space of a function or a process by time-origin change. The self-similarity property establishes a spatial proportionality relationship between the characteristics of a random function or a random process taken at two proportional times λt and t with $\lambda > 0$, with a spatial proportionality factor corresponding to a function of the time proportionality factor λ. We present a new approach to study stationary and self-similar random processes. First, we prove that the only assumptions of stationarity and self-similarity of the squared $\mathbb{L}^2$-norm of a given process in Hilbert space, without any assumption of distribution, allow obtaining the stationarity and the self-similarity of the inner product of the process, with a closed form formula for the latter only depending on power functions of exponent $\gamma \in (0, 1)$. In a second time, we show that adding the Gaussian assumption allows obtaining the stationarity and the self-similarity properties of the process not only in distribution but also in the trajectory sense (that is, in terms of equality of processes). We provide some examples of such processes, known as the Wiener process and the fractional Brownian motion (fBm). Finally, we provide an extension of the stationarity and self-similarity properties in the trajectory sense to the multidimensional Gaussian Hilbert space. We provide as an example of multidimensional self-similar Gaussian processes with stationary and correlated increments, the so-called multivariate fractional Brownian motions (mfBm). In Finance, self-similar Gaussian processes with stationary increments, called fractional Brownian motions, are useful for describing various processes exhibiting long-term (respectively short-term) dependence, such as volatility clustering, trajectories of correlated logarithmic returns, or for options pricing, portfolio optimization, high-frequency trading, or risk management, all domains in which the underlying dynamics exhibit long-range (respectively short-range) dependence. In the classical framework, price dynamics are usually described using a geometric Brownian motion (GBM), also called exponential Brownian motion. In this model, logarithmic returns are distributed as an arithmetic Brownian motion; thus they are independent of each other and do not capture long-range (respectively short-range) dependence. We wonder how this model can be improved to describe the distribution of logarithmic returns as realistically and as accurately as possible, to predict risk measures with realism and precision. We wonder how we can modify the process to capture long-range (respectively short-range) dependency. In Chapter 3, we propose to describe price trajectories using fractional geometric

Brownian motions. Logarithmic returns are then described using self-similar Gaussian processes with stationary and correlated increments, called fractional Brownian motions (fBm). This allows adding correlations between logarithmic returns to express long-range (respectively short-range) dependence.

In this context, we focus on predicting the most commonly used risk measure by regulators, called Value-at-Risk (VaR). We introduce a model that provides a conditional VaR for the Gaussian approximation of the future asset portfolio variation under fractional dynamics, given the past asset variations. Since this VaR is Gaussian, we show that such a model is based on orthogonal projections in a Gaussian Hilbert space, taking specific forms for which closed-form formulae are provided. Then, the parameters of the model are estimated under the assumptions of Gaussianity, stationarity and self-similarity of the assets log-returns. Finally, we quantify the Gaussian approximation of VaR by providing an upper bound of the error. Backtesting experiments are provided on simulated and market data to illustrate the theory. Gaussian framework is the most commonly used in Finance due to its convenient properties. However, the Gaussian distribution is thin-tailed, then it assumes that extreme events are rare, and tends to underestimate the probability of such events occurring. Then, the modeling of the log-returns distribution by a Gaussian distribution is not appropriate when extreme events more frequently occur. We are wondering how to accurately model a distribution in which extreme events occur more frequently than in a Gaussian distribution. To address such a question, we propose to model the loss distribution thanks to a heavy-tailed distribution like Pareto, that assigns higher probabilities to extreme events, which is often more realistic in Finance. However, in a context in which extreme events occur more frequently than in the Gaussian framework, VaR is not any more an efficient risk measure since it is not sensitive to the tail risk, then it fails to capture extreme events; this leads to an underestimation of the risk. Therefore, we can wonder whether there exists a risk measure more sensitive to the tail risk. We propose to replace VaR with Expected-Shortfall (ES) that addresses such shortcomings. We therefore ask how to estimate the ES as efficiently and accurately as possible in the case of a heavy-tailed distribution. Addressing such a question is the purpose of Chapter 4. Indeed, in Chapter 4, the objective is to explore robust methods for estimating the ES in heavy-tailed distributions. The Expected-Shortfall being the average of losses exceeding the VaR, an ES estimator is an estimator of the mean applied to the distribution tail beyond the VaR. Thus, we explore robust mean estimators as an alternative to simple empirical mean, in heavy-tailed distributions through the toy case of the Pareto distribution. The Pareto distribution presents interesting properties of stability by conditioning and rescaling. We recall the theory on the Expected-Shortfall and on the Pareto distribution, and we present the characteristic properties of the latter. The rescaling property establishes a proportionality relationship between two Pareto distributions with the same shape parameter but distinct scale parameters, with a proportionality factor equal to the ratio between the two scale parameters. The stability by conditioning property states that when the conditioning threshold is independent of the underlying sample, the Pareto distribution conditional on its values being above the given threshold is still a Pareto distribution with the same shape parameter but a scale

parameter equal to the conditioning threshold. Combining the stability by conditioning and rescaling properties allows establishing a proportionality relationship between the Pareto distribution conditional on its values being above the given threshold and the marginal Pareto distribution, with a proportionality factor equal to the ratio between the conditioning threshold and the scale parameter of the marginal Pareto distribution. Therefore, ES of any Pareto distribution is proportional to the expectation of the marginal distribution, with a proportionality factor equal to the theoretical Value-at-Risk at the risk level α. Consequently, the problem of estimation of the $\mathbf{ES}_\alpha$ in a Pareto distribution, when the conditioning threshold is the true $\mathbf{VaR}_\alpha$, reduces to the problem of estimation of the expectation in the marginal Pareto distribution. Moreover, owing to the rescaling property, the problem of estimation of the expectation in any Pareto distribution reduces to the estimation of the expectation in the standardized Pareto distribution. Consequently, the stability by conditioning and rescaling properties are very useful since they allow reducing the complexity of the problem of estimation of the ES. Then, several non-asymptotic mean estimators, such as the Median-of-Means, the Trimmed-Mean, or even the Lee-Valiant estimator, are presented with their characteristics and compared to the classical empirical mean. We study their bias and provide explicit formulae when possible. Moreover, we evaluate the convergence rate of the bias. Finally, we support the theoretical analysis with some experiments, and we compare the performance of the different estimators.

Abstract in layman's terms

In everyday life, the temporal and spatial properties of the processes govern numerous phenomena in various fields such as Physics, Biology, Astronomy, Telecommunications, Music, Finance and in many others. In the first part of this thesis, we focus on two of these properties, called stationarity and self-similarity. Stationarity advocates temporal and spatial invariance of the characteristics of a given process through time-translation, or time-origin change, and thus allows describing processes that repeat identically over time as they move. We find this property notably in Music to describe certain pieces, or in Finance to describe the asset return series. Self-similarity establishes a spatial proportionality relationship between the characteristics of a process taken at two proportional times, and the spatial proportionality coefficient is a function of the time proportionality factor. Thus, the self-similarity property makes it possible to describe processes whose states, taken at proportional times, reproduce proportionally with a proportionality factor corresponding to a power function of the time proportionality coefficient. Therefore, the self-similarity property describes long-range or short-range event dependencies, according to the value taken by the exponent of the power function. We find this property in nature, to describe the development of ferns, of the Romanesco broccoli, or the Koch's flake; in Biology to explain population growth and genealogy, in Physics to analyze the weather phenomena; in Astronomy to describe the stellar fragments; or in Finance to model asset return series, volatility or risk measures. In the second part of this thesis, we apply stationary and self-similar processes in Finance, and more precisely in the risk management field, in a well-known framework, called Gaus-

sian. This means that the distribution of the asset returns is modeled by a Gaussian distribution which assumes that most values taken by the processes are concentrated around their mean, and very few values go to the extreme values. Stationary and self-similar Gaussian processes are used to model the asset return series that present a long-range or short-range dependence, to accurately predict the most commonly used risk-measure by regulators, called Value-at-Risk (VaR), for the future portfolio variation conditional on the past variations of its assets. The Gaussian framework has very interesting properties, nevertheless, it is suitable for describing smooth periods as it does not take into account potential extreme events, such as financial crises. It tends to underestimate the probability such events occurring, which can be very dangerous. To address this issue, we propose replacing the Gaussian framework with a framework more conducive to the presence of such events. In this new framework, the distribution of asset returns is no longer modeled by a Gaussian distribution but by a distribution capable of assigning a higher probability to extreme events, called heavy-tailed distribution. A widely used heavy-tailed distribution in Finance for its convenient properties is the Pareto distribution. However, applying VaR in this new setting would result in a poor risk assessment. Indeed, VaR is a risk measure insensitive to extreme events. Thus, applying VaR to a heavy-tailed distribution, in which extreme events are more likely to occur, would result in an underestimation of risk. Therefore, we propose an alternative to VaR, called Expected-Shortfall (ES), capable of taking extreme events into account. In the third part of this thesis, robust methods of mean-estimation for the ES in the Pareto distribution are proposed.

Résumé

Dans un contexte financier où les effets temporels et spatiaux des processus sont décisifs pour décrire de nombreux phénomènes qui se produisent, nous nous intéressons à l'étude des effets temporels et spatiaux les plus courants nécessaires pour décrire correctement ces phénomènes. Nous nous demandons s'il existe des propriétés caractéristiques liées à ces effets temporels et spatiaux qui auraient des conséquences sur les processus. Aborder de telles questions est l'objectif de Chapter 2. En effet, dans Chapter 2, nous nous concentrons sur deux transformations temporelles : la translation temporelle, également appelée changement d'origine temporelle, et la mise à l'échelle temporelle, ainsi que sur les propriétés associées appelées stationnarité et auto-similarité. La stationnarité est l'invariance dans le temps et l'espace des caractéristiques d'une fonction aléatoire ou d'un processus aléatoire par translation temporelle. La propriété d'auto-similarité établit une relation de proportionnalité spatiale entre les caractéristiques d'une fonction aléatoire ou d'un processus aléatoire évalués en deux instants temporels proportionnels λt et t avec $\lambda > 0$, avec un coefficient de proportionnalité spatiale correspondant à une fonction du coefficient de proportionnalité temporelle λ. Nous présentons une nouvelle approche pour étudier les processus stationnaires et auto-similaires dans l'espace de Hilbert. Tout d'abord, nous prouvons que les seules hypothèses de stationnarité et d'auto-similarité du carré de la norme $\mathbb{L}^2$ d'un processus donné dans un espace de Hilbert, sans aucune hypothèse de distribution, permettent d'obtenir la stationnarité et l'auto-similarité du produit scalaire du processus, dont une formule fermée ne dépendant que de fonctions puissances d'exposant $\gamma \in (0, 1)$ est établie. Dans un second temps, nous montrons qu'en ajoutant l'hypothèse Gaussienne, il est possible d'obtenir les propriétés de stationnarité et d'auto-similarité du processus non seulement en distribution mais aussi au sens des trajectoires (c'est-à-dire, en termes d'égalité des processus). Nous fournissons quelques exemples de tels processus, connus sous le nom de processus de Wiener et de mouvement Brownien fractionnaire (fBm). Enfin, nous proposons une extension des propriétés de stationnarité et d'auto-similarité au sens des trajectoires dans l'espace de Hilbert Gaussien multidimensionnel. Nous fournissons comme exemple de processus multidimensionnel Gaussien auto-similaire avec des incréments stationnaires et corrélés, le bien-connu mouvement Brownien fractionnaire multivarié (mfBm). En finance, les processus Gaussiens auto-similaires avec des incréments stationnaires et corrélés, appelés mouvements Browniens fractionnaires, sont utiles pour décrire divers processus présentant une dépendance à long-terme (respectivement à court-terme), tels que le regroupement de la volatilité, les trajectoires de rendements logarithmiques corrélés, ou pour la tarification des options, l'optimisation de portefeuille, le trading haute fréquence ou la gestion des risques, tous des domaines dans lesquels les dynamiques sous-jacentes présentent une dépendance à long-terme (respectivement à court-terme). Dans le cadre classique, les dynamiques des prix sont généralement décrites à l'aide d'un mouvement Brownien géométrique (GBM), également appelé mouvement Brownien exponentiel. Dans ce modèle, les rendements logarithmiques sont distribués selon un mouvement

Brownien arithmétique; ils sont donc indépendants les uns des autres et ne capturent pas la dépendance à long-terme (respectivement à court-terme). Nous nous demandons comment ce modèle peut être amélioré pour décrire la distribution des rendements logarithmiques de la manière la plus réaliste et précise possible, permettant ainsi la prédiction des mesures de risque avec réalisme et précision. Nous nous demandons comment nous pouvons modifier le processus pour capturer la dépendance à long-terme (respectivement à court-terme). Dans Chapter 3, nous proposons de décrire les trajectoires de prix en utilisant des mouvements Browniens fractionnaires géométriques. Cela permet d'ajouter des corrélations entre les rendements logarithmiques pour exprimer la dépendance de long-terme ou à court-terme. Les rendements logarithmiques sont alors décrits à l'aide de processus Gaussiens auto-similaires avec des incréments stationnaires et corrélés, appelés mouvements Browniens fractionnaires (fBm). Dans ce contexte, nous nous concentrons sur la prédiction de la mesure de risque la plus couramment utilisée par les régulateurs, appelée Valeur-à-Risque (VaR). Nous introduisons un modèle qui fournit une VaR conditionnelle pour l'approximation Gaussienne de la variation future du portefeuille d'actifs sous des dynamiques fractionnaires, étant donné les variations passées des actifs du portefeuille. Étant donné que cette VaR est Gaussienne, nous montrons qu'un tel modèle est basé sur des projections orthogonales dans un espace de Hilbert Gaussien, prenant des formes spécifiques pour lesquelles des formules fermées sont fournies. Ensuite, les paramètres du modèle sont estimés sous les hypothèses de Gaussianité, de stationnarité et d'auto-similarité des log-rendements. Enfin, nous quantifions l'approximation Gaussienne de la VaR en fournissant une borne supérieure de l'erreur. Des expériences de backtesting sont fournies sur des données simulées et de marché pour illustrer la théorie.

Le cadre Gaussien est le plus couramment utilisé en finance en raison de ses propriétés pratiques. Cependant, la distribution Gaussienne a des queues fines, ce qui signifie qu'elle suppose que les événements extrêmes sont rares et tend à sous-estimer la probabilité de survenance de tels événements. Par conséquent, la modélisation de la distribution des rendements logarithmiques par une distribution Gaussienne n'est pas appropriée lorsque des événements extrêmes se produisent plus fréquemment. Nous nous demandons comment modéliser avec précision une distribution dans laquelle les événements extrêmes se produisent plus fréquemment que dans une distribution Gaussienne. Pour répondre à une telle question, nous proposons de modéliser la distribution des pertes grâce à une distribution à queues lourdes telle que la distribution de Pareto, qui attribue des probabilités plus élevées aux événements extrêmes, ce qui est souvent plus réaliste en finance. Cependant, dans un contexte où les événements extrêmes se produisent plus fréquemment que dans le cadre Gaussien, la VaR n'est plus une mesure de risque efficace car elle n'est pas sensible au risque de queue, et échoue donc à capturer les événements extrêmes, ce qui conduit à une sous-estimation du risque. Nous pouvons donc nous demander s'il existe une mesure de risque plus sensible au risque de queue. Nous proposons de remplacer la VaR par l'Espérance-de-Perte (ES) qui répond à de telles lacunes. Notre question est de savoir comment estimer l'ES de manière la plus efficace et précise possible dans une distribution à queues lourdes. Aborder une telle question est l'objectif de Chapter 4. En effet, dans Chapter 4, l'objectif est d'explorer

des méthodes robustes pour estimer l'Espérance-de-Perte (ES) dans des distributions à queues lourdes. L'Espérance de Perte étant la moyenne des pertes dépassant la VaR, un estimateur de l'ES est un estimateur de la moyenne appliqué à la queue de la distribution au-delà de la VaR. Ainsi, nous explorons des estimateurs robustes de la moyenne comme alternative à la simple moyenne empirique, dans des distributions à queues lourdes, à travers le cas jouet de la distribution de Pareto. La distribution de Pareto présente des propriétés intéressantes de stabilité par conditionnement et remise à l'échelle. En effet, la propriété de mise à l'échelle établit une relation de proportionnalité entre deux distributions de Pareto dont le paramètre de forme est le même mais les paramètres d'échelle respectifs sont distincts, avec un facteur de proportionnalité égale au ratio entre les deux paramètres d'échelle. La propriété de stabilité par conditionnement énonce que toute distribution de Pareto conditionnée à ses valeurs supérieures à un certain seuil indépendant de l'échantillon sous-jacent, est toujours une distribution de Pareto de même paramètre de forme mais de paramètre d'échelle égal au quotient entre le paramètre de conditionnement et le paramètre d'échelle de la distribution de Pareto marginale. Enfin, en combinant les propriétés de stabilité par conditionnement et de mise à l'échelle, une relation de proportionnalité est établie entre la distribution de Pareto conditionnée à ses valeurs étant supérieures à un certain seuil de conditionnement indépendant de l'échantillon sous-jacent, et la distribution de Praeto marginale, avec un coefficient de proportionnalité égale au ratio entre le paramètre de conditionnement et le paramètre d'échelle de la distribution de Pareto marginale. Ainsi, l'ES d'une distribution de Pareto est proportionnelle à l'espérance de la distribution de Pareto marginale, avec un facteur de proportionnalité égal à la VaR théorique au seuil de risque α. Ainsi, les proportiétés de stabilité par conditionnement et de mise à l'échelle sont très utiles en ce sens qu'elles permettent de réduire la complexité du problème d'estimation de l'ES. Tout d'abord, nous rappelons la théorie sur l'Espérance-de-Perte (ES) et sur la distribution de Pareto, et nous présentons les propriétés caractéristiques de cette dernière. Ensuite, plusieurs estimateurs non-asymptotiques de la moyenne, tels que la Médiane-des-Moyennes (MoM), la Moyenne-Tronqué (TM)e, ou même l'estimateur de Lee-Valiant (LV), sont présentés avec leurs caractéristiques et comparés à la moyenne empirique classique. Nous étudions leur biais et fournissons des formules explicites lorsque cela est possible. De plus, nous évaluons le taux de convergence du biais. Enfin, nous soutenons l'analyse théorique par des expériences et comparons les performances des différents estimateurs.

Résumé vulgarisé

Dans la vie quotidienne, les propriétés temporelles et spatiales des processus régissent de nombreux phénomènes dans divers domaines tels que la Physique, la Biologie, l'Astronomie, les Télécommunications, la Musique, la Finance, et bien d'autres. Dans la première partie de cette thèse, nous nous concentrons sur deux de ces propriétés, appelées stationnarité et auto-similarité. La stationnarité prône l'invariance temporelle et spatiale par translation, aussi connue sous le nom de changement d'origine temporelle, et permet ainsi de décrire des processus qui se répètent de manière identique au fil du temps à mesure qu'ils se déplacent. Nous retrouvons cette propriété notamment en

Musique pour décrire certains morceaux, ou en Finance pour décrire les séries de rendements des actifs. L'auto-similarité établit une relation de proportionnalité spatiale entre les caractéristiques d'un processus évalué à deux instants temporels proportionnels entre eux, et le coefficient de proportionnalité spatiale est une fonction du facteur de proportionnalité temporelle. Ainsi, la propriété d'auto-similarité permet de décrire des processus dont les états pris à deux instants temporels proportionnels se reproduisent spatialement de manière proportionnelle avec un facteur de proportionnalité spatiale égal à une fonction du facteur de proportionnalité temporelle. Ainsi, l'auto-similarité décrit les dépendances à long-terme ou à court-terme des événements. Nous retrouvons cette propriété dans la nature, pour décrire le développement des fougères, du chou Romanesco ou du flocon de Koch; en Biologie pour expliquer la croissance des populations et la Généalogie, en Physique pour analyser les phénomènes météorologiques; en Astronomie pour décrire les fragments stellaires; ou en Finance pour modéliser les séries de rendements des actifs, la volatilité ou les mesures de risque.

Dans la deuxième partie de cette thèse, nous appliquons des processus stationnaires et auto-similaires en Finance, et plus précisément au domaine de la gestion des risques, dans un cadre bien connu, appelé Gaussien. Cela signifie que la distribution des rendements des actifs est modélisée par une distribution Gaussienne qui suppose que la plupart des valeurs prises par les processus sont concentrées autour de leur moyenne, et très peu de valeurs atteignent les valeurs extrêmes. Les processus Gaussiens stationnaires et auto-similaires sont utilisés pour modéliser les séries de rendements des actifs qui présentent une dépendance à long-terme, afin de prédire avec précision la mesure de risque la plus couramment utilisée par les régulateurs, appelée Valeur-à-Risque (VaR), pour la variation future d'un portefeuille d'actifs. Le cadre Gaussien présente des propriétés très intéressantes, néanmoins, il est adapté pour décrire les périodes calmes car il ne prend pas en compte les potentiels événements extrêmes, tels que les crises financières. Il tend à sous-estimer la probabilité que de tels événements surviennent, ce qui peut être très dangereux. Pour résoudre ce problème, nous proposons de remplacer le cadre Gaussien par un cadre plus propice à la présence de tels événements. Dans ce nouveau cadre, la distribution des rendements des actifs n'est plus modélisée par une distribution Gaussienne mais par une distribution capable d'attribuer une probabilité plus élevée aux événements extrêmes, appelée distribution à queues lourdes. Une distribution à queues lourdes couramment utilisée en finance pour ses propriétés pratiques est la distribution de Pareto. Cependant, appliquer la VaR dans ce nouveau contexte entraînerait une mauvaise évaluation du risque. En effet, la VaR est une mesure de risque insensible aux événements extrêmes. Ainsi, appliquer la VaR à une distribution à queues lourdes, dans laquelle les événements extrêmes sont plus susceptibles de se produire, entraînerait une sous-estimation du risque. Par conséquent, nous proposons une alternative à la VaR, appelée l'Espérance-de-Perte (ES), capable de prendre en compte les événements extrêmes. Dans la troisième partie de cette thèse, des méthodes robustes d'estimation de l'ES dans la distribution de Pareto sont proposées.

Thesis outline and reading guide

Outline

This thesis is about risk measures in finance, backtesting, sensitivity and robustness. It contains a general introduction and three chapters. The general introduction is composed of an historical part, a state of the art, an overview of the contributions and the potential perspectives of the thesis. Chapter 2 deals with the self-similar and stationary Gaussian processes. Chapter 3 is about the prediction of conditional Value-at-Risk of the future variation of an assets portfolio under fractional dynamics. Chapter 4 covers the robust estimation of the Expected-Shortfall in heavy-tailed distribution. The French translation of the state of the art, of the contributions and of the perspectives is provided at the end of the manuscript.

• Chapter 2 : *Introduction to self-similar and stationary Gaussian processes* ($RQ\#1$) In Chapter 2, we focus on two time transformations: the time-translation and the time-scaling and on the related properties called the stationarity and the self-similarity. The stationarity is the invariance in time and space of a function or a process by temporal translation while the self-similarity property refers to the invariance in time and space of a function or a process via an adequate temporal scaling. We present a new approach to study stationary and self-similar processes. First, we prove that the only assumptions of stationarity and self-similarity of the $\mathbb{L}^2$-characteristics of a given process in Hilbert space, without any assumption of distribution, allow obtaining the stationarity and the self-similarity of the inner product of the process, with a closed form formula for the latter. In a second time, we show that adding the Gaussian assumption allows obtaining the stationarity and the self-similarity properties of the process not only in distribution but also in the trajectory sense (i.e. in terms of equality of processes). We provide some examples of such processes, known as the Wiener process and the fractional Brownian motion (fBm). Finally, we provide an extension of the stationarity and self-similarity properties in the trajectory sense, to the multidimensional Gaussian processes, called multivariate fractional Brownian motions (mfBm).

• Chapter 3 : *VaR prediction for asset portfolios under fractional dynamics* ($RQ\#2$, $RQ\#3$)

In Chapter 3, we propose to describe price trajectories using fractional geometric Brownian motions. This allows adding correlations between logarithmic returns to express long-range dependency. Logarithmic returns are then described using self-similar Gaussian processes with stationary and correlated increments, called fractional Brownian motions (fBm). In this context, we focus on predicting the most commonly used risk measure by regulators, called Value-at-Risk (VaR). We introduce a model that provides a Gaussian approximation of VaR for the asset portfolio under fractional dynamics. We show that such a model is based on orthogonal projections in a Gaussian Hilbert space, taking specific forms for which closed-form formulae are provided. Finally, we quantify the

Gaussian approximation of VaR by providing an upper bound of the error. Backtesting experiments are provided on simulated and market data to illustrate the theory.

• Chapter 4 : *Mean estimation of Expected-Shortfall in heavy-tailed distributions* ($\overline{RQ\#4}$, $RQ\#5$) In Chapter 4, the objective is to explore robust methods for estimating the Expected-Shortfall in heavy-tailed distributions. The Expected-Shortfall being the average of losses exceeding the VaR, an ES estimator is an estimator of the mean applied to the distribution tail beyond the VaR. Thus, we explore robust estimators of the mean as an alternative to simple empirical mean, in heavy-tailed distributions through the toy case of the Pareto distribution. Firstly, we recall the theory on the Expected-Shortfall and on the Pareto distribution, and we present the characteristic properties of the latter. Then, several non-asymptotic mean estimators, such as the Median-of-Means, the Trimmed-Mean, or even the Lee-Valiant estimator, are presented with their characteristics and compared to the classical empirical mean. We study their bias and provide explicit formulae when possible. Moreover, we evaluate the convergence rate of the bias. Finally, we support the theoretical analysis with some experiments, and we compare the performances of the different estimators.

Reading guide

The section on the history of finance has been written to set out the historical context that led to the questions addressed in the thesis. This part can be ignored by the most discerning readers.

Each chapter of the main part contains a small introduction which describes the necessary elements of context and is divided into several sections including the context, the main results, the related technical proofs and the numerical experiments. Note that each chapter can be read independently of each other.

For a quick overview of the contributions presented in this thesis, the reader is invited to focus on the summary of contributions in Section 1.3 in Chapter 1.

"The shortest path between two truths in the real domain passes through the complex domain."

(**Jacques Hadamard**, The Mathematical Intelligencer, v. 13, no. 1, Winter 1991.)

Contents

Notation

$:=$	Equal by definition
$\mathbb{N}, \mathbb{R}$	Sets of natural and real numbers
$\mathbb{R}^d$	Set of d-dimensional real-valued vectors
$\langle x, y \rangle$	Inner product of vectors $x, y \in \mathbb{R}^d$
$\|x\|_p$	ℓ_p-norm of vector $x \in \mathbb{R}^d$
$\mathbb{R}^{n \times d}$	Set of real matrices of size $n \times d$
$A^\top$	Transpose of matrix A
$\mathbb{1}_E$	Indicator function of set E
$\mathcal{B}(\mathcal{X})$	Borel σ-field on $\mathcal{X}$
Ω	Sample space: set of all possible outcomes
$\mathcal{F}$	Event set
$\mathbb{P}(\cdot)$	Probability of an event
$(\Omega, \mathcal{F}, \mathbb{P})$	Complete probability space
$X \sim P$	Random variable (r.v.) X has distribution P
$\overset{\text{i.i.d}}{\sim}$	Independent and Identically Distributed
$\perp$	Orthogonal
$\mathbb{E}[\cdot]$	Expectation of a random variable
$\mathbb{V}[\cdot]$	Variance of a random variable
$\langle\langle X, Y \rangle\rangle$	Inner product between the r.v. X and Y: $\mathbb{E}[XY]$
$\alpha \cdot X$	Scalar product between a scalar and a random vectors: $\sum_1^n \alpha_i X_i$
$\mathbb{L}^p(\mathbb{P})$	Set of p-order integrable r.v. w.r.t. measure $\mathbb{P}$: $\left\{ X : \mathbb{E}\left[\left\| X \right\|^p \right] < \infty \right\}$
$\left\| X \right\|_p$	$\mathbb{L}^p$-norm of r.v. X: $\left(\mathbb{E}\left[\left\| X \right\|^p \right] \right)^{\frac{1}{p}}$
$\|X\|_2$	$\mathbb{L}^2$-norm of r.v. X: $\sqrt{\langle\langle X, X \rangle\rangle}$
$\mathcal{H}$	Hilbert space
$\{X(\theta)\}_{\theta \in J}$	finite or infinite family of $\mathbb{L}^2$-r.v.
$\{X_h(\theta)\}_{\theta \in J}$	h-time-translated family of $\mathbb{L}^2$-r.v.: $X(\theta + h) - X(h)$
$\{X^\lambda(\theta)\}_{\theta \in J}$	λ-time-scaled family of $\mathbb{L}^2$-r.v.: $\lambda^{-\frac{1}{2}} X(\lambda t);\ \lambda > 0$
$\mathcal{H}_X$	Hilbert space spanned by the family $\{X(\theta)\}_{\theta \in J}$

$Q_X(\theta)$ or $\mathbf{Q}(X(\theta))$	Quadratic norm of r.v. X: $\|X(\theta)\|_2^2 = \mathbb{E}\left[\|X(\theta)\|^2\right] = \|\mathbb{E}[X(\theta)]\|^2 + \mathbb{V}[X(\theta)]$
$K_X(\theta, \theta')$	Inner kernel: $\langle\!\langle X(\theta), X(\theta') \rangle\!\rangle$
$\mathbf{Q}(\alpha \cdot X)$	Quadratic norm of scalar product: $\sum_{(i,j)\in[1,n]} \alpha_i \, \alpha_j \, K_X(\theta_i, \theta_j)$
$\mathcal{N}(\mu, \Sigma)$	Gaussian distribution with mean μ and covariance matrix Σ
$\mathcal{P}(x_m, \gamma)$	Pareto distribution of scaling parameter $x_m > 0$ and of shape index $\gamma > 0$
$\mathcal{L}_d^2$	Set of d-dimensional square integrable functions $L_2(\mathbb{R}^d, \mathrm{d}u) = \left\{ f : \mathbb{R} \mapsto \mathbb{R}^d \ s.t. \left\|f\right\|_{\mathcal{L}_d^2}^2 := \int_{\mathbb{R}} \|f(u)\|^2 \mathrm{d}u < +\infty \right\}$
$\langle f, g \rangle_{\mathcal{L}_d^2}$	Scalar product in $\mathbb{R}^d$: $\int_{\mathbb{R}} f(u) \cdot g(u) \mathrm{d}u$
$\mathbf{VaR}_\alpha$	Value-at-Risk at the risk level $\alpha \in (0, 1)$
$\mathbf{ES}_\alpha$	Expected-Shortfall at the risk level $\alpha \in (0, 1)$
$\widehat{\mathbf{MoM}}_n$	Median-of-Means (MoM) estimator
$\widehat{\mathbf{RTM}}_n$	Right Trimmed-Mean (RTM) estimator
$\widehat{\mathbf{LV}}_n$	Lee-Valiant (LV) estimator
$\{W(t)\}_{\mathbb{R}}$	Wiener process
$\{B(t)\}_{\mathbb{R}}$	Brownian motion
$\{B^H(t)\}_{\mathbb{R}}$	fractional Brownian motion (fBm)
$\{\boldsymbol{B^H}(t)\}_{\mathbb{R}}$	multivariate fractional Brownian motion (mfBm)

"Mathematics is the language in which God has written the universe."

(**Galileo Galilei**, The Assayer (Il Saggiatore), 1623.)

Chapter 1

General Introduction, Motivations and Contributions

Contents

This chapter is a general introduction to the thesis. The first part deals with the history of finance from the beginnings of the financial markets to the present day, including the financial crises that have punctuated history, the setting up of the banking regulation and the introduction of risk management methods such as risk measurements. In the second part, a state of the art related to each chapter is provided, followed by our contributions and the research prospects.

1.1 History of finance

1.1.1 Origin of the financial markets and financial crisis

First financial markets

The first financial markets date back to ancient times and emerged in various regions around the world. Here are some notable early financial markets. The first financial markets were mostly future / forward markets where the products traded were derivatives based on commodities such as wheat. This is highlighted by the famous quotation of (Chabardès and Delclaux, 1997): "Wheat is wonderful in that it can be bought and sold even before the flood has inundated the country and the grain has been sown".

We recall that a **derivative product** refers to a type of financial contract whose value is dependent on an underlying asset, group of assets, index, rate or benchmark. A derivative is set between two or more parties that can trade on an exchange or over-the-counter (OTC).

These contracts can be used to trade any number of assets and carry their own risks. Prices for derivatives derive from fluctuations in the underlying asset. These financial securities are commonly used to access certain markets and may be traded to hedge against risk. Derivative can be used to either mitigate risk (hedging) or assume risk

with the expectation of commensurate reward (speculation). Derivatives can move risk (and the accompanying rewards) from the risk-averse to the risk seekers.

Specifically, forward and futures contracts are derivatives that involve two parties who agree to buy or sell a specific asset at a set price by a certain date in the future. Buyers and sellers can mitigate the risks of price changes by locking them in advance.

A **forward** is made over the counter (OTC) and settles just once - at the end of the contract. Both parties involved in the agreement privately negotiate the contract's exact term. Forwards carry a default risk since the other party might not come up with the goods or the payment. Many hedgers use forward contracts to reduce the potential volatility of an asset's price. Since the terms are set when they are executed, forward contracts don't fluctuate in price. Since they are private agreements, there is a higher degree of conterparty risk, which means there may be a chance that one party will default.

Like forwards, **futures** contracts involve agreeing to buy and sell an asset at a specific price at a future date. Futures contracts, meanwhile, are standardized to trade on stock exchanges. These contracts are marked to market daily, which means that daily changes are settled daily until the end of the contract. The futures market is generally highly liquid, giving investors the ability to enter and exit whenever they choose to do so. They have far less counterpart, as they guarantee payment on the agreed-upon date.

Around **2000 Before Common Era** (BCE), **ancient Mesopotamia** (modern-day Iraq) had rudimentary financial transactions recorded on clay tablets, suggesting early forms of debt contracts and barter exchanges.

Greek city-states, particularly **Athens**, developed markets where traders could buy, sell, and exchange goods and services. The port of Piraeus was a significant commercial and financial center.

Rome had its own sophisticated financial system, including institutions like the State Bank and markets where commercial transactions took place.

Starting from the **7th** century, financial markets flourished in regions of the **Arab and Islamic empires**, including souks (markets) where goods were exchanged and caravanserais that facilitated long-distance trade.

The merchant cities of **Venice** and **Genoa**, located on trade routes between Europe and Asia, were important financial centers where bills of exchange and maritime insurance contracts were used.

The **Amsterdam Stock Exchange**, founded in **1602**, is one of the earliest stock exchanges in the world and was the center of trading in shares of the Dutch East India Company. It also introduced financial innovations such as call and put options.

The **London Stock Exchange**, established in **1801**, became a major global financial center, trading stocks, bonds, and derivatives.

These early financial markets laid the groundwork for modern financial systems and gradually evolved to include a wider range of financial instruments, regulations, and technologies.

Some important financial crisis

The "Tulip Mania" or "Tulip Craze": refers to a speculative bubble that occurred in the Netherlands during the early 17th century, specifically in the 1630s. It was not a crisis in the conventional sense, but rather a period of intense speculation and rapid price escalation followed by a sudden collapse in the market for tulip bulbs. During this time, tulip bulbs became extremely popular and were seen as a status symbol among the Dutch elite. Prices for rare tulip bulbs skyrocketed to extraordinary levels, with some bulbs reportedly selling for prices equivalent to houses or several years' worth of income. However, the bubble eventually burst in 1637, leading to a dramatic and rapid decline in tulip bulb prices. Many investors who had purchased bulbs at inflated prices suffered substantial financial losses, and the speculative frenzy came to an abrupt end. The Tulip Mania is often cited as one of the earliest recorded speculative bubbles and serves as a cautionary tale about the dangers of irrational exuberance and herd behavior in financial markets. While it did not result in a financial crisis on the scale of modern economic downturns, it remains a notable historical event due to its impact on perceptions of financial markets and the psychology of investing.

Subprime crisis: The subprime crisis, also known as the subprime mortgage crisis, was a financial crisis that emerged primarily in the United States in 2007 and had widespread repercussions globally. At its core, it was triggered by a significant increase in mortgage delinquencies and foreclosures, particularly among borrowers with subprime mortgages. The process that led to this financial crisis is the following.

Subprime mortgages: Lenders began issuing mortgages to borrowers with poor credit histories, often referred to as subprime borrowers. These mortgages typically had higher interest rates to compensate for the increased risk of default.

Housing bubble: Housing prices soared in the early to mid-2000s, driven by factors like low interest rates, lax lending standards, and speculation. This encouraged more people to invest in real estate, further driving up prices.

Securitization: Mortgage lenders bundled these subprime mortgages with other types of loans and sold them to investors as mortgage-backed securities (MBS) or collateralized debt obligations (CDOs). The complex nature of these financial instruments made it difficult to assess their true risk.

Credit ratings agencies: Rating agencies assigned high ratings to many of these mortgage-backed securities, leading investors to believe they were safe investments, despite the underlying riskiness of the mortgages.

Deterioration of Mortgage Quality: As housing prices started to decline and interest rates rose, many subprime borrowers found themselves unable to make their mortgage payments. This led to a surge in delinquencies and foreclosures.

Financial institution failures: The increasing default rates on mortgages caused significant losses for financial institutions holding these securities. Some prominent financial institutions, such as Lehman Brothers, experienced insolvency or required government intervention to prevent collapse.

Global Financial Impact: The crisis had widespread ramifications beyond the housing market, leading to a credit crunch, stock market declines, and a severe recession in many countries. Governments around the world implemented various measures to stabilize financial markets and stimulate economic growth.

The subprime crisis exposed weaknesses in the financial system, including inadequate regulation, risky lending practices, and the interconnectedness of global financial markets. It prompted significant reforms in financial regulation and oversight to prevent similar crises in the future.

COVID-19 Pandemic: The COVID-19 recession, also known as the Great Lockdown, was a global economic recession caused by the COVID-19 pandemic. The recession began in most countries in February 2020. After a year of global economic slowdown that saw stagnation of economic growth and consumer activity, the COVID-19 lockdowns and other precautions taken in early 2020 drove the global economy into crisis. Within seven months, every advanced economy had fallen to recession. The first major sign of recession was the 2020 stock market crash, which saw major indices drop 20 to 30% in late February and March. Recovery began in early April 2020; by April 2022, the GDP for most major economies had either returned to or exceeded pre-pandemic levels and many market indices recovered or even set new records by late 2020. The recession saw unusually high and rapid increases in unemployment in many countries. By October 2020, more than 10 million unemployment cases had been filed in the United States, swamping state-funded unemployment insurance computer systems and processes. The United Nations (UN) predicted in April 2020 that global unemployment would wipe out 6.7% of working hours globally in the second quarter of 2020—equivalent to 195 million full-time workers. In some countries, unemployment was expected to be around 10%, with more severely affected nations from the pandemic having higher unemployment rates. Developing countries were also affected by a drop in remittances and exacerbating COVID-19 pandemic-related famines. The recession and the accompanying 2020 Russia–Saudi Arabia oil price war led to a drop in oil prices; the collapse of tourism, the hospitality industry, and the energy industry; and a downturn in consumer activity in comparison to the previous decade. The 2021–2023 global energy crisis was driven by a global surge in demand as the world exited the early recession caused by the pandemic, particularly due to strong energy demand in Asia. This was then further exacerbated by the reaction to escalations of the Russo-Ukrainian War, culminating in the Russian invasion of Ukraine and the 2022 Russian debt default.

1.1.2 How financial markets work ?

Financial markets are organized into two categories: the regulated and the non-regulated markets.

Over-the-counter (OTC) markets

Definition of OTC markets: Historically, the first financial markets were the non-regulated markets, also known as unregulated or over-the-counter (OTC) markets, operate without significant government oversight or regulation. An over-the-counter (OTC) market is a decentralized market in which market participants trade stocks, commodities, currencies, or other instruments directly between two parties and without a central exchange or broker. Over-the-counter markets do not have physical locations; instead, trading is conducted electronically. This is very different from an auction market system.

In an OTC market, dealers act as market-makers by quoting prices at which they will buy and sell a security, currency, or other financial products. A trade can be executed between two participants in an OTC market without others being aware of the price at which the transaction was completed. In general, OTC markets are typically less transparent than exchanges and are also subject to fewer regulations. Because of this, liquidity in the OTC market may come at a premium.

Regulated markets

MIF Regulation and creation of regulated markets: Until the mid-2000s, European markets were characterized by the existence of domestic trading platforms, each in a quasi-monopoly situation. The **directive on Markets in Financial Instruments (known as MiF)**, adopted in 2004 and implemented on November 1, 2007, defined a new organization of equity markets in Europe. This aimed to promote competition by authorizing "alternative" modes of trading alongside traditional platforms (the "exchanges"). Furthermore, to ensure the quality of the price formation mechanism in a fragmented market, the MiF directive established new rules regarding pre- and post-trading transparency and based investor protection on the principle of "best execution," founded on seeking the trading system offering the best price to the client. The MiF directive achieved two of its objectives: reducing transaction costs in the equity market and the emergence of truly pan-European trading systems. However, its effects on market liquidity and transparency were more uncertain. Revision work on the MiF directive (MiF 2) began as early as the end of 2009, with the primary objective of addressing identified weaknesses in MiF but also due to the context marked by the 2008 financial crisis and the G20 "roadmap." Following a public consultation launched in December 2010, the Commission presented its proposals in October 2011. After intense and complex debate, the MiF 2 framework (composed of a directive and a regulation) was adopted in May 2014 and came into effect on January 3, 2018. It aims to restore

fair competition (level-playing-field) between regulated markets and alternative trading modes, strengthen transparency, and improve investor protection.

MiF 2 defines a regulated market as "a multilateral system operated and/or managed by a market operator, which ensures or facilitates the meeting - within itself and according to its non-discretionary rules - of multiple buying and selling interests expressed by third parties for financial instruments, in a manner that results in the conclusion of contracts for financial instruments admitted to trading under its rules and/or systems, and which is authorized and operates regularly in accordance with Title III [of the directive]."

A regulated market is thus characterized by non-discretionary execution of transactions: an order placed on the order book cannot be withdrawn and must be automatically matched with available orders in the system.

The MiF 2 framework consists of a directive (MiFID 2) and a regulation (MiFIR). However, beyond these two so-called "level 1" texts, MiF 2 also includes over forty "level 2" texts (delegated regulations and implementing regulations) adopted by the Commission based on technical standards developed by the European Securities and Markets Authority (ESMA), as well as a series of "level 3" documents (Guidelines and Questions/Answers) published by ESMA.

MiF 2 comprises two main components: market organization and investor protection.

Market organization

Expansion of the scope of financial instruments covered (which, under MiF, was limited to equities) to equity-like securities and so-called "non-equity" instruments: bonds, derivatives, structured products, and carbon quotas.

Creation of a new category of organized trading platform (limited to non-equity instruments trading): Organised Trading Facilities (OTFs).

Restriction of over-the-counter trading scope (requirement for trading in equities and certain derivatives), strengthening of systematic internalizer regime, prohibition of order-matching systems (crossing networks).

Open Acess Establishment of a principle of non-discriminatory access ("open access") of trading platforms to central counterparties (CCPs) and vice versa, as well as to benchmark indices.

Strengthening of pre-trade transparency requirements (with possible exemptions calibrated based on instrument liquidity and/or transaction size): publication of bid and ask prices and the size of positions expressed at these prices.

Exemptions for equities are further regulated introduction of a double cap limiting the proportion of transactions conducted in dark pools, without pre-trade transparency. Strengthening of post-trade transparency requirements (possible delays calibrated according to the same criteria as above), with the implementation of consolidated publication systems (Consolidated Tape Provider - CtP) and approved publication mechanisms (Approved Publication Authority - APA).

Intensification of transaction reporting to the regulator and implementation of approved reporting mechanisms (Approved Reporting Mechanism - ARM).

Regulation of algorithmic trading and high-frequency trading to prevent malfunction and market manipulation risks.

Regulation of commodity derivatives markets (position limits and declarations).

Investors protection

Strengthening of product governance through a more precise definition of the respective responsibilities of the manufacturer (who defines the product characteristics, target market, and distribution channels) and the distributor (who understands the product characteristics, also determines the target market, and ensures consistency between it and their own clientele).

Increased transparency for investors disclosure of costs and charges related to services and products upstream, and possibly downstream, of the transaction.

Introduction of the concept of "independent advice" , with a requirement for investment firms providing advice to specify whether it is independent advice or not.

Strengthening of regulation of remuneration and inducements their receipt is prohibited in the provision of independent advice or discretionary portfolio management; it is allowed for other services, provided that it aims to enhance the quality of service and that the client is clearly informed of their nature, amount, or method of calculation, prior to the provision of the service.

Implementation of a new regime for the financing of financial analysis.

Strengthening of transparency obligations under "best execution" information on transaction execution must be more detailed and easily understandable by the client.

In a nutshell, a regulated market is a market over which government bodies or, less commonly, industry or labor groups, exert a level of oversight and control. Market regulation is often controlled by the government and involves determining who can enter the market and the prices they may charge. The government body's primary function in a market economy is to regulate and monitor the financial and economic system. These regulations aim to ensure fairness, transparency, stability, and the protection of investors and consumers. Examples of regulated markets include stock exchanges, commodity exchanges, and financial markets where trading activities are closely monitored and regulated by authorities.

Alternative means of negotiation:

They can be either regulated or unregulated markets and they include Multilateral Trading Facility (MTFs), Organized Trading Facilities (OTFs), Systematic Internalizers (SIs), and Dark Pools.

Multilateral Trading Facility – MTF Under MiF 2, Multilateral Trading Facilities (MTFs), which already existed under MiF, are defined as "a multilateral system operated by an investment firm or a market operator, which ensures the meeting - within itself and according to non-discretionary rules - of multiple buying and selling interests expressed by third parties for financial instruments, in a manner that results in the conclusion of contracts in accordance with Title II [of the directive]."

On an MTF, as on a regulated market, transactions are executed in a non-discretionary manner. MTFs generally offer less expensive access than regulated markets, but limited to the most liquid securities and those on which the largest volumes are traded.

A regulated market operator may also manage MTFs alongside to meet certain specific needs of market participants. For example, Euronext also manages the Multilateral Trading Facility (MTF) Euronext Growth (formerly Alternext), dedicated to mid-cap securities, with listing rules tailored to SMEs and ETIs.

Organized Trading Facilities - OTF: OTFs are a new category of organized platform introduced by MiF 2, which defines them as "a multilateral system, other than a regulated market or an MTF, within which multiple buying and selling interests expressed by third parties for bonds, structured financial products, emission allowances, or derivatives can interact in a manner that results in the conclusion of contracts in accordance with Title II [of the directive]."

Unlike regulated markets and MTFs, the operator of an OTF has discretionary power over how transactions are executed: they can decide to place or withdraw an order on the OTF, or decide not to match a specific order with available orders in the system at a given time, which can allow orders to be executed in the best interests of clients. In return, the operator of an OTF cannot execute orders for its own account.

Systematic Internalizers - SIs: SIs, which already existed under MiF, are defined in MiF 2 as "an investment firm that, in an organized, frequent, and systematic manner, trades on its own account by executing client orders outside a regulated market or an MTF."

Unlike an OTF, an SI executes client orders by committing its own capital. In return, it is subject to strengthened prudential requirements.

Dark Pools The concept of a dark pool refers to markets (regulated markets or MTFs) benefiting from exemptions, as provided for in the regulations, from the pre-trade reporting obligation, thus allowing transactions to be executed without disclosing

orders or their prices before execution. A regulated market operator may also manage a dark pool. For instance, Euronext operates a dark pool called Euronext Block (formerly SmartPool).

Categories of regulated markets

Regulated markets can be splitted into several categories.

Stock Exchanges: Stock exchanges are primary regulated markets where equities (stocks) are bought and sold. Examples include the New York Stock Exchange (NYSE), NASDAQ, London Stock Exchange (LSE), and Tokyo Stock Exchange (TSE).

Bond markets: Bond markets are regulated markets where fixed-income securities such as government bonds, corporate bonds, and municipal bonds are traded. These markets provide a platform for issuing, buying, and selling debt securities.

Commodity Exchanges: Commodity exchanges are regulated markets where commodities such as agricultural products, energy, metals, and financial derivatives based on commodities are traded. Examples include the Chicago Mercantile Exchange (CME) and the Intercontinental Exchange (ICE).

Derivatives Exchanges: Derivatives exchanges are regulated markets where financial instruments derived from underlying assets such as stocks, bonds, commodities, currencies, or indices are traded. These instruments include options, futures, swaps, and forwards. Examples include the Chicago Board Options Exchange (CBOE) and Eurex.

Foreign Exchange (Forex) Markets: Forex markets are regulated markets where currencies are traded.

Multilateral Trading Facilities (MTFs): MTFs are regulated markets that provide alternative venues for trading securities outside of traditional stock exchanges. They match buyers and sellers electronically and may offer lower trading fees and greater anonymity.

Dark Pools: Dark pools are regulated markets that provide private venues for trading large blocks of securities away from public exchanges. They offer increased privacy and reduced market impact for institutional investors.

Categories of OTC markets

As the regulated markets, the OTC markets are also organized in different categories.

Equity OTC Markets: These markets involve the trading of stocks (equities) of public companies that are not listed on formal exchanges such as the New York Stock Exchange (NYSE) or NASDAQ. Instead, these stocks are traded directly between buyers and sellers through broker-dealers or electronic communication networks (ECNs).

Bond OTC Markets: Bond OTC markets involve the trading of fixed-income securities such as government bonds, corporate bonds, municipal bonds, and other debt instruments. Participants in these markets include institutional investors, banks, and individual traders who buy and sell bonds directly with one another or through brokers.

Forex (Foreign Exchange) OTC Markets: Forex OTC markets involve the trading of currencies, where participants buy and sell different currencies directly with each other or through electronic trading platforms. These markets operate 24 hours a day, five days a week, and are decentralized, with trading taking place over-the-counter rather than on a centralized exchange.

Derivatives OTC Markets: Derivatives OTC markets involve the trading of financial instruments whose value is derived from an underlying asset, index, or reference rate. These instruments include options, swaps, forwards, and other complex financial products. Participants in these markets include institutional investors, hedge funds, and banks, who trade derivatives directly with one another or through brokers.

Commodity OTC Markets: Commodity OTC markets involve the trading of physical commodities such as agricultural products, energy products, metals, and other raw materials. Participants in these markets include producers, consumers, traders, and speculators who buy and sell commodities directly with one another or through brokers.

Structured Product OTC Markets: Structured product OTC markets involve the trading of complex financial products that are customized to meet specific investment objectives. These products may include structured notes, asset-backed securities, and other hybrid securities. Participants in these markets include institutional investors, investment banks, and wealth managers.

1.1.3 Financial regulation

What is financial regulation ?

Financial regulation is defined to ensure that the financial markets properly work. See for instance (Kuritzkes et al., 2003). Financial regulation refers to the set of laws and rules that govern financial institutions, markets, and transactions to ensure stability, transparency, and fairness within the financial system. These regulations are implemented by governmental or international bodies and aim to protect investors, maintain market integrity, prevent fraud and market manipulation, and mitigate systemic risk.

Financial regulations typically cover various aspects of the financial industry, including banking, securities trading, insurance, and investment and risk management. They can range from specific requirements for capital reserves and liquidity ratios for banks to rules governing the issuance and trading of securities on stock exchanges.

Regulation aims to maintain the stability of the financial system by preventing excessive risk taking and speculative behavior that could lead to financial crises. Moreover, it often includes measures to protect consumers from fraud unfair practices and fairly, promoting trust among market participants. This is the goal for instance of the anti-money laundering law. Furthermore, it seeks to ensure market integrity i.e. to ensure that financial markets operate efficiently, transparently, and fairly, promoting trust among market participants. Beside, regulation aims to manage systemic risk by identifying and mitigating risks that could threaten the stability of the entire financial system, such as the failure of large financial institutions. Regulation may also promote competitions within the financial industry, preventing monopolistic practices and encouraging innovation.

A specific area of the financial regulation is the **banking regulation**.

Banking regulation

Banking regulation refers to the set of rules, laws, and guidelines that govern the activities of banks and other financial institutions. These regulations are designed to ensure the stability, integrity, and soundness of the banking system, protect depositors' funds, and promote the efficient functioning of financial markets. Banking regulations often prescribe minimum levels of capital that banks must hold to cover potential losses and protect depositors. **Capital requirements** help ensure that banks have enough reserves to absorb losses and maintain financial stability. Banks are typically required to maintain sufficient liquid assets to meet their short-term obligations. **Liquidity regulation** aims to prevent bank runs and ensure that banks can fulfill withdrawal requests from depositors without facing liquidity crises. Regulation requires banks to implement robust **risk management** practices to identify, measure, and mitigate various types of risks, including credit risk, market risk, and operational risk. This may involve establishing risk management committees, conducting stress tests, and implementing internal controls. Many countries have **deposit insurance** schemes that protect depositors' funds in the event of bank failure. Banking regulations often establish and govern these deposit insurance programs to provide confidence to depositors and maintain financial stability. Regulatory authorities, such as central banks or banking regulators, oversee banks' activities to ensure compliance with regulations and assess their financial health. Supervisory activities may include on-site examinations, off-site monitoring, and enforcement actions for non-compliance. Banking regulations include measures to combat money laundering, terrorist financing, and other illicit activities. This task is called Anti-Money Laundering (AML) and Counter-Terrorist Financing (CTF). Banks are required to implement customer due diligence procedures, report suspicious transactions, and comply with Know Your Customer (KYC) requirements. Banking regulations often include provisions to **protect consumers** from unfair or deceptive practices. These

may include disclosure requirements for financial products and services, restrictions on predatory lending practices and mechanisms for handling consumer complaints.

Regulation agreements

To insure market regulation, a series of accords have been established over time. Among these accords, we can find the **Markets in Financial Instruments (MiF)**, the International Financial Reporting Standards (IFRS), the Basel Agreements, as well as the Fundamental Review of Trading Book (FRTB).

International Financial Reporting Standards (IFRS): The **IFRS** accords have been implemented in 2001. These are a set of accounting standards developed by the International Accounting Standards Board (IASB), an independent, private-sector organization based in London, United Kingdom. IFRS aims to provide a common global language for financial reporting, ensuring that financial statements are transparent, comparable, and understandable across different countries and industries. IFRS is used by companies around the world to prepare their financial statements, including the balance sheet, income statement, cash flow statement, and statement of changes in equity. These standards govern how companies should recognize, measure, present, and disclose various elements of their financial performance and position. Key features of IFRS include principles-based accounting, which focuses on the substance of transactions rather than their legal form, and the use of fair value measurement for certain financial instruments and assets. IFRS also emphasizes the importance of providing relevant and reliable information to investors, creditors, and other stakeholders to support informed decision-making. IFRS is increasingly becoming the global standard for financial reporting, with many countries adopting or converging their national accounting standards with IFRS. While some countries, such as the United States, still use their own set of accounting standards (Generally Accepted Accounting Principles or GAAP), efforts are ongoing to achieve greater harmonization and convergence between IFRS and other national standards.

Markets in Financial Instruments (MiF) As previously mentioned, the MiF directives have been established in the early 2000s to create the first regulated markets. These are a set of European Union regulations governing the provision of investment services and activities within the European Economic Area (EEA). These directives aim to harmonize regulation across EU member states, enhance investor protection, promote competition, and increase market transparency. There have been two main iterations of the MiF directives.

MiFID I: (Markets in Financial Instruments Directive I) Adopted in 2004 and implemented in 2007, MiFID I introduced significant changes to the regulation of investment services and trading venues within the EU. It aimed to improve investor protection, increase market efficiency, and foster competition by allowing investment firms to operate throughout the EU under a single set of rules.

MiFID II and MiFIR: MiFID II (Markets in Financial Instruments Directive II) and MiFIR (Markets in Financial Instruments Regulation) are a comprehensive revision and extension of MiFID I. MiFID II, adopted in 2014 and implemented in 2018, introduced more stringent requirements for transparency, investor protection, and market structure. MiFIR, which is directly applicable in all EU member states, complements MiFID II by specifying regulatory requirements and procedures.

Together, MiFID II and MiFIR aim to address shortcomings identified in MiFID I, enhance the functioning of financial markets, and adapt regulation to new market developments, technological advancements, and changes in market structure. They cover a wide range of topics, including trading venues, investment firms, market transparency, investor protection, and conduct of business rules.

Basel Agreements: The Basel Accords are a set of international banking regulations that aim to strengthen the stability and soundness of the global banking system. They are developed by the **Basel Committee on Banking Supervision (BCBS)**, which is a forum for central banks and banking regulators from around the world to collaborate on banking supervision and regulation. The Basel Accords provide a framework of rules and guidelines for regulating banks capital adequacy, liquidity risk, and leverage.

Basel I: introduced in 1988, established the first international framework for regulating banks' capital adequacy. It introduced the concept of minimum capital requirements based on the riskiness of banks' assets. Under Basel I, the capital adequacy risk (the risk that an unexpected loss would hurt a financial institution), categorizes the assets of financial institutions into five risk categories 0%, 10%, 20%, 50% and 100%.

Under Basel I, banks were required to hold capital equivalent to at least 8% of their risk-weighted assets. This ensures banks hold a certain amount of capital to meet obligations. Basel I divided the eligible regulatory capital of a bank into two Tiers. Tier 1 capital is the most liquid and primary funding source of the bank. Tier 2 capital includes less liquid hybrid capital instruments, loan-loss, and revaluation reserves as well as undisclosed reserves.

The Cooke ratio is a bank solvency ratio recommended by the Basel Committee as part of its initial recommendations to guarantee a minimum level of equity capital to ensure the financial soundness of banks. It is an indicator specific to segment companies, unlike companies in the agricultural sector.

The Cooke ratio sets the limit for the weighted outstanding loans granted by a financial institution in relation to the bank's equity capital. It is a minimum ratio of 8% between a bank's own funds and total assets, weighted by the risk of each asset.

$$\text{Cooke Ratio} = \frac{\text{Regulatory capital}}{\text{Credit risk}} \geq 8\%. \tag{1.1}$$

An amendment in 1998 allows market risk to be taken into account using two approaches (standard and internal model).

The ratio is named after Peter Cooke, a director of the Bank of England who was one of the first to propose the creation of the Basel Committee and was its first chairman. The ratio is calculated on the basis of capital divided into three main components (Tier 1, Tier 2 and Tier 3) and outstanding loans, on-balance sheet commitments and off-balance sheet commitments weighted according to their nature. The ratio of capital to risk-weighted assets must be at least 8%, with a minimum of 4% for Tier 1.

Basel II: also called Revised Capital Framework, introduced in 2004 and revised in 2006 and 2009, aimed to address some of the limitations of Basel I and improve risk management practices in banking focusing on three main areas: minimulm capital requirements, supervisory review of an institution's capital adequacy and internal assessment process, and the effective use of disclosure as a lever to strengthen market discipline and encourage sound banking practices including supervisory review. Together, these areas of focus are known as the three pillars. Basel II introduced a more risk-sensitive approach to capital regulation, allowing banks to use internal models to calculate capital requirements based on their own assessment of credit, market, and operational risks. It also introduced new requirements for capital buffers, supervisory review, and market discipline. Basel II divided the eligible regulatory capital of a bank from two into three tiers. The higher tier, the less subordinated securities a bank is allowed to include in it. Each tier must be of a certain minimum percentage of the total regulatory capital and is used as a numerator in the calculation of regulatory capital ratios. The new tier 3 capital is defined as tertiary capital, which many banks hold to support their market risk, commodities risk, and foreign currency risk, derived from trading activities. Tier 3 capital includes a greater variety of debt than tier 1 and 2 capital but is of a much lower quality than either of the two. Under the Basel III accords, tier 3 capital was subsequently rescinded.

The McDonough ratio replaces the Cooke ratio, which set a minimum loan-to-equity ratio of 8%. The McDonough ratio takes its name from the current Chairman of the Basel Committee during the process of establishing the Accord, William J. McDonough. The McDonough ratio is more refined than the Cooke ratio because it takes into account the greater or lesser risk of the various loans granted, by dividing them into three types: credit risk, market risk and operational risk. The banks' level of commitment is thus limited by their own financial strength. The McDonough ratio is defined as follows:

$$\text{Bank's own funds} > 8\% \left(\text{Credit Risk}(85\%) + \text{Market Risk}(5\%) + \text{Operational Risk}(10\%)\right). \tag{1.2}$$

Basel III: In the wake of the Lehman Brothers collapse of 2008 and the ensuing financial crisis, the BCBS decided to update and strengthen the Accords. The BCBS considered poor governance and risk management, inappropriate incentive structures, and an over leveraged banking industry as reasons for the collapse. In

November 2010, an agreement was reached regarding the overall design of the capital and liquidity reform package. This agreement is now known as Basel III.

Basel III is a continuation of the three pillars along with additional requirements and safeguards. For example, Basel III requires banks to have a minimum amount of common equity and a minimum liquidity ratio. Basel III also includes additional requirements for what the Accord calls "systemically important banks", or those financial institutions that are considered "too big to fail". In doing so, it got rid of tier 3 capital considerations.

The Basel III reforms have now been integrated into the consolidated Basel Framework, which comprises all of the current and forthcoming standards of the Basel Committee on Banking Supervision. Basel III tier 1 has now been implemented and all but one of the 27 Committee member countries participated in the Basel III monitoring exercise held in June 2021. The final Basel III framework includes phase-in provisions for the output floor, which will start at 50% on January 1, 2023, rising in annual steps of 5% and be fully phased-in at the 72.5% level from January 2028. These 2023 onward measures have been referred to as Basel 3.1 or Basel IV.

Basel IV: is the informal name for a set of proposed banking reforms building on the international banking accords known as Basel I, Basel II, and Basel III. Basel IV builds upon the previous Basel III framework and aims to address perceived shortcomings, strengthen the resilience of banks, and improve the consistency and comparability of regulatory capital requirements across jurisdictions. Key components of Basel IV include revisions to the calculation methodologies for credit risk, operational risk, and market risk, as well as the introduction of the standardized approach for measuring counterpart credit risk (SA-CCR).

FRTB: A specific component of Basel IV, called **Fundamental Review of the Trading Book (FRTB)**, focuses on the regulation of market risk in banks' trading books. It was developed in response to weaknesses identified in the previous market risk framework under Basel II and Basel III, particularly in the wake of the 2008 financial crisis. FRTB introduces more stringent requirements for the measurement, management, and capitalization of market risk, with the aim of ensuring that banks hold adequate capital reserves to cover potential trading book losses. Key features of FRTB include the introduction of a revised standardized approach (SA) and the internal models approach (IMA) for calculating market risk capital charges, as well as enhanced risk governance and reporting requirements.

1.1.4 Financial risk and risk factors

What is financial risk ?

Financial risk is the possibility of losing money on an investment or business venture. Some more common and distinct financial risks include credit risk, liquidity risk, and operational risk. Financial risk is a type of danger that can result in the loss of capital to

interested parties. For governments, this can mean they are unable to control monetary policy and default on bonds or other debt issues. Corporations also face the possibility of default on debt they undertake but may also experience failure in an undertaking the causes a financial burden on the business.

Financial risks can be classified into diverse categories.

Different categories of risks

Financial risks for the individuals: Financial risks for individuals occur when they make sub-optimal decisions. There are several types of Individual risk factors; pure risk, liquidity risk, speculative risk, and currency risk.

Pure Risk

Pure Risk is a type of risk where the outcome cannot be controlled, and only has two outcomes which are complete loss or no loss at all. An example of pure risk for an individual would be owning an equipment, there is risk of it being stolen and there would be a loss to the individual, however, if it weren't stolen, there is no gain but only no loss for the individual.

Liquidity risk

Liquidity Risk is when securities cannot be purchased or sold fast enough to cut losses in a volatile market. An example to which an individual might experience liquidity risk would be no one willing to purchase a security you own, and the value of your security significantly drops.

Speculative risk

Speculative risks are made based on conscious choices, and results in an uncertain degree of gain or loss. An example of speculative risk is purchasing stocks, the future of the stock's price is uncertain, and both a gain or loss could occur depending on whether if the stock price rises or decreases.

Currency risk

Currency risk is when exchange rates changes will affect the profitability of when one is committed to it and the time when it is carried out. An example of currency risk would be if interest rates were higher in U.S compared to Australia, the Australian dollar would drop in comparison to the U.S.. This is due to the increase in demand for USD as investors take advantage of higher yields, thus exchange rate fluctuates and the individual is exposed to risks in the foreign exchange markets.

Financial risks for the market: Financial Risks for the market are associated with price fluctuation and volatility. Risk factors consist of interest rates, foreign currency exchange rates, commodity and stock prices, and through their non-stop fluctuations, it produces a change in the price of the financial instrument.

Market Risk (systematic risk)

Market Risk is the risk an investor experiences when the value of an investment decreases due to financial market factors. The failure of a single company or cluster of companies could lead to the entire market crashing and the way to reduce this risk is through diversification into assets that are not co-related to the market. An example is during the 2007-2008 global financial crisis, when a core sector of the market suffered, the volatile risk created effected the monetary well-being of the entire marketplace. During this time, businesses closed, there was an estimated loss of $6 trillion to $14 trillion, and governments were forced to rethink their economic policies. A similar situation is observed during the COVID-19 global pandemic crisis, where a massive economic fall-out had occurred due to the lack of economic activity. The global economy came to a halt, aggregate demand rapidly decreased, and even oil prices plummeted to almost negative $40, which meant producers paid buyers to take oil off their hands as storing oil was costly.

Financial risks for business: Financial Risk for businesses rises due to the need for funding in order to expand and grow the business, or when they sell products on credit. There are several types of financial risks in businesses, including credit risks, specific risks, and operational risks.

Credit risk

Credit risk are the dangers of default occurring when a creditor lends money to a borrower. Examples of credit risks include businesses not being able to retrieve their money when they sell products on credit and may experience a rise in costs to collect the debt. Businesses can also experience credit risk as the borrowers, as they must manage cash flows in order to pay back their accounts payabl.

Specific risk

Specific risks a.k.a. unsystematic risks are hazards that are unique and apply only to a certain asset or company. An example of an unsystematic risk is if a company has poor reputation or there are strikes among company employees, only that specific company is affected. Unsystematic risk can be avoided through diversification where investors invest in a wide variety of stocks.

Operational risk

Companies face operational risks whenever it attempts to do ordinary business activities and can also be classified as a variety of specific risk. Operational risks stem from man-made choices, thus are the risks of business operations failing due to human error. Examples of Operational risks would be keeping a subpar sales staff team as it has lower wage costs, but it comes with higher operational risks as the staff are more likely to make mistakes.

Financial risks in investing: Investing is allocating money, effort, or time into something in hopes of generating income or profit. A common investment is investing in stocks, purchasing them at a low price then reselling it later at a higher price to earn the difference as profit. Stock investing comes with very high risks as every single piece of information would cause market prices to fluctuate. Financial risks in investing include economic risk, commodity price risk, inflationary risk and interest rate risk, headline risk, obsolescence risk, and model risk.

Economic risk

One of the most obvious risk is economic risk, where the economy could go bad at any given moment, causing stock prices to plummet.

Commodity price risk

Commodity price risk is the possibility of a commodity price fluctuating, potentially causing financial losses for the buyers or producers of a commodity. As Commodity prices are basic raw materials, it creates a domino effect, affecting all products that require the commodity. For example, oil consumers often face commodity price risk, as oil is a widely used necessity product currently, many producers' profits are affected by the fluctuation of oil price.

Inflationary risk and interest rate risk

Other risks like inflationary risk and interest rate risks usually go hand in hand, as interest rates are increased in order to combat inflation, which in turn causes businesses operation cost to increase, making it harder to stay in business, which then leads to a reduction in their stock prices. Inflation on its own also destroys value of stocks and creates recessions in the market.

Headline risk

A very transparent risk is headline risk, where any stories in the media that will damage a company's reputation would hurt their business and reduce their stock prices. An example is the Fukushima nuclear crisis in 2011, which punished their stocks and caused excessive backlash against any businesses related to the story.

Obsolescence risk

A risk that arises due to technological advancement is obsolescence risk, where a process, product or technology used by a company to generate profit becomes obsolete as competitors find cheaper alternatives. An example of this are publishing companies, as computers, phones, and devices becomes more advanced, more and more people read news, magazines and books online instead of the printed form as it's cheaper and more convenient, which caused publishing companies to slowly become obsolete.

Model risk

When people rely too much on the assumptions underlying economic and business models is model risk. When the models are inaccurate, all stakeholders that relied on the financial model are exposed to risks as the quantitative information utilized are made based on insufficient information. An example of this is the Long Term Capital Management (LTCM) debacle, which caused them great financial loss because of a small error in their computer models, which was magnified by their highly leveraged trading strategy.

1.1.5 Classic risk management approaches

What is hedging ?

To hedge, in finance, is to take an offsetting position in an asset or investment that reduces the price risk of an existing position. A hedge is therefore a trade that is made with the purpose of reducing the risk of adverse price movements in another asset. Normally, a hedge consists of taking the opposite position in a related security or in a derivative security based on the asset to be hedged.

A hedge is a strategy that seeks to limit risk exposures in financial assets. Popular hedging techniques involve taking offsetting positions in derivatives that correspond to an existing position. Other types of hedges can be constructed via other means like diversification. An example could be investing in both cyclical and countercyclical stocks. Besides protecting an investor from various types of risk, it is believed that hedging makes the market run more efficiently.

Derivatives can be effective hedges against their underlying assets because the relationship between the two is more or less clearly defined. Derivatives are securities that move in correspondence to one or more underlying assets. They include options, swaps, futures, and forward contracts. The underlying assets can be stocks, bonds, commodities, currencies, indexes, or interest rates. It's possible to use derivatives to set up a trading strategy in which a loss for one investment is mitigated or offset by a gain in a comparable derivative.

Back-to-back

The aim of back to back is to net a risky position by contracting a transaction symmetrical to the transaction you wish to manage. An example of back-to-back is provided as follows.

Consider a management fund offering an investment with a guaranteed floor, the nominal amount, and an optional remuneration corresponding to x% of the nominal amount each year where the CAC will have had a positive variation. This contract can be broken down into a zero-coupon component representing the nominal amount and an optional component (a series of digitals based on changes in the CAC) representing the optional remuneration. The optional component can be managed on a back-to-back basis by selling it to a bank trading floor.

Asset / Liability Management (ALM)

What is ALM ? Asset/liability management is the process of managing the use of assets and cash flows to reduce the firm's risk of loss from not paying a liability on time. Well-managed assets and liabilities increase business profits. The asset/liability management process is typically applied to bank loan portfolios and pension plans. It also involves the economic value of equity.

The concept of asset/liability management focuses on the timing of cash flows because company managers must plan for the payment of liabilities. The process must ensure that assets are available to pay debts as they come due and that assets or earnings can be converted into cash. The asset/liability management process applies to different categories of assets on the balance sheet.

How ALM works ? ALM relies on four main components.

Asset management, that involves managing the composition, maturity, and quality of the financial institution's assets, which include loans, investments, securities, and other income-generating assets. The goal is to optimize the return on assets while managing risks such as credit risk, interest rate risk, liquidity risk, and market risk.

Liability management, that involves managing the composition, maturity, and cost of the financial institution's liabilities, which include deposits, borrowings, and other funding sources. The goal is to ensure that the institution has adequate funding to support its asset growth, manage liquidity needs, and meet regulatory requirements.

Risk management: ALM involves identifying, measuring, and managing various risks that arise from the asset-liability structure, including interest rate risk, liquidity risk, credit risk, and market risk. This may involve using hedging strategies, diversification, stress testing, and scenario analysis to mitigate risks and ensure the institution's financial stability.

Strategic planning: ALM plays a crucial role in strategic planning for financial institutions, helping them set goals, make informed decisions, and allocate resources effectively. It provides insights into the institution's risk profile, profitability, capital adequacy, and funding needs, guiding strategic initiatives and business decisions.

Overall, ALM is a dynamic and ongoing process that involves continuous monitoring, analysis, and adjustment of the asset and liability mix to achieve the financial institution's objectives, manage risks, and adapt to changing market conditions and regulatory requirements. It is essential for ensuring the long-term viability and sustainability of financial institutions in a competitive and evolving financial landscape.

Static hedging

What is static hedging ? Static hedging is a risk management technique used to hedge or offset exposure to a particular risk using a predetermined set of financial instruments or strategies. Unlike dynamic hedging, which involves continuously adjusting hedge positions in response to changes in market conditions, static hedging involves establishing a fixed hedge position and maintaining it over time without further adjustments.

Static hedging process The process of static hedging includes several steps that are described as follows.

Identification of Risk: The first step in static hedging is identifying the specific risk or exposure that needs to be hedged. This could be a financial risk such as interest rate risk, currency risk, commodity price risk, or equity risk.

Selection of hedge instruments: Once the risk is identified, the next step is to select appropriate financial instruments or strategies to hedge the risk. These instruments could include options, futures, forwards, swaps, or other derivative products that are closely correlated with the underlying risk exposure.

Establishing hedge positions: Static hedging involves establishing hedge positions using the selected instruments or strategies to offset the exposure to the identified risk. The hedge positions are typically established based on predetermined criteria, such as the size of the exposure, the desired level of protection, and the expected market conditions.

Maintaining hedge positions: Once the hedge positions are established, they are maintained without further adjustments over time. Unlike dynamic hedging, there is no active management of the hedge positions in response to changes in market conditions. The hedge positions are held until the expiration or maturity of the hedge instruments.

Monitoring and rebalancing: Although static hedging does not involve active adjustments to hedge positions, it still requires monitoring and periodic rebalancing to ensure that the hedge remains effective and aligned with the underlying risk

exposure. Rebalancing may be necessary if there are significant changes in market conditions or the underlying risk profile.

Static hedging is often used when the risk exposure is relatively stable, and there is a high degree of certainty about the future direction of the markets. It provides a straightforward and cost-effective way to hedge against specific risks while minimizing the need for ongoing management and monitoring. However, static hedging may not be suitable for all situations, particularly when market conditions are volatile or uncertain.

Dynamic hedging via Greeks

What are the Greeks ? An option's price can be influenced by a number of factors that can either help or hurt traders depending on the positions they take. Successful traders understand the factors that influence options pricing, which include the so-called *Greeks*. They are a set of risk measures named after the Greek letters that denote them, which indicate how sensitive an option is to time-value decay, changes in implied volatility, and movements in the price of its underlying security. In other words, the Greeks, are variables used to assess risk in the options market. Each Greek variable is a result of an imperfect assumption or relationship of the option with another underlying variable. Traders use different Greek values, such as delta, theta, and others, to assess options risk and manage option portfolios.

Greeks encompass many variables. These include delta, theta, gamma, vega, and rho, among others. Each one of these Greeks has a number associated with it, and that number tells traders something about how the option moves or the risk associated with that option. The primary Greeks (delta, vega, theta, gamma, and rho) are calculated each as a first partial derivative of the options pricing model (for instance, the Black-Scholes model).

The number or value associated with a Greek changes over time. Therefore, sophisticated options traders may calculate these values daily to assess any changes that may affect their positions or outlook, or simply to check if their portfolio needs to be rebalanced. Below are several of the main Greeks traders look at.

Delta: The delta represents the change in the value of an option in relation to the movement in the market price of the underlying asset. More precisely, delta (Δ) represents the rate of change between the option's price and a \$1 change in the underlying asset's price.

In other words, the price sensitivity of the option is relative to the underlying asset. The delta of a call option has a range between 0 and 1, while the delta of a put option has a range between 0 and -1. For example, assume an investor is long a call option with a delta of 0.50. Therefore, if the underlying stock increases by \$1, the option's price would theoretically increase by 50 cents.

What is delta-heding ?

Delta hedging is an options trading strategy that aims to reduce, or hedge, the directional risk associated with price movements in the underlying asset. The approach uses options to offset the risk to either a single other option holding or an entire portfolio of holdings. The investor tries to reach a delta-neutral state and not have a directional bias on the hedge.

The most basic type of delta hedging involves an investor who buys or sells options and then offsets the delta risk by buying or selling an equivalent amount of stock or exchange-traded fund (ETF) shares. Investors may want to offset their risk of moving in the option or the underlying stock by using delta hedging strategies.

More advanced options strategies seek to trade volatility through the use of delta-neutral trading strategies. Since delta hedging attempts to neutralize or reduce the extent of the move in an option's price relative to the asset's price, it requires a constant rebalancing of the hedge. Delta hedging is a complex strategy mainly used by institutional traders and investment banks.

Let's assume the options discussed have equities as their underlying security. Traders want to know an option's delta since it can tell them how much the value of the option or the premium will rise or fall with a move in the stock's price. The theoretical change in premium for each basis point or $1 change in the price of the underlying is the delta, while the relationship between the two movements is the hedge ratio.

The delta of a call option ranges between zero and one, while the delta of a put option ranges between negative one and zero. The price of a put option with a delta of -0.50 is expected to rise by 50 cents if the underlying asset falls by $1. The opposite is true, as well. For example, the price of a call option with a hedge ratio of 0.40 will rise 40% of the stock-price move if the price of the underlying stock increases by $1.

Delta is dependent on if it is In-the-money or currently profitable, At-the-money at the same price as the strike, or Out-of-the-money not currently profitable. A put option with a delta of -0.50 is considered at-the-money meaning the strike price of the option is equal to the underlying stock's price. Conversely, a call option with a 0.50 delta has a strike that's equal to the stock's price.

A less common usage of an option's delta is the current probability that the option will expire in-the-money. For instance, a 0.40 delta call option today has an implied 40% probability of finishing in-the-money.

How reaching delta-neutral ?

An option position could be hedged with options exhibiting a delta that is opposite to that of the current options holding to maintain a **delta-neutral** position. A delta-neutral position is one in which the overall delta is zero, which minimizes the options' price movements in relation to the underlying asset.

For example, assume an investor holds one call option with a delta of 0.50, which indicates the option is at-the-money and wishes to maintain a delta neutral position. The investor could purchase an at-the-money put option with a delta of -0.50 to offset the positive delta, which would make the position have a delta of zero.

What is the hedge-ratio ?

The hedge ratio compares the value of a position protected through the use of a hedge with the size of the entire position itself. A hedge ratio may also be a comparison of the value of futures contracts purchased or sold to the value of the cash commodity being hedged.

Futures contracts are essentially investment vehicles that let the investor lock in a price for a physical asset at some point in the future.

Imagine you are holding $10,000 in foreign equity, which exposes you to currency risk. You could enter into a hedge to protect against losses in this position, which can be constructed through a variety of positions to take an offsetting position to the foreign equity investment.

If you hedge $5,000 worth of the equity with a currency position, your hedge ratio is 0.5 ($5,000 / $10,000). This means that 50% of your foreign equity investment is sheltered from currency risk.

Theta Theta (Θ) represents the rate of change between the option price and time, or time sensitivity—sometimes known as an option's time decay. Theta indicates the amount an option's price would decrease as the time to expiration decreases, all else equal. For example, assume an investor is long an option with a theta of -0.50. The option's price would decrease by 50 cents every day that passes, all else being equal.

Theta increases when options are at-the-money, and decreases when options are in- and out-of-the money. Options closer to expiration also have accelerating time decay. Long calls and long puts will usually have negative theta; short calls and short puts will have positive theta. By comparison, an instrument whose value is not eroded by time, such as a stock, would have zero theta.

Gamma Gamma (Γ) represents the rate of change between an option's delta and the underlying asset's price. This is called second-order (second-derivative) price sensitivity. Gamma indicates the amount the delta would change given a $1 move in the underlying security. For example, assume an investor is long on a call option on hypothetical stock XYZ. The call option has a delta of 0.50 and a gamma of 0.10. Therefore, if stock XYZ increases or decreases by $1, the call option's delta would increase or decrease by 0.10.

Gamma is used to determine how stable an option's delta is: Higher gamma values indicate that delta could change dramatically in response to even small movements in the underlying's price. Gamma is higher for options that are at-the-money and lower for options that are in- and out-of-the-money and accelerates in magnitude as

expiration approaches. Gamma values are generally smaller the further away from the date of expiration; options with longer expirations are less sensitive to delta changes. As expiration approaches, gamma values are typically larger, as price changes have more impact on gamma.

Options traders may opt to not only hedge delta but also gamma in order to be delta-gamma neutral, meaning that as the underlying price moves, the delta will remain close to zero.

Vega Vega (ν) represents the rate of change between an option's value and the underlying asset's implied volatility. This is the option's sensitivity to volatility. Vega indicates the amount an option's price changes given a 1% change in implied volatility. For example, an option with a vega of 0.10 indicates the option's value is expected to change by 10 cents if the implied volatility changes by 1%.

Rho Rho (ρ) represents the rate of change between an option's value and a 1% change in the interest rate. This measures sensitivity to the interest rate. For example, assume a call option has a rho of 0.05 and a price of $1.25. If interest rates rise by 1%, the value of the call option would increase to $1.30, all else being equal. The opposite is true for put options. Rho is greatest for at-the-money options with long times until expiration.

Minor greeks Some other Greeks, which aren't discussed as often, are **lambda**, **epsilon**, **vomma**, **vera**, **zomma**, and **ultima**. These Greeks are second- or third-derivatives of the pricing model and affect things such as the change in delta with a change in volatility and so on. They are increasingly used in options trading strategies, as computer software can quickly compute and account for these complex and sometimes esoteric risk factors.

Implied volatility (IV) Implied volatility is not a Greek, but it is related to them. This value forecasts how volatile the stock underlying an option will be in the future. Implied volatility is theoretical, meaning it shows what is expected but is not always dependable. This value is usually reflected in the price of an option.

The term implied volatility refers to a metric that captures the market's view of the likelihood of future changes in a given security's price. Investors can use implied volatility to project future moves and supply and demand, and often employ it to price options contracts. Implied volatility isn't the same as historical volatility (also known as realized volatility or statistical volatility), which measures past market changes and their actual results.

Implied volatility is the market's forecast of a likely movement in a security's price. It is a metric used by investors to estimate future fluctuations (volatility) of a security's price based on certain predictive factors. Implied volatility is denoted by the symbol σ

(sigma). It can often be thought to be a proxy of market risk. It is commonly expressed using percentages and standard deviations over a specified time horizon. When applied to the stock market, implied volatility generally increases in bearish markets, when investors believe equity prices will decline over time. IV decreases when the market is bullish. This is when investors believe prices will rise over time. Bearish markets are considered undesirable and riskier to most equity investors. IV doesn't predict the direction in which the price change will proceed. For example, high volatility means a large price swing, but the price could swing upward (very high), downward (very low), or fluctuate between the two directions. Low volatility means that the price likely won't make broad, unpredictable changes.

Implied volatility is one of the deciding factors in the pricing of options. Buying options contracts allow the holder to buy or sell an asset at a specific price during a predetermined period. Implied volatility approximates the future value of the option, and the option's current value is also taken into consideration. Options with high implied volatility have higher premiums and vice versa.

Stress testing

What is stress testing ? Stress testing is a computer simulation technique used to test the resilience of institutions and investment portfolios against possible future financial situations. Such testing is customarily used by the financial industry to help gauge investment risk and the adequacy of assets and help evaluate internal processes and controls. In recent years, regulators have also required financial institutions to carry out stress tests to ensure their capital holdings and other assets are adequate.

Companies that manage assets and investments commonly use stress testing to determine portfolio risk, then set in place any hedging strategies necessary to mitigate against possible losses. Specifically, their portfolio managers use internal proprietary stress-testing programs to evaluate how well the assets they manage might weather certain market occurrences and external events.

Asset and liability matching stress tests are widely used, too, by companies that want to ensure they have the proper internal controls and procedures in place. Retirement and insurance portfolios are also frequently stress-tested to ensure that cash flow, payout levels, and other measures are well aligned.

Regulatory stress-testing: Following the 2008 financial crisis, regulatory reporting for the financial industry—specifically for banks—was significantly expanded, focusing on stress testing and capital adequacy, mainly due to the 2010 Dodd-Frank Act.

> Remark: The Dodd-Frank Wall Street Reform and Consumer Protection Act is legislation that was passed by the U.S. Congress in response to financial industry behavior that led to the financial crisis of 2007–2008. It sought

to make the U.S. financial system safer for consumers and taxpayers.

Named for sponsors Sen. Christopher J. Dodd (D-Conn.) and Rep. Barney Frank (D-Mass.), the act contains numerous provisions, spelled out over 848 pages, that were to be implemented over a period of several years

Following the 2008 financial crisis, regulatory reporting for the financial industry—specifically for banks—was significantly expanded, focusing on stress testing and capital adequacy, mainly due to the 2010 Dodd-Frank Act.

Beginning in 2011, new regulations in the United States required the submission of Comprehensive Capital Analysis and Review (CCAR) documentation by the banking industry. These regulations require banks to report on their internal procedures for managing capital and carry out various stress-test scenarios.

REMARK: "Too big to fail" describes a business or business sector so ingrained in a financial system or economy that its failure would be disastrous. The government will consider bailing out a corporate entity or a market sector, such as Wall Street banks or U.S. carmakers, to prevent economic disaster.

In addition to CCAR reporting, banks in the United States deemed too big to fail by the Financial Stability Board—typically those with more than $50 billion in assets—must provide stress-test reporting on planning for a bankruptcy scenario. In the government's most recent reporting review of these banks in 2018, 22 international banks and eight based in the United States were designated as too-big-to-fail.

Currently, BASEL III is also in effect for global banks. Much like the U.S. requirements, this international regulation requires documentation of banks' capital levels and the administration of stress tests for various crisis scenarios.

REMARK: Stress testing involves running computer simulations to identify hidden vulnerabilities in institutions and investment portfolios to evaluate how well they might weather adverse events and market conditions.

Different types of stress-testing: Stress testing involves running simulations to identify hidden vulnerabilities. The literature about business strategy and corporate governance identifies several approaches to these exercises. Among the most popular are stylized scenarios, hypotheticals, and historical scenarios.

Historical stress-testing

In a historical scenario, the business—or asset class, portfolio, or individual investment—is run through a simulation based on a previous crisis. Examples of historical crises include the stock market crash of October 1987, the Asian crisis of 1997, the tech bubble that burst in 1999-2000, the subprime crisis in 2008, or even the COVID-19 pandemic in 2020.

Hypotetical stress-testing

A hypothetical stress test is generally more specific, often focusing on how a particular company might weather a particular crisis. For example, a firm in California might stress-test against a hypothetical earthquake or an oil company might do so against the outbreak of war in the Middle East.

Stylized scenarios are a little more scientific in the sense that only one or a few test variables are adjusted at once. For example, the stress test might involve the Dow Jones index losing 10% of its value in a week.

Simulated stress-testing

As for the methodology for stress tests, Monte Carlo simulation is one of the most widely known. This type of stress testing can be used for modeling probabilities of various outcomes given specific variables. Factors considered in the Monte Carlo simulation, for example, often include various economic variables.

Companies can also turn to professionally managed risk management and software providers for various types of stress tests. Moody's Analytics is one example of an outsourced stress-testing program that can be used to evaluate risk in asset portfolios.

Remarks:

(i) A **Monte Carlo simulation** is used to model the probability of different outcomes in a process that cannot easily be predicted due to the intervention of random variables. It is a technique used to understand the impact of risk and uncertainty.

A Monte Carlo simulation is used to tackle a range of problems in many fields, including investing, business, physics, and engineering. It is also referred to as a multiple probability simulation.

(ii) **Moody's Analytics** is a subsidiary of Moody's Corporation that offers tools, solutions, and best practices for measuring and managing risk. It provides data analysis and financial intelligence products to help clients navigate and respond to an evolving marketplace.

Advantages and Drawbacks of Stress Testing:

Advantages: Stress tests are forward-looking analytical tools that help financial institutions and banks better understand their financial position and risks. They help managers identify what measures to take if certain events arise and what they should do to mitigate risks. As a result, they are better able to form action plans to thwart threats and prevent failure. For investment managers, they are better able to assess how well managed assets might perform during economic downturns.

Drawbacks: To perform stress tests, financial institutions need to create the framework and processes for which the tests can be performed. This restructuring is complex and is often associated with costly mistakes. For example, it's possible that the test scenario does not represent the types of risks a bank may face. This may be due to insufficient data or the test designer's inability to create a relevant test. In the end, the results of the test may lead to the creation of plans for events not likely to occur. This misrepresentation can cause institutions to ignore the risks that are possible.

Lastly, banks with unfavorable results may be barred from paying dividends to their customers and shareholders, as well as may be penalized.

Conterparty risk management

What is counterparty risk ? Counterparty risk is the likelihood or probability that one of those involved in a transaction might default on its contractual obligation. Counterparty risk can exist in credit, investment, and trading transactions.

The numerical value of a borrower's credit score reflects the level of counterparty risk to the lender or creditor. Investors must consider the company that's issuing the bond, stock, or insurance policy to assess whether there's default or counterparty risk.

Varying degrees of counterparty risk exist in all financial transactions. Counterparty risk is also known as default risk. Default risk is the chance that companies or individuals will be unable to make the required payments on their debt obligations. Lenders and investors are exposed to default risk in virtually all forms of credit extensions. Counterparty risk is a risk that both parties should consider when evaluating a contract.

Management of counterparty risk

Due diligence and credit analysis: Conduct thorough due diligence and credit analysis on counterparties before entering into transactions or agreements with them. Evaluate their financial strength, creditworthiness, reputation, and track record to assess the likelihood of default.

In retail and commercial financial transactions, credit reports are often used by creditors to determine the counterparty's credit risk. Credit scores of borrowers are analyzed and monitored to gauge the level of risk to the creditor. A credit score is a numerical value of an individual's or a company's creditworthiness, which is based on many variables.

A person's credit score ranges from 300 to 850, and the higher the score, the more financially trustworthy a person is considered to be to the creditor.

> Remark: Numerical values of credit scores are listed as follows: Excellent: 750 and above, Good: 700 to 749, Fair: 650 to 699, Poor: 550 to 649, Bad: 550 and below.

Risk premium: If one party has a higher risk of default, a premium is usually attached to the transaction to compensate the other party. The premium added due to counterparty risk is called a risk premium.

Diversification: Diversify exposure to counterparties by spreading investments or transactions across multiple counterparties. By diversifying counterparty risk, you reduce the impact of a single counterparty defaulting on your portfolio.

Collateral and security arrangements: Require counterparties to provide collateral or security for transactions to mitigate counterparty risk. Collateral can act as a buffer against potential losses in the event of default and provide a source of recovery.

Netting and offset arrangements: Use netting and offset arrangements to consolidate and offset obligations with the same counterparty. Netting allows you to reduce the overall exposure by offsetting receivables against payables or offsetting gains against losses.

Contractual protections: Include contractual provisions and protections in agreements with counterparties to mitigate counterparty risk. These provisions may include termination clauses, default remedies, cross-default provisions, and indemnification clauses.

Credit Derivatives and Insurance: Hedge or transfer counterparty risk through the use of credit derivatives such as credit default swaps (CDS) or by purchasing insurance policies against counterparty default. These instruments can provide protection and enhance risk management capabilities.

Continuous Monitoring and Surveillance: Implement robust monitoring and surveillance processes to monitor the financial health and performance of counterparties on an ongoing basis. Regularly review counterparties' financial statements, credit ratings, market conditions, and other relevant factors to identify potential risks and take timely action.

Stress Testing and Scenario Analysis: Conduct stress testing and scenario analysis to assess the potential impact of counterparty default under adverse market conditions. Identify vulnerabilities and develop contingency plans to mitigate the impact of counterparty risk events.

1.1.6 Risk measures

History of the risk measures

The history of risk measures in finance reflects the ongoing quest to develop tools and techniques for quantifying, managing, and understanding risk in financial markets. As financial markets evolve and become increasingly complex, new risk measures continue to be developed to address emerging challenges and opportunities.

Variance: One of the earliest and most fundamental risk measures in finance is variance, which measures the dispersion of returns around the mean or expected return of an investment. Variance was introduced by Harry Markowitz in his seminal work on modern portfolio theory (MPT) in the 1950s. Markowitz demonstrated how variance could be used to quantify and manage portfolio risk by diversifying investments across assets with different risk-return profiles. See for instance (Fabozzi et al., 2002), (Francis and Kim, 2013).

Standard deviation: Standard deviation is another measure of risk that is closely related to variance. It measures the volatility or variability of returns and is the square root of variance. Standard deviation provides a more intuitive measure of risk than variance, as it represents the average deviation of returns from the mean.

Beta: Beta is a measure of systematic risk or market risk, which reflects the sensitivity of an asset's returns to changes in the overall market or a benchmark index. Beta was popularized by William Sharpe in the capital asset pricing model (CAPM) in the 1960s. Assets with betas greater than 1 are considered more volatile than the market, while those with betas less than 1 are considered less volatile. A nice reference is (Sharpe, 1977).

Value-at-Risk (VaR): VaR is a measure of the maximum potential loss that a portfolio or investment may incur over a specified time horizon and confidence level. VaR gained popularity in the 1990s as a risk management tool, particularly in the banking and financial industry. VaR provides a single, summary statistic of risk that is easy to interpret and communicate but has limitations, particularly during periods of extreme market volatility.

Conditional Value-at-Risk (CVaR) or Expected-Shortfall (ES): CVaR is an extension of VaR that measures the average loss beyond VaR in the tail of the distribution of losses. CVaR provides additional information about the severity of potential losses and is often used in conjunction with VaR for risk assessment and management.

Tail Risk Measures: Tail risk measures, such as skewness and kurtosis, capture the asymmetry and fatness of the tails of the distribution of returns. These measures are important for capturing extreme events or tail risk, which may not be adequately captured by traditional risk measures like VaR. See for instance (Groeneveld and Meeden, 1984) for more details about skewness and kurtosis.

Skewness

We say that a distribution is symmetric, if it is equally balanced on both sides of the mean. This means that the frequency of observations on both sides is equal.

However, if the frequency of occurrence of observations is more in a particular direction then the distribution is asymmetric. This asymmetry of the distribution on either side of the mean is called skewness.

A symmetric distribution has no tail on either side. Such a distribution is not skewed in any direction. This is the same as a normal distribution i.e. a distribution which has zero skewness.

Suppose we are looking at a distribution of returns with a mean return of 0. If there is a large frequency of occurrence of negative returns compared to positive returns then the distribution displays a fat left tail or negative skewness. In case the frequency of positive returns exceeds that of negative returns then the distribution displays a fat right tail or positive skewness. Skewness is a measure of degree of asymmetry of a distribution. It measures the degree to which a distribution leans towards the left or the right side.

The sample skewness is given by the following formula:

$$\text{Sample Skewness} = \frac{1}{n\sigma_n^3}\sum_{i=1}^{n}(X_i - \bar{X}_n)^3. \tag{1.3}$$

Kurtosis

Kurtosis is a measure of the "peakedness" of the distribution. Both skewness and kurtosis are measured relative to a normal distribution. Just like a distribution can be negatively or positively skewed, it can be Leptokurtic or Platykurtic depending on whether the peakedness is more than or less than the normal distribution. Leptokurtic: The distribution is more peaked than a normal distribution. Platykurtic: The distribution is less peaked than a normal distribution.

Risk-Adjusted Return Measures: Risk-adjusted return measures, such as the Sharpe ratio, Treynor ratio, and information ratio, assess the return generated by an investment relative to the risk taken. These measures provide insights into the efficiency of portfolio management and help investors evaluate investment opportunities on a risk-adjusted basis.

The sample kurtosis is given by the following formula:

$$\text{Sample Kurtosis} = \frac{1}{n\sigma_n^4}\sum_{i=1}^{n}(X_i - \bar{X}_n)^4. \tag{1.4}$$

Sharpe Ratio

The Sharpe ratio compares the return of an investment with its risk. It's a mathematical expression of the insight that excess returns over a period of time may signify more volatility and risk, rather than investing skill. The Sharpe ratio's numerator is the difference over time between realized, or expected, returns and a benchmark such as the risk-free rate of return or the performance of a particular investment category. Its

denominator is the standard deviation of returns over the same period of time, a measure of volatility and risk.

Mathematically, the Sharpe-Ratio is defined as follows:

$$\text{Sharpe Ratio} = \frac{R_p - R_f}{\sigma_p} \tag{1.5}$$

where R_p is the portfolio's return, R_f is the risk-free rate, and σ_p is the standard deviation of the portfolio's excess returns.

The Sharpe ratio is one of the most widely used methods for measuring risk-adjusted relative returns. It compares a fund's historical or projected returns relative to an investment benchmark with the historical or expected variability of such returns. The **risk-free rate** was initially used in the formula to denote an investor's hypothetical minimal borrowing costs. More generally, it represents the risk premium of an investment versus a safe asset such as a Treasury bill or bond. The ratio is useful in determining to what degree excess historical returns were accompanied by excess volatility. While excess returns are measured in comparison with an investing benchmark, the standard deviation formula gauges **volatility** based on the variance of returns from their mean.

Economist William F. Sharpe proposed the Sharpe ratio in 1966 as an outgrowth of his work on the capital asset pricing model (CAPM), calling it the reward-to-variability ratio. Sharpe won the Nobel Prize in economics for his work on CAPM in 1990. For more details, see (Sharpe, 1998).

> Remark: Generally, the higher the Sharpe ratio, the more attractive the risk-adjusted return.

Sortino Ratio

The standard deviation in the Sharpe ratio's formula assumes that price movements in either direction are equally risky. In fact, the risk of an abnormally low return is very different from the possibility of an abnormally high one for most investors and analysts. A variation of the Sharpe called the Sortino ratio ignores the above-average returns to focus solely on downside deviation as a better proxy for the risk of a fund of a portfolio. The standard deviation in the denominator of a Sortino ratio measures the variance of negative returns or those below a chosen benchmark relative to the average of such returns. See (Rollinger and Hoffman, 2013) for more details.

Treynor Ratio

The Treynor ratio, also known as the reward-to-volatility ratio, is a performance metric for determining how much excess return was generated for each unit of risk taken on by a portfolio. Excess return in this sense refers to the return earned above the return that could have been earned in a risk-free investment. Although there is no true risk-free

investment, treasury bills are often used to represent the risk-free return in the Treynor ratio. Risk in the Treynor ratio refers to systematic risk as measured by a portfolio's beta. Beta measures the tendency of a portfolio's return to change in response to changes in return for the overall market.

The Treynor ratio was developed by Jack Treynor, an American economist who was one of the inventors of the Capital Asset Pricing Model (CAPM). See (Hübner, 2005).

Mathematically, the Treynor ratio is dedined by:

$$\text{Treynor Ratio} = \frac{R_p - R_f}{\beta_p} \tag{1.6}$$

where R_p is the portfolio's return, R_f is the risk-free rate, and β_p is the beta of the portfolio.

In essence, the Treynor ratio is a risk-adjusted measurement of return based on systematic risk. It indicates how much return an investment, such as a portfolio of stocks, a mutual fund, or exchange-traded fund, earned for the amount of risk the investment assumed.

> Remark: If a portfolio has a negative beta, however, the ratio result is not meaningful. A higher ratio result is more desirable and means that a given portfolio is likely a more suitable investment. Since the Treynor ratio is based on historical data, however, it's important to note this does not necessarily indicate future performance, and one ratio should not be the only factor relied upon for investing decisions.

Ultimately, the Treynor ratio attempts to measure how successful an investment is in providing compensation to investors for taking on investment risk. The Treynor ratio is reliant upon a portfolio's beta—that is, the sensitivity of the portfolio's returns to movements in the market—to judge risk.

The premise behind this ratio is that investors must be compensated for the risk inherent to the portfolio, because diversification will not remove it.

> Remark:
>
> (i) The Treynor ratio shares similarities with the Sharpe ratio, and both measure the risk and return of a portfolio. The difference between the two metrics is that the Treynor ratio utilizes a portfolio beta, or systematic risk, to measure volatility instead of adjusting portfolio returns using the portfolio's standard deviation as done with the Sharpe ratio.
>
> (ii) A main weakness of the Treynor ratio is its backward-looking nature. Investments are likely to perform and behave differently in the future than they did in the past. The accuracy of the Treynor ratio is highly dependent on the use of appropriate benchmarks to measure beta. Additionally, there are no dimensions upon which to rank the Treynor ratio. When comparing similar investments, the higher Treynor ratio is better, all else equal, but there is no definition of how much better it is than the other investments.

Information ratio

The information ratio (IR) is a measurement of portfolio returns beyond the returns of a benchmark, usually an index, compared to the volatility of those returns. The benchmark used is typically an index that represents the market or a particular sector or industry. The IR is often used as a measure of a portfolio manager's level of skill and ability to generate excess returns relative to a benchmark, but it also attempts to identify the consistency of the performance by incorporating a tracking error, or standard deviation component into the calculation. The tracking error identifies the level of consistency in which a portfolio "tracks" the performance of an index. A low tracking error means the portfolio is beating the index consistently over time. A high tracking error means that the portfolio returns are more volatile over time and not as consistent in exceeding the benchmark.

Although compared funds may be different in nature, the IR standardizes the returns by dividing the difference in their performances, known as their expected active return, by their tracking error. Mathematically, the Information Ratio is given by:

$$\text{IR} = \frac{R_p - R_B}{e_T} \tag{1.7}$$

where R_p is the portfolio's return, R_B is the Benchmark return, and e_T is the tracking error.

The information ratio identifies how much a fund has exceeded a benchmark. Higher information ratios indicate a desired level of consistency, whereas low information ratios indicate the opposite. Many investors use the information ratio when selecting exchange-traded funds (ETFs) or mutual funds based on their preferred risk profiles. Of course, past performance is not an indicator of future results, but the IR is used to determine whether a portfolio is exceeding a benchmark index fund.

The tracking error is often calculated by using the standard deviation of the difference in returns between a portfolio and the benchmark index. Standard deviation helps to measure the level of risk or volatility associated with an investment. A high standard deviation means there is more volatility and less consistency or predictability. The information ratio helps to determine how much and how often a portfolio trades in excess of its benchmark but factors in the risk that comes with achieving the excess returns.

Like the information ratio, the Sharpe ratio is an indicator of risk-adjusted returns. However, the Sharpe ratio is calculated as the difference between an asset's return and the risk-free rate of return divided by the standard deviation of the asset's returns. The risk-free rate of return would be consistent with the rate of return from a risk-free investment like a U.S. Treasury security. If a particular Treasury security paid a 3% annual yield, the Sharpe ratio would employ 3% as the risk-free rate for comparative purposes.

The IR, on the other hand, measures the risk-adjusted return in relation to a benchmark, such as the Standard & Poor's 500 Index (S&P 500), instead of a risk-free asset. The IR also measures the consistency of an investment's performance. However, the Sharpe

ratio measures how much an investment portfolio outperformed the risk-free rate of return on a risk-adjusted basis.

Both financial metrics have their usefulness but the index comparison makes the IR more appealing to investors since index funds are typically the benchmark used in comparing investment performance and the market return is usually higher than the risk-free return.

> REMARK: comparing multiple funds against a benchmark is difficult to interpret because the funds might have different securities, asset allocations for each sector, and entry points in their investments. As with any single financial ratio, it's best to look at additional types of ratios and other financial metrics to make a more comprehensive and informed investment decision.

See (Gupta et al., 1999) for more details.

Mathematical definition of a risk measure

A coherent risk measure is a function ρ that satisfies properties of monotonicity, subadditivity, homogeneity, and translation invariance.

Consider a random outcome X viewed as an element of a linear space $\mathcal{L}$ of measurable functions, defined on an appropriate probability space. A functional

$$\rho\colon \mathcal{L} \to \mathbb{R} \cup \{+\infty\} \tag{1.8}$$

is said to be coherent risk measure for $\mathcal{L}$ if it satisfies the following properties.

Normalized:

$$\rho(0) = 0. \tag{1.9}$$

That is, the risk of holding no assets is zero.

Monotonicity:

$$\text{If } Z_1, Z_2 \in \mathcal{L} \text{ and } Z_1 \leq Z_2 \text{ a.s. then } \rho(Z_1) \geq \rho(Z_2). \tag{1.10}$$

That is, if portfolio Z_2 always has better values than portfolio Z_1 under almost all scenarios then the risk of Z_2 should be less than the risk of Z_1. E.g. If Z_1 is an in the money call option (or otherwise) on a stock, and Z_2 is also an in the money call option with a lower strike price. In financial risk management, monotonicity implies a portfolio with greater future returns has less risk.

Sub-additivity:

$$\text{If } Z_1, Z_2 \in \mathcal{L}, \text{ then } \rho(Z_1 + Z_2) \leq \rho(Z_1) + \rho(Z_2). \tag{1.11}$$

Indeed, the risk of two portfolios together cannot get any worse than adding the two risks separately : this is the diversification principle. In financial risk management, sub-additivity implies diversification is beneficial. The sub-additivity principle is sometimes also seen as problematic.

Positive Homogeneity:

$$\text{If } \alpha \geq 0 \text{ and } Z \in \mathcal{L}, \text{ then } \rho(\alpha Z) = \alpha\rho(Z). \tag{1.12}$$

Loosely speaking, if you double your portfolio then you double your risk. In financial risk management, positive homogeneity implies the risk of a position is proportional to its size.

Translation Invariance: If A is a deterministic portfolio with guaranteed return a and $Z \in \mathcal{L}$ then :

$$\rho(Z + A) = \rho(Z) - a. \tag{1.13}$$

The portfolio A is just adding cash a to your portfolio Z. In particular, if $a = \rho(Z)$ then $\rho(Z + A) = 0$. In financial risk management, translation invariance implies that the addition of a sure amount of capital reduces the risk by the same amount.

Convex risk measures: The notion of coherence has been subsequently relaxed. Indeed, the notions of Sub-additivity and Positive Homogeneity can be replaced by the notion of convexity:

$$If\ Z_1, Z_2 \in \mathcal{L}\ and\ \lambda \in [0,1]\ then\ \rho(\lambda Z_1 + (1-\lambda)Z_2)) \leq \lambda\rho(Z_1) + (1-\lambda)\rho(Z_2). \tag{1.14}$$

Nice references about coherent risk measures are (Delbaen et al., 1998), (Artzner et al., 1999), (Artzner, 1997).

Value-at-Risk

A wide literature has been developed over the years. Among the numerous references, we can find (Duffie and Pan, 1997), (Linsmeier and Pearson, 2000), (Jorion, 1996), (Best, 2000), (Alexander, 2009), (Jorion, 2007), (Dowd, 1998), (Penza and Bansal, 2001).

History: The problem of risk measurement is an old one in statistics, economics and finance. Financial risk management has been a concern of regulators and financial executives for a long time as well. Retrospective analysis has found some VaR-like concepts in this history. But VaR did not emerge as a distinct concept until the late 1980s. The triggering event was the stock market **crash of 1987**. This was the first major financial

crisis in which a lot of academically-trained quants were in high enough positions to worry about firm-wide survival.

The crash was so unlikely given standard statistical models, that it called the entire basis of quant finance into question. A reconsideration of history led some quants to decide there were recurring crises, about one or two per decade, that overwhelmed the statistical assumptions embedded in models used for trading, investment management and derivative pricing. These affected many markets at once, including ones that were usually not correlated, and seldom had discernible economic cause or warning (although after-the-fact explanations were plentiful). Much later, they were named **"Black Swans"** by Nassim Taleb and the concept extended far beyond finance. If these events were included in quantitative analysis they dominated results and led to strategies that did not work day to day. If these events were excluded, the profits made in between "Black Swans" could be much smaller than the losses suffered in the crisis. Institutions could fail as a result.

VaR was developed as a **systematic** way to segregate extreme events, which are studied qualitatively over long-term history and broad market events, from everyday price movements, which are studied quantitatively using short-term data in specific markets. It was hoped that "Black Swans" would be preceded by increases in estimated VaR or increased frequency of VaR breaks, in at least some markets. The extent to which this has proven to be true is controversial. Abnormal markets and trading were excluded from the VaR estimate in order to make it observable. It is not always possible to define loss if, for example, markets are closed as after 9/11, or severely illiquid, as happened several times in 2008. Losses can also be hard to define if the risk-bearing institution fails or breaks up. A measure that depends on traders taking certain actions, and avoiding other actions, can lead to self reference. This is risk management VaR. It was well established in quantitative trading groups at several financial institutions, notably Bankers Trust, before 1990, although neither the name nor the definition had been standardized. There was no effort to aggregate VaRs across trading desks. This is risk management VaR. It was well established in quantitative trading groups at several financial institutions, notably Bankers Trust, before 1990, although neither the name nor the definition had been standardized. There was no effort to aggregate VaRs across trading desks. The financial events of the **early 1990s** found many firms in trouble because the same underlying bet had been made at many places in the firm, in non-obvious ways. Since many trading desks already computed risk management VaR, and it was the only common risk measure that could be both defined for all businesses and aggregated without strong assumptions, it was the natural choice for reporting firmwide risk. J. P. Morgan CEO Dennis Weatherstone famously called for a "4:15 report" that combined all firm risk on one page, available within 15 minutes of the market close. Risk measurement VaR was developed for this purpose. **Development was most extensive at J. P. Morgan with its system RiskMetrics**, which published the methodology and gave free access to estimates of the necessary underlying parameters in 1994. This was the first time VaR had been exposed beyond a relatively small group

of quants. Two years later, the methodology was spun off into an independent for-profit business now part of RiskMetrics Group (now part of MSCI). In 1997, the U.S. Securities and Exchange Commission ruled that public corporations must disclose quantitative information about their derivatives activity. Major banks and dealers chose to implement the rule by including VaR information in the notes to their financial statements. Worldwide adoption of the **Basel II Accord**, that began in 1999, gave further impetus to the use of VaR. **VaR is the preferred measure of market risk**, and concepts similar to VaR are used in other parts of the accord.

What is VaR ? VaR modeling determines the potential for loss in the entity being assessed and the probability that the defined loss will occur. One measures VaR by assessing the amount of potential loss, the probability of occurrence for the amount of loss, and the time frame.

For example, a financial firm may determine an asset has a 3% one-month VaR of 2%, representing a 3% chance of the asset declining in value by 2% during the one-month time frame. The conversion of the 3% chance of occurrence to a daily ratio places the odds of a 2% loss at one day per month.

Using a firm-wide VaR assessment allows for the determination of the cumulative risks from aggregated positions held by different trading desks and departments within the institution. Using the data provided by VaR modeling, financial institutions can determine whether they have sufficient capital reserves in place to cover losses or whether higher-than-acceptable risks require them to reduce concentrated holdings.

A mathematical definition of VaR is provided as follows. Let X be a random variable representing the loss, as a positive quantity, i.e. we take the convention that big losses correspond to large positive numbers, of a portfolio at some future time, with as cumulative distribution function $F_X(x) = \mathbb{P}(X \leq x); \forall x \in \mathbb{R}$. Let $\alpha \in (0, 1)$ be the risk level. Then, the Value-at-Risk at the risk level α is defined by:

$$\mathbf{VaR}_\alpha(X) := \inf\{x \in \mathbb{R}, F_X(x) \geq \alpha\}. \tag{1.15}$$

Different VaR methodologies There are three main ways of computing VaR: the historical method, the variance-covariance method, and the Monte Carlo method.

Historical VaR

The historical method looks at one's prior returns history, orders them from worst losses to greatest gains and takes the α-empirical quantile—following from the premise that past returns experience will inform future outcomes.

Variance-covariance method

Rather than assuming that the past will inform the future, the variance-covariance method, also called the parametric method, instead assumes that gains and losses are normally distributed. This way, potential losses can be framed in terms of standard deviation events from the mean.

The variance-covariance method works best for risk measurement in which the distributions are known and reliably estimated. It is less reliable if the sample size is very small.

Monte-Carlo method

A third approach to VaR is to conduct a Monte Carlo simulation. This technique uses computational models to simulate projected returns over hundreds or thousands of possible iterations. Then, it takes the chances that a loss will occur—say, 5% of the time—and reveals the impact.

The Monte Carlo method can be used with a wide range of risk measurement problems and relies upon the assumption that the probability distribution for risk factors is known.

VaR advantages: First, VaR provides a single, summary statistic of risk that is **easy to calculate, understand, and communicate**. It expresses the potential loss of a portfolio in a single number, typically in currency units or as a percentage of the portfolio's value, making it straightforward for investors, managers, and regulators to interpret.

Moreover, VaR quantifies the potential loss of a portfolio over a specified time horizon and confidence level, providing a **quantitative measure of risk exposure**. This allows investors and managers to compare different portfolios, assess risk-adjusted performance, and make informed decisions about risk management and asset allocation.

Furthermore, VaR calculations can be standardized across different portfolios, asset classes, and time horizons, allowing for consistent risk assessment and comparison. **Standardization facilitates communication and coordination** among stakeholders and promotes best practices in risk management.

Also, **VaR can be used as a basis for scenario analysis and stress testing** to assess the impact of adverse market conditions or extreme events on portfolio risk. By varying input parameters such as market conditions, correlations, and volatility, VaR can help investors and managers evaluate the robustness of their portfolios and identify potential vulnerabilities.

And, VaR serves as a key component of a broader risk management framework, complementing other risk measures and techniques such as stress testing, sensitivity analysis, and risk-adjusted performance metrics. **VaR provides a concise summary of risk**

exposure that can be integrated into decision-making processes and risk mitigation strategies.

While VaR may not fully capture the benefits of diversification, it still provides insights into the risk profile of diversified portfolios. VaR allows investors to assess the potential impact of diversification on risk reduction and portfolio efficiency, helping them optimize asset allocation and risk-return trade-offs.

VaR presents considerable advantages with respect to Greeks. In financial risk management, VaR has certainly represented a significative step forward with respect to more traditional measures mostly based on sensitivities to market variables (the "Greeks"). The strength of VaR relies in the two following points. On the one hand, VaR applies to any financial instrument and it is expressed in the same unit of measure, namely in "lost money". Greeks on the contrary are measure created ad hoc for specific instruments or risk variables and are expressed in different units. The comparison of relative riskiness between, say, an equity portfolio and a Forex portfolio is not easy with Greeks, while it is a straight comparison knowing their VaR's. On the other hand, VaR includes an estimate of future events and allows one to convert in a single number the risk of a portfolio. Greeks on the contrary essentially amount to "what if" variables. Saying for instance that one loses 1 Euro if interest rates raise of 1 Bps, one still wonders "how likely it is" that interest rates do indeed raise of 1 Bps. VaR on the contrary does exactly this job.

For all theses reasons, VaR is widely used for regulatory compliance purposes in financial institutions, including banks, investment firms, and insurance companies. Regulators often require financial institutions to calculate and report VaR as part of their risk management framework to ensure compliance with regulatory requirements and capital adequacy standards.

VaR limitations: While Value at Risk (VaR) is a widely used measure of financial risk, it has several limitations and weaknesses that should be considered when using it for risk management and decision-making purposes.

First, Value at Risk (VaR) is not considered as a coherent risk measure because it fails to satisfy the sub-additivity property, which is a key criterion for coherence. This means that the VaR of a portfolio may be higher than the sum of the VaRs of its individual components. In other words, VaR does not adequately capture the diversification benefits of combining assets in a portfolio. The consequences of this lack of sub-additivity are the following. VaR may **overestimate the risk of a diversified portfolio** by ignoring the risk-reducing effects of diversification. As a result, VaR may provide a conservative estimate of risk, leading to inefficient risk management decisions. Since VaR does not fully capture the benefits of diversification, it may **incentivize investors to take on more risk than necessary**, especially in complex or highly correlated portfolios where diversification is essential. Investors may misinterpret VaR as providing a

complete and accurate measure of risk, leading to **misunderstandings and misjudgments about the true risk exposure** of their portfolios.

Also, VaR is **unable to capture tail risk**. VaR focuses on the most likely outcomes within a specified confidence level and may not adequately capture extreme or tail risk events that fall outside the confidence interval. Extreme events, such as market crashes or financial crises, can lead to losses beyond the estimated VaR.

Moreover, VaR is of **static nature** i.e. VaR provides a **single-point estimate of risk** at a specific confidence level over a given time horizon and does not capture changes in risk over time or in response to market conditions. VaR is static and does not account for dynamic changes in portfolio composition or risk factors.

And, VaR calculations are **sensitive to model assumptions**, including the choice of time horizon, confidence level, and calculation methodology. Different VaR models may yield different results, leading to inconsistencies and uncertainties in risk assessment.

For all theses reasons, Regulators and policymakers may be hesitant to rely solely on VaR as a risk measure for regulatory compliance purposes due to its lack of coherence. Additional risk measures, such as expected shortfall or coherent risk measures, may be required to complement VaR for regulatory purposes.

To address the shortcomings of the VaR, another risk measure, called Expected-Shortfall (ES), is developed.

Expected-Shortfall

What is Expected-Shortfall (ES) ? ES also known as Conditional Value at Risk (CVaR), is a risk assessment measure that quantifies the amount of tail risk an investment portfolio has. ES is derived by taking a weighted average of the "extreme" losses in the tail of the distribution of possible returns, beyond the value at risk (VaR) cutoff point. ES is used in portfolio optimization for effective risk management. Generally speaking, if an investment has shown stability over time, then the value at risk may be sufficient for risk management in a portfolio containing that investment. However, the less stable the investment, the greater the chance that VaR will not give a full picture of the risks, as it is indifferent to anything beyond its own threshold.

Mathematical definition of ES Let X be a random variable representing the loss[1] of a portfolio at some future time, with as cumulative distribution function $F_X(x) = \mathbb{P}(X \leq x); \forall x \in \mathbb{R}$. Let $\alpha \in (0, 1)$ be the risk level. Then, based on (Tasche, 2002b,

[1] as a positive quantity, i.e. we take the convention that big losses correspond to large positive numbers

Prop 3.4, Eq 3.3), the Expected-Shortfall is defined as:

$$\mathbf{ES}_\alpha(X) = \frac{1}{1-\alpha}\int_\alpha^1 \mathbf{VaR}_\beta(X)\mathrm{d}\beta. \tag{1.16}$$

where $\mathbf{VaR}_\beta$ is the Value-at-Risk given by:

$$\mathbf{VaR}_\beta(X) := \inf\{x \in \mathbb{R}, F_X(x) \geq \beta\} \tag{1.17}$$

as defined in (Tasche, 2002b, Def 2.1, Eq 2.1a). When the distribution is continuous, an equivalent definition can be given (see (Sarykalin et al., 2008, Def.2, p.273)) by:

$$\mathbf{ES}_\alpha(X) = \mathbb{E}[X|X \geq \mathbf{VaR}_\alpha(X)]. \tag{1.18}$$

Advantages of ES: First, ES guarantees a **Better Tail Risk Measurement**. One of the primary advantages of Expected Shortfall is its ability to provide a more accurate representation of tail risk. Unlike VAR, which only considers a specific quantile of the distribution (e.g., the 1% or 5% worst outcomes), ES accounts for the entire tail of the distribution. It takes into consideration the severity of losses beyond the chosen percentile. This means that ES can provide a more robust assessment of potential losses, especially in situations with extreme market events.

Moreover, unlike VaR, ES is a **coherent Risk Measure**. Coherence is a mathematical property that ensures that combining risk measures of individual assets leads to a meaningful measure for the entire portfolio. This property makes ES a more suitable choice for diversified portfolios. In contrast, aggregating VAR values may not accurately reflect the true risk of the portfolio.

And, ES enables **mitigation of underestimation**. ES addresses the key limitation of VAR, which is its tendency to underestimate the risk of rare, extreme events. VAR assumes that asset returns follow a normal distribution, which often doesn't hold in reality, particularly during turbulent times. ES, on the other hand, doesn't rely on this assumption and is more robust in capturing the potential losses during extreme events.

Finally, ES insures transparency and regulation. ES has gained recognition and regulatory support in recent years. It is considered more transparent and informative than VAR, which is why it is favored by regulators. The Basel III banking reforms, for instance, require banks to calculate and report ES. This regulatory push ensures a higher level of risk transparency and accountability.

Expected-Shortfall drawbacks: First, calculating Expected Shortfall is **more complex** than VAR, as it requires estimating a conditional expectation, which involves evaluating the tail of the distribution. This complexity can be a barrier for smaller organizations or less mathematically-inclined professionals.

Then, ES heavily relies on historical data to estimate potential losses, making it **sensitive to the quality and quantity of the available data**. In cases of limited historical data, ES estimates may not be reliable.

Moreover, **ES fails to universality**. While ES is a coherent risk measure, its implementation can vary between different financial institutions and portfolios. This lack of standardization can lead to inconsistencies in risk assessments.

And, while regulatory support for ES is increasing, compliance can be burdensome for financial institutions. It may **require additional infrastructure and resources** to ensure accurate calculation and reporting.

Finally, ES is **not elicitable**. This last point is the subject of a major debate between VaR and ES at the level of regulators in their choice of the reference risk measure.

Value-at-Risk (VaR) versus Expected-Shortfall (ES)

As previously mentioned, VaR and ES are the two risk measures of greatest interest to regulators. These two risk measures both present strengths and weaknesses. On the one hand, the main weaknesses of the VaR are its lack of sensitivity to the tail risk and its lack of coherence due to the fact that it fails to subadditivity. On the other hand, the main weakness of the ES is its lack of elicitability. These two weaknesses have been at the heart of the debate between VaR and ES.

Understanding the lack of sensitivity of VaR to the tail risk: This problem is easily understandable through the paradox exhibited in (Acerbi et al., 2001):

> Paradox: Consider a portfolio A (made for instance of long option positions) of value 1000 Euro with a maximum downside level of 100 Euro and suppose that the worst 5% cases on a fixed time horizon T are all of maximum downside. VaR at 5% on this time horizon would then be 100 Euro. Consider now another portfolio B again of 1000 Euro which on the other hand invests also in strong short futures positions that allow for a potential unbounded maximum loss. We could easily choose B in such a way that its VaR is still 100 Euro on the time horizon T. However, in portfolio A the 5% worst case losses are all of 100 Euro. In portfolio B the 5% worst case losses range from 100 Euro to some arbitrarily high value. Which portfolio is more risky ? According to VaR 5% they bear the same risk !

This paradox is a consequence of the fact that the VaR corresponds to the minimum potential loss that a portfolio can suffer. VaR is not sensitive to the tail risk and ignore extreme events that can occur beyond VaR. VaR, in other words, is a sort of "best of worst cases scenario" and it therefore systematically underestimates the potential losses associated with the specified level of probability.

On way to solve such an issue is to consider the possibility of introducing the mean of the 5% worst cases on a given time-horizon, called the Expected-Shortfall at the risk level 5%.

Understanding the lack of coherence of VaR: The paper (Artzner et al., 1999) face the problem of defining a complete set of axioms that have to be fulfilled by a measure of risks in a generalized sense. A measure which satisfies these axioms is defined a "Coherent Measure of Risk". It is then shown that whenever a portfolio is undoubtedly riskier than another one, it will always have an higher value of risk if the measure is coherent. On the contrary, any measure which does not satisfy some of the axioms will produce paradoxical results of some kind giving a wrong assessment of relative risks.

Sometimes it's useful to decouple the risks associated to different risk drivers. VaR can be then computed "switching on" just some class of risk drivers, holding all the remaining fixed. One then speaks in this case of partial VaR's like "Interest Rate VaR" (IRVaR), "Forex VaR" (FXVaR), "Equity VaR" (EQVaR), "Credit VaR" (CVaR) and so on. In the case of complex portfolios exposed to many risk variables such as in financial institutions, the com- putation of VaR can often be a challenging aspect is due to the fact that the computation can not be split into separate sub–computations due to the two–fold non–additivity of VaR:

Non–additivity by position: given a portfolio made of two subportfolios, total VaR is not given by the sum of the two partial VaR's, with the consequence that adding a new instrument to a portfolio often make it necessary to recompute the VaR for the whole portfolio.

Non–additivity by risk variable: For a portfolio depending on multiple risk variables, VaR is not the sum of partial VaR's. So, for instance, for a convertible bond, VaR is not simply the sum of its IRVaR and EQVaR.

In both cases, **in the case of normal distributed returns** of a portfolio, one it an be proved that the "non–additivity" is actually a "sub–additivity": total VaR is always less or equal than the sum of partial (by position or by risk driver) VaR's. A little thought is enough to understand that in the Gaussian world everything is proportional to the standard deviation which in turn is subadditive. Therefore in the Gaussian world anything is subadditive and there's nothing special with VaR.

> Remark: Recall that the subadditivity of risk measure ρ is given for any two random variables X and Y by:
>
> $$\rho(X+Y) \leq \rho(X) + \rho(Y). \tag{1.19}$$
>
> Subadditivity refers to the property that the risk measure for a portfolio should not be greater than the sum of the risk measures for its individual components. In other words, if you have two portfolios and you merge them

together, the combined risk should not exceed the sum of the risks of the individual portfolios. This is called the diversification principle.

Even if non–additivity raises serious computational difficulties, it is however the direct sign of one of the most interesting aspects of VaR as an instrument for risk analysis, namely its ability to exhibit better than any Greeks the advantages due to diversification of financial instruments and risk drivers. VaR is in fact sensitive to the hedging effect of different positions and the mutual correlation effect of risk drivers. The sub–additivity of VaR in a Gaussian world embodies the common belief that diversifying lower risks.

However, VaR is not subadditive apart from the Gaussian and some other special cases. Failure to account for this lack of subadditivity can result in insufficient capital reserves being set aside to cover potential losses, which could leave financial institutions or investors vulnerable to unexpected market movements or events.

Understanding elicitability and the impact of its absence: The elicitability property is related to the backtesting. Numerous authors interested in this problem. the following works are dedicated to the study of elicitability (Brehmer, 2017), (Fissler et al., 2021), (Ziegel, 2016), (He et al., 2022), (Fissler and Ziegel, 2016), (Embrechts et al., 2021), (Resin, 2023), (Davis, 2016). Some authors defend the thesis that elicitability is necessary for backtesting, while others defend the thesis that elicitability is only useful but not necessary for backtesting. Elicitability is a mathematical property, satisfied by some risk measures, that allows for the ranking of risk models' performance. If a risk measure is elicitable, then there exists a scoring function for that risk measure that can be used for comparative tests on models. Mathematically, elicitability is defined as follows. Let $\mathcal{P}$ be a class of probability measures on $\mathbb{R}$ with the Borel sigma algebra. We consider a functional :

$$\begin{aligned} \nu : \quad & \mathcal{P} \longrightarrow 2^{\mathbb{R}} \\ & \mathbb{P} \longmapsto \nu(\mathbb{P}) \subset \mathbb{R} \end{aligned}$$

where $2^{\mathbb{R}}$ denotes the power set of $\mathbb{R}$. Often, but not always, $\nu(\mathbb{P})$ is single valued.

Definition 1.1. *A scoring function $s : \mathbb{R} \times \mathbb{R} \to [0, +\infty)$ is consistent for the functional ν relative to the class $\mathcal{P}$, if:*

$$\mathbb{E}_{\mathbb{P}} s(t, Y) \leq \mathbb{E}_{\mathbb{P}} s(x, Y) \quad \forall \mathbb{P} \in \mathcal{P}, \quad \forall t \in \nu(\mathbb{P}), \quad \forall x \in \mathbb{R}. \tag{1.20}$$

Here, Y has distribution $\mathbb{P}$. It is strictly consistent if it is consistent and equality in the above equation implies that $x \in \nu(\mathbb{P})$.

Given a consistent scoring function s for a functional ν, an optimal forecast $x^\star$ for $\nu(P)$ is given by :

$$x^\star = arg \min_x \mathbb{E}_{\mathbb{P}} s(x, Y) \tag{1.21}$$

Definition 1.2 (Elicitability)**.** *A functional ν is elicitable relative to the class $\mathcal{P}$, if there exists a scoring function s which is strictly consistent for ν relative to $\mathcal{P}$.*

The most prominent example concerning risk management may be VaR, which is essentially a quantile and as such elicitable.

Elicitability has been proven a useful property for model selection, estimation, forecast comparison and forecast ranking. The main consequences of elicitability are the following:

Transparency: Elicitability ensures that the risk measure can be expressed as a weighted average of individual scenarios or outcomes. This transparency allows stakeholders to understand how the risk measure is calculated and how it relates to underlying factors and scenarios.

Interpretability: Elicitability enhances the interpretability of the risk measure. Stakeholders can easily grasp the implications of the risk measure in terms of the likelihood and severity of potential losses, making it easier to incorporate into decision-making processes.

Comparability: Elicitability enables straightforward comparisons between different risk measures and models. This comparability is essential for evaluating the effectiveness of risk management strategies, assessing the impact of different scenarios, and benchmarking against industry standards.

Model validation: Elicitability facilitates the validation of risk models by providing clear connections between model outputs and observed data or expert judgments. This validation process helps ensure that risk models accurately capture the underlying risk factors and dynamics.

Calibration: Elicitability simplifies the calibration of risk models to historical data or expert opinions. This calibration process ensures that the risk measure accurately reflects the desired level of risk and aligns with stakeholders' risk preferences and objectives.

Regulatory compliance: In some cases, regulatory requirements mandate the use of risk measures that are elicitable and transparent. Compliance with these regulations becomes easier when risk measures have clear interpretations and can be easily explained to regulators and other stakeholders.

Therefore, the lack of elicitability of the ES leads to a reduction of the trust among practitioners who rely on transparent and interpretable risk metrics. It also implies a complexity in risk assessment as stakeholders may struggle to grasp the implications of expected shortfall values and their relationship to underlying scenarios. Another consequence is an increasing complexity in the model validation process. Moreover, the process of calibration is more difficault leading to potential misestimation of risk. And, decision-making is more challenging leading to suboptimal decisions or increased uncertainty about the true level of risk. Finally, it leads to a non-alignment with the regulatory compliance.

1.2 Issues

In a context where the temporal and spatial effects of random processes are decisive in describing many of the phenomena that occur, we ask the following questions:

Research Question #1

What are the most common temporal and spatial effects of the random processes necessary to describe these phenomena properly? Are there random processes with particular properties with respect to these temporal and spatial effects? If such random processes exist, what are their characteristic properties? What are the consequences of these properties on the random processes, that is, how these properties are expressed on the random processes?

The purpose of Chapter 2 is to address these questions. Indeed, in Chapter 2, we focus on two time transformations: time-origin change or time translation, and time-scaling; and we are interested in the related properties called stationarity and self-similarity. The stationarity property refers to the invariance in time and space of a function or process by temporal translation (time-origin change), while the self-similarity property establishes a spatial proportionality relationship between functions or processes taken at proportional times, and the spatial proportionality factor is a function of the time proportionality factor. The self-similarity property is sometimes seen as an invariance in time and space of a function or process via adequate temporal scaling. We propose a new approach for studying stationary and self-similar processes. First, we prove that the only assumptions of stationarity and self-similarity of the quadratic norm of a given process in Hilbert space, without any assumption of distribution, allow obtaining the stationarity and self-similarity of the inner product of the process, with a closed form formula for the latter. Second, we show that adding the Gaussian assumption allows obtaining the stationarity and self-similarity properties of the process not only in distribution but also in the trajectory sense (that is, in terms of equality of processes). We provide examples of such processes, known as the Wiener process and fractional Brownian motion (fBm). Finally, we provide an extension of the stationarity and self-similarity properties in the trajectory sense, to multidimensional Gaussian processes, called multivariate fractional Brownian motions (mfBm).

In the context of increasingly stringent banking regulations, various risk assessment and hedging processes have been developed over time, from the creation of regulated markets to the present day. Risk measurement, which began to flourish in the 1990s with JP Morgan's VaR, is part of the drive to find an effective and accurate method for assessing risk. The goal of a risk measure is twofold: on the one hand, it is intended to ensure that financial institutions keep aside a sufficient amount of money to cope with the risk of default, on the other hand, it pays attention not to overcharge the cash reserve. Two questions arise:

Research Question #2

What is the most appropriate risk measure for assessing risk correctly? How can a risk measure be predicted or estimated to be as accurate as possible?

In our works, we focus in a first time, on predicting the most commonly used risk measure by regulators, called Value-at-Risk (VaR). Second, we focus on Expected-Shortfall as an alternative risk measure to VaR. Since VaR at risk level α of a given portfolio is the α-quantile of the loss distribution associated with that portfolio, and since ES at risk level α is the average of the losses above VaR, these two risk measures are based on the loss distribution of the portfolio, that is, on the distribution of the negative returns. Thus, the manner in which the loss (or returns) distribution is modeled is a determining factor in the prediction or estimation of these risk measures.

Price dynamics are usually described by geometric Brownian motions (GBM) also called exponential Brownian motions. A geometric Brownian motion is a continuous-time stochastic process in which the logarithm of the randomly varying quantity follows a Brownian motion (also called a Wiener process) with drift. A stochastic process S_t is said to follow a GBM if it satisfies the following stochastic differential equation (SDE):

$$\mathrm{d}S_t = \mu S_t \mathrm{d}t + \sigma S_t \mathrm{d}W_t \tag{1.22}$$

where W_t is a Wiener process or Brownian motion, and μ (percentage drift) and σ (percentage volatility) are constants. The former parameter is used to model deterministic trends, whereas the latter parameter is used to model unpredictable events occurring during motion. For an arbitrary initial value S_0 the SDE above has an analytic solution under Itô's interpretation:

$$S_t = S_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t}. \tag{1.23}$$

The process $X_t = \ln \frac{S_t}{S_0}$ satisfying the SDE:

$$\mathrm{d}X_t = \left(\mu - \frac{\sigma^2}{2}\right)\mathrm{d}t + \sigma \mathrm{d}W_t \tag{1.24}$$

is an arithmetic Brownian motion.

For an arbitrary initial value X_0 the SDE above has an analytic solution under Ito's interpretation:

$$X_t = X_0 + \left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t.$$

where $X_0 = \log(S_0)$.

The log-price increments of size h are then given by:

$$\delta_h X_t := X_{t+h} - X_t = \left(\mu - \frac{\sigma^2}{2}\right)h + \sigma \delta_h W_t$$

where $\delta_h W_t = W_{t+h} - W_t$ is the increment of size h of the Wiener process.

According to this model, log-returns are distributed as an arithmetic Brownian motion and benefit from the properties of the Gaussian framework. However, this approach relies on strong assumptions which are sometimes unrealistic. The natural question is as follows:

Research Question #3

How can this model be improved to describe the distribution of the log-returns as realistically and accurately as possible, and thus to predict the risk measure with realism and accuracy? How can we add a long-range (resp. short-range) dependency in the processes?

In Chapter 3, we propose to describe the price trajectories with geometric fractional Brownian motions:

$$\forall i \in [\![1, d]\!],\ S_t^i = S_0^i e^{c_t^i + \sigma_i B_t^{i,H_i}} \quad \text{with} \quad c_t^i = \log\left(\frac{\mathbb{E}[S_t^i]}{S_0^i}\right) - \frac{\sigma_i^2}{2} t^{2H_i}, \tag{1.25}$$

where S_0 is fixed and known, and c_t^i is the centering parameter in the model. This allows the addition of correlations between log-returns to express long-range dependency. The log-returns are then described using stationary and self-similar Gaussian processes with correlated increments, called fractional Brownian motions (fBm).

$$X_t^i = X_0^i + c_t^i + \sigma_i B_t^{i,H_i}. \tag{1.26}$$

And for any $h > 0$, the log-price increments of length h are defined by:

$$\delta_h X_t^i := X_{t+h}^i - X_t^i = \delta_h c_t^i + \sigma_i \delta_h B_t^{i,H_i}. \tag{1.27}$$

Log-returns are affine functions with respect to fractional Brownian motion; as such, they benefit from the properties of Gaussianity, self-similarity and stationarity of the increments, while relaxing the property of independence of the increments.

We introduce a model that provides a Gaussian approximation of the VaR for asset portfolios under fractional dynamics. We demonstrate that such a model is based on orthogonal projections in a Gaussian Hilbert space, taking specific forms for which closed-form formulae are provided. Finally, we quantify the Gaussian VaR approximation by providing an upper bound for the deviation. Backtesting experiments are conducted using simulated and market data to illustrate this theory.

The Gaussian framework is most commonly used in finance due to its convenient properties. However, the Gaussian distribution is thin-tailed, and it assumes that extreme events are rare, and tends to underestimate the probability of such events occurring. Thus, modeling of the log-return distribution by a Gaussian distribution is not appropriate when extreme events occur more frequently.

Research Question #4

Thus, the following question raises: How to accurately model a distribution in which extreme events occur more frequently than in a Gaussian distribution?

To address this question, we propose to model the loss distribution owing to a heavy-tailed distribution, such as Pareto, which assigns higher probabilities to extreme events, which is often more realistic in finance. Why do we choose Pareto distribution? Pareto distribution is a power law that presents interesting scaling and stability by conditioning properties. The cumulative distribution function is defined as follows:

$$F_X(x) = \mathbb{P}(X \leq x) = \left(1 - \left(\frac{x_m}{x}\right)^{\gamma}\right) \mathbb{1}_{\{x \geq x_m\}}. \tag{1.28}$$

where $x_m > 0$ is the scaling parameter, and $\gamma > 0$ is the shape parameter (or Pareto index). However, in a context in which extreme events occur more frequently than in the Gaussian framework, VaR is no longer an efficient risk measure because it is not sensitive to tail risk and, it fails to capture extreme events that leads to an underestimation of the risk. Therefore, we can wonder whether a risk measure more sensitive to the tail risk exists. We propose replacing VaR with Expected-Shortfall (ES) that addresses such shortcomings. Once the risk measure has properly been chosen, we ask the following question:

Research Question #5

How can Expected-Shortfall be estimated accurately and efficiently in a distribution in which extreme events frequently occur, that is, in a heavy-tailed distribution?

Chapter 4 addresses this question. Indeed, in Chapter 4, the objective is to explore robust methods for estimating the Expected-Shortfall in heavy-tailed distributions. The Expected-Shortfall is the average of losses exceeding the VaR, and an estimator of the Expected-Shortfall at risk level α is a mean estimator applied to the distribution tail beyond the $\mathbf{VaR}_\alpha$. Thus, we explore robust mean estimators as an alternative to a simple empirical mean, in heavy-tailed distributions using the toy case of the Pareto distribution. First, we recall the theory on the Expected-Shortfall and on the Pareto distribution, and we present the characteristic properties of the latter. Then, several non-asymptotic mean estimators, such as the Median-of-Means, Trimmed-Mean, or Lee-Valiant estimators, are presented with their characteristics, and compared to the classical empirical mean. We study their bias and provide explicit formulae when possible. Moreover, we evaluate the convergence rate of the bias. Finally, we support the theoretical analysis with experiments, and compare the performances of the different estimators.

1.3 Contributions

1.3.1 Chapter 2 - Introduction to self-similar and stationary Gaussian processes

State of the art

In finance, understanding temporal effects in processes is a major issue. Some well-known properties called stationarity and self-similarity are related to these temporal effects. Specifically, the property of stationarity is associated with temporal translation, also known as a change of temporal origin and the property of self-similarity is related to the change of time scale also called time-scaling.

The stationarity property of a process is interesting because it states that the increments process (i.e. the time-origin changed process) and the initial process have the same characteristics. In other words, the characteristics of a stationary process are invariant by translation, or time-origin change, in time and space. The self-similarity property assumes that the time-scale changed process and the initial process have the same characteristics through an appropriate spatial scaling. The self-similarity property of a process establishes a spatial proportionality relationship between the characteristics of a process taken at two distinct time-scales λt and t with $\lambda > 0$, with a proportionality coefficient depending on the time scale λ.

Moreover, a self-similar process exhibits either long-range or short-range dependence. Depending on the assumptions made on the process, the stationarity and self-similarity properties may concern either the $\mathbb{L}^2$ characteristics of the process, its distribution (weak sense), or its trajectory (strict sense). These properties are often useful when estimating parameters from data since they insure stability of the estimators. Mandelbrot and Van Ness have been the first to introduce the a self-similar Gaussian process with stationary increments, the well-known fractional Brownian motion in 1968 (Mandelbrot and Van Ness, 1968a). They provide an integral representation of such processes. In (Taqqu, 1978), a time-indexed representation for a sequence of self- similar processes whose finite-dimensional moments have been specified is provided. The representation of stationary processes and stationary increments processes via Langevin equation and self-similar processes has been studied in (Viitasaari, 2016). These properties are not specific to the Gaussian framework, but are often exploited in the Gaussian framework and are the subject of numerous studies. Self-similar processes are studied in details in (Das and Pan, 2011, Chap. 3), (Embrechts and Maejima, 2000), (Embrechts, 2009), (Samorodnitsky, 2006), (Chaumont, 2006). In (Lamperti, 1962), authors study properties of semi-stable processes, including self-similar processes. Stationary and self-similar processes are used in various fields, as explained in (Pardo, 2007). Indeed, they can be used to model many space-time scaling random phenomena that can be observed in Finance, Physics, Biology and other fields. For instance, stellar fragments, growth and genealogy of population, option pricing in finance, various areas of image processing, climatology, environmental science are just some of the areas in which self-similar processes are used. Self-similar processes appear in various parts of probability theory,

such as Lévy processes, branching processes, statistical physics, fragmentation theory, coalescent theory, random fields. Some well known examples are: stable Lévy process, fractional Brownian motion, Feller branching diffusion, Bessel processes, self-similar fragmentation (Bertoin, 2002), Brownian sheet. Self-similar processes are also used in Telecommunications and signal processing as mentioned in (Sheluhin et al., 2007). The construction of self-similar processes is explained in (Fan et al., 2015). Such processes are also very interesting when adding additional assumptions such as the independence of their increments (Sato, 1991). Assuming the stationarity property in addition to the self-similarity of the processes, some works show that these processes can take a specific form. Indeed, in (Samorodnitsky et al., 1996, Chap. 7) or in (Barndorff-Nielsen and Pérez-Abreu, 1999), authors show that there is a unique self-similar Gaussian process with stationary increments and this process corresponds to the fractional Brownian motion. Self-similar processes with stationary increments are also studied in the chaos theory in (Maejima and Tudor, 2012). Authors prove that Gaussian processes live in the first Wiener chaos and the latter are expressed as single integrals, with a deterministic integrand, with respect to the Wiener process. Then in the first Wiener chaos the only self-similar Gaussian process with stationary increments is the fractional Brownian motion. In addition, they prove that the elements of the second Wiener chaos are double iterated stochastic integrals with respect to the Wiener process. In the second Wiener chaos, the self-similar processes with stationary increments are not Gaussian anymore, then authors prove that there exists an infinity of such processes.

Gaussian self-similar and stationary processes, such as fractional Brownian motion (fBm), have found numerous applications in finance due to their ability to model complex behaviors observed in financial markets, as demonstrated in (Burnecki and Weron, 2004). One prominent application is in modeling and forecasting volatility, which is crucial for risk management, derivative pricing, and portfolio optimization. The works of (Fernández-Martínez et al., 2013), authors show empirically that the algorithms, based on a geometrical approach (GM algorithms), are more accurate than the classical algorithms, especially with short length time series. The authors checked that GM algorithms are good when working with (fractional) Brownian motions. In particular, they prove theoretically that GM algorithms are also valid to explore long-memory in (fractional) Lévy stable motions.

In (Lavancier et al., 2009), operator self-similar (os-s) processes are studied. In this article, a p-variate stochastic process $X = \{X(t) = (X_1(t), \ldots, X_p(t)), t \in \mathbb{R}\}$ is said operator self-similar (os-s) if there exists a $p \times p$ matrix H (called the exponent of X) such that for any $\lambda > 0$, the following equality of finite-dimensional distributions (fdd) holds:

$$X(\lambda t) \overset{fdd}{=} \lambda^H X(t), \tag{1.29}$$

and the $p \times p$ matrix λ^H is defined by the power series $\lambda^H = e^{H \log(\lambda)} = \sum_{k=0}^{\infty} \frac{H^k (\log(\lambda))^k}{k!}$. Moreover, a random process $X = \{X(t), t \in \mathbb{R}\}$ has stationary increments (si) if:

$$\{X(t+T) - X(T), t \in \mathbb{R}\} = \{X(t) - X(0), t \in \mathbb{R}\}, \text{ for any } T \in \mathbb{R}. \tag{1.30}$$

Authors study Gaussian os-s process with stationary increments called operator fractional Brownian motion (ofBm). They explain that for $p = 1$ the class of ofBm coincides with fundamental class of fractional Brownian motions (fBm) as proved in (Samorodnitsky and Taqqu, 1994). They recall that a fBm with exponent $H \in (0, 1)$ can be alternatively defined as a stochastically continuous Gaussian process $X = \{X(t), t \in \mathbb{R}\}$ with zero mean and covariance:

$$\mathbb{E}\left[X(s)X(t)\right] = \frac{\sigma^2}{2}\left(|s|^{2H} + |t|^{2H} - |t-s|^{2H}\right), \quad t, s \in \mathbb{R}, \tag{1.31}$$

where $\sigma^2 = \mathbb{E}\left[X^2(1)\right]$. According to their works, the form of covariance of general ofBm seems to be unknown and may be quite complicated. The structure of ofBm and stochastic integral representations are studied in (Didier and Pipiras, 2008), in (Didier and Pipiras, 2012), (Didier and Pipiras, 2011). A particular case of os-s processes corresponds to diagonal matrix $H = diag(H_1, \ldots, H_p)$. In this case, Equation (**??**) becomes:

$$\left(X_1(\lambda t), \ldots, X_p(\lambda t)\right) \overset{fdd}{=} \left(\lambda^{H_1} X_1(t), \ldots, \lambda^{H_p} X_p(t)\right). \tag{1.32}$$

A p-variate process X satisfying Equation (**??**) for any $\lambda > 0$ is called vector self-similar (vs-s) and a stochastically continuous Gaussian vs-s process with stationary increments (si) is called a vector fractional Brownian motion (vfBm). From Equation (**??**), each component $X_i = \{X_i(t), t \in \mathbb{R}\}$, $i = 1, \ldots, p$ of vs-s process is a (scalar) self-similar process, the fact which is not true for general os-s processes. In (Lavancier et al., 2009), a general form of the (cross-)covariance function of vs-s si process X with finite variance and exponent $H = diag(H_1, \ldots, H_p)$, $0 < H_i < 1$. In other words, the paper obtains the general form of the cross-covariance function of vector fractional Brownian motion with correlated components having different self-similarity, cross-covariance function.

Let $X = \{X(t), t \in \mathbb{R}\}$ be a 2nd order process with values in $\mathbb{R}^p$. Assume that X has stationary increments, zero mean, $X(0) = 0$, and that X is vector self-similar with exponent $H = diag(H_1, \ldots, H_p)$, $0 < H_i < 1$ $(i = 1, \ldots, p)$. Moreover, assume also that for any $i, j = 1, \ldots, p$, the function $t \mapsto \mathbb{E}\left[X_i(t)X_j(1)\right]$ is continuously differentiable on $(0, 1) \cup (1, \infty)$. Let $\sigma_i^2 > 0$ denote the variance of $X_i(1), i = 1, \ldots, p$.

(i) If $i = j$, then for any $(s, t) \in \mathbb{R}^2$, we have:

$$\mathbb{E}\left[X_i(s)X_i(t)\right] = \frac{\sigma_i^2}{2}\{|s|^{2H_i} + |t|^{2H_i} - |t-s|^{2H_i}\}. \tag{1.33}$$

(ii) If $i \neq j$ and $H_i + H_j \neq 1$, then under some regularity conditions there exists $c_{ij}, c_{ji} \in \mathbb{R}$ such that for any $(s, t) \in \mathbb{R}^2$:

$$\text{Cov}\left(X_i(s), X_j(t)\right) = \frac{\sigma_i\sigma_j}{2}\{c_{ij}(s)\,|s|^{H_i+H_j} + c_{ji}(t)\,|t|^{H_i+H_j} - c_{ij}(t-s)\,|t-s|^{H_i+H_j}\} \tag{1.34}$$

where $\sigma_i^2 := \mathbb{V}[X_i(1)]$ and:

$$c_{ij}(t) = \begin{cases} c_{ij} & \text{if } t > 0 \\ c_{ji} & \text{if } t > 0. \end{cases} \tag{1.35}$$

(iii) If $i \neq j$ and $H_i + H_j = 1$, then there exists $d_{ij}, f_{ij} \in \mathbb{R}$ such that for any $(s,t) \in \mathbb{R}^2$, we have:

$$\mathbb{E}\left[X_i(s)X_j(t)\right] = \frac{\sigma_i\sigma_j}{2}\{d_{ij}(|s| + \left|t\right| - \left|s-t\right|) + f_{ij}(t\log\left|t\right| - s\log|s|) - (t-s)\log\left|t-s\right|\}. \tag{1.36}$$

(iv) The matrix $R = (R_{ij})_{i,j=1,\dots,p}$ is positive definite, where:

$$R_{ij} := \begin{cases} 1 & \text{if } i = j, \\ c_{ij} + c_{ji} & \text{if } i \neq j, H_i + H_j \neq 1, \\ d_{ij} & \text{if } i \neq j, H_i + H_j = 1. \end{cases} \tag{1.37}$$

Authors also provide a stochastic representation of vfBm and the covariance function of vfBm. The stochastic representation of vgBm is based on (Didier and Pipiras, 2008):

$$X(t) = \int_{\mathbb{R}} \left\{ \left((t-x)_+^{H-\frac{1}{2}} - (-x)_+^{H-\frac{1}{2}} \right) A_+ + \left((t-x)_-^{H-\frac{1}{2}} - (-x)_-^{H-\frac{1}{2}} \right) A_- \right\} W(\mathrm{d}x), \tag{1.38}$$

where $H - \frac{1}{2} := diag\left(H_1 - \frac{1}{2}, \dots, H_p - \frac{1}{2}\right)$, $x_+ := max(x,0)$, $x_- := max(-x,0)$, A_+, A_- are real $p \times p$ matrices and $W(\mathrm{d}x) = \left(W_1(\mathrm{d}x), \dots, W_p(\mathrm{d}x)\right)$ is a Gaussian white noise with zero mean, independent components and covariance $\mathbb{E}\left[W_i(\mathrm{d}x)W_j(\mathrm{d}x)\right] = \delta_{ij}\mathrm{d}x$. The covariance function of vfBm $X = \{X(t), t \in \mathbb{R}\}$ given by double-sided stochastic integral representation is provided. Let $a_{ij}^{++} := \sum_{k=1}^p a_{ik}^+ a_{jk}^+$, $a_{ij}^{--} := \sum_{k=1}^p a_{ik}^- a_{jk}^-$, $a_{ij}^{+-} := \sum_{k=1}^p a_{ik}^+ a_{jk}^-$, $a_{ij}^{-+} := \sum_{k=1}^p a_{ik}^- a_{jk}^+$, where $A_+ = (a_{ij}^+)$, $A_- = (a_{ij}^-)$ are the $p \times p$ matrices. Clearly, $A_+A_+^\star = (a_{ij}^{++})$, $A_-A_-^\star = (a_{ij}^{--})$, $A_+A_-^\star = (a_{ij}^{+-})$, $A_-A_+^\star = (a_{ij}^{-+})$.

Authors show that the covariance of the process defined by the double-sided stochastic integral satisfies the following properties:

(i) for any $i = 1, \dots, p$ the variance of $X_i(1)$ is:

$$\sigma_i^2 = \frac{B\left(H_i + \frac{1}{2}\right), H_i + \frac{1}{2}}{sin(H_i\pi)}\{a_{ii}^{++} + a_{ii}^{--} - 2sin(H_i\pi)a_{ii}^{+-}\}. \tag{1.39}$$

(ii) If $H_i + H_j \neq 1$ then for any $s,t \in \mathbb{R}$, the cross-covariance $\mathbb{E}\left[X_i(s)X_j(t)\right]$ of the process defined by Equation (**??**) is given by Equation (**??**) with:

$$\frac{\sigma_i\sigma_j}{2}c_{ij} := \frac{B\left(H_i + \frac{1}{2}, H_j + \frac{1}{2}\right)}{sin(\left(H_i + H_j\right)\pi)}\left\{a_{ij}^{++}cos(H_i\pi) + a_{ij}^{--}cos(H_j\pi) - a_{ij}^{+-}sin(\left(H_i + H_j\right)\pi)\right\}. \tag{1.40}$$

(iii) If $H_i + H_j = 1$ then for any $s, t \in \mathbb{R}$, the cross-covariance $\mathbb{E}\left[X_i(s)X_j(t)\right]$ of the process defined by Equation (**??**) is given by Equation (**??**) with:

$$\sigma_i\sigma_j d_{ij} := B\left(H_i + \frac{1}{2}, H_j + \frac{1}{2}\right) \times \left\{\frac{sin(H_i\pi) + sin(H_j\pi)}{2}\left(a_{ij}^{++} + a_{ij}^{--}\right) - a_{ij}^{+-} - a_{ij}^{-+}\right\} \tag{1.41}$$

$$\sigma_i\sigma_j f_{ij} := (H_j - H_i)(a_{ij}^{++} - a_{ij}^{--}). \tag{1.42}$$

In, (Coeurjolly, 2000b) (Amblard and Coeurjolly, 2011b), authors propose an approach for the identification of the multivariate fractional Brownian motion. They use a wavelet-like filtering technique. The Hurst parameters, variances, correlation and asymmetry coefficients are estimating by regressing over the empirical log-variances and log-correlations coefficients. They show that the estimator converges almost surely and satisfies a central limit theorem. The convergence is illustrated on simulation, and they apply the estimation procedure to financial data. The irregularity's analysis of data modeled by an fBm, the study of its spectral behavior, and any forecasting problem based on fbm imply the necessity to estimate the Hurst parameter. In (Coeurjolly, 2000b), authors describe the main parametric methods to estimate the self-similarity parameter H. They distinguish four approaches: spectral methods (log-periodogram, a variant of Lobato and Robinson's method), Maximum likelihood (Whittle's estimator), Time-scale methods (wavelet decomposition of the fBm), Temporal methods (number of level crossings, discrete variations). Article (Coeurjolly, 2001), develops a class of consistent estimators of the parameters of a fractional Brownian motion based on the asymptotic behavior of the $k-th$ absolute moment of discrete variations of its sampled paths over a discrete grid of the interval $[0, 1]$. Authors derive explicit convergence rates for these types of estimators, valid through the whole range $0 < H < 1$ of the self-similarity parameter. They also establish the asymptotic normality of their estimators. The effectiveness of their procedure is investigated in a simulation study. In (Coeurjolly et al., 2013), authors study the multivariate fractional Brownian motion (mfBm) viewed through the lens of the wavelet transform. They calculate the correlation structure of the wavelet transform of the mfBm. They study the asymptotic behaviour of the correlation, showing that if the analyzing wavelet has a sufficient number of null first order moments, the decomposition eliminates any possible long-range (inter)-dependence. The cross-spectral density is also considered. Its existence is proved ans its evaluation is performed using a von Bahr-Essen like representation of the function $sign(t)\left|t\right|^{\alpha}$. the behavior of the cross-spectral density of the wavelet field at the zero frequency is also developed and confirms the results provided by the asymptotic analysis of the correlation. In (Jean-franÇois Coeurjolly and Vidakovic, 2014), authors discuss estimation of a scaling parameter σ^2 when a Hurst exponent is known. To estimate σ^2, they propose three approaches based on maximum likelihood estimation, moment-matching, and concentration inequalities, respectively and discuss the theoretical characteristics of the estimators and optimal-filtering guidelines. They justify the improvement of the estimation of σ^2 when a Hurst parameter is known. Using the three approaches and a parametric bootstrap methodology in a simulation study, they compare the confidence intervals of σ^2 intervals of their lengths, coverage rates, and computational complexity

and discuss empirical attributes of the tested approaches. They found that the approach based on maximum likelihood estimation was optimal in terms of efficiency and accuracy, but computationally expensive. The moment-matching approach was found to be not only comparably efficient and accurate but also computationally fast and robust to deviations from the fractional Brownian motion model. In (Coeurjolly, 2005), authors introduce a new class of consistent estimators of the fractal dimension of locally self-similar Gaussian processes. These estimators are based on linear combinations of empirical quantiles (L-statistics) of discrete variations of a sample path over a discrete grid of the interval $[0, 1]$. They derive the almost sure convergence for these estimators and prove the asymptotic normality. The key-ingredient is a Bahadur representation for empirical quantiles of non-linear functions of Gaussian sequences with correlation function decreasing hyperbollicaly. In (Coeurjolly et al., 2010a), authors study the covariance structure of the multivariate fractional Gaussian noise. They evaluate several parameters of the model that allow to control the correlation structure at lag zero between all the components of the multivariate process. Then, they specify an algorithm that allows the exact simulation of multivariate fractional Gaussian noises and thus fractional Brownian motions. Illustrations involve the estimation of the Hurst exponents of each of the components. The works of (Coeurjolly et al., 2010b), are devoted to study some properties of an extension of the well-known fractional Brownian motion to the multivariate case. They study the covariance structure of the multivariate fractional Gaussian noise. They evaluate several parameters of the model that allow to control the correlation structure at lag zero between all the components of the multivariate process. They particularly focus on two cases for which they can relate characteristic parameters of the covariance function to parameters of the stochastic representation of the processes. These cases are the causal case, a direct multivariate generalization of (Mandelbrot and Van Ness, 1968b) representation, and the well-balanced case which adds to the previous case and anti-causal filtering of a Brownian motion. The characterization of the covariance function is then used to study the multivariate fractional Gaussian noise, defined as the increment process of the multivariate fractional Brownian motion. They study the covariance structure as well as the spectral structure of this multivariate stationary process. They exhibit the intriguing facts that two fractional Gaussian noise may be long-range interdependent when only one is long-range dependent. They then perform a wavelet analysis of the multivariate fractional Brownian motion, and show that the wavelet analysis may destroy the long-range interdependence if the wavelet is properly chosen. In (Garcin, 2019), study the inverse Lamperti transform of a fractional Brownian motion and its properties. The inverse Lamperti transform of an fbm is a stationary process. They determine the empirical Hurst exponent of such a composite process with the help of a regression of the log-absolute moments of its increments, at various scales, on the corresponding log-scales. This perceived Hurst exponent underestimates the Hurst exponent of ht underlying fBm. They encounter some time series having a perceived Hurst exponent lower than $\frac{1}{2}$, but an underlying Hurst exponent higher than $\frac{1}{2}$. This paves the way for short- and medium-term forecasting. Indeed, in such series, mean reversion predominates at high scales, whereas persistence is overriding at lower scales. They propose, a way to characterize the Hurst horizon, namely a limit scale between these opposite behaviors. They show that the delampertized fBm, which mixes per-

sistence and mean, is relevant for financial time series, in particular for high-frequency foreign exchange rates. In their sample, the empirical Hurst horizon is always above 1h and 23min. In (Laha and Rohatgi, 1981), operator self similar stochastic processes taking values in a finite dimensional Euclidean space are introduced and some of their properties are studied. In (Hudson and Mason, 1982), a general representation for an operator-self-similar process is obtained and its class of exponents is characterized. It is shown that such a process is the limit in a certain sense of an operator-normed process and any limit of an operator-normed process is operator-self-similar. The works of (Sato, 1991) propose a study of the self-similar and operator-self-similar processes with independent increments. They prove that under some additional conditions of continuity on the operator-self-similar process, the operator can be chosen as a power function with as exponent an operator that presents some special spectral properties. In (Maejima and Mason, 1994), operator-self-similar processes are studied and several examples of operator-self-similar and stable (in the ordinary sense or in the sense of operator-stable) processes are constructed. Limit theorems for such processes are also shown. In (Marinucci and Robinson, 2000), a weak convergence to a form of fractional Brownian motion is established for a wide class of non-stationary fractionally integrated processes. An extension of the classical fractional Brownian motion, called multifractional Brownian motion (mBm), whose Hurst exponent depends on time is studied in (Stoev and Taqqu, 2006). The multifractional Brownian motion (mBm) processes are locally self-similar Gaussian processes. In literature, two types of mBm processes were introduced by using time-domain and frequency domain integral representations of the fBm, respectively. In this paper, authors show that these two types of processes have different correlations structures when the function $H(t)$ is non constant. They focus on a class of mBm processes parametrized by $(a^+, a^-) \in \mathbb{R}^2$, which contains the previously introduced two types of processes as special cases. They establish the connection for their time and frequency-domain integral representations and obtain explicit expressions for their covariances. They show, that there are non-constant functions $H(t)$ for which the correlation structure of the mBm processes depends non-trivially on the value of (a^+, a^-) and hence, even for a given function $H(t)$, there are an infinite number of mBm processes with essentially different distributions.

Modeling volatility: Volatility, the standard deviation of asset returns, is not constant but exhibits clustering and persistence over time, known as volatility clustering. Gaussian self-similar processes like fBm can capture this feature effectively.

By modeling volatility as a fractional Brownian motion, analysts can incorporate long-range dependence and memory into volatility dynamics, which is observed in real financial time series data. Large changes in prices tend to cluster together, resulting in persistence of the amplitudes of price changes. Several works have been published on this topic, as instance the works of (Cheong, 2010) or (Cont, 2007) that explain the origin of this volatility clustering, by proposing agent-based models based on a feature that allows switching between low and high activity regimes with heavy-tailed durations of regimes. The works of (Gatheral et al., 2014), deal with estimation of volatility from recent high frequency data by revisiting the question of the smoothness of the volatility

process. In these works, the assumption is that the log-volatility behaves essentially as a fractional Brownian motion with Hurst exponent H of order 0.1, at any reasonable time scale. Such assumption enables to obtain improved forecasts of realized volatility. And a quantitive market microstructure-based model relating the roughness of volatility to high frequency trading and order splitting, is provided. In (Rostek, 2012), the underlying security dynamics are driven by a jump-diffusion process where the diffusion part is fractional Brownian motion while jumps exhibit a double-exponential distribution.

In (Chong et al., 2022a), (Chong et al., 2022b) and (Szymanski and Takabatake, 2023), methods of estimation of the Hurst exponent as well as convergence rates are provided. The stationarity property that guarantees the invariance by translation of the process allows obtaining a stability of the estimations based on the increments. Moreover, the self-similarity property that establishes the invariance by time-scaling of the process insures that the estimations are the same whatever the chosen time-scale.

Option pricing: Option pricing models often assume a certain stochastic process for the underlying asset's price dynamics. Since volatility significantly influences option prices, accurately modeling volatility is crucial. Gaussian self-similar processes can be employed to model volatility processes, which in turn are used in option pricing models such as the Heston model or the stochastic volatility model. The works of (Rostek, 2009), (Rostek and Rostek, 2009) use fractional Brownian motions for the purpose of option pricing. In this book, the following questions are addressed: To what extent one can draw parallels between the fractional and the classical Brownian motion framework. More precisely, as fractional Brownian motion is an extension of Brownian motion, is it possible to extend the respective theory of option pricing? Are the well-developed techniques of stochastic calculus transferable to fractional Brownian motion? Will we be faced with conceptual problems? Can we obtain closed-form solutions? In (Rostek and Schöbel, 2006), under the assumption that the market is driven by a fractional Brownian motion, formulae for fraction European options are derived using the traditional idea of conditional expectation. The works of (Rostek and Schoebel, 2010), European option prices are derived when the underlying security dynamics are driven by geometric fractional Brownian motion. Thanks to the self-similarity property of the fractional Brownian motion, and thanks to the correlation between increments allowed by a Hurst exponent different from 0.5, the latter is a parsimonious way to capture serial correlation within financial time series. They discuss a model where market participants have constant relative risk aversion and trade in discrete time. Investor's wealth and the underlying stock are assumed to be of fBm type and follow a bivariate log-normal distribution. They introduce an equilibrium condition and provide closed-form solutions for European options. The derived results are an extension of the Black-Scholes pricing formulae and contain the latter as a special case.

Portfolio Optimization: Modern portfolio theory relies on accurate estimation of asset returns and volatilities. Gaussian self-similar processes can improve the estimation of these parameters by capturing their inherent characteristics, such as long memory and self-similarity. This, in turn, leads to more robust portfolio optimization strategies

that account for the non-linear and non-Gaussian nature of financial markets. See for instance (Papenbrock, 2011), (Zlatniczki and Telcs, 2024), (Galloway and Nolder, 2008), (Lunga, 2006). The works of (Czichowsky et al., 2018), (Czichowsky and Schachermayer, 2017) propose a reconciliation of two conflicting concepts in finance: on the one hand, the notion of no arbitrage, and on the other hand, the consideration of non-semimartingale price processes, like fractional Brownian motion. Imposing (arbitrary small) proportional transaction costs and considering logarithmic utility optimisers, the existence of semimartingale, frictionless shadow price process for an exponential fractional Brownian financial market is proved. In (Jumarie, 2005), the model of optimal portfolio first proposed by Merton is considered with the additional assumption that the noises involved in the dynamics of the wealth are fractional Brownian motions (in the sense of fractional derivative of Gaussian white noises) with short-range dependence, that is to say with a Hurst parameter lower than $1/2$. The works of (Sarol et al., 2007), authors consider the classical Merton problem of finding the optimal consumption rate and the optimal portfolio in a Black-Scholes market driven by fractional Brownian motion with Hurst parameter $H > 1/2$. The integrals with respect to the fBm are in the Skorohod sense, not pathwise which is known to lead to arbitrage. An explicit form is derived for the optimal consumption rate and the optimal portfolio in such a market for an agent with logarithmic utility functions. A true self-financing portfolio is found to lead to a consumption term that is always favorable to the investor. In (HU et al., 2003), authors present a mathematical model for a Black–Scholes market driven by fractional Brownian motion with Hurst parameter. The interpretation of the integrals with respect to the fBm is in the sense of Itô (Skorohod–Wick), not pathwise (which is known to lead to arbitrage). They find explicitly the optimal consumption rate and the optimal portfolio in such a market for an agent with utility functions of power type. In (Garcin, 2022), log-prices follow an fBm, the non-Markovian nature of the fBm is used to forecast future states of the process and make statistical arbitrages. Some questions about optimizing trading strategies in the fBm framework are addressed. Which lagged increments of the fBm, observed in discrete time, are to be considered? If the predicted increment is close to zero, up to which threshold is it more profitable not to invest? In (Bauerle and Desmettre, 2020), a fractional version of the Heston volatility model is considered. Within this model portfolio optimization problems for power utility functions are treated.

High-frequency trading: In high-frequency trading (HFT), where decisions are made within milliseconds, understanding and predicting market dynamics is crucial. Gaussian self-similar processes can provide insights into short-term market behaviors by capturing the intricate patterns and correlations present in high-frequency data. See for instance (Evertsz, 1995), (Arroum, 2007), (Smith, 2010). Based on the observation that, in the high-frequency limit, conditionally expected increments of fractional Brownian motion converge to a white noise, shedding their dependence on the path history and the forecasting horizon and making dynamic optimisation problems tractable, works of (Guasoni et al., 2021) propose an explicit formula for locally mean–variance optimal strategies and their performance for an asset price that follows fractional Brownian motion. The works of (Guasoni et al., 2019) consider a a market with an asset price

described by fractional Brownian motion, which can be traded with temporary nonlinear price impact, we find asymptotically optimal strategies for the maximization of expected terminal wealth. Exploiting the autocorrelation in increments while limiting trading costs, these strategies generate an average terminal wealth that grows with a power of the horizon, the exponent depending on both the Hurst and the price-impact parameters. The resulting Sharpe ratios are bounded, insensitive to the horizon, and asymmetric with respect to the Hurst exponent. These results extend to Gaussian processes with long memory and to a class of self-similar processes. The works of (Lim and Muniandy, 2002) study some Gaussian models for anomalous diffusion, which include the time-rescaled Brownian motion, two types of fractional Brownian motion, and models associated with fractional Brownian motion based on the generalized Langevin equation. Gaussian processes associated with these models satisfy the anomalous diffusion relation which requires the mean-square displacement to vary with t^α, $0 < \alpha < 2$. However, these processes have different properties, thus indicating that the anomalous diffusion relation with a single parameter is insufficient to characterize the underlying mechanism. Although the two versions of fractional Brownian motion and time-rescaled Brownian motion all have the same probability distribution function, the Slepian theorem can be used to compare their first passage time distributions, which are different. Finally, in order to model anomalous diffusion with a variable exponent $\alpha(t)$ it is necessary to consider the multifractional extensions of these Gaussian processes.

Risk management: Understanding and managing risk is fundamental in finance. Gaussian self-similar processes allow for more accurate risk assessment by capturing the complex and correlated nature of financial time series data. By incorporating long-range dependence and self-similarity into risk models, financial institutions can better estimate Value-at-Risk (**VaR**) and Conditional Value-at-Risk CVaR, which are measures of potential losses under adverse market conditions. This is the topic of (Michna et al., 1998). Indeed, in this paper, authors are interested in collective risk theory. They observe that collective risk theory is concerned with random fluctuations of the total assets and the risk reserve of an insurance company. They consider self-similar, continuous processes with stationary increments for the renewal model in risk theory. They construct a risk model which shows a mechanism of long range dependence of claims. An approximation of the risk process by a self-similar process with drift, is provided. The ruin probability within finite time is estimated for fractional Brownian motion with drift. A similar model is applicable in queueing systems, describing long range dependence in on/off processes and associated fluid models. The works of (Wesselhöfft, 2021) are based on the fact that a self-similar process, which is ablle to account for long-memory behaviour is the fractional Brownian motion, which has a possible non-Gaussian limit under convolution of the increments. The increments of the Fractional Brownian Motion can exhibit long memory through a parameter H, the Hurst exponent. For the Fractional Brownian Motion this scaling (Hurst) exponent would be constant over different orders of moments, being unifractal. But empirically, we observe varying Hölder exponents, the continuum of Hurst exponents, which implies multifractal behavior. Authors explain the multifractal behavior through the changing alpha-stable indices from the alpha-stable distributions over sampling frequences by applying filters

for seasonality and time dependence (long-memory) over different sampling frequencies, starting at high-frequencies up to one minute. By utilizing a filter for long-memory they show, that the low-sampling frequency process, not containing the time dependence component, can be governed by the alpha-stable motion. Under the alpha-stable motion they propose a semiparameteric method called Frequency Rescaling Methodology (FRM), which allows to rescale the filtered high-frequency dataset to the lower sampling frequency. The datasets for example weekly data which are obtained by rescaling high-frequency data with the FRM are more heavy tailed than the ones observed empirically. Authors show that using a subset of the whole dataset suffices for the FRM to obtain a better forecast in terms of risk for the whole dataset. Specifically, the FRM would have been able to account for tail events of the financial crisis 2008.

Different types of self-similar processes: Fractional Brownian motion is widely used in the modeling of phenomena with power spectral density of power-law type. However, FBM has its limitation since it can only describe phenomena with monofractal structure or a uniform degree of irregularity characterized by the constant Holder exponent. For more realistic modeling, it is necessary to take into consideration the local variation of irregularity, with the Holder exponent allowed to vary with time or space. In (Muniandy and Lim, 2001), an extension of the standard fBm to multifractional Brownian motion (mBm) indexed by a Holder exponent that is a function of time, is proposed. This paper proposes an alternative generalization based on the FBM defined by the RiemannLiouville type of fractional integral. The local properties of the Riemann-Liouville MBM, RLMBM are studied and they are found to be similar to that of the standard MBM. A numerical scheme to simulate the locally self-similar sample paths of the RLMBM for various types of time-varying Holder exponents is given. The local scaling exponents are estimated based on the local growth of the variance and the wavelet scalogram methods. Finally, an example of the possible applications of RLMBM in the modeling of multifractal time series is illustrated. In (Tudor, 2013), several self-similar processes are studied: fractional, bi-fractional and sub-fractional Brownian motions. The works of (Pagnini et al., 2012), study the Master Equation approach to model anomalou diffusion. Anomalous diffusion in complex media can be described as the result of a superposition mechanism reflecting inhomogeneity and nonstationarity properties of the medium. For instance, when this superposition is applied to the time-fractional diffusion process, the resulting Master Equation emerges to be the governing equation of the Erdelyi-Kober fractional diffusion, that describes the evolution of the marginal distribution of the so-caled generalized grey Brownian motion. This motion is a parametric class of stochastic processes that provides models for both fast and slow anomalous diffusion: it is made up of self-similar processes with stationary increments and depends on two real parameters. The class includes the fractional Brownian motion, the timefractional diffusion stochastic processes, and the standard Brownian motion. In this framework, the M-Wright function known also as Mainardi function emerges as a natural generalization of the Gaussian distribution, recovering the same key role of the Gaussian density for the standard and the fractional Brownian motion. In (Pang and Taqqu, 2019), the focus is paid to shot noise processes with Poisson arrivals and nonstationary noises. The noises are conditionally

independent given the arrival times, but the distribution of each noise does depend on its arrival time. Authors establish scaling limits for such shot noise processes in two situations: (a) the conditional variance functions of the noises have a power law and (b) the conditional noise distributions are piecewise. In both cases, the limit processes are self-similar Gaussian with nonstationary increments. Motivated by these processes, they introduce new classes of self-similar Gaussian processes with nonstationary increments, via the time-domain integral representation, which are natural generalizations of fractional Brownian motions.

Fractional Brownian motion and arbitrage management: Using a fractional geometric Brownian motion to describe the price's dynamics allows being able to manage the long-range (resp. the short-range) dependency, however the fundamental market assumption of non-arbitrage opportunity is not satisfied anymore. This is the purpose of the works of (Cheridito, 2003). Authors construct arbitrage strategies for a financial market that consists of a money market account and a stock whose discounted price follows a fractional Brownian motion with drift or an exponential fractional Brownian motion with drift. Then we show how arbitrage can be excluded from these models by restricting the class of trading strategies. In (Cheridito, 2001), authors show that the sum of a Brownian motion and a non-trivial multiple of an independent fractional Brownian motion with Hurst exponent $H \in (0,1]$ is not a semimartingale if $H \in (0,1/2) \cup (1/2,3/4]$, that it is equivalent to a multiple of Brownian motion if $H = 1/2$ and equivalent to Brownian motion if $H \in (3/4,1]$. They discuss the price of a European call option on an asset driven by a linear combination of a Brownian motion and an independent fractional Brownian motion.

Stochastic integral with respect to the fractional Brownian motion: So far, the stochastic integral was defined with respect to the standard Brownian motion. Some authors have proposed extension of the notion of stochastic integral with respect to the fractional Brownian motion for different ranges of values taken by the Hurst exponent. In (Cheridito and Nualart, 2005), the stochastic integral with respect to the fractional Browian motion with Hurst exponent $H \in (0,1/2)$ is defined. This extends the divergence integral from Malliavin calculus. For this extended divergence integral, a Fubini theorem is porved and versions of the Itô and Tanaka's formula are are established for all $H \in (0,1/2)$. The works of (Carmona et al., 2003), define a stochastic integral with respect to fractional Brownian motion for every value of the Hurst index $H \in (0,1)$. This is done by approximating fractional Brownian motion by semi-martingales. Then, for $H > 1/6$, they establish an Itô's change of variables formula, which is more precise than Privault's Ito formula (1998). Stochastic calculus for fractional Brownian motion is also treated in (Biagini et al., 2008a). The works of (Duncan et al., 2000), a stochastic calculus is given for the fractional Brownian motions that have the Hurst parameter in $(1/2,1)$. A stochastic integral of Itô type is defined for a family of integrands so that the integral has zero mean and an explicit expression for the second moment. This integral uses the Wick product and a derivative in the path space. Some Itô formulae (or change of variables formulae) are given for smooth functions of a fractional Brownian

motion or some processes related to a fractional Brownian motion. A stochastic integral of Stratonovich type is defined and the two types of stochastic integrals are explicitly related. A square integrable functional of a fractional Brownian motion is expressed as an infinite series of orthogonal multiple integrals. The book (Coutin, 2007) offers an introduction to stochastic calculus with respect to the fractional Brownian motion. Different approaches have been introduced to construct stochastic integrals with respect to fBm: pathwise techniques, Malliavin calculus, approximation by Riemann sums. In (Nualart, 2006), authors describe these methods and present the corresponding change of variable formulas. Some applications will be discussed. The works of (Decreusefond, 2003), (Decreusefond and Üstünel, 1998), present a new theoretical result on the fractional Brownian motion, including different definitions (and their relationships) of the stochastic integral with respect to this process, Girsanov theorem, Clark representation formula, Itô formula and so on. A stohastic analysis of the fractional Brownian motion is also provided in (Lin, 1995). Integration questions related to fractional Brownian motion are addressed in (Pipiras and Taqqu, 2000).

Overall, Gaussian self-similar processes with stationary increments offer a powerful framework for modeling various aspects of financial markets, ranging from volatility dynamics to risk management and trading strategies. The invariance by time-scaling guaranteed by the property of self-similarity as well as the ability of self-similar processes to capture long-range or short-range dependence makes them indispensable tools for financial analysts and researchers.

In our works, we are interested in the minimal formulation of the stationarity and the self-similarity properties of the squared $\mathbb{L}^2$-norm of $\mathbb{L}^2$-random functions. We prove that the combination of the minimal formulation of the stationarity and of the self-similarity in the $\mathbb{L}^2$-spaces is enough to fully characterize the inner kernel only depending on power functions of exponent $\gamma \in (0,1)$ and to prove the stationarity and the self-similarity of the $\mathbb{L}^2$-inner kernel. Then, we introduce the Gaussian assumption and we get the stationarity and the self-similarity properties not only in distribution but also in trajectory. We provide some examples of stationary and self-similar Gaussian processes called Brownian motion and fractional Brownian motion. In this part we recover the results provided in (Taqqu, 1994), (Mandelbrot and Van Ness, 1968a), (Das and Pan, 2011, Chap. 3), (Embrechts and Maejima, 2000), (Embrechts, 2009), (Samorodnitsky, 2006), (Chaumont, 2006), (Lamperti, 1962). Finally we provide an extension to the multivariate fractional Brownian motion (mfBm) and we recover the results stated in (**?**).

Contributions

Chapter 2 focuses on the effects of time-transformations on families of random variables and random processes in $\mathbb{L}^2$-spaces. The goal of this chapter is to study characteristic properties related to these time-transformations and their consequences on the families of random variables and random processes in $\mathbb{L}^2$-spaces. The two time-transformations of interest are time-origin change or time-translation, and time-scaling, and the related characteristic properties are stationarity and self-similarity, respectively.

In Chapter 2, we consider the $\mathbb{L}^2$-subspaces spanned by families of random variables (r.v) $\{X(\theta), \theta \in J\}$, denoted as $\mathcal{H}_X$. $\mathcal{H}_X$ is the Hilbert subspace of all the finite linear combinations $\alpha \cdot X(\theta) = \sum_1^n \alpha_i X(\theta_i)$ and their limits in $\mathbb{L}^2$; which is characterized by the $\mathbb{L}^2$-characteristics of $\{X(\theta)\}$, given by the *quadratic norm* of the components (that is, the square of the norm) $Q_X(\theta) := \mathbf{Q}(X(\theta)) = \|X(\theta)\|_2^2 = \mathbb{E}(|X(\theta)|^2)$ and by the inner products $K_X(\theta, \theta') = \langle\langle X(\theta), X(\theta')\rangle\rangle := \mathbb{E}(X(\theta)\, X(\theta'))$, a dependency indicator.

Random families indexed by real time $\{X(t), t \in \mathbb{R}\}$ (or random processes) arouse particular interest. Their $\mathbb{L}^2$-characteristics $Q_X(t)$ and $K_X(t, s)$ can be complex to calculate. Therefore, additional properties are introduced to reduce their complexity. If the properties are only true "component by component", (t by t), only the quadratic norm family $\{Q_X(t)\}$ is concerned, and we are discussing the Q_X-property.

Q_X and $\mathbb{L}^2$-Stationarity and Self-similarity of random processes The *stationarity* property is related to the effects of time-origin change on the process, $\{X_h(t) = X(t+h) - X(h), (h \in \mathbb{R})\}$. $\mathbb{L}^2$-*stationarity* expresses that the processes $\{X_h(t)\}$ and $\{X(t)\}$ have the same $\mathbb{L}^2$-characteristics for any h, that is the same quadratic norm $\mathbf{Q}(X(t+h) - X(h)) = Q_X(t)$ and inner kernel $\langle\langle X(t+h) - X(h), X(s+h) - X(h)\rangle\rangle = K_X(t, s)$, whereas the Q_X-*stationarity* expresses only that for any t and h, the random variables $X(t+h) - X(h)$ and $X(t)$ have the same Q_X-norm, that is $\mathbf{Q}(X(t+h) - X(h)) = Q_X(t)$. Fortunately, this weaker hypothesis is sufficient to compute the kernel $K_X(s, t)$, and deduce the $\mathbb{L}^2$-stationarity.

More precisely, $\mathbb{L}^2$-stationarity expresses invariance by time-translation or time-origin change of the $\mathbb{L}^2$-characteristics of the process.

Definition 1.3. (i) *A process $\{X(t)\}$ is said to be $\mathbb{L}^2$-stationary if for any $h \in \mathbb{R}$, the processes $\{X_h(t) = X(t+h) - X(h)\}$ and $\{X(t)\}$ have the same $\mathbb{L}^2$-characteristics, quadratic norm, and inner kernel*

$$\mathbf{Q}[X(t+h) - X(h)] = \mathbf{Q}[X(t)] = Q_X(t), \text{ and } \langle\langle X_h(t), X_h(s)\rangle\rangle = K_X(t, s). \quad (1.43)$$

(ii) *The process $\{X(t)\}$ is said to be Q_X-stationary if only the first condition holds:*

$$\mathbf{Q}(X_h(t)) = \mathbf{Q}(X(t)) = Q_X(t) \quad .$$

The *self-similarity* property is related to a time-scale change in the process such that $\{X^\lambda(t) = \lambda^{-\frac{1}{2}} X(\lambda t), (\lambda > 0)\}$. The $\mathbb{L}^2$-*self-similarity* supposes the existence of a non-negative function $(\Theta(\lambda), \lambda > 0)$ such that the processes $\{X(\lambda t)\}$ and $\{\Theta^{\frac{1}{2}}(\lambda) X(t)\}$ have the same $\mathbb{L}^2$-characteristics, that is, the same quadratic norm $Q_X(\lambda t) = \Theta(\lambda) Q_X(t)$ and the same inner product $\langle\langle X(\lambda t), X(\lambda s)\rangle\rangle = \Theta(\lambda) K_X(t, s)$. The Q_X-self-similarity assumes only that for any t and $\lambda > 0$, $Q_X(\lambda t) = \Theta(\lambda) Q_X(t)$. Thus, these two concepts are not equivalent.

Definition 1.4. *Let $\{\Theta(\lambda), \lambda > 0\}$ be a non-negative non constant function.*
(i) *A process $\{X(t)\}$ is said to be Θ-$\mathbb{L}^2$-self-similar, if for all $\lambda > 0$, the processes $\{X(\lambda t)\}$ and $\{\Theta^{\frac{1}{2}}(\lambda) X(t)\}$ have the same $\mathbb{L}^2$-characteristics, that is,*

$$Q_X(\lambda t) = \Theta(\lambda) Q_X(t), \quad \langle\langle X(\lambda t), X(\lambda s)\rangle\rangle = \Theta(\lambda) \langle\langle X(t), X(s)\rangle\rangle. \quad (1.44)$$

(ii) $\{X(t)\}$ *is said to be* Q_X-(Θ) *self-similar if only* $Q_X(\lambda t) = \Theta(\lambda) Q_X(t), \forall t$.

More precisely, the Q_X-self-similarity establishes a spatial proportionality relationship, between the quadratic norms taken at proportional times $\{Q_X(\lambda t)\}$ and $\{Q_X(t)\}$ with $\lambda > 0$, whose spatial proportionality factor is a function of the time proportionality factor λ such that $\Theta(\lambda) > 0$. The $\mathbb{L}^2$-self-similarity establishes a spatial proportionality relationship, both between the quadratic norms $\{Q_X(\lambda t)\}$ and $\{Q_X(t)\}$, and between the inner kernels $\langle\langle X(\lambda t), X(\lambda s)\rangle\rangle$ and $\langle\langle X(t), X(s)\rangle\rangle$, taken at proportional times, whose spatial proportionality factor is a function of the time proportionality factor λ such that $\Theta(\lambda) > 0$.

With the only Q_X-self-similarity assumption, the form of function Θ can be determined, and the quadratic norm is characterized.

Proposition 1.5. *A family* $\{X(t)\}$*, with a right-continuous (rc) quadratic norm* $\{Q_X(t)\}$ *is self-similar in quadratic norm* **only** *if the function* Q_X *is a power function, with a positive exponent* γ*,* $Q_X(t) = |t|^{2\gamma} Q_X(1)$ *if* $t > 0$*,* $Q_X(t) = \left|t\right|^{2\gamma} Q_X(-1)$ *if* $t < 0$*, and* $Q_X(0) = 0$*. In other words, for all* $t \in \mathbb{R}$ $Q_X(t) = \left|t\right|^{2\gamma} Q_X(sgn(t))$ *where* $sgn(t) = 1$ *if* $t > 0$*,* $sgn(t) = -1$ *if* $t < 0$*, and* $sgn(t) = 0$ *if* $t = 0$*.*

It can be observed that the Q_X-self-similarity property cannot be extended to the inner kernel without an additional assumption.

In Chapter 2, we place at the scale of the Hilbert spaces and prove that without any assumption of distribution, the only assumptions of Q_X-stationarity and Q_X-self-similarity are sufficient to fully characterize the covariance kernel and to show that this covariance kernel is itself stationary and self-similar, that is, to obtain the $\mathbb{L}^2$-stationarity and the $\mathbb{L}^2$-self-similarity.

Combining the self-similarity and stationarity of the quadratic norm implies restricting the exponent γ of the power function $Q_X(t)$ to the interval $(0, 1)$, owing to Q_X-sub-linearity induced by the stationarity.

Theorem 1.6. (i) *A necessary and sufficient condition for an* $\mathbb{L}^2$*-rc* $\{X(t)\}$ *to be* $\mathbb{L}^2$*-self-similar and stationary, is the existence of a power function, with exponent* $0 < \gamma < 1$ *such that the* $\mathbb{L}^2$*-characteristics are,*

$$Q_X(t) = Q_X(-t) = Q_X(1) \left|t\right|^{2\gamma} \quad \text{and} \quad K_X(t,s) = \frac{Q_X(1)}{2}(|t|^{2\gamma} + |s|^{2\gamma} - |t-s|^{2\gamma}). \tag{1.45}$$

(ii) *The condition* $\gamma = 1/2$*, is equivalent to the orthogonality of the increments, defined on the disjoint intervals. In this case* $K_X(t,s) = Q_X(1)(\left|t\right| \wedge |s|)$*.*

> Remark: – The inner kernel is both stationary from Proposition **??**, and γ-self-similar, which is obvious in Equation (1.45).

Stationarity and self-similarity in Gaussian Hilbert space Adding the Gaussian assumption to the stationarity and self-similarity properties leads to express these properties first in distribution, then in trajectory.

Definition 1.7. *The Gaussian process* $\{X(t)\}$ *is* γ*-self-similar and stationary in distribution if and only if the two Gaussian processes* ($\{X(t)\}$ *and* $\{X_h(t) = X(t+h) - X(t)\}, h \in \mathbb{R}$)*, respectively* ($\{X(\lambda t)\}$ *and* $\{\lambda^\gamma X(t)\}, \lambda > 0$) *have the same Gaussian distribution, that is, the same mean and inner kernel.*

In this framework, we prove that the only assumptions of Q_X-stationarity and Q_X-self-similarity allow us to obtain the stationarity and self-similarity in distribution.

A Gaussian process stationary and self-similar in distribution is obviously an $\mathbb{L}^2$-γ-self-similar and stationary process. Thus, all the previous results hold true. In particular, its quadratic norm is the power function $Q_X(1)|t|^{2\gamma}$ with $0 < \gamma < 1$, and the inner kernel is completely specified. The only unknown property concerns the mean, which, by the stationarity property verifies $m_X(t+h) = m_X(t) + m_X(h)$. Therefore, the mean is a linear function $m_X(t) = m_X(1)t$. However, based on the self-similarity property, $m_X(\lambda) = \lambda^\gamma m_X(1)$. These two equations are contradictory, if $\gamma \neq 1$. But $\gamma \in (0,1)$ Subsequently, $m_X(t) = 0$.

Theorem 1.8. *A Gaussian process* $\{X(t)\}$ *is* γ*-self-similar and stationary in* **distribution***, if and only if* $\{X(t)\}$ *is a centered process, and is* γ*-self-similar and stationary in the* $\mathbb{L}^2$*-sense. The quadratic norm* $Q_X(t)$ *is the variance* $V_X(t) = V_X(1)|t|^{2\gamma}, \gamma \in (0,1)$.

(ii) *If* $\gamma = \frac{1}{2}$*, the self-similar and stationary Gaussian process* $\{X(t)\}$ *has independent increments and a linear variance* $V_B(t) = V_B(1)|t|$*. This process is called Brownian motion or Wiener process. In the following, this is denoted as* $\{B(t)\}$.

(iii) *For any* $\lambda > 0$ *and* $h \in \mathbb{R}$*, the processes* $\{B_h(t) = B(t+h) - B(h)\}$, $\{B^\lambda(t) = \lambda^{-1/2}B(\lambda t)\}$ *and* $\{B_h^\lambda(t) = \lambda^{-1/2}(B(\lambda(t+h)) - B(\lambda h)\}$ *are also Brownian motions.*

Some well-known examples We provide some well-known examples of self-similar Gaussian processes with stationary increments: the Brownian motion whose increments are independent, and the Path-Dependent Brownian Motion (PDBM), which is generated by a stochastic integral of a deterministic bivariate kernel against Brownian motion, and whose increments are correlated.

Brownian motion The Gaussian process with variance $|t|$ is known as Brownian motion (indexed by $\mathbb{R}$ and not $\mathbb{R}^+$ as usual) and is denoted $\{B(t)\}$. It is the best-known process of the family $\mathcal{G}_{ss}$ of self-similar and stationary Gaussian processes.

Theorem 1.9. *Let* $\{B(t)\}$ *be standard Brownian motion.*

(i) *The* $\mathbb{L}^2$*-space spanned by the Brownian motion* $\{B(t)\}$ *is the family of Gaussian stochastic integrals* $B(\phi) := \int \phi(u)\mathrm{d}B(u)$ *of deterministic functions* ϕ *in* $\mathbb{L}^2(Leb)$*, with variance* $\mathbb{V}[B(\phi)] = \int |\phi|^2 du$.

(ii) *The covariance of two stochastic integrals is* $\mathrm{Cov}\left(B(\phi), B(\psi)\right) = \int_{\mathbb{R}} \phi(u)\psi(u)\mathrm{d}u$.

Therefore, there exists an isometry between $\mathcal{H}_B$ *and* $\mathbb{L}^2(Leb)$.

We prove the stationarity and self-similarity of Brownian motion in a trajectory sense. This result is much stronger than the previous one because we obtain an equality of processes instead of a result in distribution.

Because they have the same Q_X-norms, the processes $\{B_h(t) = B(t+h) - B(t),\, h \in \mathbb{R}\}$ and $\{B^\lambda(t) = \frac{1}{\sqrt{\lambda}} B(\lambda t), \lambda > 0\}$, also belonging to $\mathcal{G}_{ss}$, are Brownian motions. The tool of stochastic integrals of deterministic functions $\phi \in \mathbb{L}^2(\text{Leb})$, $B(\phi) = \int \phi(u) dB(u)$ allows describing the Gaussian space $\mathcal{H}_B$, but also provides a change of variable formula for the pathwise comparison between the random variables $B(\phi), B_h(\phi), B^\lambda(\phi)$.

Theorem 1.10. *Recall that for all $\lambda > 0$, $h \in \mathbb{R}$ and $t \in \mathbb{R}$, $\{B_t\}$, $\{B^\lambda(t) = \lambda^{-\frac{1}{2}} B(\lambda t)\}$, and $\{B_h(t) = B(t+h) - B(h)\}$ are Brownian motions.*
For any $\phi \in \mathbb{L}^2(Leb)$, we have the following pathwise representations,

$$\int \lambda^{-\frac{1}{2}} \phi\left(\frac{u}{\lambda}\right) \mathrm{d}B(u) = \int \phi(u) \mathrm{d}B^\lambda(u) \text{ and } \int \phi\left(u-h\right) \mathrm{d}B(u) = \int \phi(u) \mathrm{d}B_h(u).$$

All these variables are centered Gaussian variables with the same variance $\int |\phi(u)|^2 du$.

Fractional Brownian motion (fBm)

In a second part, we are concerned with the construction of $\mathcal{G}_{ss}$-processes as stochastic integrals of deterministic functions $\phi(u)$ against Brownian motion, where the deterministic functions $\phi(u)$ are replaced by bivariate kernels $\kappa(t,u)$ such that $\int |\kappa(t,u)|^2 du < \infty$. The covariance kernel is defined as $K_{X^\kappa}(t,s) = \int_{\mathbb{R}} \kappa(t,u)\kappa(s,u) \mathrm{d}u$ and $\kappa(0,u) = 0$. In line with Mandelbrot and Van Ness in 1968 (Mandelbrot and Van Ness, 1968a), a Path Dependent Brownian motion (PDBM), is defined as a Gaussian process, like $X^\kappa(t) = \int \kappa(t,u) dB(u)$.

We are interested in the time-translated and time-scaled PDBM defined for all $\lambda > 0$, $h \in \mathbb{R}$ and almost all $(t,u) \in \mathbb{R}^2$ by:

$$\begin{cases} X^\kappa(t+h, B) - X^\kappa(h, B) = \int (\kappa(t+h,u) - \kappa(h,u)) \mathrm{d}B(u) \\ X^\kappa(\lambda t, B) = \int \kappa(\lambda t, u) \mathrm{d}B(u). \end{cases} \tag{1.46}$$

The variable change formula suggests transporting the time-change initially supported by the time variable t into a change on the integration variable u. Then, we prove pathwise identities on the time-transformed PDBMs.

For all $\lambda > 0$, $h \in \mathbb{R}$ and $t \in \mathbb{R}$, $\{B_t\}$, $\{B_h(t) = B(t+h) - B(h)\}$, and $\{B^\lambda(t) = \lambda^{-\frac{1}{2}} B(\lambda t)\}$ are Brownian motions. We define the square integrable kernels for almost all $(t,u) \in \mathbb{R}^2$ and for all $h \in \mathbb{R}$ and $\lambda > 0$ by:

$$\kappa(t,u), \quad \kappa_h(t,u) = \kappa(t, u-h), \quad \kappa^\lambda(t,u) = \lambda^{-1/2} \kappa(t, u/\lambda). \tag{1.47}$$

Theorem 1.11. *Let us define the PDBMs $\{Y(t) = X^\kappa(t,B)\}, \{Y^\lambda(t) = X^\kappa(t, B^\lambda)\}, \{Y_h(t) = X^\kappa(t, B_h)\}$.*

(i) *Processes $\{Y^\lambda(t)\}$ and $\{Y_h(t)\}$ have the same distribution as $\{Y(t)\}$.*

(ii) *$\{Y^\lambda(t)\}$ and $\{Y_h(t)\}$ satisfy the pathwise identities:*

$$\{Y^\lambda(t) = X^{\kappa^\lambda}(t,B)\} \quad \text{and} \quad \{Y_h(t) = X^{\kappa_h}(t,B)\}. \tag{1.48}$$

Finally, we prove the stationarity and self-similarity of PDBMs in terms of equality of processes, that is, in a trajectory sense.

Theorem 1.12. *Let $\{B(t)\}$, $\{B^\lambda(t)\}$, $\{B_h(t)\}$, be the transformed Brownian motions.*

(i) *If the time-stationarity condition of bivariate kernel $\kappa(t,u)$ is satisfied,*

$$\kappa(t+h,u)-\kappa(h,u)=\kappa_h(t,u)=\kappa(t,u-h) \tag{1.49}$$

then the following pathwise identity holds:

$$\{X^\kappa(t+h,B)-X^\kappa(h,B)=X^\kappa(t,B_h)\} \tag{1.50}$$

and the process $\{X^\kappa(t,B)\}$ is stationary.

(ii) *If the time-self-similarity condition of the bivariate kernel $\kappa(t,u)$ is satisfied,*

$$\kappa(\lambda t,u)=\eta(\lambda)\kappa\left(t,\frac{u}{\lambda}\right)=(\sqrt{\lambda}\eta(\lambda))\kappa^\lambda(t,u) \text{ with } \eta(\lambda)=\lambda^\nu\mathbb{1}_{\{\lambda>0\}},\ \nu\in\left(-\frac{1}{2},\frac{1}{2}\right) \tag{1.51}$$

then the following pathwise identity holds:

$$\{X^\kappa(\lambda t,B)=\eta(\lambda)\sqrt{\lambda}X^\kappa(t,B^\lambda)\}. \tag{1.52}$$

and the process $\{X^\kappa(t,B)\}$ is self-similar.

(iii) *If both the time-stationarity (Equation (1.49)) and time-self-similarity (Equation (1.51))) conditions are satisfied, then the following pathwise identity holds:*

$$\{X^\kappa(\lambda(t+h),B)-X^\kappa(\lambda h,B)=\eta(\lambda)\sqrt{\lambda}X^\kappa(t,B_h^\lambda)\} \tag{1.53}$$

and the process $\{X^\kappa(t,B)\}$ is both stationary and self-similar with a variance-covariance kernel defined as:

$$V_X(t)=V_X(1)\left|t\right|^{2\gamma}\quad\text{and}\quad K_X(t,s)=\frac{V_X(1)}{2}(|t|^{2\gamma}+|s|^{2\gamma}-|t-s|^{2\gamma}) \tag{1.54}$$

where $\gamma\in(0,1)$, $\eta(\lambda)=\lambda^{\gamma-\frac{1}{2}}\mathbb{1}_{\{\lambda>0\}}$.

For the PDBM process to be self-similar and stationary, it suffices that the kernel κ satisfies the system:

$$\begin{cases}\kappa(t+h,u)-\kappa(h,u)=\kappa(t,u-h)\\ \kappa(\lambda t,\lambda u)=\eta(\lambda)\kappa\left(t,u\right)\quad\text{with}\quad\eta(\lambda)=\lambda^\nu\mathbb{1}_{\{\lambda>0\}},\ \nu\in\left(-\frac{1}{2},\frac{1}{2}\right)\end{cases} \tag{1.55}$$

The last step is to find a bivariate kernel $\kappa(t,u)$, solution of the above system, and square integrable in u. The solution proposed by Mandelbrot and Van Ness ((Mandelbrot and Van Ness, 1968a)) is the fractional kernel κ^ν, which is clearly a solution of the previous system.

Proposition 1.13. *A bivariate kernel proposed by Mandelbrot and Van Ness is given by:*

$$\kappa^{\nu}(t,u) = (t-u)^{\nu}\mathbb{1}_{\{u<t\}} - (-u)^{\nu}\mathbb{1}_{\{u<0\}}, \quad \textit{with} \quad \nu = \gamma - \frac{1}{2} \in \left(-\frac{1}{2}, \frac{1}{2}\right). \tag{1.56}$$

(i) *This kernel has some problems of definition when* $u \to 0$, $u \to t$.

(ii) *This kernel satisfies System* (1.55).

(iii) *For all* $\nu \in \left(-\frac{1}{2}, \frac{1}{2}\right)$, $\kappa(t,u) \sim \nu t(-u)^{(\nu-1)}$ *is Lebesgue square integrable when* $u \to -\infty$ *because* $2(\nu-1) < -1$.

For $\nu < 0$, *when* $u \to 0$ *(resp.* $u \to t$*),* $\kappa(t,u) \sim -(-u)^{\nu}$ *(resp.* $\kappa(t,u) \sim (t-u)^{\nu}$*) which is Lebesgue square-integrable because* $2\nu + 1 > 0$.

Then $\kappa(t,u)$ *is Lebesgue u-square integrable.*

The square integrability verifies that:

– for any value of ν, for u in the neighborhood of $-\infty$ $\int \left|\kappa(t,u)\right|^2 \mathrm{d}u < \infty$,

– for $\nu < 0$, for u in the neighborhood of 0 and for u in the neighborhood of t, $\int \left|\kappa(t,u)\right|^2 \mathrm{d}u < \infty$.

Some remarks guide intuition on obtaining such a solution.

Extension to the multidimensional framework Finally, an extension to the multidimensional framework is provided, first for the d-dimensional Brownian motion whose components are correlated Brownian motions with covariance matrix of $(\rho_{i,j}(t \wedge s))_{i,j \in [\![1,d]\!]^2}$, and then for multivariate fractional Brownian motion (mfBm).

Definition 1.14. $\{\boldsymbol{X}^{\boldsymbol{\kappa}}(t,B) = (X^{\kappa^1}(t,B^1), X^{\kappa^2}(t,B^2), \ldots, X^{\kappa^d}(t,B^d))\}$*, called* d*-dimensional fractional Brownian motion, is a multidimensional Gaussian process defined as the stochastic integral of a vector of deterministic bivariate kernels against a vector of linear transformations of the Wiener process, such that for almost all* $(t,u) \in \mathbb{R}^2$*:*

$$\boldsymbol{X}^{\boldsymbol{\kappa}}(t,B) = \int \boldsymbol{\kappa}(t,u) \odot \mathrm{d}\boldsymbol{B}(u) := \left(\int \kappa^i(t,u)\mathrm{d}B^i(u)\right)_{i \in [\![1,d]\!]} \tag{1.57}$$

where $\odot$ *is the element-wise product,* $\boldsymbol{\kappa}(t,u) = (\kappa^i(t,u))_{i \in [\![1,d]\!]}$ *is the vector of kernels proposed by Mandelbrot and Van Ness in Equation* (1.56)*,* $\boldsymbol{B}(t) = (B^i(t))_{i \in [\![1,d]\!]}$ *and its covariance kernel is given by:*

$$K_{X^{\kappa}}((i,t),(j,s)) = \mathrm{Cov}\left(X^{\kappa^i}(t,B^i), X^{\kappa^j}(s,B^j)\right) = \int \kappa^i(t,u)\kappa^j(s,u)\rho_{i,j}\mathrm{d}u. \tag{1.58}$$

Each coordinate of vector $\boldsymbol{\kappa}(t,u)$ satisfies the conditions of stationarity and self-similarity in the univariate framework given in Equation (1.55). These properties are then obtained in the vector sense, that is, for all $\lambda > 0$, $h \in \mathbb{R}$ and for almost all $(t,u) \in \mathbb{R}^2$:

$$\begin{cases} \boldsymbol{\kappa}(\lambda t, \lambda u) = \boldsymbol{\eta}(\lambda)\boldsymbol{\kappa}\left(t,u\right) \quad \text{with} \quad \boldsymbol{\eta}(\lambda) = (\lambda^{\nu_i}\mathbb{1}_{\{\lambda>0\}})_{i\in[\![1,d]\!]},\ \nu_i \in \left(-\frac{1}{2},\frac{1}{2}\right) \\ \boldsymbol{\kappa}(t+h,u) - \boldsymbol{\kappa}(h,u) = \boldsymbol{\kappa}(t,u-h). \end{cases} \tag{1.59}$$

We prove that mfBm is a self-similar multidimensional Gaussian process with stationary and correlated increments.

Theorem 1.15. *Let $\{\boldsymbol{X}^{\boldsymbol{\kappa}}(t,B)\}$ be a d-dimensional fractional Brownian motion. For all $\lambda > 0$, $h \in \mathbb{R}$ and for almost all $(t,u) \in \mathbb{R}^2$:*

$$\begin{cases} \boldsymbol{X}^{\boldsymbol{\kappa}}(t+h) - \boldsymbol{X}^{\boldsymbol{\kappa}}(h) = \int \boldsymbol{\kappa}(t,u) \odot \mathrm{d}\boldsymbol{B}_{\boldsymbol{h}}(u) = \boldsymbol{X}^{\boldsymbol{\kappa}}(t,B_h) \\ \boldsymbol{X}^{\boldsymbol{\kappa}}(\lambda t) = \sqrt{\lambda}\boldsymbol{\eta}(\lambda) \odot \int \boldsymbol{\kappa}(t,u) \odot \mathrm{d}\boldsymbol{B}^{\boldsymbol{\lambda}}(u) = \sqrt{\lambda}\boldsymbol{\eta}(\lambda) \odot \boldsymbol{X}^{\boldsymbol{\kappa}}(t,B^{\lambda}). \end{cases} \tag{1.60}$$

The vector process $\{\boldsymbol{X}^{\boldsymbol{\kappa}}(t,B)\}$ is stationary and self-similar in the trajectory sense.

Moreover, its covariance kernel is completely characterized by power functions whose exponents depend on the spatial components.

Theorem 1.16. *Let $\{\boldsymbol{X}^{\boldsymbol{\kappa}}(t,B)\}$ be a d-dimensional fractional Brownian motion.*

(i) *The covariance kernel between two distinct coordinates taken at the same time $t \in \mathbb{R}\backslash\{0\}$ is completely characterized by the power function of exponent $\gamma_i+\gamma_j$ where $\gamma_i, \gamma_j \in (0,1)$:*

$$K_{X^{\kappa}}((i,t),(j,t)) = K_{X^{\kappa}}((i,1),(j,1))\left|t\right|^{\gamma^i+\gamma^j}. \tag{1.61}$$

(ii) *The symmetrized covariance kernel between two distinct coordinates taken at two distinct times $X^{\kappa_i}(t,B^i)$ and $X^{\kappa_j}(s,B^j)$ with $t,s \in \mathbb{R}\backslash\{0\}$ is completely characterized by a linear combination of power functions of exponent $\gamma_i+\gamma_j$ where $\gamma_i, \gamma_j \in (0,1)$:*

$$K_{X^{\kappa}}((i,t),(j,s)) + K_{X^{\kappa}}((i,s),(j,t)) = K_{X^{\kappa}}((i,1),(j,1))\left(\left|t\right|^{\gamma^i+\gamma^j} + \left|s\right|^{\gamma^i+\gamma^j} - \left|t-s\right|^{\gamma^i+\gamma^j}\right). \tag{1.62}$$

Change of notations: In Chapter 3, the fractional Brownian motion $\{X^{\kappa^i}(t,B^i)\}$ is denoted as $\{B^{H_i}(t)\}$, the exponent γ_i of the power function corresponds to the Hurst H_i exponent, and the bivariate kernel $\kappa^i(t,u)$ is replaced by $\psi^{H_i}_{0,t}(u)$.

1.3.2 Chapter 3 - VaR prediction for asset portfolios under fractional dynamics

State of the art

Value at risk (VaR) is a statistic that quantifies the extent of possible financial losses within a firm, portfolio, or position over a specific time frame. This metric is most commonly used by investment and commercial banks to determine the extent and prob-

abilities of potential losses in their institutional portfolios. Risk managers use VaR to measure and control the level of risk exposure. One can apply VaR calculations to specific positions or whole portfolios or use them to measure firm-wide risk exposure. Value-at-Risk (**VaR**) is the risk measure the most currently used by the regulators. The VaR at the risk level $\alpha \in (0,1)$ corresponds to the quantile at the order α of the $P\&L$ distribution.

The literature on VaR is huge, there are more than 2700 articles referenced on Google scholar. VaR has been studied in details in the well-known books of (Wipplinger, 2007, Part II. p105) and (McNeil et al., 2015, Chap.2, p37). Computing the VaR is a challenging problem as it requires knowing the distribution of the portfolio P&L, which is usually unknown. A wide range of papers were interested in the prediction of the VaR. To name a few, see for instance (Cheung and Powell, 2012) which presents a teaching study using parametric computation and Monte-Carlo simulation to compute VaR, or (Feuerverger and Wong, 2000) that explains the computation of VaR for nonlinear portfolios. Although VaR is not a perfect risk metrics (see (Delbaen et al., 1998)), it is still commonly used. It is fundamental to have an accurate estimate of the VaR: in case of under-estimation, the financial institution will take too much risk, without being prepared for this; in case of over-estimation, the amount of money to be kept aside would be too high, preventing some banking activities. In addition, its computation has to be efficient, because its evaluation is frequent and has to be made according to numerous portfolios. In (Gaivoronski and Pflug, 2005), a method of calculating the portfolio which gives the smallest V@R among those, which yield at least some specified expected return is presented. Using this approach, the complete mean-VaR efficient frontier may be calculated. The method is based on approximating the historic VaR by a smoothed V@R (SVaR) which filters out local irregularities. Value-at-Risk (VaR) and Conditional-Value-at-Risk (CVaR) are two widely-used measures in risk management. The works of (Bardou et al., 2009) deal with the problem of estimating both VaR and CVaR using stochastic approximation (with decreasing steps): a first Robbins-Monro (RM) procedure based on Rockafellar-Uryasev's identity is proposed for the CVaR. The estimator provided by the algorithm satisfies a Gaussian Central Limit Theorem. As a second step, in order to speed up the initial procedure, a recursive and adaptive importance sampling (IS) procedure which induces a significant variance reduction of both VaR and CVaR procedures is presented. In (Manganelli and Engle, 2001), a survey and an evaluation of the performance of the most popular univariate VaR methodologies is carried out paying particular attention to their underlying assumptions and to their logical flaws. The works of (Jorion, 1996) lay out the statistical methodology for analyzing estimation error in VAR and shows how to improve the accuracy of VAR estimates. Regulators require banks to employ value-at-risk (VaR) to estimate the exposure of their trading portfolios to market risk, in order to establish capital requirements.

However, portfolio-level VaR analysis is a high-dimensional problem and hence computationally intensive. The works of (Albanese et al., 2002) present two new portfolio-based approaches to reducing the dimensionality of the VaR analysis. A classical approach to predict VaR is the weighting moving average model, as presented in (Hendricks, 1996). The works of (Lucas and Zhang, 2016) present a simple methodology for model-

ing the time variation in volatilities and other higher-order moments using a recursive updating scheme that is similar to the familiar RiskMetrics approach. In (Gabrielsen et al., 2015), author provide an insight to the time-varying dynamics of the shape of the distribution of financial return series by proposing an exponential weighted moving average model that jointly estimates volatility, skewness and kurtosis over time using a modified form of the Gram-Charlier density in which skewness and kurtosis appear directly in the functional form of this density. In this setting VaR can be described as a function of the time-varying higher moments by applying the Cornish-Fisher expansion series of the first four moments. The paper (Alexander and Dakos, 2023) discusses the Value-at-Risk Contribution under the asset liability model using the EWMA approach. It is assumed that asset returns and liabilities are time series data following the Exponential Weighted Moving Average (EWMA) model. Return of surplus which is the difference between asset return and liability is analyzed using asset liability model. In this case the risk of surplus return is measured using the Value-at-Risk model. When investments are made on multiple assets, each asset will contribute to the establishment of Value-at-Risk from the investment portfolio, which can be measured using the Value-at-Risk Contribution model. Using Value-at-Risk Contribution, it can be seen how much Value-at-Risk surplus investment portfolio, and what is the proportion of Value-at-Risk contribution of each surplus of investment asset. Based on the calculation of Value-at-Risk Contribution, can be considered for investors in investing in some assets analyzed. The works of (Sukono et al., 2018) derive a combined and dynamic hedging model- exponentially weighted moving average-generalized autoregressive conditional heteroskedasticity (GARCH)(1,1)-M applicable to the real financial markets based on previous studies. The results in this paper turn out that the model built is not only excellent for the pursuit for the minimum VaR but also practical for the actual situation where the variances of financial price data are time-varying. The paper (Anantafortuna and Anggono, 2019), intends to use Value at Risk as a tool for risk assessment for trading activities based on Exponential Moving Average and Count Back Line strategy. This paper also aims to inform the level of risk investors take with a certain investment by calculating the maximum potential loss that occur every day and assess whether trading activities were performed within or beyond Value at Risk calculated. Empirical studies have shown that a large number of financial asset returns exhibit long-range dependence. Moreover, some long-memory or long-range dependence can also be observed as a stylized fact in market volatility, with significant impact on pricing and forecasting of market volatility. Since risk-measures are based on underlying asset returns dynamics, then in such cases, the classical methodology used to predict risk measures present bad performance. To address such an issue, auto-regressive models such as ARCH or GARCH have been used to predict VaR. In (Anantafortuna and Anggono, 2019), authors show that, on the one hand, models that accomodate long memory hold the promise of improved long-run volatility forecast as well as accurate pricing of long-term contracts. On the other hand, recent focus is on whether long memory can affect the measurement of market risk in the context of Value-at-Risk (VaR). In this paper, author evaluate the Value-at-Risk (VaR) and Expected Shortfall (ESF) in financial markets under such conditions. Classical VaR estimation methodology such as exponential moving average (EMA) as well as extension to cases where long

memory is an inherent characteristics of the system are investigated. In particular, two long memory models are estimated, the Fractional Integrated Asymmetric Power-ARCH and the Hyperbolic-GARCH with different error distribution assumptions. These models exhibit better performances in predicting VaR as the classical approaches that don't take into account long-range dependence. In (Aloui, 2008), the long-range memory is explored on energy markets volatility and value-at-risk (VaR). The main question is: can we estimate better the VaR if long memory exists? To investigate this question several GARCH-type processes, including the FIGARCH process, have been implemented to some main energy products'daily prices (January 1986 to July 2007). Value- at-risk was estimated for both the short and the long trading positions and at various confidence levels. Normal, Student and skewed Student distributions are suggested for divers GARCH-type processes. The GARCH-type VaR performance is assessed by estimating the failure rate of the Kupiec test statistic. Consistent with previous studies, our results show that energy price volatility exhibits a long-range memory. The VaR computed through a skewed Student-t FIGARCH process provides the best performance for both the short and the long trading positions. The works of (Youssef et al., 2015) present three long-memory-models including, FIGARCH, HYGARCH and FIAPARCH to forecast energy commodity volatility by capturing some volatility stylized fact such as long-range memory, heteroscedasticity, asymmetry and fat-tails. The paper (Tang and Shieh, 2006) investigate the long memory properties for closing prices of three stock index futures markets. The FIGARCH (1, d, 1) and HYGARCH (1, d, 1) models with normal, Student-t, and skewed Student-t distributions for S&P500, Nasdaq100, and Dow Jones daily prices are estimated first. Then the value-at-risks are calculated by the estimated models. The empirical results show that for the three stock index futures, the HYGARCH (1, d, 1) models with skewed Student-t distribution perform better based on the Kupiec LR tests. In particular, for the S&P500 and Nasdaq100 futures prices.

Contributions

Chapter 3 focuses on the construction of an accurate predictive model for the Value-at-Risk (VaR) considering long-range dependence (respectively short-range dependence). Auto-regressive models have already been developed to fulfill this task; however, such models depend on a large number of parameters that have to be estimated. Our works propose another approach capable of capturing long-range dependence (respectively short-range dependence) owing to Gaussian self-similar processes with stationary increments, called fractional Brownian motions (fBm). We consider a setting in which the price dynamics are described by fractional Black-Scholes models, depending on fractional Brownian motions parameterized by their Hurst exponent. These models have the advantage of capturing correlations in time and between assets. Compared to the usual Black-Scholes model, the flexibility in the choice of the Hurst exponent allows for a better description of price trajectories to fit reality. Thus, the predictive model for the conditional VaR at time horizon h (that is, the VaR of P&L over the next period of length h conditionally to the observations available at the computation time) will be able to predict, in a more accurate way, the amount of money to be kept aside, that

is, neither over-charge nor under-charge the regulator. The accuracy of our VaR predictions is assessed using a backtesting procedure based on (Christoffersen, 1998). The use of fractional models in finance is not new, and has become increasingly important in recent years. Fractional models have been widely used for different purposes; however, in the above-mentioned studies, the prediction of conditional VaR has not been addressed.

Overall, we make the following contributions. We design a theoretical framework for fractional models in Gaussian Hilbert space. We design a closed-form formula for the Gaussian approximation of the conditional $\mathbf{VaR}_\alpha$, at the time horizon h, of the future portfolio variation under fractional dynamics. We then provide error quantification of the Gaussian approximation. We use a robust methodology to estimate the parameters of the fractional model. We perform a backtesting procedure to assess the accuracy of our predictive approximation. We support our analysis through various experiments that illustrate the behavior of our new model.

First, we establish a theoretical Gaussian framework without any application in the financial field. In this part, we prove the specific form of the conditional expectation and variance of a linear combination of future increments of several correlated fBm, given past increments.

Theoretical framework There is no unique way to model multivariate fractional Brownian motions: the most direct way is to give their covariance function as a multivariate Gaussian process; however, this method is quite inconvenient when dealing with conditional in time computations, with various filtrations to account for the available observations. Therefore, we prefer working directly at the level of the Gaussian Hilbert space indexed by functions in

$$\mathcal{L}^2_d := L_2(\mathbb{R}^d, \mathrm{d}u) = \left\{ f : \mathbb{R} \mapsto \mathbb{R}^d \ s.t. \ |f|^2_{\mathcal{L}^2_d} := \int_{\mathbb{R}} |f(u)|^2 \mathrm{d}u < +\infty \right\} \tag{1.63}$$

where basic quantities are defined by Wiener integrals and, projections and conditional expectations are made through projections of functions in $\mathcal{L}^2_d$.

We adopt the framework of the isonormal Gaussian process associated with the Hilbert space $\mathcal{L}^2_d$, with the scalar product

$$\langle f, g \rangle_{\mathcal{L}^2_d} = \int_{\mathbb{R}} f(u) \cdot g(u) \mathrm{d}u, \tag{1.64}$$

also called Gaussian Hilbert space; here $f(u) \cdot g(u)$ is just the scalar product in $\mathbb{R}^d$ between the vector-valued functions f and g at point u. See (Janson, 1997) for a broad account on Gaussian Hilbert space. Namely, we consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ supporting a centered Gaussian family of scalar random variables

$$\mathcal{H} = \{I(f) : f \in \mathcal{L}^2_d\} \tag{1.65}$$

as defined in (Janson, 1997, Chapter 7, Thm.7.1), such that the isometry property is in force

$$\mathbb{E}\left[I(f)I(g)\right] = \langle f, g \rangle_{\mathcal{L}^2_d}. \tag{1.66}$$

One way to construct this Gaussian Hilbert space indexed by square integrable functions $f \in \mathcal{L}^2_d$ is to assume that the probability space supports a d-dimensional Brownian motion indexed by the real line, $(W_t = (W_t^1, \cdots, W_t^d)^\top : t \in \mathbb{R})$, and to define $I(f)$ by stochastic integration

$$I(f) = \int_{\mathbb{R}} f(u) \cdot \mathrm{d}W_u = \sum_{i=1}^{d} \int_{\mathbb{R}} f^i(u) \mathrm{d}W_u^i \tag{1.67}$$

(a.k.a. Wiener integral). See (Janson, 1997, Chapter 7) where integrals are restricted to $\mathbb{R}^+$, extension to $\mathbb{R}$ is immediate. As shown in Chapter 2, fractional Brownian motion (fBm) with Hurst parameter H corresponds to a specific choice of function f: which is the well-known time-representation of Mandelbrot and Van Ness (Mandelbrot and Van Ness, 1968b).

Definition-Proposition 1. *Let $H \in (0,1)$ and set, for any $t \geq 0$,*

$$B_t^H = I((\psi_{0,t}^H(.), 0, \cdots, 0)^\top) = \int_{\mathbb{R}} \psi_{0,t}^H(u) \mathrm{d}W_u^1 \quad \textit{with} \quad \psi_{s,t}^H(u) = \frac{1}{c_H} \left((t-u)_+^{H-\frac{1}{2}} - (s-u)_+^{H-\frac{1}{2}} \right), \tag{1.68}$$

where

$$c_H := \sqrt{\frac{1}{2H} \left(\frac{3}{2} - H \right) \mathfrak{B} \left(2 - 2H, H + \frac{1}{2} \right)} \tag{1.69}$$

and $\mathfrak{B}(x,y) = \int_0^1 u^{x-1}(1-u)^{y-1}\mathrm{d}u$ is the Beta function.

This defines a scalar fBm B^H, that is, a centered Gaussian process with covariance

$$\mathrm{Cov}\left(B_t^H, B_s^H \right) = \frac{1}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H}), \quad H \in (0,1). \tag{1.70}$$

FBm enjoys a self-similarity property. The self-similarity property states that for any fixed $\lambda > 0$, the process $\{\lambda^{-H} B_{\lambda t}^H\}_{t \in \mathbb{R}}$ is also an fBm with Hurst exponent $H \in (0,1)$. This is equivalent to say that for any fixed $\lambda > 0$, the processes $\{\lambda^{-H} B_{\lambda t}^H\}_{t \in \mathbb{R}}$ and $B_t^H\}_{t \in \mathbb{R}}$ have the same distribution. In other words, self-similarity property establishes a spatial proportionality relationship of factor λ^H, between the distributions of $\{B_{\lambda t}^H\}$ and $\{B_t^H\}$. This leads to obtain spatial proportionality relationships of factor λ^{2H} between the variances, respectively the covariance kernels, of the processes $\{B_{\lambda t}^H\}$ and $\{B_t^H\}$. This type of scaling property will play a role in the simplification of some subsequent formulae, as it allows changes in time and space scales in the processes. The self-similarity property is, for example, useful for calculating the covariance function of the fBm because it allows this covariance to be expressed using only power functions of exponent $H \in (0,1)$ (self-similarity functions), and the covariance at time 1.

We define the multivariate fractional Brownian motion (mfBm) whose components are correlated. To allow correlations between fBms, we add correlations to the driving Brownian motion through a linear transformation of the Wiener process W. This works as follows: let $C = (\rho_{i,j})_{1 \leq i,j \leq d}$ be a correlation matrix, and let R be a symmetric square root of C (which exists because C is non-negative symmetric), so that

$$\rho_{i,j} = R_{:,i} \cdot R_{:,j} \tag{1.71}$$

where $R_{:,i}$ is the i-th column of R. Now, we define a multivariate fractional Brownian motion, with correlated components, each having its own Hurst parameter.

Definition-Proposition 2. *Let $H_1, \cdots, H_d$ be a sequence of Hurst exponents in $(0,1)$ and set, for any $t \geq 0$ and any $i \in \{1, \cdots, d\}$,*

$$B_t^{i,H_i} := I(\psi_{0,t}^{H_i} R_{:,i}) = \int_{\mathbb{R}} \psi_{0,t}^{H_i}(u) R_{:,i} \cdot \mathrm{d}W_u, \tag{1.72}$$

where $R_{:,i}$ is the i-th column of a matrix $R \in \mathbb{R}^d$. Then, this defines a multivariate fractional Brownian motion (mfBm) with parameter $H = (H_i)_i \in (0,1)^d$ whose coordinates (B_t^{i,H_i}) are fBms with Hurst exponent H_i, and are correlated in a way that:

$$\forall s,t \in \mathbb{R}, \quad \mathrm{Cov}(B_t^{i,H_i}, B_s^{j,H_j}) + \mathrm{Cov}(B_t^{j,H_j}, B_s^{i,H_i}) \tag{1.73}$$

$$= \mathrm{Cov}(B_1^{i,H_i}, B_1^{j,H_j}) \Big(|t|^{H_i+H_j} + |s|^{H_i+H_j} - |t-s|^{H_i+H_j} \Big), \tag{1.74}$$

where $\mathrm{Cov}(B_1^{i,H_i}, B_1^{j,H_j})$ is explicitly given in Lemma 3.1. Observe that

$$B_t^i := R_{:,i} \cdot W_t, \quad i \in \{1, \cdots, d\} \tag{1.75}$$

defines a d-dimensional Brownian motion with correlation C.

In what follows, we typically need to compute the conditional expectation and variance of some stochastic integral $I(f)$ conditionally to the fBm increments on some time grid. A general result on how to compute $\mathbb{E}\Big[I(f) \mid \{I(f_l)\}_{(l \in \mathcal{I}}\Big]$ and $\mathbb{V}\Big[I(f) \mid \{I(f_l)\}_{(l \in \mathcal{I}}\Big]$, where $(f_l)_{l \in \mathcal{I}}$ is an arbitrary family of functions in $\mathcal{L}_d^2$, $\mathcal{L}'$ is the closed subspace of $\mathcal{L}_d^2$ spanned by $(f_l)_{l \in \mathcal{I}}$ and $f^\star \in \mathcal{L}'$ is the (unique) orthogonal projection of f on $\mathcal{L}'$, is provided by:

$$\mathbb{E}\Big[I(f) \mid \{I(f_l)\}_{(l \in \mathcal{I}}\Big] = I(f^\star) \quad \text{and} \quad \mathbb{V}\Big[I(f) \mid \{I(f_l)\}_{l \in \mathcal{I}}\Big] = \Big| f - f^\star \Big|_{\mathcal{L}_d^2}^2. \tag{1.76}$$

Applying the above formulae to the specific framework of fractional Brownian motion, we obtain the following result.

Theorem 1.17. *Consider the index family $\mathcal{I} = \{(i,l) \mid i \in [\![1,d]\!],\ l \in [\![1,N]\!]\}$ and the set of $\mathcal{L}_d^2$-valued functions*

$$f_l^i(\cdot) := \sigma_i R_{:,i} \psi_{t_l,t_{l+1}}^{H_i}(\cdot). \tag{1.77}$$

Let

$$f(\cdot) = \sum_{i=1}^{d} \omega_{t_N}^i \psi_{t_N,t_{N+1}}^{H_i}(\cdot) \tag{1.78}$$

be an $\mathcal{L}_d^2$ function, where each d-dimensional coefficient $\omega_{t_N}^i = \omega_i \sigma_i S_{t_N}^i R_{:,i}$ is measurable with respect to $\{I(f_l^i)\}_{(i,l) \in \mathcal{I}}$.

Then, the expectation of $I(f)$ conditional on $\{I(f_l^i)\}_{(i,l)\in\mathcal{I}}$ can be written as

$$\mathbb{E}\left[I(f)|\{I(f_l^i)\}_{(i,l)\in\mathcal{I}}\right] = \sum_{(i,l)\in\mathcal{I}} a_l^i I(f_l^i) = \sum_{i=1}^{d}\sum_{l=1}^{N} a_l^i (B_{t_{l+1}}^{H_i} - B_{t_l}^{H_i}), \tag{1.79}$$

and the conditional variance is given by:

$$\mathbb{V}\left[I(f) \mid \{I(f_l^i)\}_{(i,l)\in\mathcal{I}}\right] = \sum_{i,j=1}^{d} (\omega_{t_N}^i \cdot \omega_{t_N}^j) \int_{\mathbb{R}} \psi_{t_N,t_{N+1}}^{H_i}(u)\psi_{t_N,t_{N+1}}^{H_j}(u)\mathrm{d}u \tag{1.80}$$

$$- \sum_{i,j=1}^{N}\sum_{k,l=1}^{N} a_l^i a_k^j \rho_{ij} \int_{\mathbb{R}} \psi_{t_l,t_{l+1}}^{H_i}(u)\psi_{t_k,t_{k+1}}^{H_j}(u)\mathrm{d}u, \tag{1.81}$$

where the projection coefficients $\alpha^\star = (a_1^1,\cdots,a_N^1,a_1^2,\cdots,a_N^2,\cdots,a_1^d,\cdots,a_N^d)^\top$ are equal to:

$$\alpha^\star = \mathfrak{M}^+ y, \tag{1.82}$$

where $\mathfrak{M}^+$ is the Moore-Penrose pseudo-inverse (see (MacAusland, 2014, Section 5.5.4)) of the matrix

$$\mathfrak{M} = \left(\langle f_{k_1}^{i_1}, f_{k_2}^{i_2}\rangle_{\mathcal{L}_d^2}\right)_{(i_1-1)\times N+k_1,(i_2-1)\times N+k_2}$$

and where

$$y = \left(\langle f, f_{k_2}^{i_2}\rangle_{\mathcal{L}_d^2}\right)_{(i_2-1)\times N+k_2}.$$

Among the projection coefficients involved in the conditional expectation, the coefficient $\alpha^\star$ in Equation (1.82) is the one with the minimal norm.

2) Assume that the time-discretization is uniform, that is, $t_l = lh$ for $l = 1,\ldots,N$. The self-similarity property allows the covariance kernel to be factorized by a power of the time step h, and the remaining amount in the factorization represents the covariance kernel at time 1 (independent of h) between the two given fBms:

$$\mathfrak{M}_{(i_1-1)*N+k_1,(i_2-1)*N+k_2} = \rho_{i_1,i_2} h^{H_{i_1}+H_{i_2}} \int_{\mathbb{R}} \psi_{0,1}^{H_{i_1}}\left(v-k_1\right)\psi_{0,1}^{H_{i_2}}\left(v-k_2\right)\mathrm{d}v, \tag{1.83}$$

$$y_{(i_2-1)N+k_2} = \sum_{i=1}^{d} (\omega_{t_N}^i \cdot R_{:,i_2}) h^{H_i+H_{i_2}} \int_{\mathbb{R}} \psi_{0,1}^{H_i}\left(v-N\right)\psi_{0,1}^{H_{i_2}}\left(v-k_2\right)\mathrm{d}v. \tag{1.84}$$

The theoretical results are then applied to the framework of risk management in finance.

Financial framework We present the model used to describe the price dynamics. Usually, the latter are described using a standard Black-Scholes model based on standard Brownian motion. The approach presented in this work is quite different, because the price dynamics are described by a fractional Black-Scholes model. This choice of model is justified by the fact that the log-price increments remain Gaussian self-similar and stationary. Therefore, interesting and tractable properties can be exploited, on the one

hand, to model the trajectory of the log-price increments with accuracy and realism, by introducing correlations between them which express long-range (respectively short-range) dependence; and on the other hand, to insure stability and coherence of the estimators and to perform theoretical computations of variances-covariances. Moreover, the assets that we describe, such as FX rates, present significant correlations between the log-price increments, which can be taken into account by the fractal properties of fBm. Such an approach has already been adopted in the works of (Garcin, 2020) (in dimension $d = 1$), in which fBm is used to take into account positive correlations between returns. Here, we propose a multivariate extension – the so-called multivariate fractional Black-Scholes model – to model the price dynamics of a universe composed of several assets correlated in both time and space. Assume that we have a universe composed of d assets whose market prices at time t are denoted by $\left(S_t^1, \ldots, S_t^d\right)$ and whose market log-prices at time t are denoted by $\left(X_t^1, \ldots, X_t^d\right)$.

The price dynamics are given by the geometric fractional Brownian motion formula:

$$\forall i \in [\![1, d]\!],\ S_t^i = S_0^i e^{c_t^i + \sigma_i B_t^{i,H_i}} \quad \text{with} \quad c_t^i = \log\left(\frac{\mathbb{E}[S_t^i]}{S_0^i}\right) - \frac{\sigma_i^2}{2} t^{2H_i}, \tag{1.85}$$

where S_0 is fixed and known, and c_t^i is the centering parameter in the model. We assume that all parameters are known owing to their estimations from the market data.

Let us focus on the properties of the log-price process:

$$X_t^i = X_0^i + c_t^i + \sigma_i B_t^{i,H_i}. \tag{1.86}$$

Log-price process is an affine function with respect to fractional Brownian motion, and as such, it benefits from its properties related to the self-similarity and stationarity of the increments, while relaxing the property of independence of the increments. For any $h > 0$, the log-price increments of length h are defined by (see Equation (1.72)):

$$\delta_h X_t^i := X_{t+h}^i - X_t^i = \delta_h c_t^i + \sigma_i I(\psi_{t,t+h}^{H_i} R_{:,i}). \tag{1.87}$$

In this framework, we are interested in determining risk measures of the future portfolio price increment at time horizon h. The portfolio and its increments of size h are defined by:

$$P_t = \sum_{i=1}^{d} \omega^i S_t^i \qquad \text{and} \qquad \delta_h P_t := P_{t+h} - P_t. \tag{1.88}$$

The objective is to compute the conditional $\mathbf{VaR}_\alpha$ of the future portfolio increment $\delta_h P_{t_N} = P_{t_{N+1}} - P_{t_N}$, given the past observations $\{S_{t_0}^1, S_{t_1}^1, \ldots, S_{t_N}^1, \ldots, S_{t_0}^d, S_{t_1}^d, \ldots, S_{t_N}^d\}$ taken on a uniform partition of the time interval $[0, t]$, such that the time-step is constant and equal to h, $t_0 = 0$, $t_N = t$ and for all $l \in [\![0, N]\!]$, $t_l = lh$. Because all parameters are known, observing the prices or the log-price or the log-price increments yields the same information, modeled by sigma-algebra

$$\mathcal{G} := \sigma(\delta_h X_{t_l}^i : (i, l) \in \mathcal{I}) = \sigma(I(f_l^i) : (i, l) \in \mathcal{I}) \qquad \text{where} \qquad f_l^i(\cdot) := \sigma_i R_{:,i} \psi_{t_l, t_{l+1}}^{H_i}(\cdot) \tag{1.89}$$

and $\mathcal{I} = \{(i,l) \mid i \in [\![1,d]\!],\ l \in [\![1,N]\!]\}$ is an index family allowing both the indexing of the assets and of the time increments.

Conditional on $\mathcal{G}$, $\delta_h \hat{P}_{t_N} = I(f)$, where the form of the function f is defined by Equation (1.78) of Theorem 1.17 with the specific weighting $\omega^i_{t_N} = \omega_i \sigma_i S^i_{t_N} R_{:,i}$:

$$f(\cdot) = \sum_{i=1}^{d} \omega_i \sigma_i S^i_{t_N} R_{:,i} \psi^{H_i}_{t_N, t_{N+1}}(\cdot), \tag{1.90}$$

follows a Gaussian distribution. Specifically, conditional on $\mathcal{G}$, $\delta_h \hat{P}_{t_N}$ has a Gaussian distribution characterized by its conditional expectation and conditional variance.

VaR model For the purpose of the conditional $\mathbf{VaR}_\alpha$ of $\delta_h P_{t_N}$, we need to know the conditional distribution of the future portfolio increment, given the past observations. However, $\delta_h P_{t_N}$ is a linear combination of log-normal random variables, whose marginal and conditional distributions are not known. Therefore, we propose a Gaussian approximation of the conditional $\mathbf{VaR}_\alpha$ of $\delta_h P_{t_N}$, with the conditional $\mathbf{VaR}_\alpha$ of $\delta_h \hat{P}_t$, whose conditional distribution is Gaussian. The Gaussian **VaR** is easily determined because it relies only on the expectation and standard deviation of the Gaussian distribution.

$$\delta_h P_{t_N} = P_{t_N+h} - P_{t_N} = \sum_{i=1}^{d} \omega^i \delta_h S^i_{t_N} \simeq \sum_{i=1}^{d} \omega^i S^i_{t_N} \delta_h X^i_{t_N} =: \delta_h \hat{P}_{t_N}. \tag{1.91}$$

Theorem 1.18 (Conditionally Gaussian Value-at-Risk)**.** *If $\delta_h \hat{P}_{t_N} = I(f)$, with f given by Equation (**??**), is the future increment of the log-price portfolio, and $(\delta_h X^i_{t_l} = I(f^i_l))_{(i,l)\in\mathcal{I}}$, with f^i_l given by Equation (1.77) of Theorem 1.17, are the past log-price increments of the assets that form the universe, then the conditional Gaussian $\mathbf{VaR}_\alpha$ is given by the following formula:*

$$\mathbf{VaR}_\alpha(\delta_h \hat{P}_{t_N} | \mathcal{G}) = \sqrt{\mathbb{V}[\delta_h \hat{P}_{t_N} | \mathcal{G}]} \mathcal{N}^{-1}(\alpha) + \mathbb{E}[\delta_h \hat{P}_{t_N} | \mathcal{G}] \tag{1.92}$$

where $\mathbb{E}[\delta_h \hat{P}_{t_N} | \mathcal{G}]$ and $\mathbb{V}[\delta_h \hat{P}_{t_N} - \mathbb{E}[\delta_h \hat{P}_{t_N} | \mathcal{G}]]$ are given by Theorem 1.17.

ES model The conditional Expected-Shortfall of the Gaussian approximation of the future portfolio increment is provided in the following theorem.

Theorem 1.19 (Conditionally Gaussian Expected-Shortfall)**.** *If $\delta_h \hat{P}_{t_N} = I(f)$, with f given by Equation (**??**), is the future increment of the log-price portfolio, and $(\delta_h X^i_{t_l} = I(f^i_l))_{(i,l)\in\mathcal{I}}$, with f^i_l given by Equation (1.77) of Theorem 1.17, are the past log-price increments of the assets that form the universe, the conditional Gaussian $\mathbf{ES}_\alpha$ is given by the following formula:*

$$\mathbf{ES}_\alpha(\delta_h \hat{P}_{t_N} \mid \mathcal{G}) = \frac{\sqrt{\mathbb{V}[\delta_h \hat{P}_{t_N} | \mathcal{G}]}}{1-\alpha} n(\mathcal{N}^{-1}(\alpha)) + \mathbb{E}[\delta_h \hat{P}_{t_N} | \mathcal{G}] \tag{1.93}$$

where $n(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ and $N(x) = \int_{-\infty}^{x} n(y)\mathrm{d}y$ are respectively the standard Gaussian probability density function and cumulative distribution function.

Error quantification We provide an error quantification of our predictive approximation. More specifically, the goal is to quantify the accuracy in the approximation of $\mathbf{VaR}_\alpha(\delta_h P_{t_N} \mid \mathcal{G})$ with $\mathbf{VaR}_\alpha(\delta_h \hat{P}_{t_N} \mid \mathcal{G})$. In other words, we want to determine the upper bound of the following amount:

$$\left|\mathbf{VaR}_\alpha(\delta_h P_{t_N} \mid \mathcal{G}) - \mathbf{VaR}_\alpha(\delta_h \hat{P}_{t_N} \mid \mathcal{G})\right|.$$

Theorem 1.20. *Consider the future portfolio increment $\delta_h P_{t_N}$ and its Gaussian approximation conditionally to $\mathcal{G}$, given by the future portfolio log-return $\delta_h \hat{P}_{t_N}$. Thus, we have the following upper bound:*

$$\Delta = \left\|\delta_h P_{t_N} - \delta_h \hat{P}_{t_N}\right\|_{p,\mathcal{G}} \leq 2\sum_{i=1}^{d} \left|\omega^i\right| S^i_{t_N} \left(\sigma_i^2 h^{2H_i} C_{2pq}^{1/pq} + M_i^2\right) e^{(M_i)_+ + \frac{pr}{2}\sigma_i^2 h^{2H_i}} \tag{1.94}$$

with $p > 1$ and $\frac{1}{q} + \frac{1}{r} = 1$.

If we assume that Δ is small enough, that is, $|\Delta| < 1 - \alpha$ a.s., then:

(i) *The quantification of the approximation of the conditional $\mathbf{VaR}_\alpha$ of the future portfolio price increment $\delta_h P_{t_N}$ with the conditional Gaussian $\mathbf{VaR}_\alpha$ of the future portfolio log-price increment $\delta_h \hat{P}_{t_N}$ is given as follows:*

$$\left|\mathbf{VaR}_\alpha(\delta_h P_{t_N} \mid \mathcal{G}) - \mathbf{VaR}_\alpha(\delta_h \hat{P}_{t_N} \mid \mathcal{G})\right| \leq \left(\frac{2\sigma(\mathcal{G})}{1-\alpha-\Delta} \sum_{i=1}^{d} \left|\omega^i\right| S^i_{t_N} \right. \tag{1.95}$$

$$\left. \times \left(\sigma_i^2 h^{2H_i} C_{2pq}^{1/pq} + M_i^2\right) e^{(M_i)_+ + \frac{pr}{2}\sigma_i^2 h^{2H_i}}\right), a.s. \tag{1.96}$$

with $M_i := \mathbb{E}\left[\delta_h X^i_{t_N} \mid \mathcal{G}\right]$, $\sigma(\mathcal{G}) = \sqrt{\mathbb{V}[\delta_h \hat{P}_{t_N} | \mathcal{G}]}$, $G \sim \mathcal{N}(0,1)$, and $C_{2pq} := \mathbb{E}\left[G^{2pq}\right]$.

(ii) *The quantification of the approximation of the conditional $\mathbf{ES}_\alpha$ of the future portfolio price increment $\delta_h P_{t_N}$ with the conditional Gaussian $\mathbf{ES}_\alpha$ of the future portfolio log-price increment $\delta_h \hat{P}_{t_N}$ is given as follows:*

$$\left|\mathbf{ES}_\alpha(\delta_h P_{t_N} \mid \mathcal{G}) - \mathbf{ES}_\alpha(\delta_h \hat{P}_{t_N} \mid \mathcal{G})\right| \tag{1.97}$$

$$\leq \left[\left(1 + \frac{1}{1-\alpha}\right)\frac{2\sigma(\mathcal{G})}{1-\alpha-\Delta} + \frac{2}{1-\alpha}\right] \sum_{i=1}^{d} \left|\omega^i\right| S^i_{t_N} \left(\sigma_i^2 h^{2H_i} C_{2pq}^{\frac{1}{pq}} + M_i^2\right) e^{(M_i)_+ + \frac{pr}{2}\sigma_i^2 h^{2H_i}}. \tag{1.98}$$

Parameters estimation Our predictive model involves the following parameters: $\sigma = (\sigma_1, \ldots, \sigma_d)$ represents the vector of market volatility related to each asset, $H = (H_1, \ldots, H_d)$ is the vector of Hurst exponents controlling the smoothness of the price trajectories of each asset, and $R = \{\rho_{ij}, i, j = 1, \ldots d\}$ refers to the matrix of correlations between assets i and j. We consider the sequence of observations $(X^1_{t_0}, \ldots, X^1_{t_N}, \ldots, X^d_{t_0}, \ldots, X^d_{t_N})$ representing the log-price trajectories of the d assets. For ease of estimation, we assume

that the observations are uniformly spaced in time, such that $t_j = j\tau$, where τ is the minimal time-step. VaR model is backtested first on synthetic data, then on real data. Therefore, we estimate these parameters from synthetic data in a first time, and then from real data. The estimation methodology is based on three key assumptions: *Gaussianity*, *stationarity* and *self-similarity*. These assumptions are required to estimate the parameters.

The *Gaussian assumption* states that for each asset, the log-prices are assimilated to an fBm. The log-price increments exhibit the same behavior as the fBm increments. Thus, the log-price increments form a Gaussian vector and, any linear combination of the increments is a Gaussian variable. The approximation of the future portfolio variation $\delta_h \hat{P}_{t_N}$ with the future variation of the log-price portfolio conditionally to the past log-price increments of the assets $\delta_h \hat{P}_{t_N}$, is a Gaussian variable. Therefore, the conditional $\mathbf{VaR}_\alpha$ of $\delta_h \hat{P}_{t_N}$, given the past log-returns of the assets is the quantile of a Gaussian distribution. More precisely, $\mathbf{VaR}_\alpha(\delta_h \hat{P}_N \mid \mathcal{G})$ is an affine function of the α-quantile of the standard Gaussian distribution, whose intercept is given by the conditional expectation $\mu(\mathcal{G}) = \mathbb{E}\left[\delta_h \hat{P}_N \mid \mathcal{G}\right]$ and whose slope is the square root of the conditional variance $\sigma(\mathcal{G}) = \sqrt{\mathbb{V}[\delta_h \hat{P}_N \mid \mathcal{G}]}$.

The *stationarity assumption* establishes the invariance by time-origin change, or translation, of the properties and characteristics of the families of random variables in $\mathbb{L}^2$. Specifically in our case, the stationarity property establishes the invariance by time-origin change of the expectations, variances and covariances, of the fBm increments process and thus of the log-price increments process. The stationarity property implies that the estimators of the expectations, variances and covariances, applied to the increments process remain the same when the time-origin of the increments process changes; thus, estimations of the parameters performed on the increments process are stable and reliable.

The *self-similarity assumption* establishes a spatial proportionality relationship between the characteristics, especially the square expectations, variances and covariances, of the process itself and between the characteristics of the increments process, taken at two proportional time-scales λt and t, with as spatial proportionality factor a power function of λ: λ^{2H_i}, $H_i \in (0, 1)$. This implies that the estimators of the square expectations, variances and covariances applied to the increments process at two proportional time-scales λt and t are linked by a spatial proportionality relationship of factor λ^{2H_i}, which allows easy adaptation of the estimators to different time-scales. The self-similarity property allows obtaining reliable estimations, coherent with the different time-scales.

Under these assumptions, the method of parameter estimation of the mfBm consists, for each component of the mfBm, in computing the empirical variance of the log-price increments of several lengths, following the moment-based estimator of the fBm introduced by (Istas and Lang, 1994; Kent and Wood, 1997).

The quadratic variation of the increments of size $m\tau$ with $m \in \mathbb{N}^\star$ is respectively given for disjoint increments and for increments with overlapping, as follows:

$$V_m^i = \frac{1}{\left\lfloor N/m \right\rfloor} \sum_{j=0}^{\left\lfloor \frac{N}{m} \right\rfloor - 1} (X^i_{(j+1)m\tau} - X^i_{jm\tau})^2 \quad \text{and} \quad V_m^i = \frac{1}{N-m+1} \sum_{j=0}^{N-m} (X^i_{(j+m)\tau} - X^i_{j\tau})^2. \tag{1.99}$$

Two alternative methods of estimation, that are almost equivalent, have been studied. The first method based on the works of (Amblard and Coeurjolly, 2011a) consists in the linear regression of the log-variances versus the log-time-scales. The estimation of the Hurst exponent is given by half the slope and the volatility by the exponential of half the intercept. Another method of estimation based on the energy levels, studied in details in (Chong et al., 2022a, p.12), (Chong et al., 2022b), and (Szymanski and Takabatake, 2023), is also exploited.

Backtesting We provide a backtesting procedure to assess the performance of our model through numerical experiments.

We consider an information set $\mathfrak{I}_{t-1} = \{\mathcal{I}_{t-1}, \mathcal{I}_{t-2}, \mathcal{I}_{t-3}, \ldots, \mathcal{I}_1\}$ which consists of a sequence of indicator functions, where the hit variable $\mathcal{I}_t$ at time t is a Bernoulli variable equal to 1 if the return between times t and $t+h$ exceeds the VaR predicted at time t, and 0 otherwise. More precisely, we obtain the hit variables using a rolling window approach.

For each window of size w, we predict at time t the conditional VaR at time horizon h, given the history of the window $[t-w, t]$, and we compare it to the log-return realized between times t and $t+h$, thus generating a hit variable equal to either 1 or 0.

Following traditional literature on VaR backtesting (Christoffersen, 1998; Davis, 2016), a satisfying $\mathbf{VaR}_\alpha$ must be such that $\mathbb{E}[\mathcal{I}_t|\mathfrak{I}_{t-1}] = \alpha$. Equivalently, we evaluate two criteria on the generated VaRs: the conditional coverage and the independence of the hit variables; namely, it has to be shown that $\{\mathcal{I}_t\} \overset{iid}{\sim} \mathcal{B}ern(\alpha)$.

In practice, the coverage test consists of estimating the conditional expectation of the sequence of hit variables that should correspond to the desired risk level α. This estimation was performed by computing the empirical mean of the hit variables. The closer the empirical mean is to α, the more satisfying the conditional coverage test is.

Regarding independence, we propose a method inspired by extreme value theory, which was first proposed by (Bücher et al., 2020). Indeed, two extreme events become approximately independent if they are separated by a sufficient amount of time. To measure the degree of dependence in a time series, we use the extremal index.

The extremal index θ is an indicator that quantifies the degree of dependence in a time sequence of random variables, by counting the number of clusters of variables above a predetermined threshold, namely, extreme values. It is equal to one if the sequence is independent, which means that there is no cluster of extreme values. The closer the extremal index is to zero, the more numerous the clusters of extreme values are, and the larger the serial dependence in the sequence is.

In the context of VaR backtesting, we intend to apply this method to the sequence of hit variables, considering that an extreme value corresponds to a hit equal to one. Therefore, the extremal index quantifies the presence of clusters of **VaR** violations.

An estimator of the extremal index can be provided by performing either a block or a run declustering algorithm. More precisely, given an arbitrary block size b, we partition the sequence of n successive hit variables into $k = \left\lfloor \frac{n}{b} \right\rfloor$ non-overlapping blocks (block declustering) or $n - b$ running windows (run declustering). In the block-declustering approach, each block containing nonzero hit variables represents a cluster. In the run-declustering approach, we count the number of windows without a nonzero hit starting at the end of a cluster. With both methods, the estimator of the extremal index is then the ratio between the estimated number of clusters and the number of nonzero hit variables. Formally, if we note $M_{i,j} = \max\{\mathcal{I}_{i+1}, \ldots, \mathcal{I}_j\}$, the two estimators are defined as follows:

$$\hat{\theta}_n^B(b) = \frac{\sum_{i=1}^{k} \mathbb{1}_{\{M_{(i-1)b,ib}=1\}}}{\sum_{i=1}^{kb} \mathbb{1}_{\{\mathcal{I}_i=1\}}} \quad \text{and} \quad \hat{\theta}_n^R(b) = \frac{\sum_{i=1}^{n-b} \mathbb{1}_{\{\mathcal{I}_i=1, M_{i,i+b}=0\}}}{\sum_{i=1}^{n-b} \mathbb{1}_{\{\mathcal{I}_i=1\}}}. \tag{1.100}$$

The backtesting process allows the comparison of three VaR models: conditional VaR, Gaussian VaR (we force H to be equal to 0.5), and empirical VaR. In most cases, the VaR model stands out in terms of coverage rate and extremal index, followed by the Gaussian VaR and empirical VaR. The lowest VaR is almost always the empirical VaR but its coverage rate and extremal index are very far from the desired ones.

1.3.3 Chapter 4 - Mean Estimation of Expected-Shortfall in Heavy-Tailed Distribution

State of the art

Gaussian framework is the most commonly used in finance due its convenient properties. However, the Gaussian distribution is thin-tailed then it assumes that extreme events are rare, and tends to underestimate the probability of such events occurring. On the contrary, heavy-tailed distribution like Pareto, assign higher probabilities to extreme events, which is often more realistic in finance. One of the shortcomings of the VaR is that VaR is not sensitive to the tail risk then it fails to capture extreme events what often leads to an underestimation of the risk. When the distribution of the losses is thin-tailed, such as the Gaussian distribution, extreme events occur with a very low probability, then VaR remains efficient despite its shortcomings. However, when the losses are described thanks to a heavy-tailed distribution, extreme events occur more frequently, and VaR is not relevant any more. Unlike VaR, Expected-Shortfall provides a measure of the average loss magnitude beyond the VaR threshold, focusing on tail risk. This is crucial for institutions to understand the potential extent of losses during extreme events, which VaR may not fully capture. Consequently, when losses are described by heavy-tailed distributions, such as Pareto, ES is more appropriate than VaR to provide an accurate evaluation of the risk.

Literature on Expected-Shortfall The literature about Expected-Shortfall as a solution to address the shortcomings of Value-at-Risk is wide. In a context of reinforcement of the banking regulation, many authors have been interested in mirroring VaR and ES in order to compare their strengths and weaknesses, (Kellner and Rösch, 2016). Especially, the works of (Acerbi et al., 2001), (Acerbi and Tasche, 2002b), (Artzner et al., 1999) are nice references.

Some authors focused on the famous debate between VaR and ES. This debate comes from the fact that VaR is not a coherent risk measure due to its lack of subadditivity but VaR is elicitable whereas ES is a coherent risk measure but is not elicitable. On the one hand, the lack of subadditivity in the VaR can have significant consequences, primarily in risk management and decision-making processes within financial institutions or any organization exposed to market risks. Subadditivity refers to the property where the total risk of a portfolio is less than or equal to the sum of the risks of its individual components. However, if the VAR lacks subbadditivity, it means that the combined risk of a portfolio could be higher than the sum of its individual part's risks, which can lead to an underestimation of risk, inaccurate hedging strategies, inadequate capital reserves, misguided asset allocation or even systemic risk amplification. This is the topic of papers (Embrechts, 2000), (Tibiletti, 2008).

In (Garcia et al., 2007), some methods are proposed to tackle the lack of subadditivity of the VaR. The works of (Daníelsson et al., 2013) show that the VaR can be subadditive in some regions of heavy-tailed distributions. In order to deal with the conceptual problems caused by VaR, (Artzner et al., 1999), (Tasche, 2002a) present an alternative risk measure to VaR, called ES and highlights the advantages of ES that allows addressing the shortcomings of VaR. The article (Inui and Kijima, 2005) shows that any coherent risk measure is given by a convex combination of expected shortfalls, and an expected shortfall (ES) is optimal in the sense that it gives the minimum value among the class of plausible coherent risk measures. An extrapolation method of the ES is provided. Authors discuss properties of ES as well as its generalization to a class of coherent risk measures which can incorporate higher moment effects. The works of (Yamai and Yoshiba, 2005) illustrate how the tail risk of VaR can cause serious problems in certain cases, cases in which Expected-Shortfall can serve more aptly in its place. Specifically, they discuss two cases: concentrated credit portfolio and foreign exchange rates under market stress. They show that Expected Shortfall requires a larger sample size than VaR to provide the same level of accuracy. The works of (Acerbi and Tasche, 2002a), highlight the fact that most definitions of ES lead to the same results when applied to continuous loss distributions but differences may appear when the underlying loss distributions have discontinuities. In this case even the coherence property of ES can get lost unless one took care of the details in its definition. Authors compare some of the definitions of ES, pointing out that there is one which is robust in the sense of yielding a coherent risk measure regardless of the underlying distributions. Moreover, this ES can be estimated effectively even in cases where the usual estimators for VaR fail. Based on the observation that portfolio risk-adjusted performance measurement involves the calculation of the risk contribution for each asset it contains. The paper (Fan et al., 2012) uses multivariate Copula functions to model the dependence structure among the

assets in a portfolio, then, based on a simulation, decomposes the portfolio VaR and Expected Shortfall. Furthermore, with this approach, the risk contribution calculated using Expected Shortfall is more robust, and its estimation error can be reduced by increasing the simulation sample size. An application of Expected-Shortfall to credit risk is provided in (Fan et al., 2010).

However, Expected-Shortfall does not satisfy the elicitability property. Elicitability is a concept in the field of risk measurement and decision theory, particularly in the context of scoring rules and risk assessment. A scoring rule is a function used to assess the accuracy of a probabilistic forecast or estimate. Elicitability refers to the property of a scoring rule that dictates whether it can accurately evaluate the quality of probabilistic forecasts or estimates without providing incentives for forecasters to misrepresent their beliefs. Moreover, elicitability allows comparing different models to rank their performance thanks to the scoring rule which is convenient to backtest the models. The lack of elicitability can lead to several consequences among which we can find the incentives for misrepresentation, the difficulty in evaluation, the potential for systemic risk and the loss of trust. Opinion was divided on the need for elicitability in backtesting. According some authors, elicitability is only useful but not necessary to backtest the models whereas according others, elicitability is necessary and in this case it is better to use VaR instead of ES. This debate is for instance presented in (Acerbi and Szekely, 2014). In (Fissler et al., 2015), authors show that ES is jointly elicitable with VaR.

An important stake lies in the choice of the method of estimation of the Expected-Shortfall. Many authors interested to this topic. In (Brazauskas et al., 2008), authors develop statistical inferential tools for estimating and comparing conditional tail expectation (CTE) functions, which are of considerable interest in actuarial science. In (Chen, 2008), two non-parametric Expected-Shortfall estimators for dependent financial losses are presented. One is a sample average of excessive losses larger than a VaR. The other is a kernel smoothed version of the first estimator, hoping that more accurate estimation can be achieved by smoothing. In (Taylor, 2008) a method of estimation of VaR and ES thanks to expectiles is provided. A non-parameteric estimation of ES is proposed in (Scaillet, 2004). A robust estimation method of ES is developed in (Jadhav et al., 2009) and (Pan et al., 2019). Authors develop a tail-based normal approximation with explicit formulae is derived by matching a specific quantile and the mean excess square of the sample observations. To enhance the estimation accuracy, they propose an adjusted tail-based normal approximation based on the sample's tail weight. The adjusted expected shortfall estimator is robust and efficient in the sense that it can be applied to various heavy-tailed distributions, such as Student, lognormal, Gamma and Weibull, and the errors are all small.

Literature on heavy-tailed distributions Gaussian framework is the most commonly used in finance due its convenient properties. However, the Gaussian distribution is thin-tailed then it assumes that extreme events are rare, and tends to underestimate the probability of such events occurring. On the contrary, heavy-tailed distribution like Pareto, assign higher probabilities to extreme events, which is often more realistic in finance. Heavy-tailed distributions are not only used in economics and finance but in

many other fields such as in physics, biology, earth and planetary sciences, computer science, demography and the social sciences. Numerous authors interested in studying heavy-tailed distributions. For instance, the distributions of the sizes of cities, earthquakes, forest fires, solar flares, moon craters and people's personal fortunes all appear to follow heavy-tailed distributions. In finance, heavy-tailed distributions are dedicated to model losses or returns distributions.

In (ZINCHENKO, 2001), authors provide a survey of the main trends in theoretical investigations and practical applications of heavy-tailed models with emphasis on subexponential, Pareto-type and stable distributions. Certain problems connected with limit theorems, approximation, estimation, numerical simulation for heavy tails are treated as well as the connection with the risk theory. Fat-tailed distributions in economics and finance are studied in details in (Haas and Pigorsch, 2009), (Broda and Paolella, 2011). Based on the observation that it is of great importance for those in charge of managing risk to understand how financial asset returns are distributed. Practitioners often assume for convenience that the distribution is normal. Since the 1960s, however, empirical evidence has led many to reject this assumption in favor of various heavy-tailed alternatives. In a heavy-tailed distribution the likelihood that one encounters significant deviations from the mean is much greater than in the case of the normal distribution. It is now commonly accepted that financial asset returns are, in fact, heavy-tailed. In (Bradley and Taqqu, 2003) a survey to examine how these heavy tails affect several aspects of financial portfolio theory and risk management is carried out. Some of the methods that one can use to deal with heavy tails are presented and are illustrated using the NASDAQ composite index. In the same way, the paper of (Guo, 2017), exhibits the fact that heavy-tailed distribution, which accurately estimates the tail risk, would significantly improve quantitative risk management practice. In (Nolan, 2014) an accessible introduction to stable distributions for financial modeling is provided. There is a real need to use better models for financial returns because the normal (or bell curve/Gaussian) model does not capture the large fluctuations seen in real assets. Stable laws are a class of heavy-tailed probability distributions that can model large fluctuations and allow more general dependence structures. The works of (Peng and Qi, 2017) present methods of inference for heavy-tailed data and apply them in the context of insurance and finance. Many insurance loss data are known to be heavy-tailed. In the article (Ahn et al., 2012) the class of Log-phase-type (LogPH) distributions is studied as a parametric alternative in fitting heavy tailed data. Transformed from the popular phase-type distribution class, the LogPH introduced by Ramaswami exhibits several advantages over other parametric alternatives.

Literature on Pareto distributions The Pareto distribution is a well-known distribution that belongs to the class of the heavy-tailed distributions. The Pareto distribution is often used in the economic and financial context to model the distribution of the losses or of the returns when extreme events occur more frequently than in the Gaussian framework. Such a distribution present very convenient property such as the stability by conditioning and the scaling properties that will be detailed in the sequel. More specifically, the Pareto distribution is a power law distribution. When the prob-

ability of measuring a particular value of some quantity varies inversely as a power of that value, the quantity is said to follow a power law, also known variously as Zipf's law or the Pareto distribution. In, (Newman, 2005), authors review some of the empirical evidence for the existence of power-law forms and the theories proposed to explain them. A nice reference about the Pareto distribution is (Chattamvelli and Shanmugam, 2021). Pareto distributions and related generalizations have historically been viewed as being suitable for modeling income and wealth distributions. In this context, in (Arnold, 2014) a brief review of the history of such models, distributional properties and inference procedures are surveyed is carried out. Various related distributions, including multivariate variants, are described. The works of (Arnold, 2008) provide a survey of results related to these Pareto-like models including discussion of related distributional and inferential questions. Topics include the classical Pareto models and its generalizations, stochastic income models leading to Paretian income distributions, distributional properties of generalized Pareto distributions, related discrete distributions, inequality measures for Paretian models, inferential issues and multivariate extensions. Most of the times, this Paretianity is inferred from the observation of some plots, such as the Zipf plot and the mean excess plot. If the Zipf plot looks almost linear, then everything is ok and the parameters of the Pareto distribution are estimated. Often with OLS. Unfortunately, these heuristic graphical tools are not reliable. This is what authors show in (Cirillo, 2013). Indeed they show that only a combination of plots can give some degree of confidence about the real presence of Paretianity in the data. In (Crovelli and Barton, 1995), Fractals and the Pareto distribution are applied to petroleum accumulation-size distributions. In (Su and Furman, 2017), a new multivariate distribution possessing arbitrarily parametrized and positively dependent univariate Pareto margins is introduced.

Literature on estimation of the Pareto index The Pareto index manages the heaviness or the thinness of the distribution tail. A challenge is to estimate it accurately.

In extreme value statistics, the extreme value index is a well-known parameter to measure the tail heaviness of a distribution. Pareto-type distributions, with strictly positive extreme value index (or tail index) are considered. In (Rytgaard, 1990), different estimators of the Pareto parameter are proposed and compared to each others. First, traditional estimators as the maximum likelihood estimator and the moment estimator are deduced and their statistical properties are analyzed. It is shown that the maximum likelihood estimator is biased but it can easily be modified to an minimum-variance unbiased estimator. But still the coefficient of variance of this estimator is very large. For similar portfolios containing same types of risks the estimated values are expected to be at the same level. Therefore, credibility theory is used to obtain an alternative estimator more stable and less sensitive to random fluctuations in the observed losses. Finally, an estimator of the risk premium for an unlimited excess of loss cover is proposed. It is shown that this estimator is a minimum-variance unbiased estimator of the risk premium. This estimator of the risk premium is compared to the more traditional methods of calculating the risk premium. In (Crovella and Taqqu, 1999), authors propose a method (called the "scaling estimator") based on the scaling properties of

sums of heavy-tailed random variables. More precisely, for any random variable X, Σ_n is defined as random variable that is the sum of n independent random variables each with the same distribution as X. For heavy-tailed distributions with tail index α, limit theorems similar as the usual central limit theorems may be formulated showing that sums of such variables converge to a stable distributions with the same α. A distribution is stable in the strict sense if for each n there exists constants $c_n > 0$ such that:

$$\Sigma_n \stackrel{d}{=} c_n X. \tag{1.101}$$

In the specific case of sum of independent strictly stable variables $c_n = n^{\frac{1}{\alpha}}$. This property is the scaling property on which the method presented in this paper is based to obtain the estimator of the heavy-tailed index α. It has the advantages of being non-parametric, of being easy to apply, of yielding a single value, and of being relatively accurate on synthetic datasets. Since the method relies on the scaling of sums, it measures a property that is often one of the most important effects of heavy-tailed behavior. Most importantly, they present evidence that the scaling estimator appears to increase in accuracy as the size of the dataset grows. It is thus particularly suited for large datasets. One of the most important problems involved in the estimation of Pareto indices is the reduction of bias in case the slowly varying part of the Pareto type model disappears at a very slow rate. In other cases, when the bias problem is not so severe, the application of well-known estimators such as the Hill (1975) and the moment estimator (Dekkers et al. (1989)) still asks for an adaptive selection of the sample fraction to be used in such estimation procedures. In (Beirlant et al., 1996), (Beirlant et al., 1999), (Ocran et al., 2022), a reduced-bias estimators for the estimation of the tail index of a Pareto-type distribution. This is achieved through the use of a regularised weighted least squares with an exponential regression model for log-spacings of top order statistics. The asymptotic properties of the proposed estimators are investigated analytically and found to be asymptotically unbiased, consistent and normally distributed. The works of (Finkelstein et al., 2006), propose an estimator of the tail index of a Pareto distribution based on the use of the probability integral transform. This new estimator provides performance that is comparable to the best robust estimators, while retaining conceptual and computational simplicity. A tuning parameter in the new estimator can be adjusted to control the tradeoff between robustness and efficiency. A new generalized median type estimator is introduced in (Brazauskas and Serfling, 2000) and compared with the MLE and several well-established estimators associated with the methods of moments, trimming, least squares, quantiles, and percentile matching. A robust estimator of the tail index is proposed in (Vandewalle et al., 2007), by combining a refinement of the Pareto approximation for the conditional distribution of relative excesses over a large threshold with an integrated squared error approach on partial density component estimation. The paper (Brzezinski, 2016) investigates the small-sample properties of the most popular robust estimators for the Pareto tail index, including the optimal B-robust estimator, the weighted maximum likelihood estimator, the generalized median estimator, the partial density component estimator, and the probability integral transform statistic estimator (PITSE). Monte Carlo simulations show that the PITSE offers the desired compromise between ease of use and power to protect against outliers in the small-sample setting. In (Beirlant and Goegebeur, 2004), an approach motiv-

ated by the fact that in some applications the threshold should be allowed to change with the covariates due to significant effects on scale and location of the conditional distributions, is proposed. The approach followed is based on the technique of local polynomial maximum likelihood estimation. Using the asymptotic results they are able to derive an expression for the asymptotic mean squared error, which can be used to guide the selection of the bandwidth and the threshold. In (Goegebeur et al., 2008), the relation between goodness-of-fit testing and the optimal selection of the sample fraction for tail estimation, for instance using Hill's estimator, is examined. Authors consider this problem under a general kernel goodness-of-fit test statistic for assessing whether a sample is consistent with the Pareto-type model. Two important special cases of the kernel statistic, the Jackson and the Lewis statistic, are discussed in greater depth.

Contributions

In Chapter 4, the goal is to explore robust methods for calculating expectations in heavy-tailed distributions, as an alternative to the simple average. We consider the toy case of the Pareto distribution because such a distribution is often used to model losses in finance, and presents interesting properties. We are particularly interested in estimating the Expected-Shortfall (ES) for the Pareto distribution.

Expected-Shortfall for Pareto distribution First, we recall the theory of Expected-Shortfall and Pareto distribution and present the characteristic properties of the latter. We prove the useful properties of scaling and stability by conditioning of the Pareto distribution, and we prove the form of the ES at the risk level α ($\mathbf{ES}_\alpha$) based on these properties.

Various definitions of ES exist, all of which are equivalent. The definition considered in this chapter is the one according to which $\mathbf{ES}_\alpha$ is the conditional expectation of excess losses above the $\mathbf{VaR}_\alpha$. Indeed, based on (Tasche, 2002b, Prop 3.4, Eq 3.3),(Sarykalin et al., 2008, Def.2, p.273), for a continuous distribution, the Expected-Shortfall at the risk level α is defined as:

$$\mathbf{ES}_\alpha(X) = \mathbb{E}[X|X \geq \mathbf{VaR}_\alpha(X)] \tag{1.102}$$

where $\alpha \in (0,1)$ is the risk level, $X \in \mathcal{L}^p(\mathcal{F})$ is a random variable representing the loss (as a positive quantity, i.e. we take the convention that big losses correspond to large positive numbers), of a portfolio at some future time, with as cumulative distribution function $F_X(x) = \mathbb{P}(X \leq x); \forall x \in \mathbb{R}$.

For the sake of modeling accuracy and realism, we are interested in the estimation of ES in heavy-tailed distributions, particularly in the Pareto distribution. The choice of Pareto distribution is induced by the interesting properties of this distribution, as detailed by (Arnold, 2014). There are several reasons for focusing on this distribution.

Pareto distribution plays a crucial role in extreme value theory (EVT), a branch of statistics that deals with the statistical behavior of extreme events. EVT is particularly relevant in finance, where extreme events, such as market crashes, have significant

implications for investors and financial institutions. In finance, data often exhibit heavy-tailed behavior, meaning that extreme events, such as large price movements or financial crises, occur more severely than expected from a normal distribution. Tail risk refers to the risk of extreme events occurring in the tails of a distribution. Although rare, these events can significantly impact financial markets and portfolios. The Pareto distribution, with its ability to model heavy tails, provides a better fit to such data than traditional distributions such as the Gaussian distribution. Because Gaussian modeling assigns small weights to distribution tails, it ignores extreme events that can lead to inaccurate VaR prediction. On the contrary, the Pareto distribution attributes more weights to distribution tails, taking into account extreme events, which enables better VaR predictions, for instance when dealing with assets or portfolios that exhibit heavy-tailed behavior.

Mathematically, the Pareto distribution exhibits interesting properties. First, the distribution is relatively simple to understand and work with. It only has two parameters (scale x_m and shape γ parameters; the latter is also called Pareto index), making it easier to estimate and interpret compared to more complex heavy-tailed distributions such as the stable distribution. The probability density function and cumulative distribution function of the Pareto distribution have relatively simple analytical forms, making it easier to perform mathematical and statistical calculations compared to some other heavy-tailed distributions.

Definition 1.21 (Pareto distribution $\mathcal{P}(x_m, \gamma)$)**.** *If X is a random variable following a Pareto distribution $\mathcal{P}(x_m, \gamma)$, $(x_m > 0, \gamma > 0)$, then the probability that X is larger than some number x, that is, the survival function, also called the tail function, is given by:*

$$\bar{F}_X(x) = \mathbb{P}(X > x) = \mathbb{1}_{\{x<x_m\}} + \left(\frac{x_m}{x}\right)^{\gamma} \mathbb{1}_{\{x\geq x_m\}} \tag{1.103}$$

where x_m is the (necessarily positive) minimum possible value of X, and γ is a positive parameter. The Pareto distribution is characterized by a scale parameter x_m and a shape parameter γ, which is known as the tail index.

The cumulative distribution function (c.d.f.) of a Pareto random variable with parameters x_m and γ is:

$$F_X(x) = \left(1 - \left(\frac{x_m}{x}\right)^{\gamma}\right) \mathbb{1}_{\{x\geq x_m\}}. \tag{1.104}$$

And the probability density function (p.d.f.) is given by:

$$f_X(x) = \frac{\gamma x_m^{\gamma}}{x^{\gamma+1}} \mathbb{1}_{\{x\geq x_m\}}. \tag{1.105}$$

The Pareto distribution is a power-law and is well-adapted to fit data that display power-law behavior. Indeed, the Pareto distribution tail follows a linear function on a log-log plot. This is supplementary evidence that extreme events occur with more severe

intensity than those predicted by other heavy-tailed distributions. This is particularly useful when modeling financial phenomena that follow a power-law distribution, such as income distribution or extreme price movements.

One of the most interesting properties of Pareto distribution is its *scaling property*. Indeed, it allows establishing links between two Pareto distributions with the same shape parameter γ but different scale parameters x_m. For instance, the scaling property allows switching from the standardized Pareto distribution $\mathcal{P}(1,\gamma)$, to any non-standardized Pareto distribution $\mathcal{P}(x_m,\gamma)$, by a simple multiplication of the standardized Pareto distribution with the scaling parameter x_m of the non-standardized Pareto distribution. Conversely, the scaling property allows switching from any non-standardized Pareto distribution $\mathcal{P}(x_m,\gamma)$ to the standardized Pareto distribution $\mathcal{P}(1,\gamma)$, by a simple division of the non-standardized Pareto distribution by its scaling parameter x_m. More precisely, the scaling property establishes a proportionality relationship between the non-standardized and the standardized Pareto distributions, with a proportionality factor equal to the scale parameter x_m of the non-standardized Pareto distribution. Conversely, the scaling property allows switching from the non-standardized Pareto distribution $\mathcal{P}(x_m,\gamma)$ to the standardized Pareto distribution $\mathcal{P}(1,\gamma)$, by dividing the non-standardized Pareto distribution by its scale parameter x_m. Consequently, there exists a proportionality relationship between the standardized Pareto distribution $\mathcal{P}(1,\gamma)$ and the non-standardized Pareto distribution $\mathcal{P}(x_m,\gamma)$, with a proportionality factor equal to $\frac{1}{x_m}$. More generally, the scaling property establishes a link between any non-standardized Pareto distributions $\mathcal{P}(x_m^{(1)},\gamma)$ and $\mathcal{P}(x_m^{(2)},\gamma)$ with $x_m^{(1)}>0$ and $x_m^{(2)}>0$. Indeed, the Pareto distribution $\mathcal{P}(x_m^{(1)},\gamma)$ is proportional to the Pareto distribution $\mathcal{P}(x_m^{(2)},\gamma)$ with a proportionality factor equal to the ratio between the two scaling parameters $\frac{x_m^{(1)}}{x_m^{(2)}}$, and vice versa. The proportionality factor is a ratio whose numerator corresponds to the scale parameter of the non-standardized Pareto distribution that we want to reach $\mathcal{P}(x_m^{(1)},\gamma)$ (the target distribution), and denominator corresponds to the scale parameter of the initial non-standardized Pareto distribution $\mathcal{P}(x_m^{(2)},\gamma)$. The division of the non-standardized Pareto distribution $\mathcal{P}(x_m^{(2)},\gamma)$ by its own scaling parameter $x_m^{(2)}$ allows the standardization of the distribution, thus reaching the standardized Pareto distribution $\mathcal{P}(1,\gamma)$. Then, the multiplication of the standardized Pareto distribution by $x_m^{(1)}$ allows reaching the desired Pareto distribution $\mathcal{P}(x_m^{(1)},\gamma)$. Conversely, the Pareto distribution $\mathcal{P}(x_m^{(2)},\gamma)$ is proportional to the Pareto distribution $\mathcal{P}(x_m^{(1)},\gamma)$ with a proportionality factor equal to the ratio between the two scaling parameters $\frac{x_m^{(2)}}{x_m^{(1)}}$. The proportionality factor is a ratio whose numerator corresponds to the scale parameter of the non-standardized Pareto distribution that we want to reach $\mathcal{P}(x_m^{(2)},\gamma)$, and denominator corresponds to the scale parameter of the initial non-standardized Pareto distribution $\mathcal{P}(x_m^{(1)},\gamma)$. The division of the Pareto distribution $\mathcal{P}(x_m^{(1)},\gamma)$ by its own scaling parameter $x_m^{(1)}$ allows the standardization of the distribution, thus reaching the standardized Pareto distribution $\mathcal{P}(1,\gamma)$. Then, the multiplication of the standardized Pareto distribution by $x_m^{(2)}$ allows reaching the desired Pareto distribution $\mathcal{P}(x_m^{(2)},\gamma)$. Consequently, Pareto distribution is scale-invariant, which means that scaling a Pareto distribution using a constant parameter does not change the shape of the distribution

(γ remains the same). The new scaled distribution is still a Pareto distribution with the same shape parameter γ but a new scale parameter. The new Pareto distribution is obtained by multiplying the initial Pareto distribution by the ratio between the scaling parameter of the target Pareto distribution and that of the initial Pareto distribution.

In addition to the scaling property, the Pareto distribution satisfies the *property of stability by conditioning*. This implies that the tail of any Pareto distribution above a certain positive threshold is still a Pareto distribution with the same shape parameter but a new scaling parameter. For instance, the tail of any standardized Pareto distribution $\mathcal{P}(1,\gamma)$ above a given positive threshold $s_m^{(1)} > 0$, is still a Pareto distribution with the same shape parameter γ but a new scaling parameter equal to the conditioning parameter $s_m^{(1)}$: $\mathcal{P}(s_m^{(1)},\gamma)$. Owing to the scaling property, there exists a proportionality relationship linking the non-standardized Pareto distribution $\mathcal{P}(s_m^{(1)},\gamma)$ and the standardized Pareto distribution $\mathcal{P}(1,\gamma)$ with a proportionality factor equal to $s_m^{(1)}$. Consequently, by combining both the stability by conditioning and the scaling properties, a proportionality relationship is established between the standardized Pareto distribution conditional on its values being above the threshold $s_m^{(1)}$ and the marginal Pareto distribution $\mathcal{P}(1,\gamma)$, with a proportionality factor equal to the conditioning parameter $s_m^{(1)}$. Consequently, any standardized Pareto distribution $\mathcal{P}(1,\gamma)$ conditional on its values being above a given threshold $s_m^{(1)}$ is still a Pareto distribution with the same shape parameter γ but with a new scaling parameter equal to the conditioning parameter $s_m^{(1)}$.

More generally, the tail of any non-standardized Pareto distribution $\mathcal{P}(x_m^{(1)},\gamma)$ above a given positive threshold $s_m^{(1)} > x_m^{(1)}$ is still a Pareto distribution with the same shape parameter γ but a new scale parameter equal to the conditioning parameter $s_m^{(1)}$: $\mathcal{P}(s_m^{(1)},\gamma)$. Owing to the scaling property, there exists a proportionality relationship between the non-standardized Pareto distribution $\mathcal{P}(s_m^{(1)},\gamma)$ and the standardized Pareto distribution $\mathcal{P}(1,\gamma)$, with a proportionality factor equal to $s_m^{(1)}$. Moreover, the scaling property also establishes a proportionality relationship between the two non-standardized Pareto distributions $\mathcal{P}(s_m^{(1)},\gamma)$ and $\mathcal{P}(x_m^{(1)},\gamma)$, with a proportionality factor equal to the ratio between the two scaling parameters $\frac{s_m^{(1)}}{x_m^{(1)}}$. Consequently, by combining the stability by conditioning and the scaling properties, a proportionality relationship is established between the non-standardized Pareto distribution $\mathcal{P}(x_m^{(1)},\gamma)$ conditional on its values being above the threshold $s_m^{(1)}$ and the standardized Pareto distribution $\mathcal{P}(1,\gamma)$ with a proportionality factor equal to the conditioning parameter $s_m^{(1)}$. And, a proportionality relationship is also established between the non-standardized Pareto distribution $\mathcal{P}(x_m^{(1)},\gamma)$ conditional on its values being above the threshold $s_m^{(1)}$ and the marginal distribution $\mathcal{P}(x_m^{(1)},\gamma)$ with a proportionality factor equal to the ratio $\frac{s_m^{(1)}}{x_m^{(1)}}$ where the numerator is the conditioning parameter $s_m^{(1)}$ and denominator is the scaling parameter $x_m^{(1)}$ of the marginal distribution. Consequently, any non-standardized Pareto distribution $\mathcal{P}(x_m^{(1)},\gamma)$ conditional on its values being above the threshold $s_m^{(1)}$ is still a Pareto distribution with the same shape parameter γ but with a new scaling parameter equal to the conditioning parameter $s_m^{(1)}$.

These properties are very convenient. Moreover, they are specific to the Pareto distribution, that is, they are not shared by all heavy-tailed distributions. They are mathematically formulated in the following theorem.

Theorem 1.22 (Pareto stability by conditioning and rescaling)**.** (i) *Let X be a standard Pareto random variable $X \sim \mathcal{P}(1, \gamma)$, $\gamma > 0$. Let $x_m > 0$ be a new scaling parameter. Let Y be a non-standard Pareto random variable such that $Y \sim \mathcal{P}(x_m, \gamma)$. Then we have:*

$$\mathbb{P}(X \leq x \mid X \geq x_m) = \mathbb{P}(x_m X \leq x) = \mathbb{P}(Y \leq x). \tag{1.106}$$

Equivalently,

$$X \mid X \geq x_m \overset{d}{=} x_m X \overset{d}{=} Y \quad \text{or} \quad \frac{Y}{x_m} \overset{d}{=} \frac{X \mid X \geq x_m}{x_m} \overset{d}{=} X. \tag{1.107}$$

In other words, conditioning a standard Pareto random variable X from a given threshold x_m, amounts to scaling the standardized Pareto distribution, that is, the marginal distribution of X, with the conditioning parameter x_m. Then, the standardized Pareto distribution conditional on its values being above a certain threshold remains a Pareto distribution with the same shape parameter γ but a new scaling parameter equal to the conditioning parameter x_m.

(ii) *Let Z_1, Z_2 be two non-standard Pareto random variables such that $Z_1 \sim \mathcal{P}(x_m^{(1)}, \gamma)$ and $Z_2 \sim \mathcal{P}(x_m^{(2)}, \gamma)$ with $x_m^{(1)} > 0, x_m^{(2)} > 0, \gamma > 0$. Then we have:*

$$Z_1 \overset{d}{=} \frac{x_m^{(1)}}{x_m^{(2)}} Z_2 \overset{d}{=} x_m^{(1)} X \quad \text{or} \quad Z_2 \overset{d}{=} \frac{x_m^{(2)}}{x_m^{(1)}} Z_1 \overset{d}{=} x_m^{(2)} X. \tag{1.108}$$

In other words, any non-standardized Pareto distribution can be expressed from any other non-standardized Pareto distribution by dividing the initial non-standardized Pareto distribution by its own scaling parameter to reach the standardized Pareto distribution $\mathcal{P}(1, \gamma)$, then by scaling it with the scaling parameter of the target distribution.

(iii) *Let $s_m^{(1)}$ be a conditioning parameter such that $s_m^{(1)} > x_m^{(1)} > 0$. Then, Equations (1.107) and (1.108) lead to:*

$$Z_1 \mid Z_1 \geq s_m^{(1)} \overset{d}{=} \frac{s_m^{(1)}}{x_m^{(1)}} Z_1 \overset{d}{=} s_m^{(1)} X. \tag{1.109}$$

In other words, conditioning any non-standardized Pareto distribution $\mathcal{P}(x_m^{(1)}, \gamma)$ from a certain threshold $s_m^{(1)}$ amounts to scaling the standardized Pareto distribution $\mathcal{P}(1, \gamma)$ with the conditioning parameter $s_m^{(1)}$, or equivalently to scaling the marginal distribution $\mathcal{P}(x_m^{(1)}, \gamma)$ with the ratio between the conditioning parameter and the scaling parameter of the marginal distribution $\frac{s_m^{(1)}}{x_m^{(1)}}$.

Furthermore, the moments of the Pareto distribution exist only for certain ranges of shape parameter γ. Specifically, the k-th moment exists if and only if $\gamma > k$. This means that for small values of γ, moments may not exist, which is related to the heavy-tailed nature of the distribution. As the shape parameter γ decreases, the distribution

becomes even more heavy-tailed, giving more weight to extreme events. This leads to finite moments for only very small orders. This is a characteristic of heavy-tailed distributions, where extreme observations have a more significant impact on higher moments than distributions with lighter tails. For instance, the Pareto distribution has an infinite mean for shape parameter $\gamma \leq 1$. For $\gamma \leq 2$, the Pareto distribution has a finite mean but an infinite variance. For $\gamma > 2$, the variance of the Pareto distribution exists and is finite. This means that if the shape parameter is greater than 2, the distribution is characterized by both a finite mean and finite variance, making it more manageable in certain statistical analyses.

The scaling property of the Pareto distribution states that any non-standardized Pareto distribution $\mathcal{P}(x_m, \gamma)$ is proportional to the standardized Pareto distribution $\mathcal{P}(1, \gamma)$ with a proportionality factor equal to x_m. This implies that all the quantities (statistics) computed on the non-standardized Pareto distribution $\mathcal{P}(x_m, \gamma)$ are proportional to the ones computed on the standardized Pareto distribution $\mathcal{P}(1, \gamma)$, with a proportionality factor equal to x_m. Therefore, the calculus can be carried out on the standardized Pareto distribution, and the equivalent quantities on any non-standardized Pareto distribution $\mathcal{P}(x_m, \gamma)$, can be recovered by the multiplication of the standardized quantities with the proper scaling parameter x_m. For instance, the scaling property of the Pareto distribution allows defining any k-th moment of any non-standardized Pareto distribution $\mathcal{P}(x_m, \gamma)$ by a simple multiplication of the k-th moment of the standardized Pareto distribution $\mathcal{P}(1, \gamma)$ with the scaling parameter x_m raised to the power k.

These properties are mathematically formulated in the following lemma.

Lemma 1.23 (Moments of Pareto distribution)**.** *Let $X \sim \mathcal{P}(1, \gamma)$, $Y \sim \mathcal{P}(x_m, \gamma)$ be two Pareto random variables, with $\gamma > k$, $k \in \mathbb{N}^\star$. Subsequently, the Pareto random variables X and Y admit the finite first k moments.*

(i) *Moments of order k ($\gamma > k$):*

$$\mathbb{E}\left[X^k\right] = \frac{\gamma}{\gamma - k} \quad \textit{and} \quad \mathbb{E}\left[Y^k\right] = x_m^k \mathbb{E}\left[X^k\right]. \tag{1.110}$$

(ii) *Central moment of order k ($\gamma > k$):*

$$\mathbb{E}\left[(X - \mathbb{E}[X])^k\right] = \sum_{j=0}^{k} \binom{k}{j} \frac{\gamma^{j+1}}{(1-\gamma)^j(\gamma + j - k)} \quad \textit{and} \quad \mathbb{E}\left[(Y - \mathbb{E}[Y])^k\right] = x_m^k \mathbb{E}\left[(X - \mathbb{E}[X])^k\right]. \tag{1.111}$$

These properties hold for all statistics computed on the Pareto distribution, particularly for the $\mathbf{VaR}_\alpha$ and the $\mathbf{ES}_\alpha$. Let $X \sim \mathcal{P}(1, \gamma)$ be a standardized Pareto random variable and $Y \sim \mathcal{P}(x_m, \gamma)$, $Z_1 \sim \mathcal{P}(x_m^{(1)}, \gamma)$ and $Z_2 \sim \mathcal{P}(x_m^{(2)}, \gamma)$ be non-standardized Pareto random variables. The scaling property states that the non-standardized Pareto distribution $\mathcal{P}(x_m, \gamma)$ is proportional to the standardized Pareto distribution $\mathcal{P}(1, \gamma)$, with a proportionality factor equal to x_m. This implies that the $\mathbf{VaR}_\alpha$ of the non-standardized Pareto distribution $\mathcal{P}(x_m, \gamma)$ is proportional to the $\mathbf{VaR}_\alpha$ of the standardized Pareto

distribution with a proportionality coefficient equal to the scaling parameter x_m. Similarly, the scaling property of the Pareto distribution states that the non-standardized Pareto distribution $\mathcal{P}(x_m^{(1)}, \gamma)$ is proportional to the non-standardized Pareto distribution $\mathcal{P}(x_m^2, \gamma)$, with a proportionality factor equal to the ratio between the two scaling parameters $\frac{x_m^{(1)}}{x_m^{(2)}}$. This implies that the $\mathbf{VaR}_\alpha$ of the non-standardized Pareto distribution $\mathcal{P}(x_m^{(1)}, \gamma)$ is proportional to the $\mathbf{VaR}_\alpha$ of the non-standardized Pareto distribution $\mathcal{P}(x_m^{(2)}, \gamma)$ with a proportionality coefficient equal to $\frac{x_m^{(1)}}{x_m^{(2)}}$.

On the other hand, the property of stability by conditioning states that the standardized Pareto distribution $\mathcal{P}(1, \gamma)$ conditional on its values being above a certain threshold, is still a Pareto distribution with the same shape parameter γ but a new scale parameter equal to the conditioning parameter. Moreover, the scaling property implies that the standardized Pareto distribution $\mathcal{P}(1, \gamma)$ conditional on its values being above a certain threshold, is proportional to the standardized Pareto distribution with a proportionality factor equal to the conditioning threshold. The Expected-Shortfall of the standardized Pareto distribution is the conditional expectation given that the standard Pareto random variable X is larger than $\mathbf{VaR}_\alpha(X)$: $\mathbf{ES}_\alpha(X) = \mathbb{E}\left[X \mid X \geq \mathbf{VaR}_\alpha(X)\right]$ (defined by (4.3)). Combining the stability by conditioning and the scaling properties leads to a proportionality relationship between the standardized Pareto distribution $\mathcal{P}(1, \gamma)$ conditional on its values being above $\mathbf{VaR}_\alpha(X)$, and the marginal distribution which is the standardized Pareto distribution $\mathcal{P}(1, \gamma)$, with a proportionality factor equal to the conditioning threshold $\mathbf{VaR}_\alpha(X)$. Consequently, the Expected-Shortfall of the standardized Pareto distribution $\mathbf{ES}_\alpha(X)$ is proportional to its expectation $\mathbb{E}\left[X\right]$ with a proportionality factor equal to $\mathbf{VaR}_\alpha(X)$, which is supposed to be known. Therefore, $\mathbf{ES}_\alpha(X)$ can be easily recovered owing to the simple scaling by $\mathbf{VaR}_\alpha$ of the expectation of the standardized Pareto distribution.

Similarly, the property of stability by conditioning states that the non-standardized Pareto distribution $Z_1 \sim \mathcal{P}(x_m^{(1)}, \gamma)$ conditional on its values being above a certain threshold, is still a Pareto distribution with the same shape parameter γ but a new scale parameter equal to the conditioning parameter. Moreover, the scaling property implies that the non-standardized Pareto distribution $\mathcal{P}(x_m^{(1)}, \gamma)$ conditional on its values being above a certain threshold, is proportional to the standardized Pareto distribution $\mathcal{P}(1, \gamma)$ with a proportionality factor equal to the conditioning parameter, and is also proportional to the marginal Pareto distribution $\mathcal{P}(x_m^{(1)}, \gamma)$ with a proportionality factor equal to the ratio between the conditioning parameter and the scaling parameter $x_m^{(1)}$. Combining the stability by conditioning and the scaling properties implies that the Expected-Shortfall of the non-standardized Pareto distribution $\mathcal{P}(x_m^{(1)}, \gamma)$, which is the expectation of the distribution $\mathcal{P}(x_m^{(1)}, \gamma)$ conditional on its values being above $\mathbf{VaR}_\alpha(Z_1)$, is proportional to the expectation of the standardized Pareto distribution $\mathcal{P}(1, \gamma)$ with a proportionality factor equal to conditioning parameter $\mathbf{VaR}_\alpha(Z_1)$, and is proportional to the expectation of the marginal Pareto distribution $\mathcal{P}(x_m^{(1)}, \gamma)$ with a proportionality factor equal to the ratio $\frac{\mathbf{VaR}_\alpha(Z_1)}{x_m^{(1)}}$. Therefore, the Expected-Shortfall of any non-standardized Pareto distribution $\mathcal{P}(x_m^{(1)}, \gamma)$ can be recovered by a simple scaling of the expectation of the standardized Pareto distribution $\mathcal{P}(1, \gamma)$ with the conditioning

threshold $\mathbf{VaR}_\alpha(Z_1)$, and by a simple scaling of the expectation of the marginal Pareto distribution $\mathcal{P}(x_m^{(1)}, \gamma)$ with the ratio $\frac{\mathbf{VaR}_\alpha(Z_1)}{x_m^{(1)}}$.

All these properties are mathematically formulated in the following proposition.

Proposition 1.24 (Pareto $\mathbf{VaR}_\alpha$ and $\mathbf{ES}_\alpha$)**.** *If $X \sim \mathcal{P}(1, \gamma)$ and $Y \sim \mathcal{P}(x_m, \gamma)$, $x_m > 0, \gamma > 0$, then their respective $\mathbf{VaR}_\alpha$ are defined as follows:*

$$\begin{cases} \mathbf{VaR}_\alpha(X) & = F_X^{-1}(\alpha) = (1-\alpha)^{-\frac{1}{\gamma}}, \\ \mathbf{VaR}_\alpha(Y) & = F_Y^{-1}(\alpha) = x_m \mathbf{VaR}_\alpha(X), \end{cases} \quad \alpha \in (0,1), \tag{1.112}$$

and the respective $\mathbf{ES}_\alpha$ are given by:

$$\begin{cases} \mathbf{ES}_\alpha(X) = \mathbf{VaR}_\alpha(X)\mathbb{E}\left[X\right] = \frac{\gamma}{\gamma-1}(1-\alpha)^{-\frac{1}{\gamma}}, \\ \mathbf{ES}_\alpha(Y) = x_m \mathbf{ES}_\alpha(X), \end{cases} \quad \alpha \in (0,1). \tag{1.113}$$

As previously mentioned, owing to the scaling property of the Pareto distribution, all the statistics of any non-standardized Pareto distribution are proportional to the corresponding statistics of the standardized Pareto distribution, with a proportionality factor equal to the scaling parameter of the target Pareto distribution. Consequently, in the sequel, all the statistics will be computed using the standardized Pareto distribution. A simple scaling of these standardized statistics by the proper scaling parameter of the target Pareto distribution will allow for the recovery of the corresponding statistics of the desired non-standardized target Pareto distribution.

As previously explained, owing to the property of stability by conditioning, the Expected-Shortfall of the standardized Pareto distribution is proportional to its expectation, with a proportionality factor equal to the Value-at-Risk of the standardized Pareto distribution $\mathbf{VaR}_\alpha$, which is supposed to be known. Therefore, the main step in computing the Expected-Shortfall of the standardized Pareto distribution, is to compute its expectation. The Expected-Shortfall of the standardized Pareto distribution can be recovered thanks to a simple scaling of this expectation by the $\mathbf{VaR}_\alpha$, which is supposed to be known.

Issue We wonder how to accurately estimate the Expected-Shortfall. The presence of extreme values with low probability in heavy-tailed distributions can influence the estimator (empirical mean) of the expectation and lead to poor performance. In this context, the goal is to find an estimator of ES that is more resistant to the presence of extreme values. In this chapter, we assume that $\mathbf{VaR}_\alpha$ is known and focus on the methods of estimation of the $\mathbf{ES}_\alpha$. Different mean estimators are studied based on their concentration and fluctuation properties.

Owing to the scaling property of the Pareto distribution that establishes a proportionality relationship between the non-standardized Pareto distribution $\mathcal{P}(x_m, \gamma)$ and the standardized Pareto distribution $\mathcal{P}(1, \gamma)$ with a proportionality factor equal to x_m, it suffices to estimate the ES on the standardized Pareto distribution $\mathcal{P}(1, \gamma)$; then

we can recover the estimation of the ES on the non-standardized Pareto distribution $\mathcal{P}(x_m, \gamma)$ by a simple scaling of the standardized ES with x_m. In addition, owing to the stability by conditioning and the scaling properties of the Pareto distribution that establish a proportionality relationship between the standardized Pareto distribution $\mathcal{P}(1, \gamma)$ conditional on its values being above a certain threshold and the standardized Pareto distribution, with a proportionality factor equal to the conditioning threshold, estimating the first order moment of the standardized Pareto distribution is sufficient to obtain the estimation of the standardized $\mathbf{ES}_\alpha$, because we are able to recover the estimation of the standardized $\mathbf{ES}_\alpha$ by a simple scaling of the standardized first order moment with $\mathbf{VaR}_\alpha$.

Ideally, an estimator is expected to satisfy the following constraints: be robust to heavy tailed-distribution, that is, to the presence of extreme values, and reach a high-level of accuracy with a high-level of confidence. It is noteworthy that this criterion is non-asymptotic.

These constraints are satisfied by sub-Gaussian estimators. Despite the advantages of sub-Gaussian estimators, they also present some limitations. Indeed, proving that an estimator is sub-Gaussian is not so obvious because it relies on restrictive assumptions on the models, or it requires some tips to develop a new estimator such as Median-of-Means (see later in this chapter). Moreover, some sub-Gaussian estimators might have a higher computational complexity than simpler estimators. This can be a drawback when dealing with large datasets or in real-time applications.

For these reasons, our works propose several robust mean-estimators that are not sub-Gaussian, for which a study of their bias with the convergence rate is carried out and concentration inequalities are developed.

Mean-estimators The attention is mainly focused on three estimators: the Median-of-Means (MoM) estimator, the Trimmed-Mean (TM) estimator and the Lee-Valiant (LV) estimator. The performances of these estimators are then compared with each other and with the estimator of reference, that is, the empirical mean which reaches sub-Gaussian performance in the asymptotic framework.

Recall that the true $\mathbf{ES}_\alpha$ is the average excess loss beyond the true $\mathbf{VaR}_\alpha$ defined as follows:

$$\mathbf{ES}_\alpha(X) = \mathbb{E}[X|X \geq \mathbf{VaR}_\alpha] = \frac{\mathbb{E}[X\mathbb{1}_{\{X \geq \mathbf{VaR}_\alpha\}}]}{\mathbb{P}(X \geq \mathbf{VaR}_\alpha)}. \tag{1.114}$$

An estimator of the $\mathbf{ES}_\alpha$, called the empirical $\mathbf{ES}_\alpha$, is the average excess loss beyond the estimator of $\mathbf{VaR}_\alpha$, which is the empirical α-quantile ($\mathbf{q}_{n,\alpha}$), as stated in the following definition.

Definition 1.25 (Empirical Expected-Shortfall)**.** *Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sequence of Pareto i.i.d. random variables such that $\forall i \in [\![1, n]\!], X_i \sim \mathcal{P}(1, \gamma)$, with $\gamma > 2$. Let's denote by $\boldsymbol{X}^\star = (X_1^\star, \ldots, X_n^\star)$ the order statistics related to $\mathbf{X}$. The empirical α-quantile*

is given by $q_\alpha^n = X^\star_{\lceil n\alpha \rceil}$*, and the empirical Expected-Shortfall at the risk level* α *is defined as follows :*

$$\mathbf{ES}_\alpha^n = \frac{\frac{1}{n}\sum_{i=1}^n X_i \mathbb{1}_{\{X_i \geq q_\alpha^n\}}}{\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{X_i \geq q_\alpha^n\}}} = \frac{1}{n - \lceil n\alpha \rceil} \sum_{i=1}^n X_i \mathbb{1}_{\{X_i \geq q_\alpha^n\}}. \tag{1.115}$$

In our study we distinguish two frameworks for the the estimation of the $\mathbf{ES}_\alpha$ in the standardized Pareto distribution $\mathcal{P}(1,\gamma)$: the first is called *idealized case* and the other is called *realistic case*.

Idealized case: $q_\alpha^n = \mathbf{VaR}_\alpha$ We consider n independent and identically distributed samples $X_1, \ldots, X_n$ following a standardized Pareto distribution $\mathcal{P}(1,\gamma)$. The *idealized case* assumes that the empirical α-quantile q_α^n matches the true $\mathbf{VaR}_\alpha$: although only true in the asymptotic framework but not in practice for small sample sizes, we will use this assumption in some parts of our study to simplify the analysis. In such a case, the estimation of the $\mathbf{ES}_\alpha$ in the standardized Pareto distribution $\mathcal{P}(1,\gamma)$ corresponds to the empirical average excess loss above the true $\mathbf{VaR}_\alpha$. In other words, the estimation of the $\mathbf{ES}_\alpha$ in the standardized Pareto distribution $\mathcal{P}(1,\gamma)$ corresponds to the empirical average of the standardized Pareto distribution $\mathcal{P}(1,\gamma)$ conditional on its values being above the true $\mathbf{VaR}_\alpha$. In this case, the conditioning threshold does not depend on the underlying sample. This implies that the samples larger than the true $\mathbf{VaR}_\alpha$ are still independent and identically distributed (i.i.d.) and the stability by conditioning and scaling properties are valid. Owing to the stability by conditioning property, the standardized Pareto distribution tail above the true $\mathbf{VaR}_\alpha$ is still a Pareto distribution with the same shape parameter γ but a new scaling parameter equal to the conditioning threshold $\mathbf{VaR}_\alpha$: $\mathcal{P}(\mathbf{VaR}_\alpha, \gamma)$. Moreover, the scaling property establishes a proportionality relationship between the non-standardized Pareto distribution $\mathcal{P}(\mathbf{VaR}_\alpha, \gamma)$ and the standardized Pareto distribution $\mathcal{P}(1,\gamma)$, with a proportionality factor equal to $\mathbf{VaR}_\alpha$. Therefore, there exists a proportionality relationship between the standardized Pareto distribution conditional on its values being above the true $\mathbf{VaR}_\alpha$ and the marginal distribution $\mathcal{P}(1,\gamma)$, with a proportionality factor equal to the conditioning threshold $\mathbf{VaR}_\alpha$. This implies that the expectation of the standardized Pareto distribution conditional on its values being above the true $\mathbf{VaR}_\alpha$ (that is, the $\mathbf{ES}_\alpha$) is proportional to the expectation of the standardized Pareto distribution $\mathcal{P}(1,\gamma)$, with a proportionality factor equal to $\mathbf{VaR}_\alpha$. Similarly, the empirical mean of the standardized Pareto distribution conditional on its values being above the true $\mathbf{VaR}_\alpha$ (that is the empirical $\mathbf{ES}_\alpha$) is proportional to the empirical mean of the standardized Pareto distribution $\mathcal{P}(1,\gamma)$, with a proportionality factor equal to $\mathbf{VaR}_\alpha$. Consequently, the bias between the empirical $\mathbf{ES}_\alpha$ and the true $\mathbf{ES}_\alpha$ of the standardized Pareto distribution is proportional to the bias between the expectation and the empirical mean of the standardized Pareto distribution, with a proportionality factor equal to $\mathbf{VaR}_\alpha$. Thus, in the idealized case, it suffices to determine a closed form formula for the bias between the empirical mean and the expectation of the standardized Pareto distribution $\mathcal{P}(1,\gamma)$. The bias between the empirical $\mathbf{ES}_\alpha$ and the true $\mathbf{ES}_\alpha$ of the standardized Pareto distribution can be recovered thanks to a simple scaling of the bias between the empirical mean and

the expectation of the standardized Pareto distribution with the conditioning threshold $\mathbf{VaR}_\alpha$. The empirical α-quantile matches the true $\mathbf{VaR}_\alpha$ only in the asymptotic case (that is, when the sample size is sufficiently large), but in most cases, when the sample size is far from the asymptotic case, the empirical α-quantile does not match the true $\mathbf{VaR}_\alpha$.

Realistic case: $q_\alpha^n \neq \mathbf{VaR}_\alpha$ In the *realistic case*, the empirical α-quantile does not match the theoretical $\mathbf{VaR}_\alpha$. In this last case, the estimation of the $\mathbf{ES}_\alpha$ in the standardized Pareto distribution $\mathcal{P}(1,\gamma)$ corresponds to the empirical mean of the excess loss above the empirical α-quantile. In other words, the estimation of the $\mathbf{ES}_\alpha$ in the standardized Pareto distribution $\mathcal{P}(1,\gamma)$ corresponds to the empirical mean of the standardized Pareto distribution conditional on its values being above the empirical α-quantile. In this case, the conditioning threshold depends on the underlying sample. This implies that the samples larger than the empirical α-quantile are no longer independent and identically distributed (i.i.d.) and the stability by conditioning and scaling properties of the Pareto distribution are no longer valid. Therefore, the standardized Pareto distribution tail above the empirical α-quantile is not necessarily a Pareto distribution. Thus, the distribution of the samples larger than the empirical α-quantile is unknown and it is difficult to derive a closed form formula for the bias between the empirical $\mathbf{ES}_\alpha$ and the true $\mathbf{ES}_\alpha$. For this reason, in the realistic case, we provide an empirical study of the bias between the empirical $\mathbf{ES}_\alpha$ and the true $\mathbf{ES}_\alpha$, and the convergence rate. In the realistic case, the bias between the empirical $\mathbf{ES}_\alpha$ and the true $\mathbf{ES}_\alpha$ is larger than in the idealized case because there is an additional error term corresponding to the bias estimation between the true $\mathbf{VaR}_\alpha$ and the empirical α-quantile.

In all the chapter, the estimators of the $\mathbf{ES}_\alpha$ are studied both in the *idealized case*, in which the empirical α-quantile matches the true $\mathbf{VaR}_\alpha$, and in the *realistic case* in which the two quantities differ.

Because the bias of an estimator is given by the difference between the target value and the estimator's expectation, the challenge of computing the bias lies in the computation of the estimator's expectation.

Median-of-Means (MoM)

We attempted to derive a closed-form formula for the bias of the MoM estimator. This issue is challenging, and remains an open question. In this context, we develop a methodology to derive a closed-form formula for the bias of the MoM estimator, and we prove the intermediary results in this procedure, such as the derivation of a closed form formula for the bias of the empirical median.

The MoM estimator combines the empirical mean and median estimators. The Median-of-Means estimator requires partitioning the data into k groups of roughly equal size, computing the empirical mean in each group, and taking the median of the sequence composed by the k empirical means.

Definition 1.26 (Median-of-Means (**MoM**)). *Let $X_1, \dots, X_n$ be n independent, identically, distributed (i.i.d.) random draws from the distribution of X. Let m, k be two positive integers. We assume that n is a multiple of k such that $n = mk$. The empirical mean of each block is defined as follows:*

$$\forall j \in [\![1, k]\!], \quad \bar{\mu}_{B_j} = \frac{1}{\left|B_j\right|} \sum_{i \in B_j} X_i. \tag{1.116}$$

The **MoM** *estimator is then defined by* $\widehat{MoM}_n = M(\bar{\mu}_1, \dots, \bar{\mu}_k)$.

Owing to the way in which the estimator is constructed, the MoM estimator can be considered a good estimator of the mean.

First step in the construction of the MoM estimator: It consists of partitioning the set composed of n i.i.d. random draws into k disjoint blocks, and computing the empirical mean $\bar{\mu}_j$ of each of the k blocks. For the sake of simplicity, let us assume that the number of random draws n is a multiple of the number of blocks k such that $n = mk$, with m, k, n beging positive integers. For each of the k blocks, the empirical mean is an unbiased estimator of the expectation of the distribution. Indeed, because the random draws X_i are i.i.d. with mean $\mu = \mathbb{E}[X]$, then $\mathbb{E}[\bar{\mu}_j] = \mu$. Moreover, because the random draws are i.i.d. with variance $\mathbb{V}[X_i] = \sigma^2$ then, the standard deviation of the empirical means $\bar{\mu}_j$ is given by $\frac{\sigma}{\sqrt{m}}$. This means that, in each of the k blocks, the empirical mean does not deviate from the expectation of the distribution by more than a few units of $\frac{\sigma}{\sqrt{m}}$.

Second step it consists of taking the median of the sequence composed of the k block-wise empirical means.

Why can we say that the median of block-wise means is a good estimator of the expectation of the distribution ?

Actually the empirical median does not deviate from the empirical mean by more than the empirical standard deviation, this is the following statement.

Lemma 1.27. *Let X be a scalar random variable, with standard deviation σ and median M. Then, the following inequality holds:*

$$\left|\mathbb{E}\left[X\right] - M\right| \leq \sigma. \tag{1.117}$$

Assuming that σ is sufficiently small, we can say that the median is close enough to the expectation of the distribution to be a good estimator.

In our case, if $\mathbf{X} = (X_1, \dots, X_n)$ is a sample composed of n independent and identically distributed (i.i.d.) random variables, then $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^n X_i$ is the related empirical mean. Let us denote by $B_j, \forall j \in [\![1, k]\!]$ the k blocks composed of m random variables. The block-wise empirical mean $\bar{\mu}_{B_j} = \frac{1}{\left|B_j\right|}\sum_{i \in B_j} X_i$ is an unbiased estimator of the expectation μ of the distribution. Moreover, because the initial random draws are i.i.d.

and the blocks are disjoint, the block-wise means are i.i.d. random variables. The empirical mean of the block-wise means is equal to that of the initial sample $\hat{\mu}_n$. The empirical variance is denoted by $\hat{\sigma}_n^2 = \frac{1}{k}\sum_{j=1}^{k}(\bar{\mu}_{B_j} - \hat{\mu}_n)^2$. Then from Equation (1.117), the following concentration inequality states :

$$\left|\hat{\mu}_n - \widehat{\mathbf{MoM}}_n\right| \leq \hat{\sigma}_n. \tag{1.118}$$

If the empirical standard deviation of the sample is small, we can conclude that the Median-of-Means estimator and empirical mean estimator are close to each other, and thus they should both estimate the true expectation. Nonetheless, in general, the fluctuations of the MoM estimator are smaller than those of the EM estimator.

Advantages of the MoM estimator: This estimator allows the combination of two centrality estimators: the empirical mean and empirical median. The empirical mean is not robust to extreme values and is significantly influenced by their presence. Therefore, the k block-wise empirical means consider the extreme values; they are unbiased with respect to the expectation of the sequence, but they have a large standard deviation. The empirical median is robust to extreme values because it depends only on the central values of the sequence, and the standard deviation of the empirical median is expected to be much lower than that of the empirical mean; however, the median may exhibit a bias with respect to the expectation of the sequence. This implies that the final estimator can be sensitive to the data while not being too influenced by the extreme values of the distribution. Moreover, because the second layer of construction of the estimator relies on taking the empirical median of the sequence of block-wise empirical means, a bias is introduced with respect to the expectation of the sequence.

Concentration inequalities are provided by the works of (Lugosi and Mendelson, 2019, Theorem 2 p.7, Theorem 4 p.9), proving that under conditions that are not so restrictive, the MoM estimator is a good estimator for the mean.

After choosing the proper parameterization of the estimator, we quantify the bias of the MoM estimator. Since the MoM estimator results from the combination of the empirical mean and empirical median estimators, then we demonstrate in a first time, a closed form-formula for the bias of the empirical median.

Theorem 1.28 (Expectation of empirical median when n is odd)**.** (i) *Let $X_1, \ldots, X_{2k+1}$ be $2k+1$ i.i.d. random variables that follow the same distribution as $X \sim \mathcal{P}(1, \gamma)$. Let $X_1^\star, \ldots, X_{2k+1}^\star$ be the related sequence of ordered statistics. The density of the empirical median $X_{k+1}^\star$ is given by:*

$$f_{\hat{M}_{2k+1}^{(1)}}(x) = \frac{(2k+1)!}{k!k!}\left(1 - \frac{1}{x^\gamma}\right)^k \frac{1}{x^{k\gamma}}\left(\frac{\gamma}{x^{\gamma+1}}\right)\mathbb{1}_{\{x \geq 1\}}. \tag{1.119}$$

The expectation of the empirical median is defined as follows:

$$\mathbb{E}[\hat{M}_{2k+1}^{(1)}] = \frac{(2k+1)!}{k!k!}\mathfrak{B}\left(k+1, k+1-\frac{1}{\gamma}\right) \tag{1.120}$$

where $\mathfrak{B}(x,y) = \int_0^1 (1-t)^{x-1}t^{y-1}\mathrm{d}t = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$ *is the Beta function, and* $\Gamma(z) = \int_0^{+\infty} t^{z-1}e^{-t}\mathrm{d}t$ *is the Gamma function.*

(ii) *Let* $Y_1, \ldots, Y_{2k+1}$ *be* $2k+1$ *i.i.d. random variables that follow the same distribution as* $Y \sim \mathcal{P}(x_m, \gamma)$*. Then, from the scaling property of the Pareto distribution (Theorem 1.22), we obtain the following proportionality relationship between the distributions* $\mathcal{P}(x_m, \gamma)$ *and* $\mathcal{P}(1, \gamma)$*:*

$$(Y_1, \ldots, Y_{2k+1}) \overset{d}{=} (x_m X_1, \ldots, x_m X_{2k+1}). \tag{1.121}$$

In particular, a proportionality relationship of factor x_m *links the empirical median of the non-standardized Pareto distribution* $\mathcal{P}(x_m, \gamma)$ *and that of the standardized Pareto distribution* $\mathcal{P}(1, \gamma)$ *such that* $\hat{M}_{2k+1}^{(x_m)} \overset{d}{=} x_m \hat{M}_{2k+1}^{(1)}$*, as well as their expectations:*

$$\mathbb{E}[\hat{M}_{2k+1}^{(x_m)}] = x_m \mathbb{E}[\hat{M}_{2k+1}^{(1)}]. \tag{1.122}$$

Theorem 1.29 (Expectation of empirical median when n is even)**.** (i) *Let* $X_1, \ldots, X_{2k}$ *be* $2k$ *independent and identically distributed (i.i.d.) random variables following the same distribution as* $X \sim \mathcal{P}(1, \gamma)$*. Let* $X_1^\star, \ldots, X_{2k}^\star$ *be the related sequence of ordered statistics. We define*

$$\hat{M}_{2k} := \frac{1}{2}(X_k^\star + X_{k+1}^\star). \tag{1.123}$$

The expectation of the empirical median is given by:

$$\mathbb{E}\left[\hat{M}_{2k}\right] = \frac{1}{2}\left[\frac{(2k)!}{(k-1)!k!}\mathfrak{B}\left(k, k+1-\frac{1}{\gamma}\right) + \frac{(2k)!}{k!(k-1)!}\mathfrak{B}\left(k+1, k-\frac{1}{\gamma}\right)\right] \tag{1.124}$$

(ii) *Let* $Y_1, \ldots, Y_{2k}$ *be* $2k$ *i.i.d. random variables that follow the same distribution as* $Y \sim \mathcal{P}(x_m, \gamma)$*. Then, from the scaling property of the Pareto distribution (Theorem 1.22), we obtain the following proportionality relationship between the distributions* $\mathcal{P}(x_m, \gamma)$ *and* $\mathcal{P}(1, \gamma)$*:*

$$(Y_1, \ldots, Y_{2k+1}) \overset{d}{=} (x_m X_1, \ldots, x_m X_{2k+1}). \tag{1.125}$$

In particular, a proportionality relationship of factor x_m *links the empirical median of the non-standardized Pareto distribution* $\mathcal{P}(x_m, \gamma)$ *and that of the standardized Pareto distribution* $\mathcal{P}(1, \gamma)$ *such that* $\hat{M}_{2k}^{(x_m)} \overset{d}{=} x_m \hat{M}_{2k}^{(1)}$*, as well as their expectations:*

$$\mathbb{E}[\hat{M}_{2k}^{(x_m)}] = x_m \mathbb{E}[\hat{M}_{2k}^{(1)}]. \tag{1.126}$$

Theorem 1.30. *As* $n \to +\infty$*,*

$$\mathbb{E}\left[\hat{M}_n\right] = \sqrt[\gamma]{2}\left(1 + \frac{1}{2\gamma n} + \frac{1}{2\gamma^2 n} + o\left(\frac{1}{n}\right)\right). \tag{1.127}$$

The challenge in calculating the bias of the MoM is that the density of the estimator is required to compute the expectation of the MoM. However, this estimator is defined as the empirical median of the block-wise means. The distribution of an order statistic is based on the distribution of the underlying sample, that is, on the sample of block-wise means whose distribution is unknown. To address this issue, we propose a closed-form formula for the characteristic function of block-wise means.

Theorem 1.31 (Standard Pareto characteristic function)**.** *Let $X \sim \mathcal{P}(1,\gamma)$ with $\gamma > 1$. The characteristic function of X, given by $\phi_X(t) = \mathbb{E}\left[e^{itX}\right] = \int_1^{+\infty} \frac{\gamma}{x^{\gamma+1}} \mathrm{d}x$ is a solution of the Ordinary Differential Equation (ODE):*

$$\frac{\partial}{\partial t} J_\gamma(t) - \frac{\gamma}{t} J_\gamma(t) = -\frac{\gamma}{t} e^{it}, \tag{1.128}$$

and takes the following form:

$$\forall t \in \mathbb{R}, \phi_X(t) = \gamma(-it)^\gamma \Gamma(-\gamma, -it). \tag{1.129}$$

where $\Gamma(z) = \int_0^{+\infty} u^{z-1} e^{-u} \mathrm{d}u$, and satisfies the following relationship $\Gamma(z+1) = z\Gamma(z)$.

From Equation (4.161), the characteristic function of the block-wise mean $\bar{\mu}_j$ can be defined as follows:

$$\phi_{\bar{\mu}_j}(t) = \gamma^m \left(-i\frac{t}{m}\right)^{m\gamma} \left(\Gamma\left(-\gamma, -i\frac{t}{m}\right)\right)^m. \tag{1.130}$$

Then, the density function of the block-wise mean can be recovered using the inverse Fourier transform, as follows:

$$f_{\bar{\mu}_j}(x) = \int_{\mathbb{R}} \frac{1}{2\pi} e^{-itx} \phi_{\bar{\mu}_j}(t) \mathrm{d}t. \tag{1.131}$$

However, because the characteristic function of the block-wise mean $\bar{\mu}_j$ depends on the Gamma function, which is already an integral, it is very complicated to determine an explicit analytic form for the density of the block-wise mean. In this framework, it is difficult to obtain an analytic form of the bias between the **MoM** estimator applied to the standardized Pareto distribution and its expectation. This is still a work in progress.

A numerical study on the bias of the $\widehat{\mathbf{MoM}}_n$ estimator is conducted.

The convergence rate of the bias between the **MoM** estimator applied to the complete standardized Pareto distribution and its expectation seems to be dependent on the Pareto index γ. Indeed, the convergence rate varies between $\frac{1}{n^{0.7}}$ and $\frac{1}{n}$ and becomes closer to $\frac{1}{n}$ when the distribution tail becomes increasingly thinner, that is, when γ is increasing.

Once the evolution of the bias between the **MoM** estimator and the expectation of the entire standardized Pareto distribution $\mathcal{P}(1,\gamma)$, as well as its convergence speed, have been studied, we are interested in studying the evolution of the bias between the **MoM** estimator applied to the standardized Pareto distribution tail and the $\mathbf{ES}_\alpha$ in two cases: an *idealized case* in which the conditioning threshold corresponds to the theoretical $\mathbf{VaR}_\alpha$, and a *realistic case* in which the conditioning threshold corresponds to the empirical α-quantile.

Idealized case: In the idealized case, the empirical α-quantile matches the true $\mathbf{VaR}_\alpha$, which is known. In this case, the estimator of the $\mathbf{ES}_\alpha$ in the standardized Pareto distribution $\mathcal{P}(1,\gamma)$ is the empirical average of the standardized Pareto distribution conditional on its values being above the true $\mathbf{VaR}_\alpha$. The conditioning threshold is independent on the underlying sample. This implies that the samples larger than the true $\mathbf{VaR}_\alpha$ are independent and identically distributed (i.i.d.) and the stability by conditioning and scaling properties, as stated in Theorem 1.22, are still valid. The stability by conditioning property implies that the standardized Pareto distribution $\mathcal{P}(1,\gamma)$ conditional on its values being above $\mathbf{VaR}_\alpha$ is still a Pareto distribution, with the same shape parameter γ, but a new scaling parameter equal to the conditioning parameter $\mathbf{VaR}_\alpha$. The scaling property states that this non-standardized Pareto distribution $\mathcal{P}(\mathbf{VaR}_\alpha,\gamma)$ is proportional to the marginal Pareto distribution $\mathcal{P}(1,\gamma)$, with a proportionality factor equal to the conditioning parameter $\mathbf{VaR}_\alpha$. Therefore, the standardized Pareto distribution conditional on its values being above $\mathbf{VaR}_\alpha$ is proportional to the standardized Pareto distribution with a proportionality factor equal to $\mathbf{VaR}_\alpha$. This implies that the **MoM** estimator applied to the standardized Pareto distribution conditional on its values being above $\mathbf{VaR}_\alpha$, is proportional to the **MoM** estimator applied to the entire standardized Pareto distribution, with a proportionality factor equal to $\mathbf{VaR}_\alpha$. In the same way, the $\mathbf{ES}_\alpha$ is proportional to the expectation of the standardized Pareto distribution, with a proportionality factor equal to $\mathbf{VaR}_\alpha$. Thus, the bias between the **MoM** estimator applied to the standardized Pareto distribution conditional on its values being above $\mathbf{VaR}_\alpha$ and the $\mathbf{ES}_\alpha$ is proportional to the bias between the **MoM** estimator applied to the entire standardized Pareto distribution and the expectation, with a proportionality factor equal to $\mathbf{VaR}_\alpha$. Moreover, because the conditioning threshold does not depend on the underlying sample, the convergence rate of the bias between the **MoM** estimator applied to the standardized Pareto distribution conditional on its values being above the true $\mathbf{VaR}_\alpha$ and the true $\mathbf{ES}_\alpha$ is the same as that of the bias between the **MoM** estimator applied to the entire standardized Pareto distribution and its expectation.

Realistic case: In the realistic case, the empirical α-quantile does not match the true $\mathbf{VaR}_\alpha$. In this case, the estimator of the $\mathbf{ES}_\alpha$ in the standardized Pareto distribution $\mathcal{P}(1,\gamma)$ is the empirical average of the standardized Pareto distribution conditional on its values being above the empirical α-quantile. The conditioning threshold is an order statistic and depends on the underlying sample. This implies that the samples larger than the empirical α-quantile are no longer independent and identically distributed (i.i.d.). Therefore, the distribution of the samples larger than the empirical α-quantile is not necessarily a Pareto distribution. The stability by conditioning and scaling properties of the Pareto distribution, as stated in Theorem 1.22, are no longer valid. Thus, the distribution of the samples larger than the empirical α-quantile is unknown and it is difficult to establish an analytic closed-form formula for the bias between the **MoM** estimator applied to the standardized Pareto distribution tail above this empirical α-quantile and the true $\mathbf{ES}_\alpha$. For this reason, we provide some experimental study to give an insight about the convergence speed of the bias between the **MoM** applied to

the Pareto distribution tail above the empirical α-quantile and the Expected-Shortfall. The numerical study shows that the convergence rate is very close to that of the bias in the idealized case.

Trimmed-Mean (TM)

We provide a closed-form formula for the bias of the TM estimator in both the non-asymptotic case and asymptotic cases.

The Trimmed-Mean estimator is based on the following idea. The most natural attempt to improve the performance of the empirical mean estimator is to remove possible outliers by truncating the sample. The TM estimator is defined by removing a fraction of the sample, consisting of the ϵn largest and smallest points for some parameter $\epsilon \in (0,1)$, and then averaging over the rest.

The TM estimator is constructed as follows. We split data into two equal parts. One half is used to determine the correct truncation level. The points from the other half are averaged, except for the data points that fall outside the truncation region, which are thresholded.

For convenience of notation, we assume that the data are composed of $2n$ independent copies of random variable $X\mathcal{P}(1,\gamma)$, denoted by $X_1,\dots,X_n,Y_1,\dots,Y_n$. We respectively denote by $X_1^\star,\dots,X_n^\star$ and $Y_1^\star,\dots,Y_n^\star$ the order statistics sequences related to samples $X_1,\dots,X_n =: \mathbf{X}_n$ and $Y_1,\dots,Y_n =: \mathbf{Y}_n$. The expectation of the distribution is denoted as $\mu = \mathbb{E}[X] = \frac{\gamma}{\gamma-1}$. Truncation on both sides of the distribution is relevant when both tails of the distribution are unbounded and contain extreme values. However, the Pareto distribution presents the particularity that its left tail is bounded, whereas its right tail is not bounded. This implies that to eliminate the outliers, it is not required to threshold the left tail of the distribution, but only the right tail.

In this specific case, the truncation function is given by:

$$\phi_{\mathbf{Y}_n}(x) = \begin{cases} Y^\star_{\lceil (1-\epsilon)n \rceil} & \text{if } x > Y^\star_{\lceil (1-\epsilon)n \rceil}, \\ x & \text{if } x \in \left[1, Y^\star_{\lceil (1-\epsilon)n \rceil}\right]. \end{cases} \tag{1.132}$$

Moreover, it is important to note that the truncation thresholds are fixed on a sample that is independent of the sample on which the estimator is built. With this notation in place, the estimator is defined as follows:

(i) Given the confidence level $\delta \geq 8e^{-3n/16}$, set:

$$\epsilon_n = \frac{16\log(8/\delta)}{3n}. \tag{1.133}$$

(ii) Set:

$$\widehat{TM}_{2n} = \frac{1}{n}\sum_{i=1}^{n} \phi_{\mathbf{Y}_n}(X_i). \tag{1.134}$$

The drawback of this estimator is that, for a high confidence level, the amount of data required data to obtain a reasonable truncation threshold is very large.

In (Lugosi and Mendelson, 2019, p.14), a concentration inequality of the Trimmed-Mean estimator is given to quantify the performance of the estimator.

In this context, we provide a closed-form formula for the bias of the TM estimator in the non-asymptotic case.

We study the evolution of the bias between the Trimmed-Mean estimator applied to the standardized Pareto distribution tail and the $\mathbf{ES}_\alpha$ in two different frameworks. The first framework refers to the idealistic case where the empirical α-quantile match the theoretical $\mathbf{VaR}_\alpha$. The second framework corresponds to the realistic case where the empirical α-quantile does not match the theoretical $\mathbf{VaR}_\alpha$. To this end, we first study the bias between the Trimmed-Mean estimator applied to the entire standardized Pareto distribution $\mathcal{P}(1,\gamma)$ and the expectation.

(i) The bias between the Trimmed Mean estimator and the expectation of the Pareto distribution is defined by:

$$B_\mu(\widehat{TM}_{2n}) = \mathbb{E}\left[\widehat{TM}_{2n}\right] - \mu. \tag{1.135}$$

Proposition 1.32 (Bias of the Trimmed-Mean estimator in the standardized Pareto distribution)**.** *Let $X_1,\dots,X_n,Y_1,\dots,Y_n$ and $X_1^\star,\dots,X_n^\star,Y_1^\star,\dots,Y_n^\star$, be respectively $2n$ independent copies of the random variable X that follows a standardized Pareto distribution $\mathcal{P}(1,\gamma)$, and the related order statistics. Let $\epsilon_n \in (0,1)$ be the truncation threshold that satisfies Equation (1.133). The bias between the Trimmed-Mean estimator applied to the entire standardized Pareto distribution $\mathcal{P}(1,\gamma)$ and the expectation is then given as follows:*

$$B_\mu(\widehat{TM}_{2n}) = -\frac{1}{\gamma-1}\mathbb{E}\left[\left(Y^\star_{\lceil(1-\epsilon)n\rceil}\right)^{1-\gamma}\right] = -\frac{1}{\gamma-1}\kappa_{\epsilon,n}\mathfrak{B}\left(n-\left\lceil(1-\epsilon)n\right\rceil+2-\frac{1}{\gamma},\left\lceil(1-\epsilon)n\right\rceil\right). \tag{1.136}$$

where $\kappa_{\epsilon,n} = \frac{n!}{\left(\left\lceil(1-\epsilon)n\right\rceil-1\right)!\left(n-\left\lceil(1-\epsilon)n\right\rceil\right)!}$ *and* $\mathfrak{B}(x,y) = \int_0^1 u^{x-1}(1-u)^{y-1}\mathrm{d}u = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$.

We also provide a closed-form formula for the bias of the TM estimator in an asymptotic case.

Theorem 1.33 (Asymptotic bias of the Trimmed-Mean estimator in the standardized Pareto distribution)**.** *As $n \to +\infty$,*

$$B_\mu(\widehat{TM}_{2n}) = -\frac{1}{\gamma-1}\frac{\Gamma\left(\lfloor C_\delta\rfloor+2-\frac{1}{\gamma}\right)}{\Gamma(\lfloor C_\delta\rfloor+1)}n^{\frac{1}{\gamma}-1} + o\left(n^{\frac{1}{\gamma}-1}\right). \tag{1.137}$$

where for all $x\in\mathbb{R}$, $\Gamma(x) = \int_0^{+\infty} u^{x-1}e^{-u}\mathrm{d}u$ is the Gamma function.

A closed-form formula for the bias between the Right-Trimmed-Mean estimator applied to the entire standardized Pareto distribution $\mathcal{P}(1,\gamma)$ and the expectation has been provided. Recall that the goal is to study the bias between the Right-Trimmed-Mean estimator applied to the standardized Pareto distribution tail and the $\mathbf{ES}_\alpha$.

Idealized case: In the idealized case, the empirical α-quantile matches the true $\mathbf{VaR}_\alpha$, and the empirical $\mathbf{ES}_\alpha$ corresponds to the empirical average of the standardized Pareto distribution conditional on its values being above the true $\mathbf{VaR}_\alpha$. The conditioning threshold is independent on the underlying sample. This implies that the samples larger than the true $\mathbf{VaR}_\alpha$ are independent and identically distributed (i.i.d.) and the stability by conditioning and scaling properties are valid. Therefore, the standardized Pareto distribution conditional on its values being above the true $\mathbf{VaR}_\alpha$ is still a Pareto distribution with the same shape parameter γ but a new scale parameter equal to the conditioning threshold $\mathbf{VaR}_\alpha$: $\mathcal{P}(\mathbf{VaR}_\alpha, \gamma)$. Moreover, the scaling property establishes a proportionality relationship between the non-standardized Pareto distribution $\mathcal{P}(\mathbf{VaR}_\alpha, \gamma)$ and the standardized Pareto distribution $\mathcal{P}(1, \gamma)$ with a proportionality factor equal to the scaling parameter $\mathbf{VaR}_\alpha$. Consequently, combining the stability by conditioning and scaling properties, the standardized Pareto distribution conditional on its values being above the true $\mathbf{VaR}_\alpha$ is proportional to the marginal distribution $\mathcal{P}(1, \gamma)$ with a proportionality factor equal to $\mathbf{VaR}_\alpha$.

This implies that the Right-Trimmed-Mean estimator applied to the standardized Pareto distribution conditional on its values being above $\mathbf{VaR}_\alpha$, is proportional to the Right-Trimmed-Mean estimator applied to the entire standardized Pareto distribution, with a proportionality factor equal to $\mathbf{VaR}_\alpha$. In the same way, the $\mathbf{ES}_\alpha$ is proportional to the expectation of the standardized Pareto distribution, with a proportionality factor equal to $\mathbf{VaR}_\alpha$. Thus, the bias between the Right-Trimmed-Mean estimator applied to the standardized Pareto distribution conditional on its values being above $\mathbf{VaR}_\alpha$ and the $\mathbf{ES}_\alpha$ is proportional to the bias between the Right-Trimmed-Mean estimator applied to the entire standardized Pareto distribution and the expectation, with a proportionality factor equal to $\mathbf{VaR}_\alpha$. Moreover, because the conditioning threshold does not depend on the underlying sample, the convergence rate of the bias between the Right-Trimmed-Mean estimator applied to the standardized Pareto distribution conditional on its values being above the true $\mathbf{VaR}_\alpha$ and the true $\mathbf{ES}_\alpha$ is the same as that of the bias between the Right-Trimmed-Mean estimator applied to the entire standardized Pareto distribution and its expectation. The closed-form formula of the bias provides the convergence rate as stated in Theorem 1.33. The convergence speed varies as a function of the Pareto index γ and satisfies the rule $n^{\frac{1}{\gamma}-1}$, multiplied by the proper constant.

Realistic case: In the realistic case, the empirical α-quantile does not match the true $\mathbf{VaR}_\alpha$, and the empirical $\mathbf{ES}_\alpha$ corresponds to the empirical average of the standardized Pareto distribution conditional on its values being above the empirical α-quantile. The conditioning threshold is an order statistic and depends on the underlying sample. This implies that the samples larger than the empirical α-quantile are no longer independent and identically distributed (i.i.d.) and the stability by conditioning and scaling properties are no longer valid. Therefore, the standardized Pareto distribution conditional on its values being above the empirical α-quantile is not necessarily a Pareto distribution. Thus, the distribution of the samples larger than the empirical α-quantile is unknown and a closed-form formula for the bias between the true $\mathbf{ES}_\alpha$ and the empirical $\mathbf{ES}_\alpha$ is tricky to be established. For this reason, numerical experiments give an idea on the

convergence speed of the bias and show whether the convergence speed of the bias in the realistic case is far from or close to the convergence speed of the bias in the idealistic case. The numerical study shows that the convergence rate of the bias between the empirical $\mathbf{ES}_\alpha$ and the true $\mathbf{ES}_\alpha$ in the realistic case seems to be very close to that of the bias between the empirical $\mathbf{ES}_\alpha$ and the true $\mathbf{ES}_\alpha$ in the idealized case.

Lee-Valiant (LV)

The Lee-Valiant estimator is an improved version of the MoM estimator, which includes a correction term. Because a closed-form formula for the bias of the MoM estimator is not available, it is difficult to derive a closed-form formula for the bias of the LV estimator. For this reason, we provide an empirical study that allows us to understand the behavior of the bias between the LV estimator applied to the standardized Pareto distribution tail and the true $\mathbf{ES}_\alpha$, and determine whether the convergence rate of the bias depends on the Pareto index.

Definition 1.34 (Lee-Valiant estimator)**.** *For a given δ, define the Median-of-Mean estimator $\widehat{\mathbf{MoM}}_n = \mathbf{MoM}(X_1, \dots, X_n)$, computed on $k = log\left(\frac{1}{\delta}\right) \leq n$ blocks with $\delta \geq e^{-n}$, and k an integer. The Lee-Valiant estimator is then defined as*

$$\widehat{LV}_n = \widehat{\mathbf{MoM}}_n + \frac{1}{n}\sum_{i=1}^{n}(X_i - \widehat{\mathbf{MoM}}_n)(1 - \min(\alpha(X_i - \widehat{\mathbf{MoM}}_n)^2, 1)) \tag{1.138}$$

where the parameter α is the solution of the monotonic, piecewise-linear equation

$$\sum_{i=1}^{n}\min(\alpha(X_i - \widehat{\mathbf{MoM}}_n)^2, 1) = \frac{1}{3}\log\left(\frac{1}{\delta}\right). \tag{1.139}$$

The zero of the equation:

$$\sum_{i=1}^{n}\min(\alpha(X_i - \widehat{\mathbf{MoM}}_n)^2, 1) - \frac{1}{3}\log\left(\frac{1}{\delta}\right) = 0 \tag{1.140}$$

can be solved using a dichotomy algorithm.

In (Gobet et al., 2022, p.14, Thm. 2.5), a deviation result is provided. See (Lee and Valiant, 2022) for more details about the Lee-Valiant estimator.

The following survey aims at studying the evolution of the bias between the Lee-Valiant estimator applied to the standardized Pareto distribution tail and the $\mathbf{ES}_\alpha$ in two different frameworks. The first framework refers to the idealized case where the empirical α-quantile matches the theoretical $\mathbf{VaR}_\alpha$. The second framework corresponds to the realistic case where the empirical α-quantile does not match the theoretical $\mathbf{VaR}_\alpha$.

To this end, we first study the bias between the Lee-Valiant estimator applied to the entire standardized Pareto distribution $\mathcal{P}(1, \gamma)$ and the expectation. The numerical study shows that the convergence rate of the bias between the LV estimator applied to the entire standardized Pareto distribution $\mathcal{P}(1, \gamma)$ and its expectation is larger than $\frac{1}{n}$ and lower than $\frac{1}{\sqrt{n}}$.

Once the evolution of the bias between the $\widehat{\mathbf{LV}}_n$ estimator and the expectation of the entire standardized Pareto distribution $\mathcal{P}(1,\gamma)$, as well as its convergence speed, have been studied, we are interested in studying the evolution of the bias between the $\widehat{\mathbf{LV}}_n$ estimator applied to the standardized Pareto distribution tail and the $\mathbf{ES}_\alpha$ in two cases: an *idealized case* in which the conditioning threshold corresponds to the theoretical $\mathbf{VaR}_\alpha$, and a *realistic case* in which the conditioning threshold corresponds to the empirical α-quantile.

Idealized case: In the idealized case, the empirical α-quantile matches the true $\mathbf{VaR}_\alpha$, which is known. In this case, the estimator of the $\mathbf{ES}_\alpha$ in the standardized Pareto distribution $\mathcal{P}(1,\gamma)$ is the empirical average of the standardized Pareto distribution conditional on its values being above the true $\mathbf{VaR}_\alpha$. The conditioning threshold is independent on the underlying sample. This implies that the samples larger than the true $\mathbf{VaR}_\alpha$ are independent and identically distributed (i.i.d.) and the stability by conditioning and scaling properties, as stated in Theorem 1.22, are still valid. The stability by conditioning property implies that the standardized Pareto distribution $\mathcal{P}(1,\gamma)$ conditional on its values being above $\mathbf{VaR}_\alpha$ is still a Pareto distribution, with the same shape parameter γ, but a new scaling parameter equal to the conditioning parameter $\mathbf{VaR}_\alpha$. The scaling property states that this non-standardized Pareto distribution $\mathcal{P}(\mathbf{VaR}_\alpha,\gamma)$ is proportional to the marginal Pareto distribution $\mathcal{P}(1,\gamma)$, with a proportionality factor equal to the conditioning parameter $\mathbf{VaR}_\alpha$. Therefore, the standardized Pareto distribution conditional on its values being above $\mathbf{VaR}_\alpha$ is proportional to the standardized Pareto distribution with a proportionality factor equal to $\mathbf{VaR}_\alpha$. This implies that the $\widehat{\mathbf{LV}}_n$ estimator applied to the standardized Pareto distribution conditional on its values being above $\mathbf{VaR}_\alpha$, is proportional to the $\widehat{\mathbf{LV}}_n$ estimator applied to the entire standardized Pareto distribution, with a proportionality factor equal to $\mathbf{VaR}_\alpha$. In the same way, the $\mathbf{ES}_\alpha$ is proportional to the expectation of the standardized Pareto distribution, with a proportionality factor equal to $\mathbf{VaR}_\alpha$. Thus, the bias between the $\widehat{\mathbf{LV}}_n$ estimator applied to the standardized Pareto distribution conditional on its values being above $\mathbf{VaR}_\alpha$ and the $\mathbf{ES}_\alpha$ is proportional to the bias between the $\widehat{\mathbf{LV}}_n$ estimator applied to the entire standardized Pareto distribution and the expectation, with a proportionality factor equal to $\mathbf{VaR}_\alpha$. Moreover, because the conditioning threshold does not depend on the underlying sample, the convergence rate of the bias between the $\widehat{\mathbf{LV}}_n$ estimator applied to the standardized Pareto distribution conditional on its values being above the true $\mathbf{VaR}_\alpha$ and the true $\mathbf{ES}_\alpha$ is the same as that of the bias between the $\widehat{\mathbf{LV}}_n$ estimator applied to the entire standardized Pareto distribution and its expectation.

Realistic case: In the realistic case, the empirical α-quantile does not match the true $\mathbf{VaR}_\alpha$. In this case, the estimator of the $\mathbf{ES}_\alpha$ in the standardized Pareto distribution $\mathcal{P}(1,\gamma)$ is the empirical average of the standardized Pareto distribution conditional on its values being above the empirical α-quantile. The conditioning threshold is an order statistic and depends on the underlying sample. This implies that the samples lar-

ger than the empirical α-quantile are no longer independent and identically distributed (i.i.d.). Therefore, the distribution of the samples larger than the empirical α-quantile is not necessarily a Pareto distribution. The stability by conditioning and scaling properties of the Pareto distribution, as stated in Theorem 1.22, are no longer valid. Thus, the distribution of the samples larger than the empirical α-quantile is unknown and it is difficult to establish an analytic closed-form formula for the bias between the $\widehat{\mathbf{LV}}_n$ estimator applied to the standardized Pareto distribution tail above this empirical α-quantile and the true $\mathbf{ES}_\alpha$. For this reason, we provide some experimental study to give an insight about the convergence speed of the bias between the $\widehat{\mathbf{LV}}_n$ applied to the Pareto distribution tail above the empirical α-quantile and the Expected-Shortfall. The convergence rate of the bias between the $\widehat{\mathbf{LV}}_n$ estimator applied to the standardized Pareto distribution conditional on its values begin above the empirical α-quantile and the true $\mathbf{ES}_\alpha$ seems to be similar to that of the bias in the idealized case, that is between $\frac{1}{n}$ and $\frac{1}{\sqrt{n}}$. More precisely, tests have been performed for three Pareto distributions $\mathcal{P}(1, 2.5)$, $\mathcal{P}(1, 3.5)$ and $\mathcal{P}(1, 5)$ and the convergence rate is between $\frac{1}{n^{0.9}}$ and $\frac{1}{n^{0.7}}$.

Comparative study Finally, we support the theoretical analysis with some experiments. We also compare the performance of the different estimators.

Bias between the empirical $\mathbf{ES}_\alpha$ and true $\mathbf{ES}_\alpha$ in the realistic case (work in progress) As previously mentioned, in the realistic case, the empirical α-quantile does not match the true $\mathbf{VaR}_\alpha$. In this case, the estimator of the $\mathbf{ES}_\alpha$ in the standardized Pareto distribution $\mathcal{P}(1, \gamma)$ is the empirical average of the standardized Pareto distribution conditional on its values being above the empirical α-quantile. The conditioning threshold is an order statistic and depends on the underlying sample. This implies that the samples larger than the empirical α-quantile are no longer independent and identically distributed (i.i.d.). Therefore, the distribution of the samples larger than the empirical α-quantile is not necessarily a Pareto distribution. The stability by conditioning and scaling properties of the Pareto distribution, as stated in Theorem 1.22, are no longer valid. Thus, the distribution of the samples larger than the empirical α-quantile is unknown and it is more challenging to establish an analytic closed-form formula for the bias between the empirical $\mathbf{ES}_\alpha$ and the true $\mathbf{ES}_\alpha$ in the standardized Pareto distribution. We provide a work in progress about the expression of this bias. The empirical $\mathbf{ES}_\alpha$ formula is given as follows:

$$\mathbf{ES}_\alpha^n = \frac{1}{n - \lceil \alpha n \rceil} \sum_{i=1}^{n} X_i \mathbb{1}_{\{X_i \geq q_\alpha^n\}}. \tag{1.141}$$

In the sake of simplicity, let us denote by $r = \lceil \alpha n \rceil$ the rank of the empirical quantile. Then, the empirical **ES** corresponds to the empirical average of the samples larger than the order statistics of order r and can be rewritten as $\mathbf{ES}_r^n = \frac{1}{n-r} \sum_{i=1}^{n} X_i \mathbb{1}_{\{X_i \geq X_r^\star\}}$. The goal is to compute the bias between the empirical **ES** and the theoretical **ES** in

the standardized Pareto distribution $\mathcal{P}(1,\gamma)$.

$$B_{\mathbf{ES}_r}[\mathbf{ES}_r^n] = \mathbb{E}\left[\mathbf{ES}_r^n\right] - \mathbf{ES}_r \quad \text{where} \quad \mathbb{E}\left[\mathbf{ES}_r^n\right] = \frac{1}{n-r}\sum_{i=1}^{n}\mathbb{E}\left[X_i\mathbb{1}_{\{X_i\geq X_r^\star\}}\right] = \frac{n}{n-r}\mathbb{E}\left[X_1\mathbb{1}_{\{X_1\geq X_r^\star\}}\right] \tag{1.142}$$

since the sample is i.i.d.. Under the uniform distribution, we prove the form of the joint distribution:

$$\mathbb{P}(X_r^\star \leq u, X_i \leq v) = v\sum_{k=r}^{n-1}\binom{n-1}{k}u^k(1-u)^{n-1-k} + \min(u,v)\binom{n-1}{r-1}u^{r-1}(1-u)^{n-r}. \tag{1.143}$$

Proceeding to the following quantile transformation $\mathbf{VaR}_\alpha(X_i) = Y_i$ in Equation (4.246), we adapt the above formula for the standardized Pareto distribution:

$$\mathbb{P}(X_r^\star \leq u, X_i \leq v) = \mathbb{P}(Y_r^\star \leq \mathbf{VaR}_\alpha(u), Y_i \leq \mathbf{VaR}_\alpha(v)). \tag{1.144}$$

1.4 Perspectives

Based on our works, diverse perspectives can be envisaged.

Perspectives of Chapter 2 In Chapter 2, stationary and self-similar processes are studied in the framework of the $\mathbb{L}^2$-spaces. Especially, in Gaussian framework, we show that the kernel proposed by (Mandelbrot and Van Ness, 1968a) satisfies the assumptions of stationarity and self-similarity. An improvement would be to prove the uniqueness of this kernel.

Future works could extend the stationarity and self-similarity properties to the general framework of multidimensional $\mathbb{L}^2$-spaces based on the approaches of (Lavancier et al., 2009) and (Laha and Rohatgi, 1981).

A closed-form formula for the symmetrized cross-covariance kernel of the multivariate fractional Brownian motion has been provided. An improvement would be to provide a closed-form formula not only for the symmetrized cross-coveriance kernel but also for the cross-covariance itself, as in the works of (Lavancier et al., 2009) and (Laha and Rohatgi, 1981).

The fractional models used in our works are based on constant Hurst exponents. In order to improve the performance of the fractional models, it could be relevant to add time-dependency or time-randomness in the Hurst exponent. See for instance (Muniandy and Lim, 2001).

Moreover, the fractional models could be improved by introducing time-dependence or time-randomness in the volatility to obtain for instance local volatility or stochastic volatility to better capture the price variations. Volatility dynamics could also be described thanks to a fractional Brownian motion as in (Gatheral et al., 2017), introducing the self-similarity property of the volatility, and thus long-range (respectively short-range) dependence.

Another idea would be to integrate Hawkes processes in the fractional models in order to take into account jumps in the price dynamics. See for instance (Hainaut, 2020).

Moreover, it would be interesting to add a corrective term in the fractional Black-Scholes models so that the fractional models might become semi-martingales and satisfy the Absence of Opportunity of Arbitrage (AOA). See (Cheridito, 2003), (Cheridito, 2001), and (Rostek and Schöbel, 2013).

Perspectives on Chapter 3 In Chapter 3, we propose a predictive model for the conditional Value-at-Risk of the Gaussian approximation at the first order of the future portfolio variation, given the past variations of the assets. The model can be extended to the conditional Expected-Shortfall of the Gaussian approximation at the first order of the future portfolio variation, given the past variations of the assets. The quantification of the error of approximation can be provided based on the one of the Value-at-Risk.

A considerable improvement would be to provide a Taylor expansion instead of an upper bound of the error of approximation. The works of (Talay and Tubaro, 1990a) and (Talay and Tubaro, 1990b), (Frikha and Huang, 2015) are useful to fulfill this task. In (Frikha and Huang, 2015) authors assume that:

$$\begin{cases} \mathbb{P}(L^h \leq l) - \mathbb{P}(L \leq l) = c_1^0 h + c_2^0 h^2 + \ldots + O(h^k) \\ f_{L^h} - f_L = c_1^1 h + c_2^1 h^2 + \ldots + O(h^k) \\ \partial f_{L^h} - \partial f_L = c_1^2 h + c_2^2 h^2 + \ldots + O(h^k) \\ \vdots \\ \partial^{(M)} f_{L^h} - \partial^{(M)} f_L = c_1^{M+1} h + c_2^{M+1} h^2 + \ldots + O(h^k) \end{cases} \tag{1.145}$$

where h represents here the error of approximation.

Based on the above system of equations, it comes that:

$$\mathbf{VaR}_\alpha(L) - \mathbf{VaR}_\alpha(L^h) = \tilde{c_1} h + \tilde{c_2} h^2 + \ldots + O(h^k). \tag{1.146}$$

A Taylor expansion of the future portfolio variation $\delta_h P_{t_N}$ can be provided using Ito formula. Then, we can exploit the semi-group development as mentioned in (Bally and Rey, 2016), and Malliavin calculus to provide an expansion of VaR. Malliavin calculus is used to study expectation of functions that are not regular like the indicator functions. Then, Malliavin calculus can be used in the framework of the VaR expansion to compute the cumulative distribution function $\mathbb{P}(X \leq x) = \mathbb{E}\left[\mathbb{1}_{\{X \leq x\}}\right]$. See for instance (Bally, 2007). In the framework of the VaR expansion, we have to identify the first error term, the results of (Huang et al., 2015) are useful. Indeed, they state that if a Taylor expansion exist for the cumulative distribution functions and their derivatives, a Taylor expansion of the density exists. An expansion of the density also exists in the case of the EDS approximations (Bally and Talay, 1996). The Euler scheme provides an expansion of the density. Moreover, the results of (Huang et al., 2015) rely on the weak approximation on regular test functions. These methods are inspired by parametric expansions and Euler schemes.

Futhermore, an interesting question would be the following: How does Gaussian approximation of portfolio variation behave in the framework of option pricing ?

Also, in Chapter 3, the parameters of the VaR model H and σ are estimated thanks to first order estimators by linear regression of the log-quadratic variations versus the log-increment sizes. A method based on quadratic variations applied to other filtration (wavelets, second order) could be used to estimate with efficiency the Hurst exponent H for $H > 0.75$. See for instance (Coeurjolly, 2001) or (Szymanski and Takabatake, 2023), (Chong et al., 2022b), (Chong et al., 2022a).

In our works, the multivariate fractional Brownian motion (mfBm) is generated thanks to the Cholesky method whose complexity is in $\mathcal{O}((dN)^3)$. To improve the complexity, an interesting method based on circulating block matrix developed in (Coeurjolly and Philippe, 2010) could be used. This would reduce the computation cost from $\mathcal{O}((dN)^3)$ to $\mathcal{O}((d^2 N \log(N)))$.

The VaR model provided in Chapter 3 has been backtested at horizon equal to one day and at risk level equal to $\alpha = 0.99$, on stock indices, of length equal to one year, daily sampled. This model could be backtested in other frameworks. For instance we could consider other assets such as FX-rates, volatility indices, stocks, bonds, commodities. We could consider other frequency sampling for the data such as high-frequency data. The number of assets d in the portfolio could be increased.

An interesting perspective could be to establish asymptotic properties for the predicted VaR.

Perspectives of Chapter 4 Finally, in Chapter 4, we have studied various non-asymptotic robust estimators to estimate Expected-Shortfall in heavy-tailed distributions. Closed-form formulae have been provided for the bias of the median and of the Trimmed-Mean estimators. It would be interesting to develop closed-form formulae for the bias of Median-of-Means and Lee-Valiant estimators.

Also, we could extend the study of these robust estimators to the Generalized Pareto distribution as well as to other heavy-tailed distributions.

"Pure mathematics is, in its way, the poetry of logical ideas."

(**Albert Einstein**, The World As I See It, 1949.)

Chapter 2

Introduction to self-similar and stationary Gaussian processes

Contents

This chapter addresses research question ($RQ\#1$). In this chapter, we focus on two time-transformations: time-translation (or time-origin change) and time-scaling, and on the related properties called stationarity and self-similarity. Stationarity is the invariance in the time and space of a function or process by time-origin change (or temporal translation). Self-similarity establishes a spatial proportionality relationship between a function or a process taken at two proportional times t and λt with $\lambda > 0$, and the spatial proportionality factor is a function of λ. We present a new approach for studying stationary and self-similar processes in Hilbert space. First, we prove that the only assumptions of stationarity and self-similarity of the quadratic norm of a given process in Hilbert space, without any assumption of distribution, allow obtaining the stationarity and self-similarity of the inner product of the process, with a closed-form formula for the latter. Second, we show that adding the Gaussian assumption allows obtaining the stationarity and self-similarity properties of the process not only in distribution but also in the trajectory sense (i.e., in terms of equality of processes). We provide examples of such processes, known as the Wiener process and fractional Brownian motion (fBm). Finally, we provide an extension of the stationarity and self-similarity properties in the trajectory sense, to multidimensional Gaussian processes, called multivariate fractional Brownian motions (mfBm).

2.1 Introduction

In this chapter, we consider the $\mathbb{L}^2$-subspaces spanned by families of random variables (r.v.) $\{X(\theta), \theta \in J\}$, denoted as $\mathcal{H}_X$. $\mathcal{H}_X$ is the Hilbert subspace of all the finite linear combinations $\alpha \cdot X(\theta) = \sum_1^n \alpha_i X(\theta_i)$ and their limits in $\mathbb{L}^2$; which is characterized by the $\mathbb{L}^2$-characteristics of $\{X(\theta)\}$, given by the *quadratic norm* of the components (i.e., the square of the norm) $Q_X(\theta) := \mathbf{Q}(X(\theta)) = \|X(\theta)\|_2^2 = \mathbb{E}(|X(\theta)|^2)$ and by the inner products $K_X(\theta, \theta') = \langle\langle X(\theta), X(\theta')\rangle\rangle := \mathbb{E}(X(\theta)\, X(\theta'))$, a dependency indicator.

Random families indexed in real time $\{X(t), t \in \mathbb{R}\}$ (or random processes) arouse particular interest. Their $\mathbb{L}^2$-characteristics $Q_X(t)$ and $K_X(t,s)$ can be complex. In classical approaches, additional properties have been introduced to reduce their complexity. If the properties are only true "component by component", (t by t), only the quadratic norm family $\{Q_X(t)\}$ is concerned, and we are discussing the Q_X-property.
The classical time-transformations applied to $\{X(t)\}$ are:
TRANSLATION OF ORIGIN, $\{X_h(t) = X(t+h) - X(h), (h \in \mathbb{R})\}$. Then, $\mathbb{L}^2$-*stationarity* expresses that the processes $\{X_h(t)\}$ and $\{X(t)\}$ have the same $\mathbb{L}^2$-characteristics for any h, while the Q_X-*stationarity* expresses only that for any t and h, $\mathbf{Q}(X(t+h)) - X(h)) = Q_X(t)$. Fortunately, this weaker hypothesis is sufficient to compute the kernel $K_X(s,t)$, and to deduce the $\mathbb{L}^2$-stationarity.
CHANGE OF SCALE: $\{X(t) \to X(\lambda t), (\lambda > 0)\}$. The $\mathbb{L}^2$-*self-similarity* supposes the existence of a non-negative function $(\Theta(\lambda), \lambda \in \mathbb{R}^+)$ such that processes $\{X(\lambda t)\}$ and $\{\Theta^{\frac{1}{2}}(\lambda)X(t)\}$ have the same $\mathbb{L}^2$-characteristics. The Q_X-self-similarity only assumes that for any t and $\lambda > 0$, $Q_X(\lambda t) = \Theta(\lambda) Q_X(t)$. Thus, these two concepts are not equivalent. The Q_X-self-similarity is sufficient to show that $\Theta(\lambda)$ is a power function such that $\Theta(\lambda) = \lambda^\gamma, \gamma > 0$, and thus to fully characterize the quadratic norm owing to this power function but not the inner kernel.

When these two Q_X-properties are satisfied simultaneously, it is shown (*Theorem 1*) that the inner kernel is fully characterized as a linear combination of power functions of exponent $\gamma \in (0,1)$ (in place of positive, as above) and the process $\{X(t)\}$ is both $\mathbb{L}^2$-stationary and self-similar.

$$Q_X(t) = Q_X(1) \left|t\right|^{2\gamma} \text{ and } K_X(t,s) = \frac{Q_X(1)}{2}(|t|^{2\gamma} + |s|^{2\gamma} - |t-s|^{2\gamma}) \text{ with } \gamma \in (0,1). \tag{2.1}$$

When $\gamma = 1/2$, it becomes $Q_X(t) = Q_X(1)\left|t\right|, K_X(t,s) = Q_X(1)(|t| \wedge |s|)$.
In the next section, $\mathbb{L}^2$-families $\{X(t)\}$ are considered under the additional hypothesis of a *Gaussian distribution* for any finite linear combination. Therefore, the Hilbert space $\mathcal{H}_X$ is a Gaussian space, where the Gaussian processes are now identified by their distribution, characterized by their $\mathbb{L}^2$-characteristics **and** their expectations $\{m_X(t) = \mathbb{E}(X(t))\}$.
Nevertheless, Gaussian processes that are simultaneously self-similar and stationary are necessarily centered (*Theorem 4*). Therefore, the results obtained in the general framework of the $\mathbb{L}^2$-spaces, can be applied without modification, (with the familiar vocabulary of variance and covariance kernel). The case $\gamma = 1/2$ is specific, because the disjoint increments of the processes are $\mathbb{L}^2$-orthogonal, and therefore independent in the Gaussian framework.

The Gaussian process with variance $|t|$ is known as Brownian motion (indexed by $\mathbb{R}$ and not $\mathbb{R}^+$ as usual) and is denoted $\{B(t)\}$. It is the best-known process of the family $\mathcal{G}_{ss}$ of self-similar and stationary Gaussian processes. Because they have the same Q_X-characteristics, the processes $\{B_h(t) = B(t+h) - B(h),\ h \in \mathbb{R}\}$ and $\{B^\lambda(t) = \frac{1}{\sqrt{\lambda}} B(\lambda t), \lambda > 0\}$, which also belong to $\mathcal{G}_{ss}$, are also Brownian motions.
The tool of stochastic integrals of deterministic functions $\phi \in \mathbb{L}^2(\text{Leb})$, $B(\phi) = \int \phi(u) dB(u)$

allows describing the Gaussian space $\mathcal{H}_B$, but also gives a variable change formula for the pathwise comparison between the r.v. $B(\phi), B_h(\phi), B^\lambda(\phi)$, as

$$\left\{ \begin{array}{l} B_h(\phi) = \int \phi(u) \mathrm{d}B_h(u) = \int \phi(u-h) \mathrm{d}B(u) \\ B^\lambda(\phi) = \int \phi(u) \mathrm{d}B^\lambda(u) = \frac{1}{\sqrt{\lambda}} \int \phi(u/\lambda) \mathrm{d}B(u). \end{array} \right. \tag{2.2}$$

All these variables are Gaussian variables with the same variance $\int |\phi(u)|^2 \mathrm{d}u$.

In a second part, we are concerned with the construction of $\mathcal{G}_{ss}$-processes as stochastic integrals of deterministic functions against the Brownian motion, where the functions $\phi(u)$ are replaced by bivariate kernels $\kappa(t,u)$ such that $\int |\kappa(t,u)|^2 du < \infty$. In line with Mandelbrot and Van Ness in 1968 (Mandelbrot and Van Ness, 1968b), a Path Dependent Brownian motion (PDBM), is defined as a Gaussian process, like $X^\kappa(t) = \int \kappa(t,u) dB(u)$. The variable change formula suggests transporting the time-change initially supported by variable t into a time-change on the integration variable u. For the PDBM process to be self-similar and stationary, it suffices that the kernel κ satisfies the system:

$$\left\{ \begin{array}{l} \kappa(t+h,u) - \kappa(h,u) = \kappa(t,u-h) \\ \kappa(\lambda t, \lambda u) = \eta(\lambda) \kappa\left(t,u\right) \quad \text{with} \quad \eta(\lambda) = \lambda^\nu \mathbb{1}_{\{\lambda>0\}},\ \nu \in \left(-\frac{1}{2}, \frac{1}{2}\right) \end{array} \right. \tag{2.3}$$

since then, $X^\kappa(\lambda t, B) = \sqrt{\lambda}\eta(\lambda) X^\kappa(t, B^\lambda)$, $\quad X_h^\kappa(t,B) = X^\kappa(t, B_h)$. From Equation (1), $\lambda\eta^2(\lambda) = 2\gamma$, which explains why $\nu \in \left(-\frac{1}{2}, \frac{1}{2}\right)$.

The last step is to find a bivariate kernel $\kappa(t,u)$, solution of System (3) and square integrable in u. The solution proposed by Mandelbrot and Van Ness ((Mandelbrot and Van Ness, 1968b)) is the fractional kernel κ^ν, which is clearly a solution of the previous system,

$$\kappa^\nu(t,u) = (t-u)^\nu \mathbb{1}_{\{u<t\}} - (0-u)^\nu \mathbb{1}_{\{u<0\}}, \quad \nu \in \left(-\frac{1}{2}, \frac{1}{2}\right). \tag{2.4}$$

The square integrability verifies that:

– for any value of ν, for u in the neighborhood of $-\infty$, $\int \left|\kappa(t,u)\right|^2 \mathrm{d}u < \infty$,

– for $\nu < 0$, for u in the neighborhood of 0 and for u in the neighborhood of t, $\int \left|\kappa(t,u)\right|^2 \mathrm{d}u < \infty$.

Some remarks guide the intuition for obtaining such a solution.

Finally, an extension to the multidimensional framework is provided, first for the d-dimensional Brownian motion, whose components are correlated Brownian motions with covariance matrix of $(\rho_{i,j}(t \wedge s))$. Multidimensional fractional Brownian motion is a self-similar multidimensional Gaussian process with stationary and correlated increments:

$$\left\{ \begin{array}{l} \boldsymbol{X}^{\boldsymbol{\kappa}}(t+h) - \boldsymbol{X}^{\boldsymbol{\kappa}}(h) = \int \boldsymbol{\kappa}(t,u) \odot \mathrm{d}\boldsymbol{B}_{\boldsymbol{h}}(u) = \boldsymbol{X}^{\boldsymbol{\kappa}}(t, B_h) \\ \boldsymbol{X}^{\boldsymbol{\kappa}}(\lambda t) = \sqrt{\lambda}\boldsymbol{\eta}(\lambda) \odot \int \boldsymbol{\kappa}(t,u) \odot \mathrm{d}\boldsymbol{B}^{\boldsymbol{\lambda}}(u) = \sqrt{\lambda}\boldsymbol{\eta}(\lambda) \odot \boldsymbol{X}^{\boldsymbol{\kappa}}(t, B^\lambda) \end{array} \right. \tag{2.5}$$

of variance-covariance kernel:

$$\begin{cases} K_{X^\kappa}((i,t),(j,t)) = K_{X^\kappa}((i,1),(j,1))\left|t\right|^{\gamma^i+\gamma^j}. \\ K_{X^\kappa}((i,t),(j,s)) + K_{X^\kappa}((i,s),(j,t)) = K_{X^\kappa}((i,1),(i,1))\left(\left|t\right|^{\gamma^i+\gamma^j} + \left|s\right|^{\gamma^i+\gamma^j} - \left|t-s\right|^{\gamma^i+\gamma^j}\right). \end{cases} \tag{2.6}$$

2.2 General background

Let us introduce the general background of the study.

2.2.1 The framework

The working framework is composed of a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$, (the $\mathbb{P}$-negligible sets belong to $\mathcal{F}$), and a measured space $(E, \mathcal{E}, \mu)$. A random variable (r.v.) X is by definition an $\mathcal{E}$-measurable application from Ω to E, such that for all $B \in \mathcal{E}$, $\{X \in B\} \in \mathcal{F}$. When $E = \mathbb{R}$, $\mathcal{E} := \mathcal{B}(\mathbb{R})$ is the Borelian σ-field and the r.v. X is called real random variable (r.v.r.).

The *distribution* of X is the probability measure on $(E, \mathcal{E})$ defined by $\mu_X(B) := \mathbb{P}(X \in B)$. Note that two r.v. X and Y almost surely equal, have the same distribution.

By extension, for any non-negative measurable application $\phi : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$:

$$\mathbb{E}\left[\phi(X)\right] := \int_\Omega \phi(X(\omega))\mathrm{d}\mathbb{P}(\omega) = \int_E \phi(x)\mathrm{d}\mu_X(x). \tag{2.7}$$

$\mathbb{L}^p(\mathbb{P})$-spaces.

In the real case, $E = \mathbb{R}$, the convex power functions $\phi_p(x) = |x|^p$, $(p \geq 1)$ allow defining the p-th-order moments of X, $\mathbb{E}\left[\left|X\right|^p\right] = \int_E |x|^p \,\mathrm{d}\mu_X(x)$. The set of equivalence classes of r.v. X such that $\mathbb{E}\left[\left|X\right|^p\right] < +\infty$, called $\mathbb{L}^p(\mathbb{P})$-space, ($\mathbb{L}^p$ when there is no ambiguity) is a Banach space for the norm $\left\|X\right\|_p$:

$$\mathbb{L}^p(\mathbb{P}) := \left\{X : \mathbb{E}\left[\left|X\right|^p\right] < \infty\right\} \quad \text{equipped with} \quad \left\|X\right\|_p \overset{def}{=} \left(\mathbb{E}\left[\left|X\right|^p\right]\right)^{\frac{1}{p}}. \tag{2.8}$$

As it is well-known, the norm function $X \to \left\|X\right\|_p$ is continuous in $\mathbb{L}^p$.
Owing to convexity inequalities, for $p \geq 1$, $p \mapsto \left\|X\right\|_p$ is non-decreasing and $\mathbb{L}^1$ is included in all $\mathbb{L}^p$-spaces. Then, when a sequence (X_n) converges to X in $\mathbb{L}^p$, (X_n) converges to X in $\mathbb{L}^r$, for $1 \leq r \leq p$, and the different norms also converge.

2.2.2 The $\mathbb{L}^2$-space

Generalities on the $\mathbb{L}^2$-space

– Special attention is given to the $\mathbb{L}^2$-space which is not only a Banach space, but also a Hilbert space, that is, a normed complete vector space, equipped with the *inner product* $\langle\!\langle X, Y \rangle\!\rangle = \mathbb{E}\left[X\,Y\right]$, and the norm $\|X\|_2 = \sqrt{\langle\!\langle X, X \rangle\!\rangle}$.

– Since $\mathbb{L}^1 \subset \mathbb{L}^2$, any $\mathbb{L}^2$-r.v. X is integrable, with expectation $m_X = \mathbb{E}\left[X\right]$ and variance $V_X = \mathbb{V}[X] = \mathbb{E}\left[(X - m_X)^2\right]$. The subset of centered r.v. $(m_X = 0)$ is denoted $\mathbb{L}^2_c$, with $\mathbb{L}^2$-norm $\sqrt{V_X}$. The two applications $X \mapsto m_X$ and $X \mapsto V_X$ are continuous in $\mathbb{L}^2$.

– Generally, it is convenient to replace the variance with the *quadratic norm*, defined as the square of the norm $\mathbf{Q}(X) = \|X\|_2^2 = \mathbb{E}\left[|X|^2\right] = |m_X|^2 + V_X$.

As in any vector space, most of the properties concern linear combinations of variables, $\alpha \cdot X = \sum_1^n \alpha_i X_i$, (with $\alpha = (\alpha_i)_{i=1}^n \in \mathbb{R}^n$ and $X = (X_i)_{i=1}^n$). Their quadratic norm, $\mathbf{Q}(\alpha \cdot X) = \sum_{(i,j)\in[1,n]} \alpha_i\,\alpha_j\,\langle\!\langle X_i,\, X_j \rangle\!\rangle$ is a quadratic function of $\alpha = (\alpha_i)_{i=1}^n \in \mathbb{R}^n$.

– *Orthogonality* plays a central role in Hilbert spaces. As usual, the $\mathbb{L}^2$-orthogonality of X and Y is defined by $\langle\!\langle X, Y \rangle\!\rangle = 0$. Generally, the $\mathbb{L}^2$-orthogonality of two $\mathbb{L}^2$-sub-Hilbert spaces $\mathcal{H}^1$ and $\mathcal{H}^2$ is defined by,

$$\mathcal{H}^1 \perp \mathcal{H}^2 \qquad \text{if} \quad \langle\!\langle X_1, X_2 \rangle\!\rangle = 0, \quad \forall X_1 \in \mathcal{H}^1, \quad X_2 \in \mathcal{H}^2 \tag{2.9}$$

Then, if the r.v. (X_i) are orthogonal, $\mathbf{Q}(\alpha \cdot X) = \sum_1^n \alpha_i^2\,\mathbf{Q}(X_i)$.

$\mathbb{L}^2$-space spanned by a family of random variables

Often, the data are a finite or infinite family of $\mathbb{L}^2$-r.v. $\{X(\theta)\}_{\theta\in J}$, with quadratic norms $Q_X(\theta) \overset{def}{=} \mathbf{Q}(X(\theta)) = \|X(\theta)\|_2^2$. This quadratic norm family is not sufficient to know the quadratic norms of linear combinations of such variables, and it is also convenient to define a dependency indicator, the *Inner kernel* $K_X(\theta, \theta') = \langle\!\langle X(\theta), X(\theta') \rangle\!\rangle$. Subsequently, the quadratic norm of $\alpha \cdot X$ becomes $\mathbf{Q}(\alpha \cdot X) = \sum_{(i,j)\in[1,n]} \alpha_i\,\alpha_j\,K_X(\theta_i, \theta_j)$.

Definition 2.1. *The $\mathbb{L}^2$-subspace, $\mathcal{H}_X$, spanned by the family $\{X(\theta)\}_{\theta\in J}$ is the smallest Hilbert space containing the family $\{X(\theta)\}$. $\mathcal{H}_X$ is the space of all the finite linear combinations of r.v. $X(\theta)$ and their limits. The $\mathbb{L}^2$-characteristics of $\mathcal{H}_X$ are the family $\{Q_X(\theta), K_X(\theta, \theta'), \theta, \theta' \in J\}$.*

2.2.3 Self-similar and stationary $\mathbb{L}^2$- Random process

In the sequel, special attention will be given to families of $\mathbb{L}^2$-random variables, indexed by real-time ($t \in \mathbb{R}$), also called random processes. It is an infinite dimension family, which implies a high complexity of its $\mathbb{L}^2$-characteristics (quadratic norm $Q_X(t)$, and inner kernel $K_X(s,t)$)). Therefore, we are interested in studying additional properties, that would reduce their complexity. However, sometimes, properties are only formulated on the components of the family, that is, t by t. In this case, only the quadratic norm family $\{Q_X(t)\}$ is considered. We are talking of Q_X-property.

Two main time-transformations will be at the heart of the study:

– the *time-origin change* corresponding to a translation of the origin, for any (t, h), $(X(t) \mapsto (X_h(t) = X(t+h) - X(h))$. $X_h(t)$ is the increment (variation) of X on the interval $I_{t,h} = (h, t+h)$,
– the *timescale change* corresponding to a homothety, $(X(t) \mapsto X(\lambda t),\ \lambda > 0)$.

Q_X and $\mathbb{L}^2$-Stationarity and Self-Similarity of random processes

Classical stability assumptions are made. The simplest is the stationarity.

$\mathbb{L}^2$-Stationarity

The stationarity property is related to the effects of time-origin change on the process. More precisely, it expresses invariance by the time-origin change (or time-translation) of the characteristics of the process.

Definition 2.2. (i) *A process $\{X(t)\}$ is said to be $\mathbb{L}^2$-stationary if for any $h \in \mathbb{R}$, the processes $\{X_h(t) = X(t+h) - X(h)\}$ and $\{X(t)\}$ have the same $\mathbb{L}^2$-characteristics, quadratic norm, and inner kernel*

$$\mathbf{Q}[X(t+h) - X(h)] = \mathbf{Q}[X(t)] = Q_X(t), \text{ and } \quad \langle\!\langle X_h(t), X_h(s) \rangle\!\rangle = K_X(t,s). \tag{2.10}$$

(ii) *The process $\{X(t)\}$ is said to be Q_X-stationary if only the first condition holds:*

$$\mathbf{Q}(X_h(t)) = \mathbf{Q}(X(t)) = Q_X(t) \quad .$$

The Q_X-stationarity implies some immediate properties of the random process:

Remark: (i) Taking $t = 0$ in the Equation (2.10). Since $\mathbf{Q}(0) = 0$, $\mathbf{Q}(X(0)) = Q_X(0) = 0$ and then $X(0) = 0, a.s.$.

(ii) The quadratic norm function is even, $Q_X(t) = Q_X(-t)$, because $\mathbf{Q}(X(t)) = \mathbf{Q}(X(t+h) - X(h)) = \mathbf{Q}(X(h) - X(t+h)) = \mathbf{Q}(X(-t))$.

The main result is given by the following proposition.

Proposition 2.3. *Any Q_X-stationary process is $\mathbb{L}^2$-stationary and thus, the processes $\{X_h(t)\}$ are also $\mathbb{L}^2$-stationary with the same $\mathbb{L}^2$-characteristics.*

The proof contains interesting results detailed in the next lemma.

Lemma 2.4. *Let $\{X(t)\}$ be Q_X-stationary, then the Inner kernel is explained by the quadratic norm Q_X as, for any $(t, h) \in \mathbb{R}^2$, and more generally, the inner kernel is also stationary, that is, for all $(t, s, h) \in \mathbb{R}$,*

$$\begin{cases} K_X(t,s) = \langle\!\langle X(t), X(s) \rangle\!\rangle = \frac{1}{2}(Q_X(t) - Q_X(t-s) + Q_X(s)) \\ \langle\!\langle X(t+h) - X(h), X(s+h) - X(h) \rangle\!\rangle = \langle\!\langle X(t), X(s) \rangle\!\rangle. \end{cases} \tag{2.11}$$

Proof The proof is based on the elementary formula, $2xy = x^2 + y^2 - (x-y)^2$, applied to $X(t)$ and $X(s)$ for the first equality, and to the Q_X-stationarity property $\mathbf{Q}[X(t) - X(s)] = Q_X(t-s)$. The second equation is obtained using the same method applied to the two variables, $X(t+h) - X(h)$ and $X(s+h) - X(h)$ whose difference is $X(t+h) - X(s+h)$ with $\mathbf{Q}[X(t+h) - X(s+h)] = Q_X(t-s)$. ■

$\mathbb{L}^2$- Self-similarity

The self-similarity property is related to the effects of time-scale change on the quadratic norm function $\{Q_X(t)\}$. This property states that for all $\lambda > 0$, there exists a function $\Theta(\lambda)$ such that the processes $\{X(\lambda t)\}$ and $\{\Theta^{\frac{1}{2}}(\lambda)X(t)\}$ have the same quadratic norms. More precisely, the self-similarity property establishes a spatial proportionality relationship between the quadratic norms $\{Q_X(\lambda t)\}$ and $\{Q_X(t)\}$ taken at two proportional times λt and t, whose spatial proportionality factor is a function of the time proportionality factor λ such that $\Theta(\lambda) > 0$.

Definition 2.5. *Let $\{\Theta(\lambda), \lambda > 0\}$ be a non-negative non-constant function.*
(i) *A process $\{X(t)\}$ is said to be Θ-$\mathbb{L}^2$-self-similar, if for all $\lambda > 0$, the processes $\{X(\lambda t)\}$ and $\{\Theta^{\frac{1}{2}}(\lambda)X(t)\}$ have the same $\mathbb{L}^2$-characteristics, that is,*

$$Q_X(\lambda t) = \Theta(\lambda)Q_X(t), \quad \langle\!\langle X(\lambda t), X(\lambda s)\rangle\!\rangle = \Theta(\lambda)\langle\!\langle X(t), X(s)\rangle\!\rangle. \tag{2.12}$$

(ii) *$\{X(t)\}$ is said to be Q_X-(Θ)-self-similar if only $Q_X(\lambda t) = \Theta(\lambda)Q_X(t), \forall t$.*

It can be observed that the Q_X-self-similarity property cannot be extended to the inner kernel without an additional assumption.

Remark: (i) Under the Q_X-self-similarity assumption, (including Θ non-constant), $X(0) = 0$, *a.s.*, because $Q_X(\lambda.0) = \Theta(\lambda)Q_X(0) = Q_X(0) = 0$.

(ii) For $\lambda = 0$, $\Theta(0) = 0$ because, for any t, $Q_X(0) = \Theta(0)Q_X(t)$.

(iii) The self-similarity function $\Theta(\lambda)$ corresponds to the proportionality factor in the proportionality relationship linking the quadratic norms $Q_X(\lambda t)$ and $Q_X(t)$ taken at two proportional times λt and t for all $t \in \mathbb{R}$. Therefore, the function $\Theta(\lambda)$ can be interpreted as a normalized quadratic norm because for all $\lambda > 0$, $\Theta(\lambda) = Q_X(\lambda)/Q_X(1) = Q_X(-\lambda)/Q_X(-1)$. Thus, the function Θ is symmetrical.

Thus, Θ can be extended to negative values by symmetry, $\Theta(\lambda) = \Theta(|\lambda|)$.

Therefore, for all $t \in \mathbb{R}$, the function $\Theta(|t|) = Q_X(|t|)/Q_X(1)$ satisfies the multiplicative relationship, $\Theta(\lambda|t|) = \Theta(\lambda)\Theta(|t|)$. Such functions are well-known, under the additional assumption of right-continuity (rc).

Lemma 2.6. *Consider the multiplicative equation on* $\mathbb{R}_+$, $\phi(u\,v) = \phi(u)\phi(v)$, $\phi(0) = 0$. *Any non-negative right-continuous solution* ϕ *is a power function* $\phi(u) = u^{2\gamma}$, $\gamma \geq 0$.

Proof First, observe that $\phi(1) = 1$, because $\phi(1) = \phi(1)^2$. To solve the multiplicative functional equation $\phi(uv) = \phi(u)\phi(v)$, we substitute $x = \ln(u)$ and $y = \ln(v)$. Thus, we obtain $\phi(e^{x+y}) = \phi(e^x)\phi(e^y)$. Taking the logarithm, we find that the new function $g(x) = \ln(\phi(e^x))$ is a solution of the additive equation $g(x + y) = g(x) + g(y)$ on $\mathbb{R}$ satisfying $g(0) = 0$. The assumption $\phi(0) = 0$ leads to $\lim_{x\to-\infty} g(x) = -\infty$ which implies that $\kappa > 0$. The only rc-solutions are linear, for all $x \in \mathbb{R}$, $g(x) = \kappa x$ and $\kappa = g(1)$. Returning to the multiplicative equation, $\phi(e^x) = e^{\kappa x} = (e^x)^\kappa = u^{2\gamma}$ where $2\gamma = \kappa$, with $\kappa > 0$ because $\phi(0) = 0$. ■

With the only Q_X-self-similarity assumption, the form of function Θ can be determined, and the quadratic norm is characterized.

Proposition 2.7. *A family* $\{X(t)\}$, *with a right-continuous (rc) quadratic norm* $\{Q_X(t)\}$ *is self-similar in quadratic norm* **only** *if the function* Q_X *is a power function, with a positive exponent* γ, $Q_X(t) = |t|^{2\gamma}Q_X(1)$ *if* $t > 0$, $Q_X(t) = \big|t\big|^{2\gamma} Q_X(-1)$ *if* $t < 0$, *and* $Q_X(0) = 0$. *In other words, for all* $t \in \mathbb{R}$ $Q_X(t) = \big|t\big|^{2\gamma} Q_X(sgn(t))$ *where* $sgn(t) = 1$ *if* $t > 0$, $sgn(t) = -1$ *if* $t < 0$, *and* $sgn(t) = 0$ *if* $t = 0$.

$\mathbb{L}^2(\mathbb{P})$-Stationarity and self-similarity

Combining the self-similarity and stationarity of the quadratic norm implies restricting the exponent γ of the power function $Q_X(t)$ to the interval $(0, 1)$, owing to Q_X-sub-linearity induced by the stationarity.

Theorem 2.8. (i) *A necessary and sufficient condition for an* $\mathbb{L}^2$*-rc* $\{X(t)\}$ *to be* $\mathbb{L}^2$*-self-similar and stationary, is the existence of a power function, with exponent* $0 < \gamma < 1$ *such that the* $\mathbb{L}^2$*-characteristics are,*

$$Q_X(t) = Q_X(-t) = Q_X(1)\,\big|t\big|^{2\gamma} \quad \textit{and} \quad K_X(t,s) = \frac{Q_X(1)}{2}(|t|^{2\gamma} + |s|^{2\gamma} - |t-s|^{2\gamma}). \tag{2.13}$$

(ii) *The condition* $\gamma = 1/2$, *is equivalent to the orthogonality of the increments, defined on the disjoint intervals. In this case* $K_X(t,s) = Q_X(1)(\big|t\big| \wedge |s|)$.

> Remark: – The Inner kernel is both stationary from Proposition **??**, and γ-self-similar, which is obvious in Equation (2.13).

Proof (i) The only property to be proved is that $\gamma < 1$.

This is due to the fact that the $\mathbb{L}^2$-norm $\sqrt{\mathbf{Q}(X(t))} = \sqrt{Q_X(t)}$ is sublinear, that is $|\sqrt{Q_X(t)} - \sqrt{Q_X(s)}| \leq \sqrt{\mathbf{Q}(X(t) - X(s))} = \sqrt{Q_X(t-s)}$, where the last equality comes from the stationarity assumption.

By replacing the quadratic norm function with its explicit power form (up to the multiplicative coefficient) $\big|t\big|^{2\gamma}$, the power functions must be sublinear, which is true only if $0 < \gamma < 1$.

(ii) If $\gamma = \frac{1}{2}$, when t and h have the same sign, $|t+h| = |t| + |h|$, and $\langle\langle X(t+h) - X(h), X(h)\rangle\rangle = \frac{Q_X(1)}{2}(|t+h| - |t| - |h|) = 0$ and the disjoint increments are orthogonal. When $t > 0 > h$, the intervals $[h,t]$ and $[h,0]$ are not disjoint, so their increments are not orthogonal, but $[h,0)$ and $[0,t]$ are disjoint intervals, and $\langle\langle X(t), X(h)\rangle\rangle = \frac{Q_X(1)}{2}(|t|+|h|-|t-h|)$ with $|t|+|h|-|t-h| = t-h-(t-h) = 0$ since $t > 0 > h$. Be careful, having increments defined on disjoint intervals is required to get the orthogonality of the increments.

(iii) If $\gamma = \frac{1}{2}$ then $K_X(t,s) = \frac{Q_X(1)}{2}(|t| + |s| - |t-s|)$. If $t, s > 0$ and $t \geq s$ then $K_X(t,s) = \frac{Q_X(1)}{2}(t + s - (t-s)) = Q_X(1)s = Q_X(1)(t \wedge s)$. If $t \geq 0$ and $s \leq 0$, then $K_X(t,s) = \frac{Q_X(1)}{2}(t - |s| - (t + |s|)) = Q_X(1)(-|s|) = Q_X(1)s = Q_X(1)(|s| \wedge |t|)$. ■

The following formulation is another form of the self-similarity and stationarity properties, which is useful in the study of Gaussian processes.

Proposition 2.9. *Let $\{X(t)\}$ be a family of $\mathcal{H}_X$, with an rc-quadratic norm Q_X.*
(i) *For all $\lambda > 0$, $h \in \mathbb{R}$ and $\gamma \in (0,1)$:*

- *the time-scaled family is $\left\{X^{\lambda,\gamma}(t) = \lambda^{-\gamma}X(\lambda t)\right\}$ with $\lambda > 0$,*
- *the time-translated family is defined as $\left\{X_h(t) = X(t+h) - X(h)\right\}$.*

When $\gamma = \frac{1}{2}$, the parameter γ will not be mentioned in the operators $\boldsymbol{X}^{\boldsymbol{\lambda}}$.
(ii) *Process $\{X(t)\}$ is γ-self-similar and stationary if and only if the processes $\{X_h(t)\}$ and $\{X^{\lambda,\gamma}(t)\}$ have the same $\mathbb{L}^2$-characteristics as $\{X(t)\}$, inner product $K_X(t,s) = \frac{Q_X(1)}{2}(|t|^{2\gamma} + |s|^{2\gamma} - |t-s|^{2\gamma})$, and quadratic norm $Q_X(t) = Q_X(-t) = Q_X(1)|t|^{\gamma}$.*

Proof All these processes have the same stationary and self-similar quadratic norms. Thus, by a direct application of Lemma 2.4 and of Theorem 2.8, they have the same $\mathbb{L}^2$-characteristics. ■

2.3 Gaussian spaces spanned by Gaussian families

In the general framework of the Hilbert space, results have been developed on the $\mathbb{L}^2$-characteristics of the $\mathbb{L}^2$-random families. Now we are interested in adding an assumption of distribution, and more specifically, in the Gaussian distribution, which presents very useful properties. We want to study the properties of the elements of the space $\mathcal{H}_X$ spanned by Gaussian families.

2.3.1 Gaussian variables and Gaussian families

Let us define the distribution of Gaussian variables and of a Gaussian family through their Fourier transform.

Gaussian random variables

The distribution of any variable is characterized by its Fourier transform.

Definition 2.10. *The distribution of a Gaussian variable X with mean m_X and variance V_X is characterized by its Fourier transform, as follows:*

$$\phi_X(u) := \mathbb{E}\left[e^{iuX}\right] = e^{iu\, m_X - \frac{1}{2}u^2 V_X},\ \forall u \in \mathbb{R}. \tag{2.14}$$

Recall that the quadratic norm of X is defined by $Q_X = m_X^2 + V_X$.

The main property of the Gaussian variables is that all quantities, mean m_X, variance V_X, and Fourier transform $\phi_X(u)$ are continuous operators in $\mathbb{L}^2$. (The first two properties have been recalled at the beginning of Section 2.2.2; the Fourier continuity results from the Lipschitz property of the Fourier transform). Any limit in the $\mathbb{L}^2$-space of the Gaussian variables is then a Gaussian variable.

The properties established in Section 2.2.2, can be generalized to the Gaussian framework. In particular, we consider orthogonality between the Gaussian variables.

Theorem 2.11. *Let X, Y be two Gaussian variables, such that any linear combination $\alpha X + \beta Y$ is Gaussian, with $(\alpha, \beta) \in \mathbb{R}^2$. Variables X and Y are independent, if and only if X and Y are $\mathbb{L}^2$-orthogonal, that is $\langle\langle X, Y\rangle\rangle = 0$.*

To verify the independence property between the two variables, it is sufficient to prove that their Fourier transforms are multiplicative.

Proof Let X, Y be two Gaussian variables, such that any linear combination $\alpha X + \beta Y$ is Gaussian, with $(\alpha, \beta) \in \mathbb{R}^2$.

Because $\alpha X + \beta Y$ is a Gaussian variable, its Fourier transform is given by $\phi_{\alpha X+\beta Y}(u) = \mathbb{E}\left[e^{iu(\alpha X+\beta Y)}\right] = e^{iu\mathbb{E}[\alpha X+\beta Y]-\frac{u^2}{2}\mathbb{V}[\alpha X+\beta Y]}$.

From Equation (2.9), if $X \perp Y$, then $\langle\langle X, Y\rangle\rangle = 0$ and $\mathbb{V}[\alpha X + \beta Y] = \alpha^2\mathbb{V}[X] + \beta^2\mathbb{V}[Y]$.

Therefore, we get: $\phi_{\alpha X+\beta Y}(u) = e^{iu\alpha\mathbb{E}[X]-\frac{u^2}{2}\alpha^2\mathbb{V}[X]} e^{iu\beta\mathbb{E}[Y]-\frac{u^2}{2}\alpha^2\mathbb{V}[Y]} = \phi_{\alpha X}(u)\phi_{\beta Y}(u)$. Consequently, X and Y are independent. ■

> Remark: Recall that for any non-negative bounded Borelian functions f, g, the random variables X and Y are said to be independent if: $\mathbb{E}\left[f(X)g(Y)\right] = \mathbb{E}\left[f(X)\right]\mathbb{E}\left[g(Y)\right]$.

Gaussian family and their Gaussian space

In this section, the notations introduced in the general framework of the $\mathbb{L}^2$-spaces are reused.

Definition 2.12. *Let $\{X(\theta)\}_{\theta\in J}$ be a $\mathbb{L}^2$-family. $\{X(\theta)\}_{\theta\in J}$ is a Gaussian family, if any finite linear combination $\alpha \cdot X(\theta) = \sum_1^n \alpha_i X(\theta_i)$ is a Gaussian variable.*

Space $\mathcal{H}_X$ contains only Gaussian variables, and is called the Gaussian Hilbert space.

REMARK: (i) Be careful, the variables of a Gaussian family are Gaussian variables, but the reciprocal is incorrect.

(ii) Any subfamily of a Gaussian family remains a Gaussian family,

REMARK: From Theorem 2.11, if two Gaussian subspaces $\mathcal{H}_X^1$ and $\mathcal{H}_X^2$ are orthogonal in the sense of the $\mathbb{L}^2$-orthogonality, then any linear combination of $\mathcal{H}_X^1$ is independent of any linear combination of $\mathcal{H}_X^2$. Any Borelian function of the variables of $\mathcal{H}_X^1$ is independent of any Borelian function of the variables of $\mathcal{H}_X^2$. Consequently, any σ-algebra of $\mathcal{H}_X^1$ is independent of any σ-algebra of $\mathcal{H}_X^2$.

2.3.2 Self-similar and stationary Gaussian process

The Gaussian family is now an $\mathbb{L}^2$-Gaussian function (or process) indexed in real time $\{X(t)\}_{t\in\mathbb{R}}$ for which we continue to exploit the results obtained in Section 2.2.2. $\mathbb{L}^2$-random variables were identified using their quadratic norm. Here, Gaussian variables (family) are identified by their Gaussian distribution, depending on both the covariance kernel and, the mean (expectation) of the variables. However, under self-similarity and stationarity assumptions, the process is necessarily centered, then the $\mathbb{L}^2$-properties and distributional properties coincide. To compare two Gaussian families, we compare their $\mathbb{L}^2$-characteristics **and** their mean functions $m_X(t) = \mathbb{E}(X(t))$.

Definition 2.13. *The Gaussian process $\{X(t)\}$ is γ-self-similar and stationary in distribution if and only if the two Gaussian processes ($\{X(t)\}$ and $\{X_h(t) = X(t+h) - X(t)\}, h \in \mathbb{R}$), respectively ($\{X(\lambda t)\}$ and $\{\lambda^\gamma X(t)\}, \lambda > 0$) have the same Gaussian distribution, that is, the same mean and inner kernel.*

A Gaussian process stationary and self-similar in distribution is obviously an $\mathbb{L}^2$-γ-self-similar and stationary process. Thus, all the previous results hold true. In particular, its quadratic norm is the power function $Q_X(t) = Q_X(1)|t|^{2\gamma}$ with $0 < \gamma < 1$, and the inner kernel is completely specified as a linear combination of these power functions $K_X(t,s) = \frac{Q_X(1)}{2}(|t|^{2\gamma}+|s|^{2\gamma}-|t-s|^{2\gamma})$. The only unknown property concerns the mean, which, by the stationarity property verifies $m_X(t+h) = m_X(t) + m_X(h)$. Therefore, the mean is a linear function and takes the following form $m_X(t) = m_X(1)t$. However, based on the self-similarity property, for all $\lambda > 0$, $m_X(\lambda) = \lambda^\gamma m_X(1)$. These two equations are contradictory, if $\gamma \neq 1$. But $\gamma \in (0,1)$. Subsequently, $m_X(t) = 0$.

Theorem 2.14. *A Gaussian process $\{X(t)\}$ is γ-self-similar and stationary in* **distribution**, *if and only if $\{X(t)\}$ is a centered process, and is γ-self-similar and stationary in the $\mathbb{L}^2$-sense. The quadratic norm $Q_X(t)$ is the variance $V_X(t) = V_X(1)|t|^{2\gamma}, \gamma \in (0,1)$.*

(ii) *If $\gamma = \frac{1}{2}$, the self-similar and stationary Gaussian process $\{X(t)\}$ has independent increments and a linear variance $V_B(t) = V_B(1)|t|$. This process is called Brownian motion or Wiener process. In the following, this is denoted as $\{B(t)\}$.*

(iii) *For any* $\lambda > 0$ *and* $h \in \mathbb{R}$, *the processes* $\{B_h(t) = B(t+h) - B(h)\}$, $\{B^\lambda(t) = \lambda^{-1/2}B(\lambda t)\}$ *and* $\{B_h^\lambda(t) = \lambda^{-1/2}(B(\lambda(t+h)) - B(\lambda h))\}$ *are also Brownian motions.*

2.3.3 Wiener-Brown space and stochastic integral

The Gaussian space spanned by Brownian motion is described with the help of the stochastic integral of deterministic functions against Brownian motion, which correspond to centered Gaussian variables with variance easy to characterize.

The results regarding the self-similarity and stationarity of the Gaussian families obtained in Section 2.3.2 are exploited in the framework of the stochastic integral.

$\mathbb{L}^2$-space $\mathcal{H}_B$ spanned by the Brownian motion

Recall that $\mathcal{H}_B$ is a Hilbert space containing all the finite linear combinations of Brownian motions and their limits.

(i) For our purposes, it is more convenient to describe these linear combinations based on the family of increments defined on disjoint intervals. Let us consider an increasing sequence of times $(t_0, t_1, t_2, \ldots, t_n)$ such that $t_0 < t_1 < t_2, \ldots, < t_n$. The finite linear combination $\alpha \cdot B = \sum_{i=1}^n \alpha_i(B(t_{i+1}) - B(t_i))$ is a Gaussian variable with variance $\mathbb{V}(\alpha \cdot B) = \sum_{i=1}^n \alpha_i^2(t_{i+1} - t_i)$ because the increments $B(t_{i+1}) - B(t_i)$ are independent.

Analogous to finite variation calculus, it is convenient to describe the random variable $\sum_{i=1}^n \alpha_i(B(t_{i+1}) - B(t_i))$ as the Brownian integral of the step function $\phi^e(u) = \sum_{i=1}^n \alpha_i \mathbb{1}_{(t_i, t_{i+1})}(u)$, and its variance as the Lebesgue integral of the square of the function. Then, for $\phi^e(u) = \sum_{i=1}^n \alpha_i \mathbb{1}_{(t_i, t_{i+1})}(u)$,

$$B(\phi^e) := \sum_{i=1}^n \alpha_i(B(t_{i+1}) - B(t_i)) = \int_{\mathbb{R}} \phi^e(u)\mathrm{d}B(u) \text{ with variance } \mathbb{V}(B(\phi^e)) = \int |\phi^e(u)|^2 du.$$

(iii) We wonder how to interpret the limit of the finite linear combinations of the increments of Brownian motion on disjoint intervals. Because the step functions are dense in $\mathbb{L}^2(Leb)$ space, if $(\phi_n^e)_{n\in\mathbb{N}}$ is a Cauchy sequence of step functions in $\mathbb{L}^2(Leb)$ such that $\lim_n \phi_n^e = \phi$, then the sequence of stochastic integrals $(B(\phi_n^e))$ converges in $\mathcal{H}_B$ to an r.v. denoted $B(\phi) = \int \phi(u)\mathrm{d}B(u)$ with variance $\int_{\mathbb{R}} |\phi(u)|^2 du$.

Theorem 2.15. *Let* $\{B(t)\}$ *be the standard Brownian motion.*

(i) *The* $\mathbb{L}^2$*-space spanned by the Brownian motion* $\{B(t)\}$ *is the family of Gaussian stochastic integrals* $B(\phi) := \int \phi(u)\mathrm{d}B(u)$ *of deterministic functions* ϕ *in* $\mathbb{L}^2(Leb)$, *with variance* $\mathbb{V}[B(\phi)] = \int |\phi|^2 du$.

(ii) *The covariance of two stochastic integrals is* $\mathrm{Cov}\left(B(\phi), B(\psi)\right) = \int_{\mathbb{R}} \phi(u)\psi(u)\mathrm{d}u$.

Therefore, there exists an isometry between $\mathcal{H}_B$ *and* $\mathbb{L}^2(Leb)$.

Change of Variable formula

Let us consider the stochastic integral of an affine time-changed Borelian function against the standard Brownian motion. Owing to a variable change, we prove that the time-change, which initially holds on the Borelian function ϕ, is transferred to the Brownian of integration under the form of the inverse time-change.

Theorem 2.16. *Recall that for all $\lambda > 0$, $h \in \mathbb{R}$ and $t \in \mathbb{R}$, $\{B_t\}$, $\{B^\lambda(t) = \lambda^{-\frac{1}{2}} B(\lambda t)\}$, and $\{B_h(t) = B(t+h) - B(h)\}$ are Brownian motions.*
For any $\phi \in \mathbb{L}^2(Leb)$, we have the following pathwise representation,

$$\int \lambda^{-\frac{1}{2}} \phi\left(\frac{u}{\lambda}\right) \mathrm{d}B(u) = \int \phi(u) \mathrm{d}B^\lambda(u) \text{ and } \int \phi\left(u-h\right) \mathrm{d}B(u) = \int \phi(u) \mathrm{d}B_h(u).$$

All these variables are Gaussian variables with the same variance $\int |\phi(u)|^2 du$.

> Remark: This result, (provided in Theorem 2.16), is much stronger than the previous one because we obtain an equality of processes (that is, in trajectory) instead of a result in distribution.

Proof Let $\phi \in \mathbb{L}^2(Leb)$. Let $\{B(t)\}$ be a standard Brownian motion.

Let us prove the trajectory identity on the indicator function $\mathbb{1}_{I_\alpha}(u) = \mathbb{1}_{(s_\alpha, t_\alpha)}(u)$. The extension to any step function and subsequently to any deterministic function is straightforward.

(i) Let us proceed to substitution in $\int_{\mathbb{R}} \mathbb{1}_{(s_\alpha,t_\alpha)}\left(\frac{u}{\lambda}\right) \mathrm{d}B(u)$, taking $v = \frac{u}{\lambda}$. The integration bounds then become λs_α and λt_α. Therefore, we get:
$\int_{\mathbb{R}} \mathbb{1}_{(s_\alpha,t_\alpha)}\left(\frac{u}{\lambda}\right) \mathrm{d}B(u) = \int_{\mathbb{R}} \mathbb{1}_{(\lambda s_\alpha, \lambda t_\alpha)}(u)\, \mathrm{d}B(u) = (B(\lambda s_\alpha) - B(\lambda t_\alpha))$. Because $(B(\lambda s_\alpha) - B(\lambda t_\alpha)) = \sqrt{\lambda} \int_{\mathbb{R}} \mathbb{1}_{(s_\alpha,t_\alpha)}(v) \mathrm{d}B^\lambda(v)$, the first equality is proved.

(ii) We proceed to a substitution in $\int_{\mathbb{R}} \mathbb{1}_{(s_\alpha,t_\alpha)}\left(u-h\right) \mathrm{d}B(u)$, taking $v = u - h$. Then, the integration bounds become $s_\alpha + h$ and $t_\alpha + h$. Therefore, we obtain:
$\int_{\mathbb{R}} \mathbb{1}_{(s_\alpha,t_\alpha)}\left(u-h\right) \mathrm{d}B(u) = \int_{\mathbb{R}} \mathbb{1}_{(s_\alpha+h, t_\alpha+h)}(u)\, \mathrm{d}B(u) = B(t_\alpha+h) - B(s_\alpha+h)$. This last term is the increment of $B_h(v)$ on (s_α, t_α) also written $\int_{\mathbb{R}} \mathbb{1}_{(s_\alpha,t_\alpha)}(v) \mathrm{d}B_h(v)$. Then,

$$\int_{\mathbb{R}} \mathbb{1}_{(s_\alpha,t_\alpha)}\left(u-h\right) \mathrm{d}B(u) = \int_{\mathbb{R}} \mathbb{1}_{(s_\alpha,t_\alpha)}(v) \mathrm{d}B_h(v).$$

By density of the step functions in $\mathbb{L}^2(Leb)$, this result can be extended to any square integrable Borelian function ϕ. ■

2.3.4 Existence and construction of self-similar and stationary Path-Dependent Brownian Motion (PDBM)

So far, the existence of self-similar and stationary Gaussian processes has been assumed, but has never been proven. The question of the existence of such processes is now addressed.

Existence of the Wiener process

The problem of the "mathematical existence" of such process has been solved for the first time in 1923 by Wiener, who was the first to give a mathematical construction based on ideas from signal theory. It is known as Karhunen-Loeve construction.

Proposition 2.17 (Existence of Wiener process)**.** *Let* $\{Z_k\}$ *be an infinite sequence of independent, normalized Gaussian random variables, used as the orthogonal basis of a Gaussian space. Similarly, let* $\left\{ e_k(t) = \alpha_k \sqrt{2} \sin\left(\frac{(2k-1)\pi}{2} t \right) \right\}$ *be the Karhunen-Loeve orthonormal basis of* $\mathbb{L}^2(\text{Leb})$*.*

Then, the process $\mathbf{B} = \{B(t)\}$*, defined by the family of* $\mathbb{L}^2(\mathbb{P})$*-random series*

$$B(t) = \sum_{k=1}^{+\infty} e_k(t)\ Z_k \tag{2.15}$$

is a stationary Gaussian process with independent increments.
It can also be shown that this process has continuous paths.

(See (Giambartolomei, 2015, Thm.5.14, p.48) for more details.)

Definition and properties of Path-Dependent Brownian Motions

Stochastic integrals of any Borel function $\phi \in \mathbb{L}^2(Leb)$ against Brownian motion allow us to obtain Gaussian variables. The challenge is to generate Gaussian processes. For this purpose, we generate stochastic integrals of a family of bivariate kernels $\{\kappa(t,u), (t,u) \in \mathbb{R}^2\}$ against Brownian motion, where t is a fixed time parameter, u is an integration variable and $\kappa(t,u)$ is a square-integrable with respect to u. In this sense, we are in line with the work of Mandelbrot and Van Ness in 1968 (Mandelbrot and Van Ness, 1968b). They formally define a path-dependent generalization of Brownian motion, called the *Path-Dependent Brownian Motion (PDBM)*, as follows:

$$X^\kappa(t,B) := \int \kappa(t,u) dB(u), \quad \text{with} \quad \int_{\mathbb{R}} \left| \kappa(t,u) \right|^2 \mathrm{d}u < \infty. \tag{2.16}$$

The covariance kernel is defined as $K_{X^\kappa}(t,s) = \int_{\mathbb{R}} \kappa(t,u)\kappa(s,u)\mathrm{d}u$ and $\kappa(0,u) = 0$.

We are interested in the time-translated and time-scaled PDBM defined for all $\lambda > 0$, $h \in \mathbb{R}$ and almost all $(t,u) \in \mathbb{R}^2$ by:

$$\begin{cases} X^\kappa(t+h,B) - X^\kappa(h,B) = \int (\kappa(t+h,u) - \kappa(h,u)) \mathrm{d}B(u) \\ X^\kappa(\lambda t, B) = \int \kappa(\lambda t, u) \mathrm{d}B(u). \end{cases} \tag{2.17}$$

Transformations of Path-Dependent Brownian Motion

Recall that in all the sequels, λ is positive and $\{B(t)\}$, $\{B_h(t) = B(t+h) - B(h)\}$, and $\{B^\lambda(t) = \lambda^{-\frac{1}{2}} B(\lambda t)\}$ are Brownian motions. We define the square integrable kernels

for almost all $(t,u) \in \mathbb{R}^2$ and for all $h \in \mathbb{R}$ and $\lambda > 0$ by:

$$\kappa(t,u), \quad \kappa_h(t,u) = \kappa(t,u-h), \quad \kappa^\lambda(t,u) = \lambda^{-1/2}\kappa(t,u/\lambda). \tag{2.18}$$

Theorem 2.18. *Let us define the PDBMs* $\{Y(t) = X^\kappa(t,B)\}, \{Y^\lambda(t) = X^\kappa(t,B^\lambda)\}, \{Y_h(t) = X^\kappa(t,B_h)\}$.

(i) *Processes* $\{Y^\lambda(t)\}$ *and* $\{Y_h(t)\}$ *have the same distribution as* $\{Y(t)\}$.

(ii) $\{Y^\lambda(t)\}$ *and* $\{Y_h(t)\}$ *satisfy the pathwise identities:*

$$\{Y^\lambda(t) = X^{\kappa^\lambda}(t,B)\} \quad \textit{and} \quad \{Y_h(t) = X^{\kappa_h}(t,B)\}. \tag{2.19}$$

Proof This Theorem is the kernel version of the variable change formula stated in Theorem (2.16).

(i) Because processes $\{B(t)\}$, $\{B_h(t)\}$ and $\{B^\lambda(t)\}$ are all Brownian motions, processes $\{Y(t)\}$, $\{Y_h(t)\}$ and $\{Y^\lambda(t)\}$ are all stochastic integrals of the same deterministic kernel $\kappa(t,u)$ against Brownian motion. Consequently, $\{Y(t)\}$, $\{Y_h(t)\}$ and $\{Y^\lambda(t)\}$ have the same Gaussian distribution, with a null expectation and a variance equal to $\int(\kappa(t,u))^2 \mathrm{d}u$.

(ii) From Equation (2.18), $X^{\kappa_h}(t,B) = \int \kappa(t,u-h)\mathrm{d}B_u$. From Theorem 2.16, the variable change $v = u - h$ transfers the time-translation from the integration variable u to the Brownian of integration, which becomes $B_h(u)$. We then obtain $X^{\kappa_h}(t,B) = \int \kappa(t,v)\mathrm{d}B_h(v) = X^\kappa(t,B_h)$.

From Equation (2.18), $X^{\kappa^\lambda}(t,B) = \lambda^{-\frac{1}{2}} \int \kappa\left(t, \frac{u}{\lambda}\right) \mathrm{d}B(u)$. From Theorem 2.16, the variable change $w = \frac{u}{\lambda}$ transfers the time-scaling from the integration variable u to the Brownian of integration, which becomes $B^\lambda(w) = \lambda^{-\frac{1}{2}} B(\lambda w)$. Then we get, $X^{\kappa^\lambda}(t,B) = \int \kappa(t,w)\mathrm{d}B^\lambda(v) = X^\kappa(t,B^\lambda)$.

■

The interpretation in terms of stationarity and self-similarity of the PDBM requires the transfer of these properties obtained in u to time t.

Theorem 2.19. *Let* $\{B(t)\}$, $\{B^\lambda(t)\}$, $\{B_h(t)\}$, *be the transformed Brownian motions.*

(i) *If the time-stationarity condition of bivariate kernel* $\kappa(t,u)$ *is satisfied,*

$$\kappa(t+h,u) - \kappa(h,u) = \kappa_h(t,u) = \kappa(t,u-h) \tag{2.20}$$

then the following pathwise identity holds:

$$\{X^\kappa(t+h,B) - X^\kappa(h,B) = X^\kappa(t,B_h)\} \tag{2.21}$$

and the process $\{X^\kappa(t,B)\}$ *is stationary in the trajectory sense.*

(ii) *If the time-self-similarity condition of the bivariate kernel* $\kappa(t,u)$ *is satisfied,*

$$\kappa(\lambda t,u) = \eta(\lambda)\kappa\left(t,\frac{u}{\lambda}\right) = (\sqrt{\lambda}\eta(\lambda))\kappa^\lambda(t,u) \textit{ with } \eta(\lambda) = \lambda^\nu \mathbb{1}_{\{\lambda>0\}},\ \nu \in \left(-\frac{1}{2},\frac{1}{2}\right) \tag{2.22}$$

then the following pathwise identity holds:

$$\{X^\kappa(\lambda t, B) = \eta(\lambda)\sqrt{\lambda}X^\kappa(t, B^\lambda)\}. \tag{2.23}$$

and the process $\{X^\kappa(t, B)\}$ *is self-similar in the trajectory sense.*

(iii) *If both the time-stationarity (Equation (2.20)) and time-self-similarity (Equation (2.22))) conditions are satisfied, then the following pathwise identity holds:*

$$\{X^\kappa(\lambda(t+h), B) - X^\kappa(\lambda h, B) = \eta(\lambda)\sqrt{\lambda}X^\kappa(t, B_h^\lambda)\} \tag{2.24}$$

and the process $\{X^\kappa(t, B)\}$ *is both stationary and self-similar, in the trajectory sense, with variance-covariance kernel:*

$$V_X(t) = V_X(1)\left|t\right|^{2\gamma} \quad \textit{and} \quad K_X(t,s) = \frac{V_X(1)}{2}(|t|^{2\gamma} + |s|^{2\gamma} - |t-s|^{2\gamma}) \tag{2.25}$$

with $\gamma \in (0,1)$, $\eta(\lambda) = \lambda^{\gamma-\frac{1}{2}}\mathbb{1}_{\{\lambda>0\}}$.

Proof (i) From Equation (2.17), $X^\kappa(t+h,B) - X^\kappa(h,B) = \int(\kappa(t+h,u) - \kappa(h,u))\mathrm{d}B(u)$. If the time-stationarity property of the bivariate kernel $\kappa(t,u)$ given in Equation (2.20) is satisfied, then $\kappa(t+h,u) - \kappa(h,u) = \kappa(t,u-h)$, and we get $X^\kappa(t+h,B) - X^\kappa(h,B) = \int \kappa(t,u-h)\mathrm{d}B(u)$. From Theorem 2.16, the time-translation is transferred from integration variable u to the Brownian of integration, and we obtain: $X^\kappa(t+h,B) - X^\kappa(h,B) = \int \kappa(t,u)\mathrm{d}B_h(u) = X^\kappa(t,B_h)$. This proves that the process $\{X^\kappa(t,B)\}$ is stationary in the trajectory sense.

(ii) From Equation (2.17), $X^\kappa(\lambda t, B) = \int \kappa(\lambda t, u)\mathrm{d}B(u)$.

If the time-self-similarity property of the bivariate kernel $\kappa(t,u)$ given in Equation (2.22) is satisfied, then $\kappa(\lambda t, u) = \eta(\lambda)\kappa\left(t, \frac{u}{\lambda}\right) = (\sqrt{\lambda}\eta(\lambda))\kappa^\lambda(t,u)$ with $\eta(\lambda) = \lambda^\nu \mathbb{1}_{\{\lambda>0\}}$, $\nu \in \left(-\frac{1}{2}, \frac{1}{2}\right)$, and we get $X^\kappa(\lambda t, B) = \int \eta(\lambda)\kappa\left(t, \frac{u}{\lambda}\right)\mathrm{d}B(u)$. From Theorem 2.16, the time-scaling is transferred from the integration variable u to the Brownian of integration, and we obtain: $X^\kappa(\lambda t, B) = \sqrt{\lambda}\eta(\lambda)\int \kappa\left(t,u\right)\mathrm{d}B^\lambda(u) = \sqrt{\lambda}\eta(\lambda)X^\kappa(t,B^\lambda)$. This proves that the process $\{X^\kappa(t,B)\}$ is self-similar in the trajectory sense.

(iii) Gathering time-translation and time-scaling, $X^\kappa(\lambda(t+h),B) - X^\kappa(\lambda h, B) = \int(\kappa(\lambda(t+h),u) - \kappa(\lambda h, u))\mathrm{d}B(u)$. If both the time-stationarity and time-self-similarity properties of the bivariate kernel $\kappa(t,u)$ are satisfied, then $\kappa(\lambda(t+h),u) - \kappa(\lambda h, u) = \eta(\lambda)\kappa\left(t, \frac{u}{\lambda} - h\right)$.

Then we get, $X^\kappa(\lambda(t+h),B) - X^\kappa(\lambda h, B) = \int \eta(\lambda)\kappa\left(t, \frac{u}{\lambda} - h\right)\mathrm{d}B(u)$. Applying twice Theorem 2.16, once for the time-translation and once for the time-scaling, it allows transferring the time-transformations from the integration variable to the Brownian of integration, and we get: $X^\kappa(\lambda(t+h),B) - X^\kappa(\lambda h, B) = \sqrt{\lambda}\eta(\lambda)\int \kappa\left(t,u\right)\mathrm{d}B_h^\lambda(u)$. ■

A reference dealing with self-similar Gaussian processes with stationary increments is (Taqqu, 1994, Chap.7 - Def.7.1.7 p.314, Cor.7.2.3 p.320). Authors proved that a Gaussian process is self-similar and with stationary increments if and only if it is an fBm whose auto-covariance function is given by Equation (2.13). In their approach, distributional assumptions are made. Stationarity and self-similarity properties are expressed in terms of distributions. That is what distinguishes their approach from ours. Indeed, in our approach, we assume the stationarity and self-similarity properties of the quadratic norm in the $\mathbb{L}^2$-spaces, and we prove that these assumptions alone are sufficient to fully characterize the covariance kernel and to deduce that the latter also satisfies the stationarity and self-similarity properties.

Kernel identification

From Theorem 2.19, now, the issue is to study the following system:

$$\begin{cases} \kappa(t+h,u) - \kappa(h,u) = \kappa(t,u-h) \\ \kappa(\lambda t, \lambda u) = \eta(\lambda)\kappa\left(t,u\right) \quad \text{with} \quad \eta(\lambda) = \lambda^{\nu}\mathbb{1}_{\{\lambda>0\}},\ \nu \in \left(-\frac{1}{2},\frac{1}{2}\right). \end{cases} \tag{2.26}$$

Remarks:

(i) The above system is defined for almost all $(t,u) \in \mathbb{R}^2$ and $\lambda > 0$.

(ii) The exponent ν can take negative values when it belongs to the interval $\left(-\frac{1}{2},0\right)$. In this case, the kernel can encounter some problems of definition in the neighborhood of 0.

(iii) We look for a kernel that is almost surely Lebesgue.

We highlight some useful properties induced by the self-similarity and stationarity conditions of the bivariate kernel stated in Equation (2.26).

(i) On the one hand, the stationarity property of $\kappa(t,u)$ establishes invariance by the time-translation of the bivariate kernel: $\kappa(t+h,u) - \kappa(h,u) = \kappa(t,u-h)$.

(ii) On the other hand, the self-similarity condition establishes a spatial proportionality relationship between two bivariate kernels whose time variables are linked by a proportionality factor $\lambda > 0$, and the spatial proportionality factor is a function of the time-proportionality factor λ, $\eta(\lambda)$. Therefore, the self-similarity condition given in Equation (2.26) can be rewritten as follows:

$$\kappa(\lambda t,u) = \eta(\lambda)\kappa\left(t,\frac{u}{\lambda}\right) \quad \text{with} \quad \eta(\lambda) = \lambda^{\nu}\mathbb{1}_{\{\lambda>0\}},\ \nu \in \left(-\frac{1}{2},\frac{1}{2}\right). \tag{2.27}$$

This last form is less intuitive than the form given in Equation (2.26) but is the form used to prove the self-similarity of the PDBMs.

The self-similarity property of the bivariate kernel $\{\kappa(t,u)\}$ with respect to the time-component t, with scaling factor λ and self-similarity function $\eta(\lambda)$ as stated in Equation (2.27) is equivalent to the self-similarity property of the bivariate kernel $\{\kappa(t,u)\}$ with

respect to the integration variable u, with inverse scaling factor $\frac{1}{\lambda}$, and inverse self-similarity function $\frac{1}{\eta(\lambda)}$. So that the variance of the PDBM is a power function, $\eta(\lambda)$ is necessarily a power function, then $\frac{1}{\eta(\lambda)} = \eta\left(\frac{1}{\lambda}\right)$.

$$\kappa\left(t, \frac{u}{\lambda}\right) = \frac{1}{\eta(\lambda)}\kappa\left(\frac{t}{1/\lambda}, u\right) = \eta\left(\frac{1}{\lambda}\right)\kappa\left(\lambda t, u\right) \quad \text{with} \quad \eta(\lambda) = \lambda^{-\nu}\mathbb{1}_{\{\lambda>0\}},\ \nu \in \left(-\frac{1}{2}, \frac{1}{2}\right). \tag{2.28}$$

The self-similarity property of the bivariate kernel with respect to the time-variable given by Equation (2.27), and the self-similarity property of the bivariate kernel with respect to the integration variable given by Equation (2.28), allow going from a positive exponent ν to a negative exponent and vice versa. These properties are useful to solve the problem of definition of the bivariate kernel in the neighborhood of 0 when $\nu < 0$.

(iii) The self-similarity property of the bivariate kernel $\{\kappa(t,u)\}$ with respect to the time-variable allows proving the symmetry property of the ratio between couples of bivariate kernels taken at two proportional times. Indeed, the self-similarity property of $\{\kappa(t,u)\}$ with respect to the time-variable establishes a spatial proportionality relationship between the bivariate kernels taken at two proportional times of factor λ, and the spatial proportionality factor is a function of the time proportionality factor $\eta(\lambda)$. This implies that the ratio between couples of bivariate kernels taken at two proportional times of factor λ is equal to the spatial proportionality factor $\eta(\lambda)$ which is constant and independent of time.

$$\eta(\lambda) = \frac{\kappa(\lambda t, \lambda u)}{\kappa(t,u)} = \frac{\kappa(\lambda t, u)}{\kappa\left(t, \frac{u}{\lambda}\right)}. \tag{2.29}$$

Specifically, the self-similarity property of $\{\kappa(t,u)\}$ leads to a symmetry property:

$$\eta(\lambda) = \frac{\kappa(-\lambda t, -\lambda u)}{\kappa(-t,-u)} = \frac{\kappa(\lambda t, \lambda u)}{\kappa(t,u)} \text{ or equivalently } \eta(\lambda) = \frac{\kappa(-\lambda t, -u)}{\kappa\left(-t, -\frac{u}{\lambda}\right)} = \frac{\kappa(\lambda t, u)}{\kappa\left(t, \frac{u}{\lambda}\right)}. \tag{2.30}$$

In this context, the goal is to find a solution to System (2.26), which, in addition, is right-continuous with respect to the time-variable t and Lebesgue square-integrable with respect to the integration variable u. The set of these solutions is not empty because Mandelbrot and Van Ness proposed a solution.

Remarks:

(i) The solution proposed by Mandelbrot and Van Ness in Equation (2.32) can be rewritten as follows:

$$\kappa(t,u) = \eta(t-u) - \eta(0-u) \quad \text{with} \quad \eta(u) = u^{\nu}\mathbb{1}_{\{u>0\}}. \tag{2.31}$$

In other words, $\kappa(t,u)$ is the increment between two times $0-u$ and $t-u$ of the power function η. This specific form of kernel has two consequences:

– Because $\kappa(t,u)$ can be expressed as the difference in a function η taken at two distinct times, then $\kappa(t,u)$ satisfies the stationarity assumption given in Equation (2.26).

– Because $\eta(u) = u^\nu \mathbb{1}_{\{u>0\}}$ is a power function, it is multiplicative, and since $\kappa(t,u)$ is fully characterized by the function η, then $\kappa(t,u)$ satisfies the self-similarity assumption given in Equation (2.26).

(ii) We want that $\kappa(t,u)$ to be square-integrable. Because the kernel is defined on $(-\infty,t)$, the square integrability when $u \to +\infty$ is not a problem. In particular, we must check whether $\kappa(t,u)$ is square integrable in the neighborhood of 0, t and $-\infty$.

Proposition 2.20. *A bivariate kernel proposed by Mandelbrot and Van Ness is given by:*

$$\kappa^\nu(t,u) = (t-u)^\nu \mathbb{1}_{\{u<t\}} - (-u)^\nu \mathbb{1}_{\{u<0\}}, \quad \textit{with} \quad \nu = \gamma - \frac{1}{2} \in \left(-\frac{1}{2}, \frac{1}{2}\right). \tag{2.32}$$

(i) *This kernel has some definition problems when $u \to 0$, $u \to t$.*

(ii) *This kernel satisfies System (2.26).*

(iii) *For all $\nu \in \left(-\frac{1}{2}, \frac{1}{2}\right)$, $\kappa(t,u) \sim \nu t(-u)^{(\nu-1)}$ is Lebesgue square-integrable when $u \to -\infty$ because $2(\nu-1) < -1$.*

For $\nu < 0$, when $u \to 0$ (resp. $u \to t$), $\kappa(t,u) \sim -(-u)^\nu$ (resp. $\kappa(t,u) \sim (t-u)^\nu$) which is Lebesgue square-integrable because $2\nu + 1 > 0$.

Then $\kappa(t,u)$ is Lebesgue u-square integrable.

Proof (ii) Stationarity condition: For all $h \in \mathbb{R}$, $\kappa(t+h,u) - \kappa(h,u) = (t+h-u)^\nu \mathbb{1}_{\{u<t+h\}} - (h-u)^\nu \mathbb{1}_{\{u<h\}}$. This kernel can be rewritten as $\kappa(t+h,u) - \kappa(h,u) = (t-(u-h))^\nu \mathbb{1}_{\{u-h<t\}} - (-(u-h))^\nu \mathbb{1}_{\{u-h<0\}} = \kappa(t,u-h)$.

Self-similarity condition: For all $\lambda > 0$, $\kappa(\lambda t,u) = (\lambda t - u)^\nu \mathbb{1}_{\{u<\lambda t\}} - (-u)^\nu \mathbb{1}_{\{u<0\}}$. This kernel can be rewritten as $\kappa(\lambda t,u) = \lambda^\nu \left(\left(t - \frac{u}{\lambda}\right)^\nu \mathbb{1}_{\{\frac{u}{\lambda}<t\}} - \left(-\frac{u}{\lambda}\right)^\nu \mathbb{1}_{\{\frac{u}{\lambda}<0\}} \right) = \lambda^\nu \kappa\left(t, \frac{u}{\lambda}\right)$.
Consequently, the kernel proposed by Mandelbrot and Van Ness is a solution to the System (2.26).

(iii) We study the square integrability of the path-dependent kernel.

For all $\nu \in \left(-\frac{1}{2}, \frac{1}{2}\right)$, when $u \to -\infty$, $\frac{t}{u} \to 0$, an equivalent of $\kappa(t,u)$ is given as follows: $\kappa(t,u) = (t-u)^\nu - (-u)^\nu = (-u)^\nu \left(\left(1 - \frac{t}{u}\right)^\nu - 1 \right) \sim \nu t(-u)^{\nu-1}$. Because $2\nu - 2 < -1$, then $\kappa(t,u)$ is u-square integrable in $-\infty$.

For $\nu < 0$, when $u \to 0$ resp. $u \to t$, the problem of u-square-integrability of the bivariate kernel is symmetrical owing to the form of the kernel.

When $u \to 0$ (resp. $u \to t$), $\kappa(t,u) \sim -(-u)^{\nu}$ (resp. $\kappa(t,u) \sim (t-u)^{\nu}$). Because $2\nu + 1 > 0$, then $\kappa(t,u)$ is Lebesgue u-square integrable in the neighborhood of 0 resp. in the neighborhood of t. ■

> REMARK: With additional efforts, it is possible to prove the uniqueness of the solution proposed by Mandelbrot and Van Ness.

The variance of the process $\{X^{\kappa}(t,B)\}$ is self-similar and stationary. Then from Theorem 2.8, $V_X(t) = V_X(1)\left|t\right|^{2\gamma}$ and $K_X(t,s) = \frac{V_X(1)}{2}(|t|^{2\gamma} + |s|^{2\gamma} - |t-s|^{2\gamma})$ with $0 < \gamma < 1$. To completely determine the variance function and covariance kernel, it remains to compute the variance of $\{X^{\kappa}(t,B)\}$ at time 1.

Theorem 2.21. *Let $\{X^{\kappa}(t,B)\}$ be a PDBM with a bivariate kernel defined by $\kappa(t,u) = \left((t-u)_+^{\gamma-\frac{1}{2}} - (-u)_+^{\gamma-\frac{1}{2}}\right)$ with $\gamma \in (0,1)$. The standardization constant that allows obtaining a variance equal to 1 at time 1 is given by:*

$$V_{X^{\kappa}}(1) = \frac{1}{2\gamma}\left(\frac{3}{2} - \gamma\right)\mathfrak{B}\left(2-2\gamma, \gamma + \frac{1}{2}\right) \tag{2.33}$$

where $\mathfrak{B}$ refers to the function beta and takes the form of $\mathfrak{B}(\alpha,\beta) = \int_0^1 u^{\alpha-1}(1-u)^{\beta-1}du$.

A proof of Theorem 2.21 will be provided in Chapter 3.

2.3.5 Multidimensional framework

The definition and properties of Brownian motion and Path-Dependent Brownian Motion (PDBM) can be extended to the multidimensional framework as a generalization of the unidimensional framework.

Multivariate Wiener process

The extension of the univariate Wiener process to the multivariate framework with independent coordinates is called d-dimensional Wiener process.

Definition 2.22. *A d-dimensional Wiener process is defined by $\{\boldsymbol{W}(t) = (W^1(t), \ldots, W^d(t))\}_{t\in\mathbb{R}}$ and satisfies the following statements:*

(i) *Each coordinate $\{W^i(t)\}$ is a univariate Wiener process,*

(ii) *For all $(i,j) \in [\![1,d]\!]^2$, the coordinates $\{W^i(t)\}$ and $\{W^j(t)\}$ are independent, that is,* $\text{Cov}\left(W^i(t), W^j(s)\right) = 0$.

REMARK: Because the coordinates of the d-dimensional Wiener process are independent of each other, the properties satisfied by the coordinates are extended directly to the d-dimensional process. Each coordinate $W^i(t)$ is a univariate Wiener process thus, is Gaussian self-similar with stationary and independent increments. Therefore, the d-dimensional Wiener process is Gaussian self-similar with stationary and independent increments in vector sense.

Proposition 2.23. *Let* $\left\{ \boldsymbol{W}(t) \right\}$ *be a* d*-dimensional Wiener process.*

(i) *For all* $\lambda > 0$, $\{\boldsymbol{W^{\lambda}}(t) = \lambda^{-\frac{1}{2}} \boldsymbol{W}(\lambda t)\}$ *and for all* $h \in \mathbb{R}$, $\{\boldsymbol{W_h}(t) = \boldsymbol{W}(t+h) - \boldsymbol{W}(h)\}$ *have the same distribution as* $\left\{ \boldsymbol{W}(t) \right\}$.

(ii) *The* d*-dimensional Wiener process* $\left\{ \boldsymbol{W}(t) \right\}$ *is self-similar with stationary and independent increments in vector sense.*

Multivariate Brownian motion

Now, we construct the correlated Brownian motion. For this purpose, as already done in the univariate framework, we create correlated variables from independent variables. The d-dimensional Wiener process with correlated coordinates is called d-dimensional Brownian motion and is denoted by $\{\boldsymbol{B}(t) = (B^1(t), \ldots, B^d(t))\}_{t \in \mathbb{R}}$.

For all $i \in [\![1, d]\!]$, the variables $B^i(t)$ are correlated owing to the linear transformation of the independent variables.

Let $R = (r_k^i) \in M_d(\mathbb{R})$ be a square matrix of dimension d where the rows are vectors $\boldsymbol{r^i} = (r_k^i)_{k \in [\![1,d]\!]}$. Then, RR^T is a symmetrical matrix, whose coefficients $(\rho_{i,j})_{(i,j) \in [\![1,d]\!]^2}$ are the results of the inner products between row vectors $\boldsymbol{r^i}$ and $\boldsymbol{r^j}$.

Moreover, we want the variance of each coordinate $B^i(t)$ to be equal to t which is also the variance of each coordinate in the d-dimensional Wiener process.

Therefore, the matrix of the coefficients of RR^T might be equal to the inner product between the row vectors of R, and so that the variance of each coordinate $B^i(t)$ might be equal to t, matrix R must satisfy the following conditions.

Conditions on the matrix of transformation:

(i) The diagonal coefficients are given by $\rho_{i,i} = \sum_k (r_k^i)^2 = 1$ for all $i \in [\![1, d]\!]$,

(ii) The non-diagonal coefficients are given by $\rho_{i,j} = \sum_k r_k^i r_k^j \in (0, 1)$,

(iii) For all $(i, j) \in [\![1, d]\!]^2$, $\rho_{i,j} = \rho_{j,i}$.

Let us take: $\{\mathbf{B}(t) = R\mathbf{W}(t)\}$ where for all $i \in [\![1, d]\!]$, $B^i(t) = \sum_{k=1}^d r_k^i W^k(t)$.

Owing to Condition (i) , the variance of each coordinate $B^i(t)$ is equal to t.

Definition 2.24. *$\{R\boldsymbol{W}(t)\}$, where R is the matrix satisfying Conditions 2.3.5, is a d-dimensional Brownian motion. More generally the process $\{\boldsymbol{B}(t) = (B^1(t), \dots, B^d(t))\}_{t\in\mathbb{R}}$ whose distribution is the same as $\{R\boldsymbol{W}(t)\}$ is a d-dimensional Brownian motion.*

(i) *Each coordinate $\{B^i(t)\}_{t\in\mathbb{R}}$ is a univariate Brownian motion,*

(iii) *The coordinates of the d-dimensional Brownian motion are correlated through a correlation matrix $RR^T = (\rho_{i,j}) \in M_d([0,1])$ independent of time, and the covariance kernel is provided by: $K_B((i,t),(j,s)) = \mathrm{Cov}\left(B^i(t), B^j(s)\right) = \rho_{i,j}(t \wedge s)$.*

> REMARK: Each coordinate $B^i(t)$ is a linear combination of independent variables $\{\mathbf{B}(t) = R\mathbf{W}(t)\}$ with for all $i \in [\![1,d]\!]$, $B^i(t) = \sum_{k=1}^d r_k^i W^k(t)$ where the coefficients of the linear combination are constants.
>
> Then, the properties of the independent variables are recovered at the level of linear combination. Consequently, each coordinate $B^i(t)$ is self-similar, with stationary and independent increments.
>
> Moreover, owing to Condition (i) on matrix R, the variance of each coordinate is equal to t. Consequently, each coordinate $B^i(t)$ is a Brownian motion.

Proposition 2.25. *Let $\{\boldsymbol{B}(t)\}$ be a d-dimensional Brownian motion. For all $\lambda > 0$, $\{\boldsymbol{B}^{\boldsymbol{\lambda}}(t) = \lambda^{-\frac{1}{2}}\boldsymbol{B}(\lambda t)\}$ and for all $h \in \mathbb{R}$, $\{\boldsymbol{B}_{\boldsymbol{h}}(t) = \boldsymbol{B}(t+h) - \boldsymbol{B}(h)\}$ where R satisfies Conditions 2.3.5.*

(i) *$\{\boldsymbol{B}^{\boldsymbol{\lambda}}(t)\}$ and $\{\boldsymbol{B}_{\boldsymbol{h}}(t)\}$ have the same distribution as $\{\boldsymbol{B}(t)\}$.*

(ii) *$\{\boldsymbol{B}(t)\}$ is a d-dimensional self-similar Gaussian process with stationary and independent increments.*

Proof $\{\mathbf{B}(t) = R\mathbf{W}(t)\}$ where for all $i \in [\![1,d]\!]$, $B^i(t) = \sum_{k=1}^d r_k^i W^k(t)$.

As R is a matrix of constant coefficients, it is sufficient to prove the properties on the d-dimensional Wiener process.

From Proposition 2.23, for all $\lambda > 0$, $\{\boldsymbol{W}^{\boldsymbol{\lambda}}(t) = \lambda^{-\frac{1}{2}}\boldsymbol{W}(\lambda t)\}$ and for all $h \in \mathbb{R}$, $\{\boldsymbol{W}_{\boldsymbol{h}}(t) = \boldsymbol{W}(t+h) - \boldsymbol{W}(h)\}$ have the same distribution as $\left\{\mathbf{W}(t)\right\}$. Thus, $\{\boldsymbol{W}(t)\}$ is self-similar and stationary. This implies that $\{\boldsymbol{B}^{\boldsymbol{\lambda}}(t) = R\boldsymbol{W}^{\lambda}(t)\}$ and $\{\boldsymbol{B}_{\boldsymbol{h}}(t) = R\boldsymbol{W}_h(t)\}$ have the same distribution as $\{R\boldsymbol{W}(t) = \boldsymbol{B}(t)\}$. Consequently, $\{\boldsymbol{B}(t)\}$ is self-similar and stationary. ■

Multivariate fractional Brownian motion

We generate Gaussian processes as stochastic integrals of Mandelbrot's kernels defined in Equation (2.32) against correlated Brownian motion, such that these processes might satisfy the desired properties, that is, self-similarity and stationarity properties.

Definition 2.26. *$\{\boldsymbol{X}^{\boldsymbol{\kappa}}(t,B) = (X^{\kappa^1}(t,B^1), X^{\kappa^2}(t,B^2), \dots, X^{\kappa^d}(t,B^d))\}$, called d-dimensional fractional Brownian motion, is a multidimensional Gaussian process defined as the stochastic integral of a vector of deterministic bivariate kernels against a vector of linear transformations of the Wiener process, such that for almost all $(t,u) \in \mathbb{R}^2$:*

$$\boldsymbol{X}^{\boldsymbol{\kappa}}(t,B) = \int \boldsymbol{\kappa}(t,u) \odot \mathrm{d}\boldsymbol{B}(u) := \left(\int \kappa^i(t,u) \mathrm{d}B^i(u) \right)_{i \in [\![1,d]\!]} \tag{2.34}$$

where $\odot$ is the element-wise product, $\boldsymbol{\kappa}(t,u) = (\kappa^i(t,u))_{i\in[\![1,d]\!]}$ is the vector of bivariate kernels proposed by Mandelbrot and Van Ness in Equation (2.32), $\boldsymbol{B}(t) = (B^i(t))_{i\in[\![1,d]\!]}$ is the d-dimensional Brownian motion. The covariance kernel of the d-dimensional fractional Brownian motion $\{\boldsymbol{X}^{\boldsymbol{\kappa}}(t,B)\}$ is given by:

$$K_{X^\kappa}((i,t),(j,s)) = \mathrm{Cov}\left(X^{\kappa^i}(t,B^i), X^{\kappa^j}(s,B^j) \right) = \int \kappa^i(t,u)\kappa^j(s,u)\rho_{i,j}\mathrm{d}u. \tag{2.35}$$

Each coordinate of the vector $\boldsymbol{\kappa}(t,u)$ satisfies the conditions of stationarity and self-similarity in the univariate framework given in Equation (2.26). These properties are then obtained in the vector sense, that is, for all $\lambda > 0$, $h \in \mathbb{R}$ and for almost all $(t,u) \in \mathbb{R}^2$:

$$\begin{cases} \boldsymbol{\kappa}(\lambda t, \lambda u) = \boldsymbol{\eta}(\lambda)\boldsymbol{\kappa}\left(t,u\right) \quad \text{with} \quad \boldsymbol{\eta}(\lambda) = (\lambda^{\nu_i}\mathbb{1}_{\{\lambda>0\}})_{i\in[\![1,d]\!]},\ \nu_i \in \left(-\frac{1}{2},\frac{1}{2}\right) \\ \boldsymbol{\kappa}(t+h,u) - \boldsymbol{\kappa}(h,u) = \boldsymbol{\kappa}(t,u-h). \end{cases} \tag{2.36}$$

Because the properties of stationarity and self-similarity of the bivariate kernel are satisfied in the vector sense, the time-translation and time-scaling carried by the time variable of the vector bivariate kernel are transferred to its integration variable. From Theorem 2.16, a simple variable change allows the transfer of the time-transformations from the integration variable to the d-dimensional Brownian of integration.

From Proposition 2.25, the d-dimensional Brownian motion satisfies the properties of stationarity and self-similarity, and these properties still hold in the stochastic integral (because the kernel is deterministic). Therefore, multivariate fractional Brownian motion is stationary and self-similar in the vector sense.

Theorem 2.27. *Let $\{\boldsymbol{X}^{\boldsymbol{\kappa}}(t,B)\}$ be a d-dimensional fractional Brownian motion. For all $\lambda > 0$, $h \in \mathbb{R}$ and for almost all $(t,u) \in \mathbb{R}^2$:*

$$\begin{cases} \boldsymbol{X}^{\boldsymbol{\kappa}}(t+h) - \boldsymbol{X}^{\boldsymbol{\kappa}}(h) = \int \boldsymbol{\kappa}(t,u) \odot \mathrm{d}\boldsymbol{B_h}(u) = \boldsymbol{X}^{\boldsymbol{\kappa}}(t,B_h) \\ \boldsymbol{X}^{\boldsymbol{\kappa}}(\lambda t) = \sqrt{\lambda}\boldsymbol{\eta}(\lambda) \odot \int \boldsymbol{\kappa}(t,u) \odot \mathrm{d}\boldsymbol{B}^{\boldsymbol{\lambda}}(u) = \sqrt{\lambda}\boldsymbol{\eta}(\lambda) \odot \boldsymbol{X}^{\boldsymbol{\kappa}}(t,B^\lambda). \end{cases} \tag{2.37}$$

$\{\boldsymbol{X}^{\boldsymbol{\kappa}}(t,B)\}$ is stationary and self-similar in the vector sense and in the trajectory sense.

> Remark: Recall that, in the univariate framework, the only assumptions of stationarity and self-similarity of the quadratic norm $Q_{X^\kappa}(t)$ are equivalent to the stationarity and self-similarity of the covariance kernel $K_{X^\kappa}(t,s)$ and then of the process itself. In the multivariate framework, the analog

of the quadratic form is defined by the spatial covariance kernel, that is, the covariance kernel between two distinct spatial components taken at the same time $K_{X^\kappa}((i,t),(j,t)) = \int \kappa^i(t,u)\kappa^j(t,u)\rho_{i,j}\mathrm{d}u$, and the analog of the univariate covariance kernel $K_{X^\kappa}(t,s)$ is the cross-covariance kernel, that is, the covariance kernel between two distinct spatial components taken at distinct times, $\{K_{X^\kappa}((i,t),(j,s)) = \int \kappa^i(t,u)\kappa^j(s,u)\rho_{i,j}\mathrm{d}u\}$.

The stationarity and self-similarity properties allow the complete characterization of the covariance kernel between two distinct spatial components taken at the same time, $K_{X^\kappa}((i,t),(j,t))$, as a product between a power function and the spatial covariance-kernel at time 1.

However, the characterization of the covariance kernel between two distinct spatial components taken at two distinct times $K_{X^\kappa}((i,t),(j,s))$, is more challenging. Indeed, the self-similarity property of the bivariate kernels would have to be applied to two distinct time-components $|t|$ and $|s|$ and the self-similarity functions would depend on the distinct spatial exponents: $\kappa^i(t,u)\kappa^j(s,u) = \eta^i(|t|)\eta^j(|s|)\kappa^i\left(sgn(t), \frac{u}{|t|}\right)\kappa^j\left(sgn(s), \frac{u}{|s|}\right)$. As previously mentioned, the self-similarity property allows the transfer of the the scaling of factor $|t|$ (resp. $|s|$) from the time-variable to the integration variable in the form of an inverse scaling $\frac{1}{|t|}$ (resp. $\frac{1}{|s|}$). This raises two issues.

(i) On the one hand, because the self-similarity functions are not applied to the same time-component, we cannot obtain a power function of exponent $\gamma_i + \gamma_j$.

(ii) On the other hand, the variable change required to transfer the time-transformation from the integration variable to the Brownian of integration is not the same.

Consequently, we cannot directly compute the covariance kernel between two distinct spatial components taken at two distinct times, only using the stationarity and self-similarity properties. However, we are able to characterize the symmetrized covariance kernel using the stationarity and self-similarity properties: $K_{X^\kappa}((i,t),(j,s)) + K_{X^\kappa}((i,s),(j,t))$.

Theorem 2.28. *Let $\{\boldsymbol{X}^{\boldsymbol{\kappa}}(t,B)\}$ be a d-dimensional fractional Brownian motion.*

(i) *The covariance kernel between two distinct coordinates taken at the same time $t \in \mathbb{R}\backslash\{0\}$ is completely characterized by the power function of exponent $\gamma_i+\gamma_j$ where $\gamma_i, \gamma_j \in (0,1)$:*

$$K_{X^\kappa}((i,t),(j,t)) = K_{X^\kappa}((i,1),(j,1))\,|t|^{\gamma^i+\gamma^j}. \tag{2.38}$$

(ii) *The symmetrized covariance kernel between two distinct coordinates taken at two distinct times $X^{\kappa_i}(t,B^i)$ and $X^{\kappa_j}(s,B^j)$ with $t,s \in \mathbb{R}\backslash\{0\}$ is completely characterized by a linear combination of power functions of exponent $\gamma_i+\gamma_j$ where $\gamma_i, \gamma_j \in (0,1)$:*

$$K_{X^\kappa}((i,t),(j,s)) + K_{X^\kappa}((i,s),(j,t)) = K_{X^\kappa}((i,1),(j,1))\left(|t|^{\gamma^i+\gamma^j} + |s|^{\gamma^i+\gamma^j} - |t-s|^{\gamma^i+\gamma^j}\right). \tag{2.39}$$

Proof (i) The self-similarity property of the bivariate kernel, given by Equation (2.26), states that for all $i \in [\![1, d]\!]^2$ and for all $t > 0$:

$$\kappa^i(t, u) = \eta^i(t)\kappa^i\left(1, \frac{u}{t}\right) \quad \text{where} \quad \eta^i(t) = t^{\nu^i} \quad \text{and} \quad \nu^i \in \left(-\frac{1}{2}, \frac{1}{2}\right).$$

Under this assumption of self-similarity, the product between two bivariate kernels $\kappa^i(t, u)\kappa^j(t, u)$ becomes:

$$\kappa^i(t, u)\kappa^j(t, u) = t^{\nu^i+\nu^j}\kappa^i\left(1, \frac{u}{t}\right)\kappa^j\left(1, \frac{u}{t}\right).$$

When the covariance kernel is taken between two spatial components taken at the same time t, the variable change to use in the integral is the same for both kernels and is given taking $v = \frac{u}{t}$:

$$K_{X^\kappa}((i, t), (j, t)) = t^{\nu^i+\nu^j+1}\int \kappa^i(1, v)\kappa^j(1, v)\rho_{i,j}\mathrm{d}v = t^{\gamma^i+\gamma^j}K_{X^\kappa}((i, 1), (j, 1)).$$

If $t < 0$, then $-t > 0$, the self-simliarity property states:

$$\kappa^i(t, u) = \eta^i(-t)\kappa^i\left(-1, -\frac{u}{t}\right) \quad \text{where} \quad \eta^i(-t) = (-t)^{\nu^i} \quad \text{and} \quad \nu^i \in \left(-\frac{1}{2}, \frac{1}{2}\right).$$

Then it comes that:

$$\kappa^i(t, u)\kappa^j(t, u) = (-t)^{\nu^i+\nu^j}\kappa^i\left(-1, -\frac{u}{t}\right)\kappa^j\left(-1, -\frac{u}{t}\right).$$

Taking $v = -\frac{u}{t}$, we get:

$$K_{X^\kappa}((i, t), (j, t)) = (-t)^{\nu^i+\nu^j+1}\int \kappa^i(1, v)\kappa^j(1, v)\rho_{i,j}\mathrm{d}v = (-t)^{\gamma^i+\gamma^j}K_{X^\kappa}((i, -1), (j, -1)).$$

Without any distinction on the sign of t, the covariance kernel can be written as follows:

$$K_{X^\kappa}((i, t), (j, t)) = \left|t\right|^{\gamma^i+\gamma^j} K_{X^\kappa}((i, sgn(t)), (j, sgn(t)))$$

where $sgn(t) = 1$ if $t \geq 0$ and $sgn(t) = -1$ i $t < 0$.

Now, let us simplify the expression of $K_{X^\kappa}((i, sgn(t)), (j, sgn(t)))$. Using the stationarity property of the covariance kernel, for all $h \in \mathbb{R}$, we obtain:

$$\mathrm{Cov}\left(X^{\kappa^i}(t + h, B^i) - X^{\kappa^i}(t, B^i), X^{\kappa^j}(s + h, B^j) - X^{\kappa^j}(h, B^j)\right) = K_{X^\kappa}((i, t), (j, s)).$$

Moreover, developing the covariance of the synchronized increments, we obtain:

$$\begin{aligned}&\mathrm{Cov}\left(X^{\kappa^i}(t + h, B^i) - X^{\kappa^i}(t, B^i), X^{\kappa^j}(s + h, B^j) - X^{\kappa^j}(h, B^j)\right)\\ &= K_{X^\kappa}((i, t + h), (j, s + h)) + K_{X^\kappa}((i, h), (j, h)) - K_{X^\kappa}((i, t + h), (j, h)) - K_{X^\kappa}((i, h), (j, s + h)).\end{aligned}$$

Finally, we have:

$$K_{X^\kappa}((i,t),(j,s))$$
$$= K_{X^\kappa}((i,t+h),(j,s+h)) + K_{X^\kappa}((i,h),(j,h)) - K_{X^\kappa}((i,t+h),(j,h)) - K_{X^\kappa}((i,h),(j,s+h)).$$

Taking $s = t = -1$ and $h = 1$, we obtain:

$$K_{X^\kappa}((i,-1),(j,-1)) = K_{X^\kappa}((i,0),(j,0)) + K_{X^\kappa}((i,1),(j,1)) - K_{X^\kappa}((i,0),(j,1)) - K_{X^\kappa}((i,1),(j,0)).$$

Using $K_{X^\kappa}((i,0),(j,0)) = K_{X^\kappa}((i,1),(j,0)) = K_{X^\kappa}((i,0),(j,1)) = 0$, we obtain the symmetry of the covariance kernel:

$$K_{X^\kappa}((i,-1),(j,-1)) = K_{X^\kappa}((i,1),(j,1)).$$

More generally, $K_{X^\kappa}((i,-1),(j,-1)) = K_{X^\kappa}((i,sgn(t)),(j,sgn(t))) = K_{X^\kappa}((i,sgn(s)),(j,sgn(s))) = K_{X^\kappa}((i,1),(j,1))$.

Finally, we obtain:

$$K_{X^\kappa}((i,t),(j,t)) = |t|^{\gamma^i+\gamma^j} K_{X^\kappa}((i,1),(j,1)).$$

(ii) In the univariate framework, the covariance kernel between two distinct times is established from the quadratic norm of the increment of the process taken at the two given times. Similarly, in the multivariate framework, the covariance kernel between two distinct spatial components taken at two distinct times is based on the covariance kernel between two synchronized increments of the two distinct spatial components. This covariance kernel depends both on the covariance kernel between the two distinct spatial components taken at the same time and on the symmetrized covariance kernel between the two distinct spatial components taken at two distinct times:

$$\mathrm{Cov}\left(X^{\kappa^i}(t,B^i) - X^{\kappa^i}(s,B^i), X^{\kappa^j}(t,B^j) - X^{\kappa^j}(s,B^j)\right)$$
$$= K_{X^\kappa}((i,t),(j,t)) + K_{X^\kappa}((i,s),(j,s)) - K_{X^\kappa}((i,t),(j,s)) - K_{X^\kappa}((j,t),(i,s)).$$

Then, the symmetrized covariance kernel is given by:

$$K_{X^\kappa}((i,t),(j,s)) + K_{X^\kappa}((j,t),(i,s))$$
$$= K_{X^\kappa}((i,t),(j,t)) + K_{X^\kappa}((i,s),(j,s))$$
$$- \mathrm{Cov}\left(X^{\kappa^i}(t,B^i) - X^{\kappa^i}(s,B^i), X^{\kappa^j}(t,B^j) - X^{\kappa^j}(s,B^j)\right).$$

From Theorem 2.27, $\{\boldsymbol{X}^{\boldsymbol{\kappa}}(t,B)\}$ is stationary, then:

$$\mathrm{Cov}\left(X^{\kappa^i}(t,B^i) - X^{\kappa^i}(s,B^i), X^{\kappa^j}(t,B^j) - X^{\kappa^j}(s,B^j)\right) = \mathrm{Cov}\left(X^{\kappa^i}(t-s,B^i), X^{\kappa^j}(t-s,B^j)\right).$$

Moreover by symmetry of the covariance kernel and from the stationarity we get:

$$\mathrm{Cov}\left(X^{\kappa^i}(t,B^i) - X^{\kappa^i}(s,B^i), X^{\kappa^j}(t,B^j) - X^{\kappa^j}(s,B^j)\right)$$
$$= \mathrm{Cov}\left(X^{\kappa^i}(s,B^i) - X^{\kappa^i}(t,B^i), X^{\kappa^j}(s,B^j) - X^{\kappa^j}(t,B^j)\right)$$
$$= \mathrm{Cov}\left(X^{\kappa^i}(s-t,B^i), X^{\kappa^j}(s-t,B^j)\right).$$

Finally, we are able to express the symmetrized covariance kernel as the sum of three covariance kernels between two distinct components taken at the same time (t, s, and t - s).

$$\begin{aligned}&K_{X^\kappa}((i,t),(j,s)) + K_{X^\kappa}((j,t),(i,s))\\&= K_{X^\kappa}((i,t),(j,t)) + K_{X^\kappa}((i,s),(j,s)) - K_{X^\kappa}((i,t-s),(j,t-s)).\end{aligned}$$

Thus, the self-similarity property can be applied to each of these three covariance kernels and allows expressing each of them as a power function with an exponent depending on the two spatial components, multiplied by a covariance kernel taken at time 1. Therefore, the symmetrized covariance kernel can be expressed as a linear combination of power functions with exponents depending on the two spatial components, multiplied by the covariance kernel at time 1.

$$\begin{aligned}&K_{X^\kappa}((i,t),(j,s)) + K_{X^\kappa}((j,t),(i,s))\\&= |t|^{\gamma_i+\gamma_j} K_{X^\kappa}((i,sgn(t)),(j,sgn(t))) + |t|^{\gamma_i+\gamma_j} K_{X^\kappa}((i,sgn(s)),(j,sgn(s)))\\&- |t-s|^{\gamma_i+\gamma_j} K_{X^\kappa}((i,sgn(t-s)),(j,sgn(t-s))).\end{aligned}$$

From the symmetry property of the covariance kernel at time 1 and -1, we have $K_{X^\kappa}((i,sgn(t)),(j,sgn(t))) = K_{X^\kappa}((i,sgn(s)),(j,sgn(s))) = K_{X^\kappa}((i,sgn(t-s)),(j,sgn(t-s))) = K_{X^\kappa}((i,1),(j,1))$.

Therefore, we obtain the following formula for the symmetrized covariance kernel:

$$K_{X^\kappa}((i,t),(j,s)) + K_{X^\kappa}((j,t),(i,s)) = K_{X^\kappa}((i,1),(j,1))(|t|^{\gamma_i+\gamma_j} + |s|^{\gamma_i+\gamma_j} - |t-s|^{\gamma_i+\gamma_j}).$$

■

The explicit formula for the cross-covariance kernel at time 1, $(K_{X^\kappa}((i,1),(j,1)))_{i,j\in[\![1,d]\!]}$ is provided in the following lemma.

Lemma 2.29 (Covariance at time 1). *Let $\gamma^i, \gamma^j \in (0,1)$, then:*

$$K_{X^\kappa}((i,1),(j,1)) = \rho_{i,j}\left[\frac{\left(\gamma^i - \frac{1}{2}\right)\left(\frac{3}{2} - \gamma^i\right)}{(\gamma^i+\gamma^j-1)(\gamma^i+\gamma^j)}\mathfrak{B}\left(2-\gamma^i-\gamma^j, \gamma^j+\frac{1}{2}\right)\right. \tag{2.40}$$

$$\left. + \frac{\left(\gamma^j - \frac{1}{2}\right)\left(\frac{3}{2} - \gamma^j\right)}{(\gamma^i+\gamma^j-1)(\gamma^i+\gamma^j)}\mathfrak{B}\left(2-\gamma^i-\gamma^j, \gamma^i+\frac{1}{2}\right)\right] \tag{2.41}$$

where $\mathfrak{B}$ refers to the function beta and takes the form of $\mathfrak{B}(\alpha,\beta) = \int_0^1 u^{\alpha-1}(1-u)^{\beta-1}du$.

A proof will be provided in Chapter 3.

With some additional efforts, we can provide a closed-form formula for the covariance kernel between two distinct spatial components taken two distinct times, and not only for the symmetrized covariance kernel, as in the works of (Lavancier et al., 2009). Indeed, we can adapt to the $\mathbb{L}^2$-spaces framework, the works of (Lavancier et al., 2009)

that study multidimensional processes that satisfy the operator self-similarity and stationarity properties in terms of finite dimensional distributions. In other words, we can adapt the operator self-similarity and stationarity properties of the processes in terms of $\mathbb{L}^2$-norm and inner product.

Some words about the works of (Lavancier et al., 2009): Authors deal with the operator self-similarity (Os-s) for finite dimensional distributions. For all $\lambda > 0$, $t \in \mathbb{R}$:

$$\boldsymbol{X}(\lambda t) \overset{fdd}{=} \boldsymbol{\lambda^H} \boldsymbol{X}(t)$$

where $\overset{fdd}{=}$ means equality of finite-dimensional distributions, and the $d \times d$ matrix $\boldsymbol{\lambda^H}$ is defined by the power series:

$$\boldsymbol{\lambda^H} = e^{H \log(\lambda)} = \sum_{k=0}^{+\infty} \frac{H^k (\log(\lambda))^k}{k!}. \tag{2.42}$$

In these works, a Gaussian operator self-similar process (Os-s) with stationary increments (si) is called operator fractional Brownian motion (ofBm). Authors show that for $d = 1$ the class of ofBm coincides with the fundamental class of fractional Brownian motions (fBm) as defined in (Taqqu, 1994). In other words, they show that the class of ofBm coincides with stochastically continuous self-similar Gaussian process with stationary increments, zero mean and covariance kernel:

$$K_X(t,s) = \frac{V_X(1)}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H}), \quad t,s \in \mathbb{R}. \tag{2.43}$$

The form of covariance of general ofBm seems to be unknown and may be quite complicated. Authors focus on a particular case of operator self-similar processes that corresponds to diagonal matrix $H = diag(H_1, \ldots, H_d)$. In this case, the operator self-similarity is expressed as follows:

$$(X_1(\lambda t), \ldots, X_d(\lambda t)) \overset{fdd}{=} (\lambda^{H_1} X_1(t), \ldots, \lambda^{H_d} X_d(t)).$$

A d-variate process X satisfying the above property of operator self-similarity (Os-s) with diagonal matrix is called vector self-similar (vs-s) and a stochastically continuous Gaussian vs-s process with stationary increments (si) is called a vector fractional Brownian motion (vfBm). Each component $X_i = \{X_i(t), t \in \mathbb{R}\}$ of a vector self-similar process is a scalar self-similar process, the fact which is not true for general operator self-similar process:

$$X_i(\lambda t) \overset{fdd}{=} \lambda^{H_i} X_i(t), \quad \lambda > 0, t \in \mathbb{R}, i \in [\![1, d]\!]. \tag{2.44}$$

In their works, the main result concerns the closed-form formula of the covariance kernel of vfBm given by (Lavancier et al., 2009, Theorem 2.1., p.3.).

Theorem 2.30 ((Lavancier et al., 2009)). *Let $\boldsymbol{X} = \{X(t), t \in \mathbb{R}\}$ be a 2nd order process with values in $\mathbb{R}^d$. Assume that $\boldsymbol{X}$ has stationary increments, zero mean, $\boldsymbol{X}(0) = \boldsymbol{0_d}$, and that $\boldsymbol{X}$ is vector self-similar with exponent $\boldsymbol{H} = (H_1, \ldots, H_d)$, $0 < H_i < 1$,*

$(i = 1, \dots, d)$. Moreover, assume also that for any $i, j = 1, \dots, d$ the function $t \mapsto K_X((i,t),(j,1))$ is continuously differentiable on $(0,1) \cup (1, +\infty)$. Let $\sigma_i^2 > 0$ denote the variance of $X_i(1)$, $i = 1, \dots, d$.

(i) *If $i = j$, then for any $(s,t) \in \mathbb{R}^2$, we have:*

$$K_X((i,s),(i,t)) = \frac{\sigma_i^2}{2}(|s|^{2H_i} + |t|^{2H_i} - |t-s|^{2H_i}). \tag{2.45}$$

(ii) *If $i \neq j$ and $H_i + H_j \neq 1$, then there exists $c_{ij}, c_{ji} \in \mathbb{R}$ such that for any $(s,t) \in \mathbb{R}^2$,*

$$K_X((i,s),(j,t)) = \frac{\sigma_i \sigma_j}{2}(c_{ij}(s)\,|s|^{H_i+H_j} + c_{ji}(t)\,|t|^{H_i+H_j} - c_{ji}(t-s)\,|t-s|^{H_i+H_j}) \tag{2.46}$$

where:

$$c_{ij}(t) = \begin{cases} c_{ij} & \text{if } t > 0 \\ c_{ji} & \text{if } t < 0. \end{cases} \tag{2.47}$$

(iii) *If $i \neq j$ and $H_i + H_j = 1$, then there exists $d_{ij}, f_{ij} \in \mathbb{R}$ such that for any $(s,t) \in \mathbb{R}^2$ we have:*

$$K_X((i,s),(j,t)) = \frac{\sigma_i \sigma_j}{2}(d_{ij}(|s| + |t| - |s-t|) + f_{ij}(t\log(|t|) - s\log(|s|) - (t-s)\log(|t-s|))). \tag{2.48}$$

(iv) *The matrix $R = (R_{ij})_{i,j=1,\dots,d}$ is positive definite, where:*

$$R_{ij} := \begin{cases} 1 & \text{if } i = j \\ c_{ij} + c_{ji} & \text{if } i \neq j, H_i + H_j \neq 1, \\ d_{ij} & \text{if } i \neq j, H_i + H_k = 1. \end{cases} \tag{2.49}$$

The notion of operator self-similarity is also formulated in the works of (Laha and Rohatgi, 1981) in a more abstract way. In further works, we will extend their approach of operator self-similarity of the processes to the $\mathbb{L}^2$-spaces in terms of quadratic norm and inner product.

Some words about the works of (Laha and Rohatgi, 1981): In their works, operator self-similar stochastic processes taking values in a finite dimensional Euclidean space are introduced and some of their properties are studied. They provide a more abstract definition of operator self-similarity.

Definition 2.31 ((Laha and Rohatgi, 1981))**.** *An $\mathbb{R}^d$-valued stochastic process $\{\boldsymbol{X}(t), t > 0\}$ is said to be operator self-similar if it satisfies the weak convergence condition:*

$$\boldsymbol{X}(t+h) \overset{d}{\Rightarrow} \boldsymbol{X} \quad as \quad h \to 0, \quad (t + h > 0) \tag{2.50}$$

for every $t > 0$ and for every $\lambda > 0$ there exists $\boldsymbol{B}(\lambda) > 0$ (linear operator in $\mathbb{R}^d$ is invertible, self-adjoint and positive) and an element $\boldsymbol{c}(\lambda) \in \mathbb{R}^d$ such that $\{\boldsymbol{X}(\lambda t)\}$ and $\{\boldsymbol{B}(\lambda)\boldsymbol{X}(t) + \boldsymbol{c}(\lambda)\}$ have the same finite dimensional distributions. More precisely, for

every finite set of positive real numbers $t_1, t_2, \dots t_k$ *the random vectors* $(X(\lambda t_1), \dots, X(\lambda t_k))$ *and* $(B(\lambda)X(t_1) + c(\lambda), \dots, B(\lambda)X(t_k) + c(\lambda))$ *have the same probability distributions. In this case we shall write the operator self-similarity in its strong form:*

$$\boldsymbol{X}(\lambda \cdot) \stackrel{d}{=} \boldsymbol{B}(\lambda)\boldsymbol{X}(\cdot) + \boldsymbol{c}(\lambda) \cdot . \tag{2.51}$$

The assumption $\boldsymbol{B}(\lambda) > 0$ *is a natural one to make in view of the work of (Lamperti, 1962) where authors considered the case* $\boldsymbol{B}(\lambda) = b(\lambda)\boldsymbol{I}$. *Here* $b(\lambda) > 0$ *and* $\boldsymbol{I}$ *is the identity operator in* $\mathbb{R}^d$. *We note that in most of the applications of assumption (2.51) we shall be using the weaker form:*

$$\boldsymbol{X}(\lambda t) \stackrel{d}{=} \boldsymbol{B}(\lambda)\boldsymbol{X}(t) + \boldsymbol{c}(\lambda) \tag{2.52}$$

where the equality is in distribution of random variables.

They prove the form of the self-similar operator $\boldsymbol{B}(\lambda)$ in (Laha and Rohatgi, 1981, Theorem 2, p.78).

Theorem 2.32 ((Laha and Rohatgi, 1981)). *Let* $\{\boldsymbol{X}(t), t > 0\}$ *be proper operator self-similar. Then there exists a self-adjoint operator* $\boldsymbol{H} \in \mathcal{M}(d, \mathbb{R}_1)$ *such that* $\boldsymbol{B}(\lambda) = \boldsymbol{\lambda}^{\boldsymbol{H}}, \lambda > 0$, *in the sense that* $\boldsymbol{\lambda}^{\boldsymbol{H}} = e^{ln(\lambda)\cdot H}$. *If* $\boldsymbol{H} = \boldsymbol{0}$, *then for all* $\lambda > 0$, $\boldsymbol{B}(\lambda) = \boldsymbol{I}$, $\boldsymbol{c}(\lambda) = (ln(\lambda))\boldsymbol{\omega}$ *for some* $\boldsymbol{\omega} \in \mathbb{R}^d$ *and* $\boldsymbol{X}(\lambda \cdot) \stackrel{d}{=} \boldsymbol{X}(\cdot) + ln(\lambda)\boldsymbol{\omega}$. *If* $\boldsymbol{H}$ *is invertible, then there exists an element* $\boldsymbol{\omega} \in \mathbb{R}^d$ *such that* $\boldsymbol{c}(\lambda) = (\boldsymbol{I} - \boldsymbol{\lambda}^{\boldsymbol{H}})\boldsymbol{\omega}$ *for all* $\lambda > 0$.

Theorem 2.33 ((Laha and Rohatgi, 1981)). *With the same notations and assumptions as in Theorem 2.32 suppose further that the self-similarity property holds also for* $t = 0$. *Then, the following assertions hold.*

(i) *If* $\boldsymbol{H} = \boldsymbol{0}$, *then* $\boldsymbol{c}(\lambda) = \boldsymbol{0_d} \in \mathbb{R}^d$ *and* $\boldsymbol{X}(\lambda) \stackrel{d}{=} \boldsymbol{X}(0)$ *for all* $\lambda \geq 0$.

(ii) *If* $\boldsymbol{H} \neq \boldsymbol{0}$, *then the distribution* μ_0 *of* $\boldsymbol{X}(0)$ *cannot be full. In particular, if* $\boldsymbol{H}$ *is invertible, then* $\boldsymbol{X}(0)$ *is degenerate at some* $\boldsymbol{\omega} \in \mathbb{R}^d$ *and* $\boldsymbol{c}(\lambda) = (\boldsymbol{I} - \boldsymbol{\lambda}^{\boldsymbol{H}})\boldsymbol{\omega}$ *for all* $\lambda \geq 0$.

Theorem 2.34 ((Laha and Rohatgi, 1981)). *An* $\mathbb{R}^d$*-valued process* $\{\boldsymbol{X}(t), t \geq 0\}$ *with stationary independent increments and* $\boldsymbol{X}(0) = \boldsymbol{0}$ *a.s. is proper operator self-similar if and only if the increments have distributions which are full and proper operator stable.*

> REMARK: In the following chapter the fractional Brownian motion $\{X^{\kappa^i}(t, B^i)\}$ will be denoted as $\{B^{H_i}(t)\}$, the exponent γ_i of the power function will correspond to the Hurst H_i exponent and the bivariate kernel $\kappa^i(t, u)$ will be replaced by $\psi^{H_i}_{0,t}(u)$.

2.4 Conclusion

This chapter provides an original approach for the theoretical study of the processes that satisfy the properties of self-similarity and stationarity of the quadratic norm in $\mathbb{L}^2$-spaces. In this chapter, we have proven that under the assumptions of square integrability, Q_X-stationarity, Q_X-self-similarity, Q_X-right-continuity with $Q_X(0) = 0$, we are able to:

(i) characterize the Q_X-norm owing to power functions,

(i) fully characterize the inner product thanks to the Q_X-norm,

(ii) obtain the $\mathbb{L}^2$-stationarity and the $\mathbb{L}^2$-self-similarity,

without any assumption of regularity on the processes.

Then, we are interested in the distributions characterized by the $\mathbb{L}^2$-characteristics. One of these distributions is the well-known Gaussian distribution. Adding the Gaussian assumption, we have proven that the previous assumptions were sufficient to obtain the stationarity and the self-similarity in distribution. And for well-known Gaussian processes such as Brownian motion and fractional Brownian motion, we are able to obtain the pathwise stationarity and self-similarity (that is, in the trajectory sense). Finally, we extend the properties of stationarity and self-similarity in distribution and in trajectory to the multivariate framework, especially to obtain the multivariate fractional Brownian motion. We show that the multivariate fractional Brownian motion is a Gaussian multidimensional self-similar process with stationary increments the vector sense and in the trajectory sense. A closed form formula for its symmetrized covariance kernel is provided as a linear combination of power functions. With some additional efforts, a closed-form formula for the covariance kernel can be provided as mentioned in the works of (Lavancier et al., 2009) and (Laha and Rohatgi, 1981). These works study operator self-similar and stationary processes in terms of finite dimensional distributions. Based on the works of (Lavancier et al., 2009) and (Laha and Rohatgi, 1981), further works will adapt the operator self-similarity and stationarity properties of the processes to the framework of the $\mathbb{L}^2$-spaces in terms of quadratic norm and inner kernel. However, the link between these processes and data has not yet been addressed. This is the subject of the following chapter.

"Clouds are not spheres, mountains are not cones, coastlines are not circles, and bark is not smooth, nor does lightning travel in a straight line. Fractal geometry is not just a chapter of mathematics, but one that helps every man to see the same world differently."

(**Benoît Mandelbrot**, The Fractal Geometry of Nature, 1982.)

Chapter 3

VaR prediction for asset portfolios under fractional dynamics

Contents

This chapter addresses research questions ($RQ\#2$, ($RQ\#3$)). In this chapter, we describe the price trajectories using fractional geometric Brownian motions. This allows the addition of correlations between logarithmic returns to express long-range dependency. Logarithmic returns are then described using self-similar Gaussian processes with stationary and correlated increments, called fractional Brownian motion (fBm). In this context, we focus on predicting the most commonly used risk measure by regulators, called Value-at-Risk (VaR). We introduce a model that provides a Gaussian approximation of the VaR for the asset portfolio under fractional dynamics. We demonstrate that such a model is based on orthogonal projections in a Gaussian Hilbert space, taking specific forms for which closed-form formulae are provided. Finally, we quantify the Gaussian approximation of VaR by providing an upper bound for the error. Backtesting experiments are conducted using simulated and market data to illustrate this theory.

3.1 Introduction

In recent decades, the reinforcement of banking regulation agreements to ensure the solvency of financial institutions, has highlighted the importance of an accurate evaluation of financial risks. The goal of this regulation is twofold: on the one hand, it is intended to ensure that financial institutions keep aside a sufficient amount of money to cope with the risk of default, on the other hand, it pays attention not to overcharge the cash reserve. Risk measures are dedicated to evaluating financial risks incurred by a financial institution. Therefore, it is crucial to develop efficient and accurate predictive models for these risk measures. In this chapter, we focus on the market Value-at-Risk (**VaR**) of portfolio Profit and Loss (P&L) of a bank, as it is currently

the most used risk measure by regulators. The **VaR** at risk level $\alpha \in (0,1)$ corresponds to the quantile at order α of the $P\&L$ distribution, which is mathematically given by $\mathbf{VaR}_\alpha(X) := \inf\{x \in \mathbb{R}, F_X(x) \geq \alpha\}$ where F_X is the cumulative distribution function related to the random variable X representing the P&L. For more details about **VaR**, we refer the reader to the well-known books of (Wipplinger, 2007, Part II. p105) and (McNeil et al., 2015, Chap.2, p37). Computing **VaR** is a challenging problem because it requires knowing the distribution of the portfolio P&L, which is usually unknown.

The literature on **VaR** is extensive, with more than 2700 articles referenced on Google Scholar. A wide range of papers has focused on the prediction of the $\mathbf{VaR}_\alpha$. To name a few, see for instance Cheung and Powell (2012) which presents a teaching study using parametric computation and Monte-Carlo simulation to compute **VaR**, and Feuerverger and Wong (2000) that explains the computation of **VaR** for nonlinear portfolios. Although **VaR** is not a perfect risk metric (see Delbaen et al. (1998)), it is still commonly used. It is fundamental to have an accurate estimate of the **VaR**: in case of under-estimation, the financial institution will take too much risk, without being prepared for this; in case of over-estimation, the amount of money to be kept aside would be too high, preventing some banking activities. In addition, its computation must be efficient, because its evaluation is frequent and must be performed according to numerous portfolios.

In this work, we contribute to the development of smart models capable of computing **VaR** as accurately as possible, and we provide backtesting experiments. We consider a challenging setting in which the price dynamics are described by fractional Black-Scholes models, depending on fractional Brownian motions parameterized by their Hurst exponent. These models have the advantage of capturing both correlations in time and between assets. Compared to the usual Black-Scholes model, the flexibility in the choice of the Hurst exponent allows for a better description of price trajectories to fit reality. Thus, the predictive model for the conditional **VaR** at time horizon h (i.e., the **VaR** of P&L over the next period of length h conditionally to the observations available at the computation time) will be able to predict in a more accurate way, the amount of money to keep aside, that is, neither over-charge nor under-charge the regulator. The accuracy of our **VaR** predictions is assessed using backtesting procedure based on Christoffersen (1998).

The use of fractional models in finance is not new, it has become increasingly important in recent years. Fractional models have been widely used for different purposes, such as modeling and pricing in foreign exchange markets Shokrollahi et al. (2014), predicting rough volatility Gatheral et al. (2017), and modeling stock prices Adamu (2017). For stochastic calculus tools for fractional Brownian motions, see Biagini et al. (2008b). However, these studies do not address, the prediction of conditional **VaR**.

Overall, we make the following contributions. We design a theoretical framework for fractional models in Gaussian Hilbert space. We design a closed-form formula for the Gaussian approximation of the conditional $\mathbf{VaR}_\alpha$, at the time horizon h, of the future portfolio's variation under fractional dynamics. We then provide error quantification of the Gaussian approximation. We use a robust methodology to estimate the parameters

of the fractional model. We perform a backtesting procedure to assess the accuracy of our predictive approximation. We support our analysis through various experiments that illustrate the behavior of our new model.

Furthermore, we support our analysis through various experiments that illustrate the behavior of our new algorithm. The remainder of this chapter is organized as follows. In Section 3.2, we discuss the prerequisites for this study. This section is divided into three parts.

First, fractional models are introduced via the Gaussian Hilbert space. Then, we define the conditional expectation and the conditional variance of a future fBm increment, owing to orthogonal projections in the Gaussian Hilbert space spanned by the past fBm increments. In this theoretical framework, the market model is introduced in Subsection 3.2.3.

Section 3.3 provides a model for the conditional **VaR** of a Gaussian approximation of the future portfolio increment, quantification of the approximation error via error bounds and a proof of the error bounds.

We then present experiments that support this theory. The theoretical results related to the algorithms used to implement the models are successively presented. In Section 3.4, the methodology used to simulate fBm and mfBm is presented. In Section 3.5, we present the estimation method used to estimate the parameters of the **VaR** model. In Section 3.6, we present the backtesting process used to assess the performance of the **VaR** model. In Section 3.7, an empirical study is carried out with graphs of convergence of the estimators, graphs of the **VaR** trajectories, extremal indices, and statistics tables to assess the performance of the models. A comparative study between different models is also presented. Finally, in Section 3.8, **VaR** models are applied to real market data, and the same graphs and tables are provided to assess the performance of the models. Again, a comparative study between the models is carried out.

3.2 Models and theoretical backgrounds

In this section, we collect fundamental notions about the fractional models required in the sequel, and state some instrumental results. There is no unique way to model multivariate fractional Brownian motions: the most direct way is to give their covariance functions as a multivariate Gaussian process; however, this method is quite inconvenient when dealing with conditional in time computations, with various filtrations to account for the available observations. Therefore, we prefer to work directly at the level of the Gaussian Hilbert space indexed by the functions in

$$\mathcal{L}_d^2 := L_2(\mathbb{R}^d, \mathrm{d}u) = \left\{ f : \mathbb{R} \mapsto \mathbb{R}^d \ s.t. \ \left|f\right|_{\mathcal{L}_d^2}^2 := \int_{\mathbb{R}} |f(u)|^2 \mathrm{d}u < +\infty \right\} \tag{3.1}$$

where the basic quantities are defined by Wiener integrals, and where projections and conditional expectations are made through projections of functions in $\mathcal{L}_d^2$ (see Lemma 3.3 later). Here, integer d is the number of assets in the portfolio.

3.2.1 Fractional models via Gaussian Hilbert space

We adopt the framework of the isonormal Gaussian process associated with the Hilbert space $\mathcal{L}^2_d$, with the scalar product

$$\langle f, g\rangle_{\mathcal{L}^2_d} = \int_{\mathbb{R}} f(u)\cdot g(u)\mathrm{d}u, \tag{3.2}$$

also called Gaussian Hilbert space; here $f(u)\cdot g(u)$ is just the scalar product in $\mathbb{R}^d$ between the vector-valued functions f and g at point u. See (Janson, 1997) for a broad account on Gaussian Hilbert space. Namely, we consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ supporting a centered Gaussian family of scalar random variables

$$\mathcal{H} = \{I(f) : f \in \mathcal{L}^2_d\} \tag{3.3}$$

as defined in (Janson, 1997, Chapter 7, Thm.7.1), such that the isometry property is in force

$$\mathbb{E}\left[I(f)I(g)\right] = \langle f, g\rangle_{\mathcal{L}^2_d}. \tag{3.4}$$

One way to realize this Gaussian Hilbert space indexed by square integrable functions $f \in \mathcal{L}^2_d$ is to assume that the probability space supports a d-dimensional Brownian motion indexed by the real line, $(W_t = (W^1_t, \cdots, W^d_t)^\top : t \in \mathbb{R})$, and to define $I(f)$ by stochastic integration

$$I(f) = \int_{\mathbb{R}} f(u)\cdot \mathrm{d}W_u = \sum_{i=1}^{d}\int_{\mathbb{R}} f^i(u)\mathrm{d}W^i_u \tag{3.5}$$

(a.k.a. Wiener integral). See (Janson, 1997, Chapter 7) where integrals are restricted to $\mathbb{R}^+$, extension to $\mathbb{R}$ is immediate. The fractional Brownian motion (fBm) with Hurst parameter H corresponds to a specific choice of function f: this is the well-known time-representation of Mandelbrot and Van Ness (Mandelbrot and Van Ness, 1968b).

Definition-Proposition 3. *Let $H \in (0,1)$ and set, for any $t \geq 0$,*

$$B^H_t = I((\psi^H_{0,t}(.), 0, \cdots, 0)^\top) = \int_{\mathbb{R}} \psi^H_{0,t}(u)\mathrm{d}W^1_u \quad \textit{with} \quad \psi^H_{s,t}(u) = \frac{1}{c_H}\left((t-u)_+^{H-\frac{1}{2}} - (s-u)_+^{H-\frac{1}{2}}\right), \tag{3.6}$$

where

$$c_H := \sqrt{\frac{1}{2H}\left(\frac{3}{2}-H\right)\mathfrak{B}\left(2-2H, H+\frac{1}{2}\right)} \tag{3.7}$$

and $\mathfrak{B}(x,y) = \int_0^1 u^{x-1}(1-u)^{y-1}\mathrm{d}u$ is the Beta function.

This defines a scalar fBm B^H, that is, a centered Gaussian process with covariance

$$\mathrm{Cov}\left(B^H_t, B^H_s\right) = \frac{1}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H}), \quad H \in (0,1). \tag{3.8}$$

Remarks on the integrability of kernel $u \mapsto \psi^H_{0,t}(u)$. In the above, the adopted convention is that $(-v)^\alpha_+ = 0$ for any value $\alpha \in \mathbb{R}$ when v is non-negative. This ensures $\psi^H_{0,t}(u) = 0$ for large positive values of u, thereby removing any integrability issues around $+\infty$.

In addition, a direct Taylor expansion shows that $\psi^H_{0,t}(u) \sim \gamma_t u^{H-\frac{3}{2}}$ as $u \to -\infty$ for some constant γ_t, which shows that the kernel $\psi^H_{0,t}(\cdot)$ is square integrable at $-\infty$ because $2H - 3 < -1$.

Proof First, the fact that B^H is a centered Gaussian process stems directly from the usual properties of the Wiener integral. Second, we compute the variance of B^H_1 as follows:

$$\mathbb{V}\left[B^H_1\right] = \mathbb{E}\left[(B^H_1)^2\right] = \int_{\mathbb{R}} (\psi^H_{0,1}(u))^2 \mathrm{d}u = \frac{1}{c_H^2} \int_{\mathbb{R}} \left((1-u)_+^{H-\frac{1}{2}} - (-u)_+^{H-\frac{1}{2}} \right)^2 \mathrm{d}u = 1 \tag{3.9}$$

where, at the last equality, we have used that

$$\int_{\mathbb{R}} \left((1-u)_+^{H-\frac{1}{2}} - (-u)_+^{H-\frac{1}{2}} \right)^2 \mathrm{d}u = c_H^2. \tag{3.10}$$

This equality is justified at the end. Third, using stationarity and self-similarity properties of Mandelbrot's kernel, the variance of $B^H_t - B^h_s$ (for $t > s$) is

$$\mathbb{E}\left[\left(B^H_t - B^H_s \right)^2 \right] = \frac{1}{c_H^2} \int_{\mathbb{R}} \left((t-u)_+^{H-\frac{1}{2}} - (s-u)_+^{H-\frac{1}{2}} \right)^2 \mathrm{d}u \tag{3.11}$$

$$\overset{(3.10)}{=} \frac{(t-s)^{2H}}{c_H^2} \int_{\mathbb{R}} \left((1-z)_+^{H-\frac{1}{2}} - (-z)_+^{H-\frac{1}{2}} \right)^2 \mathrm{d}z \tag{3.12}$$

$$= (t-s)^{2H}, \tag{3.13}$$

using the change of variable $u - s = z(t-s)$ at the second equality.

Last, a direct computation shows that

$$\mathrm{Cov}\left(B^H_s, B^H_t \right) = \frac{1}{2} \left(\mathbb{E}\left[(B^H_t)^2\right] + \mathbb{E}\left[(B^H_s)^2\right] - \mathbb{E}\left[(B^H_t - B^H_s)^2\right] \right), \quad \forall s, t \geq 0. \tag{3.14}$$

Applying the identity (3.13) with different times s and t, we get

$$\mathrm{Cov}\left(B^H_s, B^H_t \right) = \frac{1}{2} \left(|t|^{2H} + |s|^{2H} - |t-s|^{2H} \right) \tag{3.15}$$

as announced in (3.8).

However, Equation (3.10) remains to be proven. Let us denote by $\mathcal{I}$ the integral on the left-hand side of Equation (3.10) and let us break it down into an integral on $[0,1]$ and another on $]-\infty, 0]$:

$$\mathcal{I} = \int_0^1 (1-u)^{2H-1} \mathrm{d}u + \int_0^{+\infty} \left((1+u)^{H-\frac{1}{2}} - (u)^{H-\frac{1}{2}} \right)^2 \mathrm{d}u := \mathcal{I}_1 + \mathcal{I}_2. \tag{3.16}$$

Obviously, $\mathcal{I}_1 = \frac{1}{2H}$. Making $\mathcal{I}_2$ explicit is more delicate. Set

$$\phi(x) := \int_0^{+\infty} \left((x+u)^{H-\frac{1}{2}} - u^{H-\frac{1}{2}} \right)^2 \mathrm{d}u, \quad x > 0. \tag{3.17}$$

We proceed as in (Picard, 2011, p.55). Differentiating twice gives

$$\phi'(x) = \big(2H-1\big) \int_0^{+\infty} (x+u)^{H-\frac{3}{2}} \left((x+u)^{H-\frac{1}{2}} - u^{H-\frac{1}{2}} \right) \mathrm{d}u \tag{3.18}$$

$$\phi''(x) = (2H-1)(2H-2) \int_0^{+\infty} (x+u)^{2H-3} \mathrm{d}u - (2H-1)\left(H-\frac{3}{2}\right) \int_0^{+\infty} (x+u)^{H-\frac{5}{2}} u^{H-\frac{1}{2}} \mathrm{d}u \tag{3.19}$$

$$= -(2H-1)x^{2H-2} - (2H-1)\left(H-\frac{3}{2}\right) x^{2H-2} \int_1^{+\infty} y^{H-\frac{5}{2}} (y-1)^{H-\frac{1}{2}} \mathrm{d}y \tag{3.20}$$

using the change of variable $u = x(y-1)$ in the second integral. Setting $y = \frac{1}{w}$ yields

$$\int_1^{+\infty} y^{H-\frac{5}{2}} (y-1)^{H-\frac{1}{2}} \mathrm{d}y = \int_0^1 w^{1-2H} (1-w)^{H-\frac{1}{2}} \mathrm{d}w = B\left(2-2H, H+\frac{1}{2}\right) \tag{3.21}$$

using the Beta function. We have established

$$\phi''(x) = (2H-1)x^{2H-2} \left(-1 - \left(H-\frac{3}{2}\right) \mathfrak{B}\left(2-2H, H+\frac{1}{2}\right) \right). \tag{3.22}$$

Integrate twice gives

$$\phi(x) = -\frac{x^{2H}}{2H} \left(1 + \left(H-\frac{3}{2}\right) \mathfrak{B}\left(2-2H, H+\frac{1}{2}\right) \right) + ax + b. \tag{3.23}$$

Parameters a and b are zero because $\phi'(x)$ and $\phi(x)$ are respectively proportional to x^{2H-1} and x^{2H}. We deduce that

$$\mathcal{I}_2 = \phi(1) = \frac{1}{2H} \left(\left(\frac{3}{2} - H\right) \mathfrak{B}\left(2-2H, H+\frac{1}{2}\right) - 1 \right).$$

All in all, this proves that

$$\mathcal{I} = \frac{1}{2H} \left(\frac{3}{2} - H\right) \mathfrak{B}\left(2-2H, H+\frac{1}{2}\right). \tag{3.24}$$

■

FBm enjoys a self-similarity property. Self-similarity property states that for any fixed positive λ, $\{\lambda^{-H} B^H_{\lambda t}\}_{t\in\mathbb{R}}$ is also an fBm with Hurst exponent $H \in (0,1)$. In other words, self-similarity property establishes a spatial proportionality relationship of factor λ^H,

between the distributions of $\{B^H_{\lambda t}\}$ and $\{B^H_t\}$. This leads to obtain spatial proportionality relationships of factor λ^{2H} between the variances, respectively the covariance kernels, of the processes $\{B^H_{\lambda t}\}$ and $\{B^H_t\}$. This type of scaling property will play a role in the simplification of some subsequent formulae, as it allows changes in time and space scales in the processes. The self-similarity property is, for example, useful for calculating the covariance function of the fBm because it allows this covariance to be expressed using only power functions of exponent $H \in (0,1)$ (self-similarity functions), and the covariance at time 1.

The choice of defining the above fBm B^H, by choosing a function f with only the first coordinate non-zero is arbitrary; this non-zero coordinate could be any of the other coordinates, and at the end, this would define d fBms. Because the coordinates of W are independent, the resulting fBms would be independent; for financial modeling purposes, this is too restrictive.
To allow correlations between fBms, we add correlations to the driving Brownian motion through a linear transformation of W. This works as follows: let $C = (\rho_{i,j})_{1\leq i,j\leq d}$ be a correlation matrix, and let R be a symmetric square root of C (which exists because C is non-negative symmetric), so that

$$\rho_{i,j} = R_{:,i} \cdot R_{:,j} \tag{3.25}$$

where $R_{:,i}$ is the i-th column of R. Now, we define a multivariate fractional Brownian motion, with correlated components, each having its own Hurst parameter.

Definition-Proposition 4. *Let $H_1, \cdots, H_d$ be a sequence of Hurst exponents in $(0,1)$ and set, for any $t \geq 0$ and any $i \in \{1, \cdots, d\}$,*

$$B^{i,H_i}_t := I(\psi^{H_i}_{0,t} R_{:,i}) = \int_{\mathbb{R}} \psi^{H_i}_{0,t}(u) R_{:,i} \cdot \mathrm{d}W_u, \tag{3.26}$$

where $R_{:,i}$ is the i-th column of R. Then, this defines a multivariate fractional Brownian motion (mfBm) with parameter $H = (H_i)_i \in (0,1)^d$ whose coordinates (B^{i,H_i}_t) are fBms with Hurst exponent H_i, and are correlated in a way that:

$$\forall s,t \in \mathbb{R}, \quad \mathrm{Cov}(B^{i,H_i}_t, B^{j,H_j}_s) + \mathrm{Cov}(B^{j,H_j}_t, B^{i,H_i}_s) \tag{3.27}$$

$$= \mathrm{Cov}(B^{i,H_i}_1, B^{j,H_j}_1)\Big(|t|^{H_i+H_j} + |s|^{H_i+H_j} - |t-s|^{H_i+H_j}\Big), \tag{3.28}$$

where $\mathrm{Cov}(B^{i,H_i}_1, B^{j,H_j}_1)$ is explicitly given in Lemma 3.1. Observe that

$$B^i_t := R_{:,i} \cdot W_t, \quad i \in \{1, \cdots, d\} \tag{3.29}$$

defines a d-dimensional Brownian motion with correlation C.

It is an open question whether $\mathrm{Cov}(B^{H_i}_1, B^{H_j}_1) = \rho_{i,j}$ in full generality; we can only justify the inequality $|\mathrm{Cov}(B^{H_i}_1, B^{H_j}_1)| \leq |\rho_{ij}|$ by using the Cauchy-Schwarz inequality in Equation (3.33). For $i \neq j$ and $H_i = H_j$, we have $\mathrm{Cov}(B^{H_i}_1, B^{H_j}_1) = \rho_{i,j}$.

In addition, we have not been able to derive a closed-form expression for the covariance $\mathrm{Cov}(B_t^{i,H_i}, B_s^{j,H_j})$ but only for a symmetrized version $\mathrm{Cov}(B_t^{i,H_i}, B_s^{j,H_j}) + \mathrm{Cov}(B_t^{j,H_j}, B_s^{i,H_i})$. When $H_i = H_j =: H$, by invoking symmetry arguments, we obtain a formula like (3.8), i.e.

$$\mathrm{Cov}\left(B_t^{i,H}, B_s^{j,H}\right) = \frac{\rho_{i,j}}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H}). \tag{3.30}$$

Proof The fact that each B^{i,H_i} is an fBm with Hurst parameter H_i directly stems from the observation that $R_{:,i} \cdot W_\cdot$ is again a standard Brownian motion; therefore, the results of Definition-Proposition 3 apply.

Only the correlations between these fBms must be specified. We follow the arguments in the proof of Definition-Proposition 3. Let $t > s > 0$. Using Itô's isometry combined with Equation (3.25), self-similarity property, and the variable change $u - s = v(t-s)$, we obtain:

$$\mathbb{E}\left[\left(B_t^{i,H_i} - B_s^{i,H_i}\right)\left(B_t^{j,H_j} - B_s^{j,H_j}\right)\right] \tag{3.31}$$

$$= \frac{\rho_{i,j}}{c_{H_i}c_{H_j}} \int_{\mathbb{R}} \left((t-u)_+^{H_i-\frac{1}{2}} - (s-u)_+^{H_i-\frac{1}{2}}\right)\left((t-u)_+^{H_j-\frac{1}{2}} - (s-u)_+^{H_j-\frac{1}{2}}\right) \mathrm{d}u \tag{3.32}$$

$$= |t-s|^{H_i+H_j} \frac{\rho_{i,j}}{c_{H_i}c_{H_j}} \int_{\mathbb{R}} \left((1-v)_+^{H_i-\frac{1}{2}} - (-v)_+^{H_i-\frac{1}{2}}\right)\left((1-v)_+^{H_j-\frac{1}{2}} - (-v)_+^{H_j-\frac{1}{2}}\right) \mathrm{d}v \tag{3.33}$$

$$= |t-s|^{H_i+H_j} \mathrm{Cov}(B_1^{i,H_i}, B_1^{j,H_j}), \tag{3.34}$$

which highlights the time-scaling property of the fBm. The explicit computation of the covariance between the fBms at time 1 is given in Lemma (3.1) below.

To obtain the covariance (3.28), write

$$\mathrm{Cov}(B_t^{i,H_i}, B_s^{j,H_j}) + \mathrm{Cov}(B_t^{j,H_j}, B_s^{i,H_i}) = \mathrm{Cov}(B_t^{i,H_i}, B_t^{j,H_j}) + \mathrm{Cov}(B_s^{i,H_i}, B_s^{j,H_j}) \tag{3.35}$$

$$- \mathbb{E}\left[\left(B_t^{i,H_i} - B_s^{i,H_i}\right)\left(B_t^{j,H_j} - B_s^{j,H_j}\right)\right] \tag{3.36}$$

and apply Equality (3.33) several times with $(t,s) = (t,0)$ and with $(t,s) = (0,s)$. We are done with the proof of Definition-Proposition (**??**).

Lemma 3.1 (Covariance at time 1)**.** *Let H_i, H_j be two Hurst exponents, as given above:*

$$\mathrm{Cov}(B_1^{i,H_i}, B_1^{j,H_j}) = \frac{\rho_{i,j}}{c_{H_i}c_{H_j}} \left[\frac{\left(H_i - \frac{1}{2}\right)\left(\frac{3}{2} - H_i\right)}{(H_i + H_j - 1)(H_i + H_j)} \mathfrak{B}(2 - H_i - H_j, H_j + \frac{1}{2}) \right. \tag{3.37}$$

$$\left. + \frac{\left(H_j - \frac{1}{2}\right)\left(\frac{3}{2} - H_j\right)}{(H_i + H_j - 1)(H_i + H_j)} \mathfrak{B}(2 - H_i - H_j, H_i + \frac{1}{2})\right] \tag{3.38}$$

where $\mathfrak{B}$ is the beta function, $\rho_{i,j}$ is the correlation coefficient defined in Equation (3.25), c_{H_i} and c_{H_j} are defined in Equation (3.7) as the coefficients for variance normalization. If $H_i = H_j$, then $\mathrm{Cov}(B_1^{i,H_i}, B_1^{j,H_j}) = \rho_{i,j}$.

Proof [Proof of Lemma 3.1] In view of Equation (3.33), we have $\mathrm{Cov}(B_1^{i,H_i}, B_1^{j,H_j}) = \frac{\rho_{i,j}}{c_{H_i} c_{H_j}} \varphi(1)$ where

$$\varphi(\zeta) := \int_{\mathbb{R}} \left((\zeta - v)_+^{H_i - \frac{1}{2}} - (-v)_+^{H_i - \frac{1}{2}} \right) \left((\zeta - v)_+^{H_j - \frac{1}{2}} - (-v)_+^{H_j - \frac{1}{2}} \right) \mathrm{d}v, \quad \zeta > 0. \tag{3.39}$$

Splitting the integral on negative and positive real numbers gives

$$\varphi(\zeta) = \int_0^{\zeta} (\zeta - v)^{H_i + H_j - 1} \mathrm{d}v + \int_{-\infty}^{0} \left((\zeta - v)_+^{H_i - \frac{1}{2}} - (-v)_+^{H_i - \frac{1}{2}} \right) \left((\zeta - v)_+^{H_j - \frac{1}{2}} - (-v)_+^{H_j - \frac{1}{2}} \right) \mathrm{d}v \tag{3.40}$$

$$= \frac{\zeta^{H_i + H_j}}{H_i + H_j} + \int_0^{+\infty} \left((\zeta + w)^{H_i - \frac{1}{2}} - w^{H_i - \frac{1}{2}} \right) \left((\zeta + w)^{H_j - \frac{1}{2}} - w^{H_j - \frac{1}{2}} \right) \mathrm{d}w \tag{3.41}$$

$$=: \frac{\zeta^{H_i + H_j}}{H_i + H_j} + \varphi_2(\zeta). \tag{3.42}$$

We can easily justify that one can differentiate under the integral sign, which gives

$$\varphi_2'(\zeta) = \int_0^{+\infty} \Bigg\{ (H_i + H_j - 1)(\zeta + w)^{H_i + H_j - 2} \tag{3.43}$$

$$- \left(H_i - \frac{1}{2} \right) w^{H_j - \frac{1}{2}} (\zeta + w)^{H_i - \frac{3}{2}} \tag{3.44}$$

$$- \left(H_j - \frac{1}{2} \right) w^{H_i - \frac{1}{2}} (\zeta + w)^{H_j - \frac{3}{2}} \Bigg\} \mathrm{d}w, \tag{3.45}$$

then

$$\varphi_2''(\zeta) = (H_i + H_j - 1)(H_i + H_j - 2) \int_0^{+\infty} (\zeta + w)^{H_i + H_j - 3} \mathrm{d}w \tag{3.46}$$

$$- \left(H_i - \frac{1}{2} \right) \left(H_i - \frac{3}{2} \right) \int_0^{+\infty} w^{H_j - \frac{1}{2}} (\zeta + w)^{H_i - \frac{5}{2}} \mathrm{d}w \tag{3.47}$$

$$- \left(H_j - \frac{1}{2} \right) \left(H_j - \frac{3}{2} \right) \int_0^{+\infty} w^{H_i - \frac{1}{2}} (\zeta + w)^{H_j - \frac{5}{2}} \mathrm{d}w. \tag{3.48}$$

The variable change $w = \zeta(\frac{1}{y} - 1)$ gives

$$\int_0^{+\infty} w^{H_j - \frac{1}{2}} (\zeta + w)^{H_i - \frac{5}{2}} \mathrm{d}w = \zeta^{H_i + H_j - 2} \int_0^1 y^{1 - H_i - H_j} (1 - y)^{H_j - \frac{1}{2}} \mathrm{d}y \tag{3.49}$$

$$= \zeta^{H_i + H_j - 2} B \left(2 - H_i - H_j, H_j + \frac{1}{2} \right). \tag{3.50}$$

We deduce that

$$\varphi_2''(\zeta) = -\zeta^{H_i+H_j-2}(H_i+H_j-1) - \left(H_i-\frac{1}{2}\right)\left(H_i-\frac{3}{2}\right)\zeta^{H_i+H_j-2}B\left(2-H_i-H_j, H_j+\frac{1}{2}\right) \tag{3.51}$$

$$-\left(H_j-\frac{1}{2}\right)\left(H_j-\frac{3}{2}\right)\zeta^{H_i+H_j-2}B\left(2-H_i-H_j, H_i+\frac{1}{2}\right). \tag{3.52}$$

Integrating twice gives

$$\varphi_2(\zeta) = \zeta^{H_i+H_j}\left(-\frac{1}{H_i+H_j} - \frac{\left(H_i-\frac{1}{2}\right)\left(H_i-\frac{3}{2}\right)}{(H_i+H_j-1)(H_i+H_j)}B\left(2-H_i-H_j, H_j+\frac{1}{2}\right)\right. \tag{3.53}$$

$$\left. -\frac{\left(H_j-\frac{1}{2}\right)\left(H_j-\frac{3}{2}\right)}{(H_i+H_j-1)(H_i+H_j)}B\left(2-H_i-H_j, H_i+\frac{1}{2}\right)\right) + \text{affine term in } \zeta. \tag{3.54}$$

Assume that for a while, $H_i + H_j \neq 1$. Then, the affine term must be zero because from its definition, it follows $\varphi_2(\zeta) = \varphi_2(1)\zeta^{H_i+H_j}$ (coming again from scaling arguments).Overall, with Equation (3.42), we obtain

$$\varphi(\zeta) = \zeta^{H_i+H_j}\left(-\frac{\left(H_i-\frac{1}{2}\right)\left(H_i-\frac{3}{2}\right)}{(H_i+H_j-1)(H_i+H_j)}B(2-H_i-H_j, H_j+\frac{1}{2}) \right. \tag{3.55}$$

$$\left. -\frac{\left(H_j-\frac{1}{2}\right)\left(H_j-\frac{3}{2}\right)}{(H_i+H_j-1)(H_i+H_j)}B(2-H_i-H_j, H_i+\frac{1}{2})\right). \tag{3.56}$$

Thus far, the formula is valid under the condition $H_i + H_j \neq 1$. If $H_i + H_j = 1$, it is sufficient to take a sequence of $(H_i^n, H_j^n)_n$ converging to (H_i, H_j) and such that $H_i^n + H_j^n \neq 1$: because the value $\varphi(\zeta)$ depends continuously on the H parameters (this is clear from expression (3.42)), the obtained formula (3.56) is valid for $H_i, H_j \in (0,1)$. We conclude the proof of the covariance formula in Lemma 3.1 by evaluating ϕ in $\zeta = 1$.

The last statement of the lemma for the case $H_i = H_j$ can be easily verified, from the above formula and recalling formula (3.7) for c_H. ■

■

Remark 3.2. *In (Lavancier et al., 2009), authors study multivariate fractional Brownian motion (mfBm) as an operator-self-similar Gaussian process with stationary increments. An operator self-similar process is defined for any $\lambda > 0$ as:*

$$\boldsymbol{X}(\lambda t) \overset{fdd}{=} \lambda^H \boldsymbol{X}(t) \tag{3.57}$$

where $\overset{fdd}{=}$ *means equality of finite-dimensional distributions, and the* $d \times d$ *matrix* λ^H *is defined by the power series* $\lambda^H = e^{H\log(\lambda)} = \sum_{k=0}^{\infty} H^k \frac{\log(\lambda)^k}{k!}$. *The properties of stationarity and self-similarity are expressed in terms of distributions. Authors show that the multivariate fractional Brownian motion (mfBm) is a special case of operator-self-similar Gaussian process with stationary increments, where the matrix of exponents is diagonal. Moreover, they provide a characterization of the covariance kernel thanks to power functions.*

3.2.2 Conditional expectation and projection

In what follows, we typically need to compute the conditional expectation and variance of some stochastic integral $I(f)$ conditionally to the increments of fBm on some time grid. The following Lemma gives a more general result about how to compute $\mathbb{E}\left[I(f) \mid \{I(f_l)\}_{(l\in\mathcal{I})}\right]$ and $\mathbb{V}\left[I(f) \mid \{I(f_l)\}_{(l\in\mathcal{I})}\right]$, where $(f_l)_{l\in\mathcal{I}}$ is an arbitrary family of functions in $\mathcal{L}^2_d$.

Lemma 3.3 (Orthogonal projections and conditional expectations on Gaussian Hilbert spaces)**.** *Let* $f \in \mathcal{L}^2_d$ *and* $(f_l)_{l\in\mathcal{I}}$ *be an arbitrary family of functions in* $\mathcal{L}^2_d$. *Let* $\mathcal{L}'$ *be the closed subspace of* $\mathcal{L}^2_d$ *spanned by* $(f_l)_{l\in\mathcal{I}}$ *and define* $f^\star \in \mathcal{L}'$ *as the (unique) orthogonal projection of* f *on* $\mathcal{L}'$. *Then*

$$\mathbb{E}\left[I(f) \mid \{I(f_l)\}_{(l\in\mathcal{I})}\right] = I(f^\star), \tag{3.58}$$

$$\mathbb{V}\left[I(f) \mid \{I(f_l)\}_{(l\in\mathcal{I})}\right] = \left|f - f^\star\right|^2_{\mathcal{L}^2_d}. \tag{3.59}$$

Proof The formula for the conditional expectation follows directly from (Janson, 1997, Chapter 9, Thm.9.1). In addition, denote by $\mathcal{H}'$ the closed subspace of $\mathcal{H}$ generated by $(I(f_l))_{(l\in\mathcal{I})}$, and write the orthogonal decomposition of $\mathcal{H}'$ as:

$$\mathcal{H} = \mathcal{H}' \oplus \mathcal{H}''. \tag{3.60}$$

Obviously, the projection residual $I(f) - I(f^\star)$ is in $\mathcal{H}''$, meaning that $I(f) - I(f^\star)$ is independent of $\{I(f_l)\}_{l\in\mathcal{I}}$:

$$\mathbb{V}\left[I(f) \mid \{I(f_l)\}_{l\in\mathcal{I}}\right] = \mathbb{V}\left[I(f) - I(f^\star)\right]. \tag{3.61}$$

We conclude using the isometry property. ■

In the following, the conditioning is made according to the asset values observed along a time grid, that is, according to $(S^i_{t_l}), i = 1, \cdots, d, l = 1, \cdots N$, or equivalently according to $(B^{H_i}_{t_{l+1}} - B^{H_i}_{t_l})$ which corresponds to choosing $f^i_l(\cdot) = \sigma_i R_{:,i} \psi^{H_i}_{t_l,t_{l+1}}(\cdot)$. In addition, the function f to consider is related to the P&L of the portfolio increments. As such, we prove (see Equation (**??**)) that $f(\cdot) = \sum_{i=1}^d \omega_i \sigma_i S^i_{t_N} R_{:,i} \psi^{H_i}_{t_N,t_{N+1}}(\cdot)$. In the next statement, we take a slightly more general form consistent with this future choice.

Theorem 3.4. *Consider the index family* $\mathcal{I} = \{(i,l) \mid i \in [\![1,d]\!],\ l \in [\![1,N]\!]\}$ *and the set of* $\mathcal{L}_d^2$*-valued functions*

$$f_l^i(\cdot) := \sigma_i R_{:,i} \psi_{t_l,t_{l+1}}^{H_i}(\cdot). \tag{3.62}$$

Let

$$f(\cdot) = \sum_{i=1}^{d} \omega_{t_N}^i \psi_{t_N,t_{N+1}}^{H_i}(\cdot) \tag{3.63}$$

be an $\mathcal{L}_d^2$ *function, where each* d*-dimensional coefficient* $\omega_{t_N}^i = \omega_i \sigma_i S_{t_N}^i R_{:,i}$ *is measurable with respect to* $\{I(f_l^i)\}_{(i,l)\in\mathcal{I}}$*.*

Then, the expectation of $I(f)$ *conditional on* $\{I(f_l^i)\}_{(i,l)\in\mathcal{I}}$ *can be written as*

$$\mathbb{E}\left[I(f)|\{I(f_l^i)\}_{(i,l)\in\mathcal{I}}\right] = \sum_{(i,l)\in\mathcal{I}} a_l^i I(f_l^i) = \sum_{i=1}^{d}\sum_{l=1}^{N} a_l^i (B_{t_{l+1}}^{H_i} - B_{t_l}^{H_i}), \tag{3.64}$$

and the conditional variance is given by:

$$\mathbb{V}\left[I(f) \mid \{I(f_l^i)\}_{(i,l)\in\mathcal{I}}\right] = \sum_{i,j=1}^{d} (\omega_{t_N}^i \cdot \omega_{t_N}^j) \int_{\mathbb{R}} \psi_{t_N,t_{N+1}}^{H_i}(u)\psi_{t_N,t_{N+1}}^{H_j}(u)\mathrm{d}u \tag{3.65}$$

$$- \sum_{i,j=1}^{N}\sum_{k,l=1}^{N} a_l^i a_k^j \rho_{ij} \int_{\mathbb{R}} \psi_{t_l,t_{l+1}}^{H_i}(u)\psi_{t_k,t_{k+1}}^{H_j}(u)\mathrm{d}u, \tag{3.66}$$

where the projection coefficients $\alpha^\star = (a_1^1,\cdots,a_N^1,a_1^2,\cdots,a_N^2,\cdots,a_1^d,\cdots,a_N^d)^\top$ *are equal to:*

$$\alpha^\star = \mathfrak{M}^+ y, \tag{3.67}$$

where $\mathfrak{M}^+$ *is the Moore-Penrose pseudo-inverse (see (MacAusland, 2014, Section 5.5.4)) of the matrix*

$$\mathfrak{M} = \left(\langle f_{k_1}^{i_1}, f_{k_2}^{i_2}\rangle_{\mathcal{L}_d^2}\right)_{(i_1-1)\times N+k_1,(i_2-1)\times N+k_2}$$

and where

$$y = \left(\langle f, f_{k_2}^{i_2}\rangle_{\mathcal{L}_d^2}\right)_{(i_2-1)\times N+k_2}.$$

Among the projection coefficients involved in the conditional expectation, the coefficient $\alpha^\star$ *in* (3.67) *is the one with the minimal norm.*

2) Assume that the time-discretization is uniform, that is, $t_l = lh$ *for* $l = 1,\ldots,N$*. The self-similarity property allows the covariance kernel to be factorized by a power of the time step* h*, and the remaining amount in the factorization represents the covariance kernel at time* 1 *(independent of* h*) between the two given fBms:*

$$\mathfrak{M}_{(i_1-1)*N+k_1,(i_2-1)*N+k_2} = \rho_{i_1,i_2} h^{H_{i_1}+H_{i_2}} \int_{\mathbb{R}} \psi_{0,1}^{H_{i_1}}\left(v-k_1\right)\psi_{0,1}^{H_{i_2}}\left(v-k_2\right)\mathrm{d}v, \tag{3.68}$$

$$y_{(i_2-1)N+k_2} = \sum_{i=1}^{d} (\omega_{t_N}^i \cdot R_{:,i_2}) h^{H_i+H_{i_2}} \int_{\mathbb{R}} \psi_{0,1}^{H_i}\left(v-N\right)\psi_{0,1}^{H_{i_2}}\left(v-k_2\right)\mathrm{d}v. \tag{3.69}$$

The simplification in the case of constant time-step (item 2)) has mainly a computational interest: the formulae can be directly adapted if we change the time-step, without recomputing all the integrals necessary to the conditional expectation and variance.

Proof 1) In view of Lemma 3.3, the goal is to minimize the following Ordinary Least Squares (OLS) problem:

$$\min_{g\in \mathrm{Span}(f_1,\cdots,f_n)} \left|f-g\right|_{\mathcal{L}^2_d} = \min_{\alpha\in\mathbb{R}^n} \left|f-\sum_{k=1}^n \alpha_k f_k\right|_{\mathcal{L}^2_d} \tag{3.70}$$

where $n = d\times N$, $f_{(i-1)N+l} := f^i_l$ and $\alpha_{(i-1)N+l} := a^i_l$ for $i = 1,\cdots,d$, $l = 1,\cdots,N$.

If the norm is the usual Euclidean norm in $\mathbb{R}^n$, it is well-known that the set of solutions is an affine space, and the minimal norm element is given by the Moore-Penrose pseudo-inverse of some matrix applied to some vector, see (MacAusland, 2014, Section 5), (Barata and Hussein, 2012), (Ben-Israel and Greville, 2006) for details. In our setting of the non-Euclidean norm, the result is similar, but we have not been able to find a ready-to-cite reference. Thus we provide main arguments below. Consider the covariance matrix $\mathfrak{M}$ (symmetric and non-negative definite), together with its diagonalized representation:

$$\mathfrak{M} = (\langle f_k, f_l\rangle_{\mathcal{L}^2_d})_{k,l} = U\begin{pmatrix} D_L & 0\\ 0 & 0\end{pmatrix}U^\top \tag{3.71}$$

where U is an orthogonal matrix and, D_L is a diagonal matrix (of size $L\leq n$) with positive elements. Define n new functions in $\mathcal{L}^2_d$ by the formula $g_{1:n} := (g_1,\cdots,g_n)^\top = U^\top(f_1,\cdots,f_n)^\top = U^\top f_{1:n}$: it is an easy exercise to check that

$$(\langle g_k, g_l\rangle_{\mathcal{L}^2_d})_{k,l} = U^T\mathfrak{M}U = \begin{pmatrix} D_L & 0\\ 0 & 0\end{pmatrix}. \tag{3.72}$$

This implies that $g_k = 0$ if $k > L$. Set $\beta := U^\top\alpha\in\mathbb{R}^n$: then, leveraging the above properties, we get

$$\left|f-\sum_{k=1}^n \alpha_k f_k\right|_{\mathcal{L}^2_d} = \left|f-\beta^\top g_{1:n}\right|_{\mathcal{L}^2_d} = \left|f-\sum_{k=1}^L \beta_k g_k\right|_{\mathcal{L}^2_d}. \tag{3.73}$$

Hence, we can arbitrarily choose β_k for $k > L$, without changing the minimum value of the above quantity. Since the $(g_k)_{k=1,\cdots,L}$ are orthogonal, the (unique) minimizer $\beta_{1:L}$ is equal to

$$\arg\min_{\beta_{1:L}} \left|f-\sum_{k=1}^L \beta_k g_k\right|_{\mathcal{L}^2_d} = D_L^{-1}\begin{bmatrix}\langle f, g_1\rangle_{\mathcal{L}^2_d}\\ \vdots \\ \langle f, g_L\rangle_{\mathcal{L}^2_d}\end{bmatrix}. \tag{3.74}$$

Going back to α, the optimal coefficients are given by

$$\alpha^\star = U \begin{pmatrix} D_L^{-1} \begin{bmatrix} \langle f, g_1 \rangle_{\mathcal{L}_d^2} \\ \vdots \\ \langle f, g_L \rangle_{\mathcal{L}_d^2} \end{bmatrix} \\ \beta_{L+1} \\ \vdots \\ \beta_N \end{pmatrix}. \tag{3.75}$$

Because the multiplication by an orthogonal matrix U does not modify the norm, the solution with the minimal norm is obtained through the choice $\beta_{L+1:N} = 0$. The obtained solution can be rewritten using the Moore-Penrose pseudo-inverse of $\mathfrak{M}$:

$$\alpha^\star = U \begin{pmatrix} D_L^{-1} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \langle f, g_1 \rangle_{\mathcal{L}_d^2} \\ \vdots \\ \langle f, g_L \rangle_{\mathcal{L}_d^2} \\ \langle f, g_{L+1} \rangle_{\mathcal{L}_d^2} \\ \vdots \\ \langle f, g_n \rangle_{\mathcal{L}_d^2} \end{pmatrix} = U \begin{pmatrix} D_L^{-1} & 0 \\ 0 & 0 \end{pmatrix} U^\top \langle f, f_{1:N} \rangle_{\mathcal{L}_d^2} = \mathfrak{M}^+ \langle f, f_{1:N} \rangle_{\mathcal{L}_d^2}. \tag{3.76}$$

We are done with the proof of (3.64) and (3.67).

To prove Equation (3.65), we use the properties of Lemma 3.3 about orthogonal projections, which gives:

$$\mathbb{V}\left[I(f) \mid \{I(f_l^i)\}_{(i,l)\in\mathcal{I}} \right] = \left| f - f^\star \right|_{\mathcal{L}_d^2}^2 = \left| f \right|_{\mathcal{L}_d^2}^2 - \left| f^\star \right|_{\mathcal{L}_d^2}^2 \tag{3.77}$$

$$= \sum_{i,j=1}^{d} (\omega_{t_N}^i \cdot \omega_{t_N}^j) \int_{\mathbb{R}} \psi_{t_N,t_{N+1}}^{H_i}(u) \psi_{t_N,t_{N+1}}^{H_j}(u) \mathrm{d}u - \sum_{i,j=1}^{N} \sum_{k,l=1}^{N} a_l^i a_k^j \rho_{ij} \int_{\mathbb{R}} \psi_{t_l,t_{l+1}}^{H_i}(u) \psi_{t_k,t_{k+1}}^{H_j}(u) \mathrm{d}u. \tag{3.78}$$

2) Under the assumption of uniform time-partition, the self-similarity of the path-dependent kernels $\psi_{s,t}^H(.)$ implies that they can be rewritten as a product between the power of h and the kernel between 0 and 1 evaluated at the scaled and translated variable $\frac{u}{h} - k$:

$$\psi_{t_k,t_{k+1}}^H(u) = h^{H-\frac{1}{2}} \psi_{0,1}^H \left(\frac{u}{h} - k \right). \tag{3.79}$$

Integrating the above kernel with the change of variable $v = \frac{u}{h}$, we obtain

$$\int_{\mathbb{R}} \psi_{t_k,t_{k+1}}^{H_i}(u) \psi_{t_l,t_{l+1}}^{H_j}(u) \mathrm{d}u = h^{H_i+H_j} \int_{\mathbb{R}} \psi_{0,1}^{H_i}(v-k) \psi_{0,1}^{H_j}(v-l) \mathrm{d}v. \tag{3.80}$$

Using the above relation, we obtain Equations (3.68) and (3.69). ■

3.2.3 Market model

In this section, we present the model used to describe the price dynamics. Usually, the latter are described using a standard Black-Scholes model based on standard Brownian motion. The approach presented in this work is quite different, because the price dynamics are described by a fractional Black-Scholes model. This choice of model is justified by the fact that the log-price increments remain Gaussian self-similar and stationary. Therefore, interesting and tractable properties can be exploited, on the one hand to model the trajectory of the log-price increments with accuracy and realism by introducing correlations between them which express long-range (respectively short-range) dependence; on the other hand, to insure stability and coherence of the estimators and to perform theoretical computations. Moreover, the assets that we describe, such as FX rates, present significant correlations between the log-price increments, which can be taken into account by the fractal properties of fBm. Such an approach has already been adopted in the works of (Garcin, 2020) (in dimension $d = 1$), in which fBm is used to take into account positive correlations between returns. Here, we propose a multivariate extension – the so-called multivariate fractional Black-Scholes model – to model the price dynamics of a universe composed of several assets correlated in both time and space.

Assume that we have a universe composed of d assets, whose market prices at time t are denoted by $\left(S_t^1, \ldots, S_t^d\right)$ and whose market log-prices at time t are denoted by $\left(X_t^1, \ldots, X_t^d\right)$. The price dynamics are given by the fractional Geometric Brownian motion formula:

$$\forall i \in [\![1, d]\!],\ S_t^i = S_0^i e^{c_t^i + \sigma_i B_t^{i,H_i}} \quad \text{with} \quad c_t^i = \log\left(\frac{\mathbb{E}[S_t^i]}{S_0^i}\right) - \frac{\sigma_i^2}{2} t^{2H_i}, \tag{3.81}$$

where S_0 is fixed and known, and c_t^i is a centering parameter in the model. We assume that all parameters are known; see Section 3.5 about their estimation from market data.

Let us focus on the properties of the log-price process:

$$X_t^i = X_0^i + c_t^i + \sigma_i B_t^{i,H_i}. \tag{3.82}$$

It is affine with respect to fractional Brownian motion, and as such, it benefits from its properties related to Gaussianity, self-similarity and stationarity of increments, while relaxing the property of independence of increments. For any $h > 0$, the log-price increments of length h are defined by (see Equation (3.26)):

$$\delta_h X_t^i := X_{t+h}^i - X_t^i = \delta_h c_t^i + \sigma_i I(\psi_{t,t+h}^{H_i} R_{:,i}). \tag{3.83}$$

In this framework, we are interested in determining risk measures of the future portfolio price increment at time horizon h. The portfolio and its increments of size h are defined by:

$$P_t = \sum_{i=1}^{d} \omega^i S_t^i \qquad \text{and} \qquad \delta_h P_t := P_{t+h} - P_t. \tag{3.84}$$

3.3 Value-at-Risk approximation

3.3.1 Portfolio expansion and Gaussian approximation of the VaR

The objective is to compute the conditional $\mathbf{VaR}_\alpha$ of the future portfolio's increment $\delta_h P_{t_N} = P_{t_{N+1}} - P_{t_N}$ given the past observations $\{S^1_{t_0}, S^1_{t_1}, \dots, S^1_{t_N}, \dots, S^d_{t_0}, S^d_{t_1}, \dots, S^d_{t_N}\}$ taken on a uniform partition of the time interval $[0, t]$, such that the time-step is constant and equal to h, $t_0 = 0$ and $t_N = t$. Because all parameters are known, observing the prices or the log-price or the log-price increments yields the same information, modeled by sigma-algebra

$$\mathcal{G} := \sigma(\delta_h X^i_{t_l} : (i,l) \in \mathcal{I}) = \sigma(I(f^i_l) : (i,l) \in \mathcal{I}) \qquad \text{where} \qquad f^i_l(\cdot) := \sigma_i R_{:,i} \psi^{H_i}_{t_l, t_{l+1}}(\cdot) \tag{3.85}$$

and $\mathcal{I} = \{(i,l) \mid i \in [\![1,d]\!],\ l \in [\![1,N]\!]\}$ is an index family indexing assets and time increments.

For the purpose of $\mathbf{VaR}_\alpha$ of $\delta_h P_{t_N}$, we need to know the conditional distribution of the future portfolio's increment. However, $\delta_h P_{t_N}$ is a linear combination of log-normal random variables, whose marginal and conditional distributions are not known. Therefore, we propose a Gaussian approximation of the conditional $\mathbf{VaR}_\alpha$ of $\delta_h P_{t_N}$ with the conditional $\mathbf{VaR}_\alpha$ of $\delta_h \hat{P}_t$ whose conditional distribution is Gaussian:

$$\delta_h P_{t_N} = P_{t_N+h} - P_{t_N} = \sum_{i=1}^d \omega^i \delta_h S^i_{t_N} \simeq \sum_{i=1}^d \omega^i S^i_{t_N} \delta_h X^i_{t_N} =: \delta_h \hat{P}_{t_N}. \tag{3.86}$$

Let $\alpha \in (0,1)$ be the risk level, and recall, as an immediate extension of the unconditional case, that the $\mathbf{VaR}_\alpha$ of a scalar random variable X conditionally to a sigma-field $\mathcal{G}$, is defined by

$$\mathbf{VaR}_\alpha(X \mid \mathcal{G}) := \inf\{x \in \mathbb{R} : \mathbb{P}\Big(X \le x \mid \mathcal{G}\Big) \ge \alpha\}. \tag{3.87}$$

In our study, X represents the P&L of $\delta_h P_{t_N}$, and we focus on the specific case of $\mathbf{VaR}_\alpha$ of the Gaussian approximation $\delta_h \hat{P}_{t_N}$. The Gaussian **VaR** is easily determined because it relies only on the expectation and standard deviation of the Gaussian distribution.

Proposition 3.5 (Conditionally Gaussian Value-at-Risk (VaR))**.** *Let $\alpha \in (0,1)$. The $\mathbf{VaR}_\alpha$ of a $\mathcal{G}$-conditional Gaussian distribution $\mathcal{N}(\mu(\mathcal{G}), \sigma^2(\mathcal{G}))$ with $\sigma(\mathcal{G}) > 0$, is defined by:*

$$\mathbf{VaR}_\alpha(X \mid \mathcal{G}) = \sigma(\mathcal{G})\mathcal{N}^{-1}(\alpha) + \mu(\mathcal{G}). \tag{3.88}$$

In addition, conditionally to $\mathcal{G}$, $\delta_h \hat{P}_{t_N} = I(f)$ where the form of the function f is defined by Equation (3.63) of Theorem 3.4, with the specific weighting $\omega^i_{t_N} = \omega_i \sigma_i S^i_{t_N} R_{:,i}$:

$$f(\cdot) = \sum_{i=1}^d \omega_i \sigma_i S^i_{t_N} R_{:,i} \psi^{H_i}_{t_N, t_{N+1}}(\cdot). \tag{3.89}$$

follows a Gaussian distribution. More specifically, conditionally to $\mathcal{G}$, $\delta_h \hat{P}_{t_N}$ has a Gaussian distribution (see Lemma 3.3), characterized by its conditional mean and conditional variance.

Thus, we obtain the following theorem.

Theorem 3.6 (Conditionally Gaussian Value-at-Risk)**.** *If $\delta_h \hat{P}_{t_N} = I(f)$, with f given by Equation (3.89), is the future increment of the log-price portfolio, and $(\delta_h X^i_{t_l} = I(f^i_l))_{(i,l)\in\mathcal{I}}$, with f^i_l given by Equation (3.62) of Theorem 3.4, are the past log-price increments of the assets that form the universe, then the conditional Gaussian $\mathbf{VaR}_\alpha$ is given by the following formula:*

$$\mathbf{VaR}_\alpha(\delta_h \hat{P}_{t_N}|\mathcal{G}) = \sqrt{\mathbb{V}[\delta_h \hat{P}_{t_N}|\mathcal{G}]}\mathcal{N}^{-1}(\alpha) + \mathbb{E}[\delta_h \hat{P}_{t_N}|\mathcal{G}] \tag{3.90}$$

where $\mathbb{E}[\delta_h \hat{P}_{t_N}|\mathcal{G}]$ and $\mathbb{V}[\delta_h \hat{P}_{t_N} - \mathbb{E}[\delta_h \hat{P}_{t_N}|\mathcal{G}]]$ are given by Theorem 3.4.

The conditional Expected-Shortfall of the Gaussian approximation of the future portfolio increment is provided by the following theorem.

Theorem 3.7 (Conditionally Gaussian Expected-Shortfall)**.** *If $\delta_h \hat{P}_{t_N} = I(f)$, with f given by Equation (3.89), is the future increment of the log-price portfolio, and $(\delta_h X^i_{t_l} = I(f^i_l))_{(i,l)\in\mathcal{I}}$, with f^i_l given by Equation (3.62) of Theorem 3.4, are the past log-price increments of the assets that form the universe, the conditional Gaussian $\mathbf{ES}_\alpha$ is given by the following formula:*

$$\mathbf{ES}_\alpha(\delta_h \hat{P}_{t_N} \mid \mathcal{G}) = \frac{\sqrt{\mathbb{V}[\delta_h \hat{P}_{t_N}|\mathcal{G}]}}{1-\alpha} n(\mathcal{N}^{-1}(\alpha)) + \mathbb{E}[\delta_h \hat{P}_{t_N}|\mathcal{G}] \tag{3.91}$$

where $n(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ and $N(x) = \int_{-\infty}^{x} n(y)\mathrm{d}y$ are respectively the standard Gaussian probability density function and cumulative distribution function.

Proof Recall that the conditional $\mathbf{VaR}_\alpha$ of $\delta_h \hat{P}_{t_N}$ is given by:

$$\mathbf{VaR}_\alpha(\delta_h \hat{P}_{t_N} \mid \mathcal{G}) = \sigma(\mathcal{G})\mathcal{N}^{-1}(\alpha) + \mu(\mathcal{G}). \tag{3.92}$$

where $\mu(\mathcal{G}) = \mathbb{E}[\delta_h \hat{P}_{t_N}|\mathcal{G}]$ and $\sigma(\mathcal{G}) = \sqrt{\mathbb{V}[\delta_h \hat{P}_{t_N}|\mathcal{G}]}$. The conditional $\mathbf{ES}_\alpha$ is given by:

$$\begin{aligned}
\mathbf{ES}_\alpha(\delta_h \hat{P}_{t_N} \mid \mathcal{G}) &= \frac{1}{1-\alpha}\int_\alpha^1 \mathbf{VaR}_a(\delta_h \hat{P}_{t_N} \mid \mathcal{G})\mathrm{d}a &(3.93)\\
&= \frac{\sigma(\mathcal{G})}{1-\alpha}\int_\alpha^1 \mathcal{N}^{-1}(a)\mathrm{d}a + \mu(\mathcal{G}). &(3.94)
\end{aligned}$$

Let us proceed to the following substitution: $x = \mathcal{N}^{-1}(a)$ then, $a = \mathcal{N}(x)$ $\frac{\mathrm{d}a}{\mathrm{d}x} = n(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. Moreover, when $a = \alpha$ then $x = \mathcal{N}^{-1}(\alpha)$ and when $a = 1$ then $x \to \mathcal{N}^{-1}(1) = +\infty$. Thus, it comes that:

$$\mathbf{ES}_\alpha(\delta_h \hat{P}_{t_N} \mid \mathcal{G}) = \frac{1}{1-\alpha}\int_{\mathcal{N}^{-1}(\alpha)}^{+\infty} x n(x)\mathrm{d}x. \tag{3.95}$$

But $xn(x) = -\frac{\mathrm{d}n(x)}{\mathrm{d}x}$, then:

$$\mathbf{ES}_\alpha(\delta_h \hat{P}_{t_N} \mid \mathcal{G}) = \frac{\sigma(\mathcal{G})}{1-\alpha}\Big[-n(x)\Big]_{\mathcal{N}^{-1}(\alpha)}^{+\infty} + \mu(\mathcal{G}) = \frac{\sigma(\mathcal{G})}{1-\alpha} n(\mathcal{N}^{-1}(\alpha)) + \mu(\mathcal{G}). \quad (3.96)$$

The Expected-Shortfall at risk level α can also be characterized as the solution to a minimization problem, as explained in the works of (Bardou et al., 2016) and (Rockafellar et al., 2000):

$$\mathbf{ES}_\alpha(X \mid \mathcal{G}) = \min_{\xi \in \mathbb{L}^0(\mathcal{G})} \left\{ \xi + \frac{1}{1-\alpha}\mathbb{E}\Big[(X-\xi)_+ \mid \mathcal{G}\Big] \right\} \quad (3.97)$$

$$= \mathbf{VaR}_\alpha(X \mid \mathcal{G}) + \frac{1}{1-\alpha}\mathbb{E}\Big[(X - \mathbf{VaR}_\alpha(X \mid \mathcal{G}))_+ \mid \mathcal{G}\Big] \quad (3.98)$$

where $\mathbb{L}^0(\mathcal{G})$ is the set of a.s. finite and $\mathcal{G}$-measurable random variables. ■

3.3.2 Results on VaR and ES error bounds

In this section, the goal is to quantify the accuracy in the approximation of $\mathbf{VaR}_\alpha(\delta_h P_{t_N} \mid \mathcal{G})$ with $\mathbf{VaR}_\alpha(\delta_h \hat{P}_{t_N} \mid \mathcal{G})$. In other words, we want to determine an upper bound of the following amount:

$$\Big|\mathbf{VaR}_\alpha(\delta_h P_{t_N} \mid \mathcal{G}) - \mathbf{VaR}_\alpha(\delta_h \hat{P}_{t_N} \mid \mathcal{G})\Big|.$$

Our main result is the following:

Theorem 3.8. *Consider the future portfolio increment $\delta_h P_{t_N}$ and its Gaussian approximation conditionally to $\mathcal{G}$, given by the future portfolio log-return $\delta_h \hat{P}_{t_N}$. Thus, we have the following upper bound:*

$$\Delta = \Big\|\delta_h P_{t_N} - \delta_h \hat{P}_{t_N}\Big\|_{p,\mathcal{G}} \leq 2\sum_{i=1}^{d} \Big|\omega^i\Big| S_{t_N}^i \left(\sigma_i^2 h^{2H_i} C_{2pq}^{1/pq} + M_i^2\right) e^{(M_i)_+ + \frac{pr}{2}\sigma_i^2 h^{2H_i}} \quad (3.99)$$

with $p > 1$ and $\frac{1}{q} + \frac{1}{r} = 1$.

If we assume that Δ is small enough, that is, $|\Delta| < 1-\alpha$ a.s., then:

(i) *The quantification of the approximation of the conditional $\mathbf{VaR}_\alpha$ of the future portfolio price increment $\delta_h P_{t_N}$ with the conditional Gaussian $\mathbf{VaR}_\alpha$ of the future portfolio log-price increment $\delta_h \hat{P}_{t_N}$ is given as follows:*

$$\Big|\mathbf{VaR}_\alpha(\delta_h P_{t_N} \mid \mathcal{G}) - \mathbf{VaR}_\alpha(\delta_h \hat{P}_{t_N} \mid \mathcal{G})\Big| \leq \left(\frac{2\sigma(\mathcal{G})}{1-\alpha-\Delta}\sum_{i=1}^{d}\Big|\omega^i\Big| S_{t_N}^i \right. \quad (3.100)$$

$$\left. \times \left(\sigma_i^2 h^{2H_i} C_{2pq}^{1/pq} + M_i^2\right) e^{(M_i)_+ + \frac{pr}{2}\sigma_i^2 h^{2H_i}} \right), a.s. \quad (3.101)$$

with $M_i := \mathbb{E}\left[\delta_h X^i_{t_N} \mid \mathcal{G}\right]$, $\sigma(\mathcal{G}) = \sqrt{\mathbb{V}[\delta_h \hat{P}_{t_N}|\mathcal{G}]}$, $G \sim \mathcal{N}(0,1)$, *and* $C_{2pq} := \mathbb{E}\left[G^{2pq}\right]$.

(ii) *The quantification of the approximation of the conditional* $\mathbf{ES}_\alpha$ *of the future portfolio price increment* $\delta_h P_{t_N}$ *with the conditional Gaussian* $\mathbf{ES}_\alpha$ *of the future portfolio log-price increment* $\delta_h \hat{P}_{t_N}$ *is given as follows:*

$$\left|\mathbf{ES}_\alpha(\delta_h P_{t_N} \mid \mathcal{G}) - \mathbf{ES}_\alpha(\delta_h \hat{P}_{t_N} \mid \mathcal{G})\right| \tag{3.102}$$

$$\leq \left[\left(1 + \frac{1}{1-\alpha}\right)\frac{2\sigma(\mathcal{G})}{1-\alpha-\Delta} + \frac{2}{1-\alpha}\right]\sum_{i=1}^{d}\left|\omega^i\right| S^i_{t_N}\left(\sigma_i^2 h^{2H_i} C_{2pq}^{\frac{1}{pq}} + M_i^2\right) e^{(M_i)_+ + \frac{pr}{2}\sigma_i^2 h^{2H_i}}. \tag{3.103}$$

3.3.3 Proof of Theorem 3.8

The analysis was conducted in three steps. The first step corresponds to the approximation of the price increment of an asset with its log-price increment, which is a Gaussian r.v.. The second step consists of approximating the portfolio price increments, with the portfolio log-price increments, which is a linear combination of Gaussian r.v.. The third step relates to the error bound between the two conditional $\mathbf{VaR}_\alpha$.

Lemma 3.9. *The price increment* $\delta_h S^i_{t_N}$ *and log-price increment* $\delta_h X^i_{t_N}$ *are as follows:*

$$\delta_h S^i_{t_N} - S^i_{t_N}\delta_h X^i_{t_N} = S^i_{t_N}(\delta_h X^i_{t_N})^2 \int_0^1 (1-v)e^{v\delta_h X^i_{t_N}}\mathrm{d}v. \tag{3.104}$$

This indicates that they are close to each other when the price increments are small.

Proof This is a consequence of the Taylor-Young expansion with integral remainder. Let $f : [a,b] \mapsto \mathbb{R}$ be a function of class $\mathcal{C}^2$. We have:

$$f(b) = f(a) + (b-a)f'(a) + (b-a)^2\int_0^1 (1-v)f''(a+v(b-a))\mathrm{d}v. \tag{3.105}$$

The above formula applied with $f(x) = e^x$, $b = X^i_{t_N+h} - X^i_{t_N} = \delta_h X^i_{t_N}$, $a = X^i_{t_N} - X^i_{t_N} = 0$ gives the announced formula. ∎

Proposition 3.10. *Let* $p \geq 1$. *The conditional* $\mathcal{L}_p$*-norm, defined by* $\left\|X\right\|_{p,\mathcal{G}} := \mathbb{E}\left[\left|X\right|^p \mid \mathcal{G}\right]^{\frac{1}{p}}$, *between* $\delta_h S^i_{t_N}$ *and* $S^i_{t_N}\delta_h X^i_{t_N}$ *is bounded as follows:*

$$\left\|\delta_h S^i_{t_N} - S^i_{t_N}\delta_h X^i_{t_N}\right\|_{p,\mathcal{G}} \leq 2S^i_{t_N}(\sigma_i^2 h^{2H_i} C_{2pq}^{1/pq} + M_i^2)e^{(M_i)_+ + \frac{pr}{2}\sigma_i^2 h^{2H_i}} \tag{3.106}$$

with $M_i := \mathbb{E}\left[\delta_h X^i_{t_N} \mid \mathcal{G}\right]$, $G \sim \mathcal{N}(0,1)$, *and* $C_{2pq} := \mathbb{E}\left[G^{2pq}\right]$.

Proof To alleviate the notation, we set $t = t_N$. From Lemma 3.9, and using Holder's inequality, we obtain:

$$\left\|\delta_h S_t^i - S_t^i \delta_h X_t^i\right\|_{p,\mathcal{G}} \leq \left\|S_t^i (\delta_h X_t^i)^2\right\|_{pq,\mathcal{G}} \left\|\int_0^1 (1-v) e^{v\delta_h X_t^i} \mathrm{d}v\right\|_{pr,\mathcal{G}} \text{ with } \frac{1}{q} + \frac{1}{r} = 1. \tag{3.107}$$

Using the Minkowski inequality and because S_t^i is $\mathcal{G}$-measurable and positive, we obtain the following formula:

$$\left\|\delta_h S_t^i - S_t^i \delta_h X_t^i\right\|_{p,\mathcal{G}} \leq S_t^i \left\|\delta_h X_t^i\right\|_{2pq,\mathcal{G}}^2 \sup_{v\in[0,1]} \left\|e^{v\delta_h X_t^i}\right\|_{pr,\mathcal{G}}. \tag{3.108}$$

At this stage, we have to upper bound the two following amounts $\left\|\delta_h X_t^i\right\|_{2pq,\mathcal{G}}$ and $\left\|e^{v\delta_h X_t^i}\right\|_{pr,\mathcal{G}}$. Note that $\delta_h X_t^i$ conditionally to $\mathcal{G}$ is distributed as $\mathcal{N}(M_i, v_i^2)$ where $M_i := \mathbb{E}\left[\delta_h X_t^i \mid \mathcal{G}\right]$ and $v_i^2 := \mathbb{V}\left[\delta_h X_t^i \mid \mathcal{G}\right]$. Thus,

$$\left\|\delta_h X_t^i\right\|_{2pq,\mathcal{G}} = \left\|v_i G + M_i\right\|_{2pq,\mathcal{G}} \leq \left\|v_i G\right\|_{2pq,\mathcal{G}} + \left\|M_i\right\|_{2pq,\mathcal{G}} \leq v_i C_{2pq}^{1/2pq} + \left|M_i\right|, \tag{3.109}$$

$$\left\|\delta_h X_t^i\right\|_{2pq,\mathcal{G}}^2 \leq 2\left(v_i^2 C_{2pq}^{1/pq} + M_i^2\right) \tag{3.110}$$

where G is an independent standard Gaussian r.v..

On the other hand, regarding $\left\|e^{v\delta_h X_t^i}\right\|_{pr,\mathcal{G}} = \mathbb{E}\left[e^{vpr\delta_h X_t^i} \mid \mathcal{G}\right]^{\frac{1}{pr}}$, consider the same Gaussian decomposition as before, which readily yields

$$\left\|e^{v\delta_h X_t^i}\right\|_{pr,\mathcal{G}} = \mathbb{E}\left[e^{vpr(v_i G + M_i)} \mid \mathcal{G}\right]^{\frac{1}{pr}} = e^{vM_i} \mathbb{E}\left[e^{vprv_i G} \mid \mathcal{G}\right]^{\frac{1}{pr}} = e^{vM_i + \frac{v^2 pr}{2} v_i^2}. \tag{3.111}$$

Taking the sup on $v \in [0,1]$, we get:

$$\sup_{v\in[0,1]} \left\|e^{v\delta_h X_t^i}\right\|_{pr,\mathcal{G}} \leq e^{(M_i)_+ + \frac{pr}{2} v_i^2}. \tag{3.112}$$

Consequently, gathering the two parts, we fall onto:

$$\left\|\delta_h S_t^i - S_t^i \delta_h X_t^i\right\|_{p,\mathcal{G}} \leq 2S_t^i (v_i^2 C_{2pq}^{1/pq} + M_i^2) e^{(M_i)_+ + \frac{pr}{2} v_i^2}. \tag{3.113}$$

Now, it remains to provide a control on the conditional variance v_i^2: using Lemma 3.3 and denoting by $\tilde{f}$ the representative in $\mathcal{L}_d^2$ of the stochastic integral part in $\delta_h X_t^i$, we get $\mathbb{V}\left[\delta_h X_t^i \mid \mathcal{G}\right]$:

$$\mathbb{V}\left[\delta_h X_t^i \mid \mathcal{G}\right] = \left|\tilde{f} - \tilde{f}^\star\right|_{\mathcal{L}_d^2}^2 \leq \left|\tilde{f}\right|_{\mathcal{L}_d^2}^2 = \mathbb{V}\left[\delta_h X_t^i\right]. \tag{3.114}$$

Finally, from Equation (3.13) we obtain $\mathbb{V}\left[\delta_h X_t^i\right] = \sigma_i^2 h^{2H_i}$. Collecting all bounds leads to the announced statement. ■

Corollary 3.11. *The conditional $\mathcal{L}_p$-norm between the future portfolio price increment and its Gaussian approximation is such that*

$$\left\|\delta_h P_{t_N} - \delta_h \hat{P}_{t_N}\right\|_{p,\mathcal{G}} \leq 2\sum_{i=1}^{d}\left|\omega^i\right| S^i_{t_N}\left(\sigma_i^2 h^{2H_i} C_{2pq}^{1/pq} + M_i^2\right) e^{(M_i)_+ + \frac{pr}{2}\sigma_i^2 h^{2H_i}}, \quad (3.115)$$

with $M_i := \mathbb{E}\left[\delta_h X_i^t \mid \mathcal{G}\right]$, $G \sim \mathcal{N}(0,1)$, and $C_{2pq} = \mathbb{E}\left[G^{2pq}\right]$.

Proof Start from Equation (3.86) to write

$$\delta_h P_{t_N} - \delta_h \hat{P}_{t_N} = \sum_{i=1}^{d}\omega^i\left(\delta_h S^i_{t_N} - S^i_{t_N}\delta_h X^i_{t_N}\right),$$

and then apply the triangular inequality on conditional $\mathcal{L}_p$-norm to get the expected formula. ■

Lemma 3.12. *Let f be the density of the standard Gaussian r.v. $\mathcal{N}(0,1)$, F the related c.d.f., and F^{-1} the quantile function. For any $\alpha \in [0.9, 1)$, we have:*

$$1-\alpha \leq f(F^{-1}(\alpha)) \leq 2(1-\alpha)\sqrt{\log\left(\frac{1}{8\pi(1-\alpha)^2}\right)}. \quad (3.116)$$

Proof VaR error bounds

Step 1. Bounds on the Lambert W function. The Lambert W function is defined as the inverse of the continuous and strictly increasing function $\mathbb{R}^+ \ni w \mapsto we^w \in \mathbb{R}^+$:

$$\mathrm{LW}(x)e^{\mathrm{LW}(x)} = x, \qquad \forall x \geq 0. \quad (3.117)$$

Invoking the implicit function theorem, it is easy to see that LW(.) is C^∞ on $\mathbb{R}^+$. For a deeper study, see *Robert M. Corless, David J. Jeffrey, and Donald E. Knuth. Sequence of series for Lambert W function. In Proceedings of the 1997 International Symposium on Symbolic and Algebraic Computation (Kihei, HI), pages 197–204 (electronic), New York, 1997. ACM.* In addition, LW(.) is strictly increasing on $\mathbb{R}^+$, $\mathrm{LW}(e) = 1$, $\mathrm{LW}(ne^n) = n$ for $n \geq 1$.

We claim that

$$\mathrm{LW}(x) \leq \log(x), \qquad \forall x \geq e, \quad (3.118)$$

$$\log(x) - \log(\log(x)) \leq \mathrm{LW}(x), \qquad \forall x \geq e. \quad (3.119)$$

Proof Set $w = \log(x)$, then $we^w = x\log(x) \geq x$ because $x \geq e$. This shows $we^w \geq \mathrm{LW}(x)e^{\mathrm{LW}(x)}$, which implies that $w \geq \mathrm{LW}(x)$ since $y \mapsto ye^y$ is strictly increasing. Equation (3.118) is proven.
Similarly, if $w = \log(x) - \log(\log(x))$, $we^w = [\log(x) - \log(\log(x))]\frac{x}{\log(x)} \leq x$ because $\log(\log(x)) \geq 0$ for $x \geq e$. Using the same arguments as before, we conclude that $w \leq \mathrm{LW}(x)$ as stated in (3.119). ■

Step 2. Using Mill's ratio bounds. They are given by:

$$\frac{x}{x^2+1}\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \leq F(-x) \leq \frac{1}{x}\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}, \qquad \forall x > 0. \tag{3.120}$$

Note that $\alpha \geq 0.85$ implies $x = F^{-1}(\alpha) \geq 1$. The inequality (3.120) becomes

$$\frac{1}{F^{-1}(\alpha)}\frac{F^{-1}(\alpha)^2}{F^{-1}(\alpha)^2+1}\frac{e^{-\frac{F^{-1}(\alpha)^2}{2}}}{\sqrt{2\pi}} \leq F(-F^{-1}(\alpha)) \leq \frac{1}{F^{-1}(\alpha)}\frac{e^{-\frac{F^{-1}(\alpha)^2}{2}}}{\sqrt{2\pi}}, \qquad \forall \alpha \geq 0.85. \tag{3.121}$$

Observe that $F(-F^{-1}(\alpha)) = 1-\alpha$ and $\frac{1}{2} \leq \frac{y^2}{y^2+1}$ for $y \geq 1$. Therefore

$$\frac{1}{2F^{-1}(\alpha)e^{\frac{F^{-1}(\alpha)^2}{2}}} \leq \sqrt{2\pi}(1-\alpha) \leq \frac{1}{F^{-1}(\alpha)e^{\frac{F^{-1}(\alpha)^2}{2}}}, \qquad \forall \alpha \geq 0.85. \tag{3.122}$$

Taking the square and the inverse gives

$$F^{-1}(\alpha)^2 e^{F^{-1}(\alpha)^2} \leq (\sqrt{2\pi}(1-\alpha))^{-2} \leq 4F^{-1}(\alpha)^2 e^{F^{-1}(\alpha)^2}, \qquad \forall \alpha \geq 0.85. \tag{3.123}$$

This implies

$$F^{-1}(\alpha)^2 \leq \mathrm{LW}((\sqrt{2\pi}(1-\alpha))^{-2}) \quad \text{and} \quad F^{-1}(\alpha)^2 \geq \mathrm{LW}((2\sqrt{2\pi}(1-\alpha))^{-2}), \qquad \forall \alpha \geq 0.85. \tag{3.124}$$

For $\alpha \geq 0.9$, note that $(\sqrt{2\pi}(1-\alpha))^{-2} \geq (\sqrt{2\pi} \times 0.1)^{-2}) = 15.9\cdots \geq e$ and $(2\sqrt{2\pi}(1-\alpha))^{-2} \geq 15.9\ldots/4 \geq e$. Therefore, inequalities (3.118) and (3.119) can be used: this yields

$$F^{-1}(\alpha)^2 \leq \log((\sqrt{2\pi}(1-\alpha))^{-2}) \text{ and } F^{-1}(\alpha)^2 \geq [\log(.) - \log(\log(.))]((2\sqrt{2\pi}(1-\alpha))^{-2}), \ \forall \alpha \geq 0.9. \tag{3.125}$$

Step 3. Final bounds. We are now in a position to bound

$$f(F^{-1}(\alpha)) = \frac{1}{\sqrt{2\pi}}e^{-\frac{F^{-1}(\alpha)^2}{2}}.$$

- Lower bound.

$$f(F^{-1}(\alpha)) \geq \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\log((\sqrt{2\pi}(1-\alpha))^{-2})} = 1-\alpha. \tag{3.126}$$

- Upper bound.

$$f(F^{-1}(\alpha)) \leq \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\log((2\sqrt{2\pi}(1-\alpha))^{-2})+\frac{1}{2}\log(\log((2\sqrt{2\pi}(1-\alpha))^{-2}))} \tag{3.127}$$

$$\leq 2(1-\alpha)\sqrt{\log((2\sqrt{2\pi}(1-\alpha))^{-2})}. \tag{3.128}$$

■

Finally, we determine the upper bounds for the Gaussian approximation of the $\mathbf{VaR}_\alpha$. $\delta_h\hat{P}_t \mid \mathcal{G}$ is a Gaussian r.v. and $\delta_h P_t \mid \mathcal{G}$ is a linear combination of log-normal r.v.. These two r.v. have a continuous cumulative distribution function (c.d.f.). Then, for any $\alpha \in (0,1)$ and for any $\mathbf{VaR}_\alpha(\delta_h P_t \mid \mathcal{G})$ we have, from (Dominici*, 2003, p. 12):

$$\mathbf{VaR}_{\alpha-\Delta}(\delta_h\hat{P}_t \mid \mathcal{G}) \leq \mathbf{VaR}_\alpha(\delta_h P_t \mid \mathcal{G}) \leq \mathbf{VaR}_{\alpha+\Delta}(\delta_h\hat{P}_t \mid \mathcal{G}) \tag{3.129}$$

where $\Delta = d_{Kol}(\delta_h\hat{P}_t \mid \mathcal{G}, \delta_h P_t \mid \mathcal{G})$ is the distance of Kolmogorov between the two distributions related to the r.v. $\delta_h\hat{P}_t \mid \mathcal{G}$ and $\delta_h P_t \mid \mathcal{G}$.

Then, the approximation of $\mathbf{VaR}_{\alpha+\Delta}(\delta_h\hat{P}_t \mid \mathcal{G})$ with $\mathbf{VaR}_{\alpha-\Delta}(\delta_h\hat{P}_t \mid \mathcal{G})$ is given as follows:

$$\left|\mathbf{VaR}_\alpha(\delta_h P_t \mid \mathcal{G}) - \mathbf{VaR}_\alpha(\delta_h\hat{P}_t \mid \mathcal{G})\right| \leq \sup_{x\in[\alpha-\Delta,\alpha+\Delta]} |f(F^{-1}(x))|^{-1}\Delta \tag{3.130}$$

where f is the Gaussian density related to $\delta_h\hat{P}_t \mid \mathcal{G}$ and F is its related c.d.f. $\delta_h\hat{P}_t|\mathcal{G} \sim \mathcal{N}(\mu(\mathcal{G}), \sigma^2(\mathcal{G}))$, thus $\forall y \in \mathbb{R}$:

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma(\mathcal{G})} e^{-\frac{1}{2}\left(\frac{y-\mu(\mathcal{G})}{\sigma(\mathcal{G})}\right)^2}, \quad F(y) = \int_{-\infty}^{y} f(u)\mathrm{d}u. \tag{3.131}$$

Proof Goal: Find an upper bound of the following quantity:

$$\left|\mathbf{VaR}_\alpha(\delta_h P_{t_N}|\mathcal{G}) - \mathbf{VaR}_\alpha(\delta_h\hat{P}_{t_N}|\mathcal{G})\right| \tag{3.132}$$

where $\delta_h P_{t_N} = \sum_{i=1}^d \omega^i \delta_h S^i_{t_N}$ and $\delta_h\hat{P}_{t_N} = \sum_{i=1}^d \omega^i_{t_N}\delta_h X^i_{t_N}$ with $\omega^i_{t_N} = \omega^i\sigma_i S^i_{t_N} R_{:i}$. Assuming that the following inequality holds:

$$\mathbf{VaR}_{\alpha-\Delta}(\delta_h\hat{P}_t|\mathcal{G}) \leq \mathbf{VaR}_\alpha(\delta_h P_t|\mathcal{G}) \leq \mathbf{VaR}_{\alpha+\Delta}(\delta_h\hat{P}_t|\mathcal{G}) \tag{3.133}$$

and since:

$$\mathbf{VaR}_{\alpha-\Delta}(\delta_h\hat{P}_t|\mathcal{G}) \leq \mathbf{VaR}_\alpha(\delta_h\hat{P}_t|\mathcal{G}) \leq \mathbf{VaR}_{\alpha+\Delta}(\delta_h\hat{P}_t|\mathcal{G}) \tag{3.134}$$

Hence:

$$\left|\mathbf{VaR}_\alpha(\delta_h P_t|\mathcal{G}) - \mathbf{VaR}_\alpha(\delta_h\hat{P}_t|\mathcal{G})\right| \leq \left|\mathbf{VaR}_{\alpha+\Delta}(\delta_h\hat{P}_t|\mathcal{G}) - \mathbf{VaR}_{\alpha-\Delta}(\delta_h\hat{P}_t|\mathcal{G})\right| \tag{3.135}$$

Let us denote by F the cumulative distribution function of the Gaussian distribution $\mathcal{N}(\mu(\mathcal{G}), \sigma^2(\mathcal{G}))$ where $\mu(\mathcal{G}) = \mathbb{E}[\delta_h\hat{P}_t|\mathcal{G}]$ and $\sigma^2(\mathcal{G}) = \mathbb{V}[\delta_h\hat{P}_t|\mathcal{G}]$. We can then rewrite the inequality above as follows:

$$\left|\mathbf{VaR}_\alpha(\delta_h P_t|\mathcal{G}) - \mathbf{VaR}_\alpha(\delta_h\hat{P}_t|\mathcal{G})\right| \leq \left|F^{-1}(\alpha+\Delta) - F^{-1}(\alpha-\Delta)\right|. \tag{3.136}$$

From the mean value theorem we get:

$$\left|F^{-1}(\alpha+\Delta)-F^{-1}(\alpha-\Delta)\right| \leq \sup_{x\in[\alpha-\Delta,\alpha+\Delta]}\left|\frac{d}{dx}F^{-1}(x)\right| 2\Delta \tag{3.137}$$

$$\leq \sup_{x\in[\alpha-\Delta,\alpha+\Delta]}\left|f(F^{-1}(x))\right|^{-1} 2\Delta \tag{3.138}$$

Consequently, we have:

$$\left|\mathbf{VaR}_\alpha(\delta_h P_t|\mathcal{G})-\mathbf{VaR}_\alpha(\delta_h \hat{P}_t|\mathcal{G})\right| \leq 2\sup_{x\in[\alpha-\Delta,\alpha+\Delta]}\left|f(F^{-1}(x))\right|^{-1}\Delta. \tag{3.139}$$

Maximizing $|x\sqrt{-\ln(2\pi\sigma^2(\mathcal{G})x^2)}|^{-1}$ on $[\alpha-\Delta,\alpha+\Delta]$ amounts to maximizing $|x\sqrt{-\ln(2\pi\sigma^2(\mathcal{G})x^2)}|^{-1}$ on $[\alpha-UB(\Delta),\alpha+UB(\Delta)]$. However Δ is a distance thus, $\Delta\geq 0$, therefore, $|x\sqrt{-\ln(2\pi\sigma^2(\mathcal{G})x^2)}|^{-1}$ is maximal when $|x\sqrt{-\ln(2\pi\sigma^2(\mathcal{G})x^2)}|$ is minimal. Moreover, the function $x\mapsto x\sqrt{-\ln(2\pi\sigma^2(\mathcal{G})x^2)}$ is decreasing and, its minimum is reached in $\alpha+UB(\Delta)$.

Thus, we obtain:

$$\sup_{x\in[\alpha-\Delta,\alpha+\Delta]}|x\sqrt{-\ln(2\pi\sigma^2(\mathcal{G})x^2)}|^{-1}\Delta = |(\alpha+UB(\Delta))\sqrt{-\ln(2\pi\sigma^2(\mathcal{G})(\alpha+UB(\Delta))^2)}|\Delta \tag{3.140}$$

Finally we fall on:

$$\left|\mathbf{VaR}_\alpha(\delta_h P_t \mid \mathcal{G})-\mathbf{VaR}_\alpha(\delta_h \hat{P}_t \mid \mathcal{G})\right| \leq \left|\alpha+UB(\Delta)\right|\sqrt{-\ln(2\pi\sigma^2(\mathcal{G})2(\alpha+UB(\Delta))^2)}\Delta \tag{3.141}$$

with $\Delta = \left\|\delta_h\hat{P}_t-\delta_h P_t\right\|_{p,\mathcal{G}}$ and $UB(\Delta)=2\sum_{i=1}^d\left|S^i_{t_N}\right|\left(\sigma_i^2h^{2H_i}C_{2pq}^{1/pq}+\left|M_i\right|^2\right)e^{M_i+\frac{pr}{2}\sigma_i^2h^{2H_i}}$.

■

Proof ES error bounds

Now, let us quantify the error of approximation of the conditional $\mathbf{ES}_\alpha$ of the future variation of the true portfolio using the conditional $\mathbf{ES}_\alpha$ of the Gaussian approximation of the future portfolio variation:

$$\left|\mathbf{ES}_\alpha(\delta_h P_{t_N} \mid \mathcal{G})-\mathbf{ES}_\alpha(\delta_h \hat{P}_{t_N} \mid \mathcal{G})\right| \tag{3.142}$$

$$= \Bigg|\mathbf{VaR}_\alpha(\delta_h P_{t_N} \mid \mathcal{G})-\mathbf{VaR}_\alpha(\delta_h \hat{P}_{t_N} \mid \mathcal{G}) \tag{3.143}$$

$$+\frac{1}{1-\alpha}\left(\mathbb{E}\left[(\delta_h P_{t_N}-\mathbf{VaR}_\alpha(\delta_h P_{t_N} \mid \mathcal{G}))_+ \mid \mathcal{G}\right]-\mathbb{E}\left[(\delta_h \hat{P}_{t_N}-\mathbf{VaR}_\alpha(\delta_h \hat{P}_{t_N} \mid \mathcal{G}))_+ \mid \mathcal{G}\right]\right)\Bigg| \tag{3.144}$$

$$= \left|\mathbf{VaR}_\alpha(\delta_h P_{t_N} \mid \mathcal{G})-\mathbf{VaR}_\alpha(\delta_h \hat{P}_{t_N} \mid \mathcal{G})\right| \tag{3.145}$$

$$+\frac{1}{1-\alpha}\left|\mathbb{E}\left[(\delta_h P_{t_N}-\mathbf{VaR}_\alpha(\delta_h P_{t_N} \mid \mathcal{G}))_+ \mid \mathcal{G}\right]-\mathbb{E}\left[(\delta_h \hat{P}_{t_N}-\mathbf{VaR}_\alpha(\delta_h \hat{P}_{t_N} \mid \mathcal{G}))_+ \mid \mathcal{G}\right]\right|. \tag{3.146}$$

But the function $x \mapsto x_+$ is 1-Lipschitz and $\mathbf{VaR}_\alpha(\delta_h P_{t_N} \mid \mathcal{G})$ and $\mathbf{VaR}_\alpha(\delta_h \hat{P}_{t_N} \mid \mathcal{G})$ are $\mathcal{G}$-measurable, then it comes that:

$$\left|\mathbb{E}\left[(\delta_h P_{t_N} - \mathbf{VaR}_\alpha(\delta_h P_{t_N} \mid \mathcal{G}))_+ \mid \mathcal{G}\right] - \mathbb{E}\left[(\delta_h \hat{P}_{t_N} - \mathbf{VaR}_\alpha(\delta_h \hat{P}_{t_N} \mid \mathcal{G}))_+ \mid \mathcal{G}\right]\right| \tag{3.147}$$

$$\leq \left|\mathbb{E}\left[\left|\delta_h P_{t_N} - \delta_h \hat{P}_{t_N}\right| \mid \mathcal{G}\right]\right| + \left|\mathbf{VaR}_\alpha(\delta_h P_{t_N} \mid \mathcal{G}) - \mathbf{VaR}_\alpha(\delta_h \hat{P}_{t_N} \mid \mathcal{G})\right|. \tag{3.148}$$

But $\left|\mathbb{E}\left[\left|\delta_h P_{t_N} - \delta_h \hat{P}_{t_N}\right| \mid \mathcal{G}\right]\right| = \left\|\delta_h \hat{P}_t - \delta_h P_t\right\|_{1,\mathcal{G}}$. Thanks to convexity inequalities, for $p \geq 1$, $p \mapsto \|X\|_{p,\mathcal{G}}$ is non-decreasing so that $\|X\|_{1,\mathcal{G}} \leq \|X\|_{p,\mathcal{G}}$ for any $p \geq 1$. We thus get

$$\left|\mathbb{E}\left[(\delta_h P_{t_N} - \mathbf{VaR}_\alpha(\delta_h P_{t_N} \mid \mathcal{G}))_+ \mid \mathcal{G}\right] - \mathbb{E}\left[(\delta_h \hat{P}_{t_N} - \mathbf{VaR}_\alpha(\delta_h \hat{P}_{t_N} \mid \mathcal{G}))_+ \mid \mathcal{G}\right]\right| \tag{3.149}$$

$$\leq \left\|\delta_h \hat{P}_t - \delta_h P_t\right\|_{p,\mathcal{G}} + \left|\mathbf{VaR}_\alpha(\delta_h P_{t_N} \mid \mathcal{G}) - \mathbf{VaR}_\alpha(\delta_h \hat{P}_{t_N} \mid \mathcal{G})\right|. \tag{3.150}$$

Finally, we fall onto the following upper bound:

$$\left|\mathbf{ES}_\alpha(\delta_h P_{t_N} \mid \mathcal{G}) - \mathbf{ES}_\alpha(\delta_h \hat{P}_{t_N} \mid \mathcal{G})\right| \tag{3.151}$$

$$\leq \left(1 + \frac{1}{1-\alpha}\right)\left|\mathbf{VaR}_\alpha(\delta_h P_{t_N} \mid \mathcal{G}) - \mathbf{VaR}_\alpha(\delta_h \hat{P}_{t_N} \mid \mathcal{G})\right| + \frac{1}{1-\alpha}\left\|\delta_h \hat{P}_t - \delta_h P_t\right\|_{p,\mathcal{G}} \quad \text{for all } p \geq 1. \tag{3.152}$$

From Theorem 3.8 and the above upper-bound, we thus obtain:

$$\left|\mathbf{ES}_\alpha(\delta_h P_{t_N} \mid \mathcal{G}) - \mathbf{ES}_\alpha(\delta_h \hat{P}_{t_N} \mid \mathcal{G})\right| \tag{3.153}$$

$$\leq \left[\left(1 + \frac{1}{1-\alpha}\right)\frac{2\sigma(\mathcal{G})}{1-\alpha-\Delta} + \frac{2}{1-\alpha}\right]\sum_{i=1}^{d}\left|\omega^i\right| S^i_{t_N}\left(\sigma_i^2 h^{2H_i} C_{2pq}^{\frac{1}{pq}} + M_i^2\right) e^{(M_i)_+ + \frac{pr}{2}\sigma_i^2 h^{2H_i}}. \tag{3.154}$$

■

3.4 Simulation of fBm and mfBm

Different simulation methods exist, and exhibit different complexities. We used the Cholesky method in our simulations. Below, we discuss its rationale and compare it with other well-known methods.

Let K_X be the covariance kernel related to fBm. Because K_X is a symmetric definite positive matrix, it admits a Cholesky decomposition, $K_X = L^T L$, where L is a lower triangular matrix. Thus, simulating a sample of fBm at times $\frac{i}{N}, i = 1, \ldots, N$, is

equivalent to generating a vector Z composed of N independent standard Gaussian variables and applying the product LZ. Indeed, LZ is a centered Gaussian vector and $\mathbb{E}[(LZ)(LZ)^T] = K_X$. This method exactly simulates an fBm, because there exists a unique Gaussian process with the given theoretical expectation and variance. (See (Bardet et al., 2003) for more details.)

Owing to the computational complexity of the order $\mathcal{O}((N)^3)$ and the fact that K_X is ill-conditioned, it is of interest to derive methods that are computationally less demanding (Coeurjolly, 2000a). Some of these methods are also exact, insofar as they reproduce the covariance of the desired process. They reach a lower complexity than the Cholesky method because they exploit simple properties of the covariance matrix. For instance, the Levinson method relies on the Toeplitz property of the covariance matrix of fractional Gaussian noise (fGn) (Coeurjolly, 2000a) and has a complexity in $\mathcal{O}(N^2 \log(N))$. The Wood and Chan algorithm (Chan and Wood, 1999) obtains a complexity of $\mathcal{O}(N \log(N))$ by embedding the covariance of an fGn in a circulant matrix.

Even if these methods are interesting from the point of view of their algorithmic complexity, it is not easy to generalize them to the multivariate framework because the covariance matrix of an mfBm is neither Toeplitz nor circulant. In contrast, the Cholesky method is suitable for the multivariate framework as soon as we know the corresponding covariance matrix, including the covariance between distinct components of mfBm.

A considerable improvement is given in (Amblard et al., 2010). In this paper, authors propose an algorithm to perfectly simulate mfBm. The approach is based on embedding the covariance matrix G of fractional Gaussian noise, which is a Toeplitz block matrix of size $Nd \times Nd$ in a matrix $C = circ\{C(j)\}_{j=0,\dots,m-1}$ where m is a power of 2 greater than $2(N-1)$ which is a symmetric matrix with a nested block circulant structure composed of m Hermitian matrices. Therefore, the computation of $C^{1/2}$ is much less expensive than the computation of the $G^{1/2}$ because of the use of the Fast Fourier Transform (FFT) as in (Maejima and Mason, 1994) which considerably reduces the cost:

$$\mathcal{O}\left(\frac{d(d+1)}{2} m \log(m)\right) + \mathcal{O}(md^3) + \mathcal{O}(dm \log(m)). \tag{3.155}$$

The crucial point of this algorithm is the non-negativity of the eigen-values of Hermitian matrices.

In addition to these methods, efficient methods exist from the point of view of algorithmic complexity, based on an approximation. Among these, we can cite recent efforts in this direction based on neural networks (Allouche et al., 2022).

3.5 Estimation

3.5.1 Relationship between the theoretical model and the implementation

First, we recall the conditional $\mathbf{VaR}_\alpha$ models for which we are interested in estimating the parameters. The filtration considered is given in Equation (3.85) by $\mathcal{G} := \sigma(\delta_h X^i_{t_l} : (i,l) \in \mathcal{I}) = \sigma(I(f^i_l) : (i,l) \in \mathcal{I})$ where $f^i_l(\cdot) := \sigma_i R_{:,i} \psi^{H_i}_{t_l,t_{l+1}}(\cdot)$.

The Gaussian approximation of the conditional $\mathbf{VaR}_\alpha$ corresponds to an affine function of the α-quantile of the standard Gaussian distribution, where the intercept is the conditional expectation of $\delta_h \hat{P}_N$ given the filtration $\mathcal{G}$ and the slope is the square root of the conditional variance of $\delta_h \hat{P}_N$ given $\mathcal{G}$, i.e. $\mathbf{VaR}_\alpha(\delta_h \hat{P}_N \mid \mathcal{G}) = \sigma(\mathcal{G})\mathcal{N}^{-1}(\alpha) + \mu(\mathcal{G})$, as mentioned in Equation (3.88).

We recall that $\delta_h \hat{P}_N = I(f)$ where $f(\cdot) = \sum_{i=1}^d \omega_i \sigma_i S^i_{t_N} R_{:,i} \psi^{H_i}_{t_N,t_{N+1}}(\cdot)$..

The conditional expectation of $\delta_h \hat{P}_N$ given $\mathcal{G}$ corresponds to the orthogonal projection of $\delta_h \hat{P}_N$ on the Hilbert space spanned by filtration $\mathcal{G}$ and the conditional variance of $\delta_h \hat{P}_N$ corresponds to the projection residual. These two quantities are defined in Theorem 3.4 by Equations (3.64):

$$\mu(\mathcal{G}) = \mathbb{E}\left[\delta_h \hat{P}_N \mid \mathcal{G}\right] = \mathbb{E}\left[I(f) | \{I(f^i_l)\}_{(i,l)\in\mathcal{I}}\right] \tag{3.156}$$

$$= \sum_{(i,l)\in\mathcal{I}} a^i_l I(f^i_l) = \sum_{i=1}^d \sum_{l=0}^{N-1} a^i_l (B^{H_i}_{t_{l+1}} - B^{H_i}_{t_l}) = \left(m_1 + \boldsymbol{K_{12}} \boldsymbol{K_{22}}^{-1} (\boldsymbol{Y_2} - \boldsymbol{m_2})\right). \tag{3.157}$$

and

$$\sigma(\mathcal{G}) = \sqrt{\mathbb{V}[\delta_h \hat{P}_N \mid \mathcal{G}]} = \mathbb{V}\left[I(f) \mid \{I(f^i_l)\}_{(i,l)\in\mathcal{I}}\right] \tag{3.158}$$

$$= \sum_{i,j=1}^d (\omega^i_{t_N} \cdot \omega^j_{t_N}) \int_{\mathbb{R}} \psi^{H_i}_{t_N,t_{N+1}}(u) \psi^{H_j}_{t_N,t_{N+1}}(u) \mathrm{d}u \tag{3.159}$$

$$- \sum_{i,j=0}^{N-1} \sum_{k,l=0}^{N-1} a^i_l a^j_k \rho_{ij} \int_{\mathbb{R}} \psi^{H_i}_{t_l,t_{l+1}}(u) \psi^{H_j}_{t_k,t_{k+1}}(u) \mathrm{d}u \tag{3.160}$$

$$= (K_{11} - \boldsymbol{K_{12}} \boldsymbol{K_{22}}^{-1} \boldsymbol{K_{21}})^{\frac{1}{2}}. \tag{3.161}$$

where:

$$\begin{cases} m_1 = \mathbb{E}\left[\delta_h \hat{P}_N\right] = \sum_{i=1}^d w^i \mathbb{E}\left[S^i_{t_N} \delta_h X^i_{t_N}\right] \\ \boldsymbol{m_2} = \mathbb{E}\left[\left(\delta_h X^i_{t_l}\right)^T_{(i,l)\in\mathcal{I}}\right] = \mathbf{0}_{N\times d} \\ \boldsymbol{K_{12}} = \left(\mathrm{Cov}(\delta_h \hat{P}_{t_N}, \delta_h X^i_{t_l})\right)_{(i,l)\in\mathcal{I}} = \left(\sum_{j=1}^d w^j \mathrm{Cov}(S^j_{t_N}\delta_h X^j_{t_N}, \delta_h X^i_{t_l})\right)_{(i,l)\in\mathcal{I}} \\ \boldsymbol{K_{21}} = \boldsymbol{K_{12}}^T \\ \boldsymbol{K_{22}} = \left(\mathrm{Cov}(\delta_h X^i_{t_l}, \delta_h X^j_{t_k})\right)_{(i,l)\times(j,k)\in\mathcal{I}^2}. \end{cases}$$

These quantities can be explicitly expressed. For all $(i,j) \in [\![1,d]\!]^2$ and for all $(k,l) \in [\![0,N-1]\!]^2$:

$$\mathrm{Cov}(\delta_h X^i_{t_l}, \delta_h X^j_{t_k}) = \frac{\sigma_i\sigma_j\rho_{ij}}{2}\Bigg(\left|t_{l+1}-t_k\right|^{H_i+H_j} + \left|t_l - t_{k+1}\right|^{H_i+H_j} \tag{3.162}$$

$$- \left|t_{l+1}-t_{k+1}\right|^{H_i+H_j} - \left|t_l - t_k\right|^{H_i+H_j}\Bigg). \tag{3.163}$$

In practice, we assume that any time t_k is a multiple of the smallest time-step τ such that $t_k = k\tau$. The covariance kernel can be simplified as follows:

$$\mathrm{Cov}(\delta_h X^i_{t_l}, \delta_h X^j_{t_k}) = \frac{\sigma_i\sigma_j\rho_{ij}\,|\tau|^{H_i+H_j}}{2}\left(\left|l-k+1\right|^{H_i+H_j} + \left|l-k-1\right|^{H_i+H_j} - 2\left|l-k\right|^{H_i+H_j}\right).$$

Moreover, the quantities $\mathrm{Cov}(S^j_{t_N}\delta_h X^j_{t_N}, \delta_h X^i_{t_l})$ and $\mathbb{E}\left[S^i_{t_N}\delta_h X^i_{t_N}\right]$ can be computed using the partial derivatives of the Laplace Transform of the linear combination of log-price increments $L_X(\boldsymbol{\alpha}) = \mathbb{E}\left[e^{\sum_{i=1}^d \sum_{k=0}^N \alpha^i_k \delta_h X^i_{t_k}}\right]$.

3.5.2 Essential model properties for parameter estimation

It is important to recall that three assumptions are formulated on the model: *Gaussianity*, *stationarity* and *self-similarity*. These assumptions are required to estimate the parameters.

Gaussian Assumption

Our predictive model provides the conditional $\mathbf{VaR}_\alpha$ of the Gaussian approximation of future portfolio variation given the past log-price increments of the assets.

The Gaussian assumption states that for each asset, the log-prices are assimilated to an fBm. Then, the log-price increments exhibit the same behavior as the fBm increments. Moreover, the log-price increments form a Gaussian vector and, any linear combination of the increments is a Gaussian variable. The approximation of future portfolio variation $\delta_h \hat{P}_{t_N}$ is a Gaussian variable conditionally to the past log-price increments of the assets. Therefore, the conditional $\mathbf{VaR}_\alpha$ of $\delta_h \hat{P}_{t_N}$ given the past log-returns of the assets is

the quantile of a Gaussian distribution. More precisely, $\mathbf{VaR}_\alpha(\delta_h \hat{P}_N \mid \mathcal{G})$ is an affine function of the α-quantile of the standard Gaussian distribution, whose intercept is given by the conditional expectation $\mu(\mathcal{G}) = \mathbb{E}\left[\delta_h \hat{P}_N \mid \mathcal{G}\right]$ and whose slope is the square root of the conditional variance $\sigma(\mathcal{G}) = \sqrt{\mathbb{V}[\delta_h \hat{P}_N \mid \mathcal{G}]}$.

Consequently, the Gaussian assumption is very important because it allows the assimilation of log-price increments to fBm increments, which results in a Gaussian **VaR** with a convenient form.

Stationarity assumption

We are interested in the stationarity property of the increments of the process $\{X_t^i\}$. An increment is characterized by two parameters, its *origin* and its *duration*, so that we can define two types of processes based on increments: the process of the increments with the same origin and various durations $\{X_{t+m}^i - X_m^i\}_{t\in\mathbb{R}}$ and the process of the increments with various origins and the same duration $\{X_{t+m}^i - X_t^i\}_{t\in\mathbb{R}}$.

On the one hand, because the log-price process $\{X_t^i\}$ is assimilated to an fBm, Chapter 2 shows that the process $\{X_{t+m}^i - X_m^i\}_{t\in\mathbb{R}}$ is stationary. The weak-sense stationarity of this process, refers to the invariance "t by t" of the expectation and variance by time-origin change:

$$\mathbb{E}\left[X_{t+m}^i - X_m^i\right] = \mathbb{E}\left[X_t^i\right] \quad \text{and} \quad \mathbb{V}[X_{t+m}^i - X_m^i] = \mathbb{V}[X_t^i], \tag{3.164}$$

and is enough both to completely determine the covariance kernel of the process $\{X_t^i\}$,

$$\mathrm{Cov}(X_t^i, X_s^i) = \frac{1}{2}\left(\mathbb{V}[X_t^i] + \mathbb{V}[X_s^i] - \mathbb{V}[X_{t-s}^i]\right), \tag{3.165}$$

and to establish the strict-sense stationarity of the Gaussian process $\{X_{t+m}^i - X_m^i\}_{t\in\mathbb{R}}$:

$$\mathrm{Cov}(X_{t+m}^i - X_m^i, X_{s+m}^i - X_m^i) = \mathrm{Cov}(X_t^i, X_s^i). \tag{3.166}$$

However, to estimate the parameters of the model, we need to obtain the stationarity of the increments of same duration m and various origins $\{X_{t+m}^i - X_t^i\}_{t\in\mathbb{R}}$. This Gaussian process is stationary in the strict sense because, using Equation (3.165),

$$\begin{aligned}
&\mathrm{Cov}(X_{t+m}^i - X_t^i, X_{s+m}^i - X_s^i) && (3.167)\\
&= \mathrm{Cov}(X_{t+m}^i, X_{s+m}^i) + \mathrm{Cov}(X_t^i, X_s^i) - \mathrm{Cov}(X_t^i, X_{s+m}^i) - \mathrm{Cov}(X_{t+m}^i, X_s^i) && (3.168)\\
&= \frac{1}{2}(\mathbb{V}[X_{t+m}^i] + \mathbb{V}[X_{s+m}^i] - \mathbb{V}[X_{t-s}^i] + \mathbb{V}[X_t^i] + \mathbb{V}[X_s^i] - \mathbb{V}[X_{t-s}^i] && (3.169)\\
&\qquad - \mathbb{V}[X_t^i] - \mathbb{V}[X_{s+m}^i] + \mathbb{V}[X_{t-s-m}^i] - \mathbb{V}[X_{t+m}^i] - \mathbb{V}[X_s^i] + \mathbb{V}[X_{t-s+m}^i]) && (3.170)\\
&= \frac{1}{2}\left(\mathbb{V}[X_{t-s-m}^i] + \mathbb{V}[X_{t-s+m}^i] - 2\mathbb{V}[X_{t-s}^i]\right) && (3.171)\\
&= \mathrm{Cov}(X_{t+h+m}^i - X_{t+h}^i, X_{s+h+m}^i - X_{s+h}^i). && (3.172)
\end{aligned}$$

The covariance kernel of the increments process $\{X_{t+m}^i - X_t^i\}_{t\in\mathbb{R}}$ only depends on the duration m of the increments and on the time-gap $t-s$ between the time-origins of the two increments. The stationarity of the increments process $\{X_{t+m}^i - X_t^i\}_{t\in\mathbb{R}}$ expresses

the invariance by time-origin change of the characteristics of the process, especially of its expectation and variance-covariance kernel.

The weak-sense stationarity assumption of increments of the same size with various origins $\{X^i_{t+m} - X^i_t\}_{t\in\mathbb{R}}$ is sufficient to obtain the strict-sense stationarity of the increments process and is required to ensure the stability of the estimators on time-origin translated data. The weak-sense stationarity of the increments process guarantees the invariance by time-origin change of the estimators of the expectation and variance of the process. The strict-sense stationarity of the increments process insures the invariance by time-origin change of the covariance of the process. Indeed, if the increments process was not stationary, then the increments process would not be invariant by time-origin change, thus it would not exhibit the same characteristics and properties when the data is translated. Therefore, the estimations of the parameters computed from the data would not be reliable because they would change with the translation of the data. Owing to the stationarity property of the increments, the characteristics and properties of the increments process are invariant by time-origin change; thus, the estimators provide stable and reliable estimations of the parameters computed on the increments.

Self-similarity assumption

The self-similarity property is related to a time-scaling of the process $\{X^i_t\}$, and allows the establishment of a proportionality relationship between the characteristics of the processes $\{X^i_{\lambda t}\}$ and $\{X^i_t\}$, for all $\lambda > 0$. Because the log-price process $\{X^i_t\}$ is assimilated to an fBm, $\{X^i_t\}$ is self-similar with a self-similarity factor $\Theta^i(\lambda) = |\lambda|^{2H_i}$, where $H_i \in (0,1)$, as introduced in Chapter 2. The weak-sense self-similarity property establishes a spatial proportionality relationship "t by t" between the square of the expectations and between the variances of the processes $\{X^i_{\lambda t}\}$ and $\{X^i_t\}$, with as proportionality factor a function depending on λ, $\Theta^i(\lambda)$:

$$\left(\mathbb{E}\left[X^i_{\lambda t}\right]\right)^2 = \Theta^i(\lambda)\left(\mathbb{E}\left[X^i_t\right]\right)^2 \quad \text{and} \quad \mathbb{V}[X^i_{\lambda t}] = \Theta^i(\lambda)\mathbb{V}[X^i_t]. \tag{3.173}$$

Specifically, for all $t \in \mathbb{R}$:

$$\left(\mathbb{E}\left[X^i_t\right]\right)^2 = |t|^{2H_i}\left(\mathbb{E}\left[X^i_1\right]\right)^2 \quad \text{and} \quad \mathbb{V}[X^i_t] = |t|^{2H_i}\,\mathbb{V}[X^i_1]. \tag{3.174}$$

Thus, weak-sense self-similarity fully characterizes the square of the expectation and variance of the process, based on a power function.

The weak-sense self-similarity property of the data is required to ensure the stability of the estimators computed on the time-scaled process. Indeed, the self-similarity property of the process allows the establishment of a spatial proportionality relationship between the characteristics (squared expectation and variance) of the process $\{X^i_t\}$ taken at proportional times λt and t, with a proportionality coefficient $\Theta^i(\lambda)$. Therefore, estimators applied to processes taken at two proportional time-scales λt and t can be linked thanks to this spatial proportionality relationship of factor $\Theta^i(\lambda)$ which allows easy adaptation of the estimators to different time-scales. Thus, we obtain reliable estimations, coherent with the different time-scales. If the self-similarity property was not valid, it would

not be possible to establish a relationship between the characteristics of the process $\{X_t^i\}$ taken at different scales. This would imply that there would be no relationship between the estimators of the expectation and variance applied to the process $\{X_t^i\}$ taken at different scales, therefore estimations performed on a process at a given time-scale could not be adapted to another time-scale and would have to be redone. The weak-sense self-similarity property is then pivotal to obtain reliable estimations, coherent with the different time-scales. However, it is important to note that the weak-sense self-similarity property is neither sufficient to obtain the self-similarity of increments of the same length with different time-origins, nor sufficient to obtain the strict-sense self-similarity property. An additional assumption, called weak-sense stationarity, is required to obtain both the self-similarity of increments of the same length with different time-origins, and the strict-sense self-similarity of the process.

Combination of stationarity and self-similarity assumptions

From Chapter 2, we proved that when weak-sense stationarity and self-similarity properties are combined under the Gaussian assumption, if the self-similarity function Θ^i is not constant, then the expectation of the process is necessarily null. We then work with centered Gaussian processes.

Combining the weak-sense stationarity of the increments process with the same origin and various durations $\{X_{t+m}^i - X_m^i\}_{t\in\mathbb{R}}$, and the weak-sense self-similarity of the process $\{X_t^i\}$, we can extend the property of self-similarity to the increments process $\{X_{t+m}^i - X_m^i\}_{t\in\mathbb{R}}$. Indeed, using successively the weak-sense stationarity and weak-sense self-similarity, we obtain the weak-sense self-similarity of the process $\{X_{t+m}^i - X_m^i\}_{t\in\mathbb{R}}$:

$$\mathbb{V}[X_{t+m}^i - X_m^i] = \mathbb{V}[X_t^i] = \left|t\right|^{2H_i}\mathbb{V}[X_1^i] = \left|t\right|^{2H_i}\mathbb{V}[X_{1+m}^i - X_m^i], \tag{3.175}$$

or equivalently for all $\lambda > 0$:

$$\mathbb{V}[X_{\lambda t+m}^i - X_m^i] = \mathbb{V}[X_{\lambda t}^i] = \lambda^{2H_i}\mathbb{V}[X_t^i] = \lambda^{2H_i}\mathbb{V}[X_{t+m}^i - X_m^i]. \tag{3.176}$$

Combining Equation (3.165), which stems from the weak-sense stationarity of $\{X_{t+m}^i - X_m^i\}_{t\in\mathbb{R}}$, and the self-similarity of $\{X_t^i\}$ leads to the full determination of the covariance kernel of process $\{X_t^i\}$:

$$\mathrm{Cov}(X_t^i, X_s^i) = \frac{\mathbb{V}[X_1^i]}{2}\left(\left|t\right|^{2H_i} + \left|s\right|^{2H_i} - \left|t-s\right|^{2H_i}\right). \tag{3.177}$$

From this equation, we can obtain the strict-sense self-similarity of the Gaussian process $\{X_t^i\}$: $\mathrm{Cov}(X_{\lambda t}^i, X_{\lambda s}^i) = \left|\lambda\right|^{2H_i}\mathrm{Cov}(X_t^i, X_s^i)$. We note that, unlike the stationarity property, the only weak form of self-similarity is not sufficient to obtain its strict form, and an additional assumption, namely weak-form stationarity, is required.

Moreover, the strict-sense stationarity of the increments process and strict-sense self-similarity of the process itself allows obtaining the strict-sense self-similarity of the increments process:

$$\mathrm{Cov}(X_{\lambda t+m}^i - X_m^i, X_{\lambda s+m}^i - X_m^i) = \mathrm{Cov}(X_{\lambda t}^i, X_{\lambda s}^i) = \lambda^{2H_i}\mathrm{Cov}(X_t^i, X_s^i),\ \forall\lambda > 0. \tag{3.178}$$

The strict-sense self-similarity of the increments process establishes a spatial proportionality relationship of factor $\Theta^i(\lambda) = \lambda^{2H_i}, (\lambda > 0)$ between the characteristics, especially the variances and covariances, of the increments process taken at two proportional time-scales λt and t. Therefore, estimators of variances and covariances applied to the increments process taken at two proportional time-scales λt and t are linked by a spatial proportionality relationship of factor $\Theta^i(\lambda) = \lambda^{2H_i}$, which allows easy adaptation of the estimators to different time-scales. Thus, we obtain reliable estimations, coherent with the different time-scales.

As a practical application, the self-similarity property allows changing the time-scale in the variance of the increments process $\{X^i_{t+m} - X^i_m\}_{t\in\mathbb{R}}$. Indeed, let us assume that we have knowledge of the daily volatility of the increments process $\{X^i_{t+1} - X^i_t\}$ (where the time unit is the day) defined by $\sqrt{\mathbb{V}[X^i_{t+1} - X^i_t]}$ for all $t \in \mathbb{R}$, and that we want to obtain the annual volatility. Based on the assumption that the market is open 256 days a year, the self-similarity property establishes a proportionality relationship of factor $\Theta^i(256) = 256^{2H_i}$ between the annual variance and the daily variance of the Gaussian process $\{X^i_{t+m} - X^i_t\}$:

$$\sqrt{\mathbb{V}[X^i_{t+256} - X^i_t]} = 256^{H_i}\sqrt{\mathbb{V}[X^i_{t+1} - X^i_t]}. \tag{3.179}$$

A reference dealing with self-similar Gaussian processes with stationary increments is (Taqqu, 1994, Chap.7 - Def.7.1.7 p.314, Cor.7.2.3 p.320). Authors proved that a Gaussian process is self-similar with stationary increments if and only if it is an fBm whose auto-covariance function is given by Equation (3.177).

Consequently, the combination of weak-sense stationarity and weak-sense self-similarity is required to ensure the stability, reliability and coherence of the estimators on both time-origin translated and time-scaled data. Indeed, the combination of the weak-sense stationarity of the increments process with the weak-sense self-similarity of the process itself allows us to obtain both the strict-sense stationarity and the strict-sense self-similarity of the increments process. Combining the weak-sense stationarity and weak-sense self-similarity allows obtaining the invariance by time-origin change of the characteristics of the increments process and establishing a spatial proportionality relationship between the characteristics of the increments process taken at two proportional time-scales λt and t with a proportionality coefficient $\Theta^i(\lambda)$. Consequently, the characteristics of the increments process remain the same when the time-origin of the process changes, which insures the stability and the reliability of the estimators applied to the increments process. Moreover, the characteristics (and estimators) of the increments process taken at two proportional time-scales λt and t are linked by a spatial proportionality relationship of factor $\Theta^i(\lambda) = \lambda^{2H_i}$, then the estimations can easily be adapted to the time-scale and remain coherent with the time-scale.

3.5.3 Estimation method

Our predictive model involves the following parameters: $\sigma = (\sigma_1, \dots, \sigma_d)$ represents the vector of market volatility related to each asset, $H = (H_1, \dots, H_d)$ is the vector of Hurst exponents controlling the smoothness of the price trajectories of each asset, and $R = \{\rho_{ij}, i, j = 1, \dots d\}$ refers to the matrix of correlations between assets i and j.

Let us consider the sequence of observations $(X^1_{t_0}, \dots, X^1_{t_N}, \dots, X^d_{t_0}, \dots, X^d_{t_N})$ representing the log-price trajectories of the d assets. For ease of estimation, we assume that the observations are uniformly spaced in time such that $t_j = j\tau$, where τ is the minimal time step. We suppose that the log-price trajectories, and not the price trajectories, follow an mfBm. Consequently, log-price increments are Gaussian, self-similar and stationary. These properties are pivotal to insure the stability, reliability and coherence of the estimation method of mfBm parameters proposed by (Amblard and Coeurjolly, 2011a), that we apply to log-price trajectories. More precisely, we need the stationarity of the process $\{X^i_{t+m} - X^i_t\}_{t\in\mathbb{R}}$ and the self-similarity of the other increments process $\{X^i_{t+m} - X^i_m\}_{t\in\mathbb{R}}$.

The method of parameter estimation of mfBm parameters consists of computing the empirical variance of the log-price increments of several lengths for each component of mfBm, following the moment-based estimator of fBm introduced by (Istas and Lang, 1994; Kent and Wood, 1997). If the increments are disjoint and of duration $m\tau$ with $m \in \mathbb{N}^\star$, then their variance is defined as

$$V^i_m = \frac{1}{\left\lfloor N/m \right\rfloor} \sum_{j=0}^{\left\lfloor \frac{N}{m} \right\rfloor - 1} (X^i_{(j+1)m\tau} - X^i_{jm\tau})^2 \tag{3.180}$$

and if the increments are overlapping, then their variance is given by

$$V^i_m = \frac{1}{N-m+1} \sum_{j=0}^{N-m} (X^i_{(j+m)\tau} - X^i_{j\tau})^2. \tag{3.181}$$

The empirical variance of the log-price increments of duration $m\tau$ almost surely converges to the variance of the related fBm increments of duration $m\tau$ (Coeurjolly, 2001, 2000a):

$$V^i_m \longrightarrow \mathbb{V}[\delta_{m\tau} X^i_t] = (m\tau)^{2H_i} \mathbb{V}[X^i_1] \quad \text{a.s..} \tag{3.182}$$

The stationarity and self-similarity properties of the log-price process $\{X^i_t\}$ are pivotal for obtaining reliable and coherent parameter estimations. On the one hand, the stationarity property guarantees that log-price increments of the same duration might have the same distributional properties. This is a necessary condition to guarantee the stability of the estimators on time-origin translated data. On the other hand, the self-similarity property enables the establishment of a link between the variance of the increments at diverse time-scales. More precisely, the self-similarity property establishes a spatial proportionality relationship between the characteristics of the processes $\{X^i_{\lambda t}\}$ and $\{X^i_t\}$ for all $\lambda > 0$, with a proportionality factor $\Theta^i(\lambda) = \lambda^{2H_i}$. Therefore, the

self-similarity property ensures the reliability and the coherence of the estimators on data at different time-scales. Owing to self-similarity, regardless of the considered time scales, the estimations of the Hurst exponent and volatility remain the same, which means that the self-similarity property guarantees the stability of the estimators. If we consider the increments $\{\delta_{m_1\tau}X_t^i\}$ and $\{\delta_{m_2\tau}X_t^i\}$ of respective durations $m_1\tau$ and $m_2\tau$ with $m_1 > m_2$, then the almost sure convergence of the empirical variance leads to:

$$V_{m_1}^i \longrightarrow \mathbb{V}[\delta_{m_1\tau}X_t^i] = (m_1\tau)^{2H_i}\mathbb{V}[X_1^i] \quad \text{and} \quad V_{m_2\tau}^i \longrightarrow \mathbb{V}[\delta_{m_2\tau}X_t^i] = (m_2\tau)^{2H_i}\mathbb{V}[X_1^i] \quad \text{a.s..} \tag{3.183}$$

The self-similarity property allows us to establish the following relationship between the variance of the increments $\{\delta_{m_1\tau}X_t^i\}$ and $\{\delta_{m_2\tau}X_t^i\}$:

$$\mathbb{V}[\delta_{m_1\tau}X_t^i] = \left(\frac{m_1}{m_2}\right)^{2H_i} \mathbb{V}[\delta_{m_2\tau}X_t^i]. \tag{3.184}$$

Then, by the continuous mapping theorem, we have

$$\log\left(\frac{V_{m_1}^i}{V_{m_2}^i}\right) \longrightarrow 2H_i \log\left(\frac{m_1}{m_2}\right) \quad \text{a.s..} \tag{3.185}$$

Finally, we get a strongly consistent estimator for the Hurst exponent:

$$\hat{H}_i = \frac{1}{2\log\left(\frac{m_1}{m_2}\right)} \log\left(\frac{V_{m_1}^i}{V_{m_2}^i}\right). \tag{3.186}$$

In this two-scale framework, the volatility at time-scale m_1 is estimated using the square-root of $V_{m_1}^i$. Owing to the self-similarity property of the variance, annual volatility is estimated by:

$$\hat{\sigma}_i = \left(\frac{256}{m_1}\right)^{\hat{H}_i} \sqrt{V_{m_1}^i}. \tag{3.187}$$

This method, called the method of energy levels, is in fact a linear regression based on only two time-scales.

Alternatively to this method based on the observation of increments at two time scales $m_1\tau$ and $m_2\tau$, we can introduce an estimator based on a higher number of scales: $m_1\tau < ... < m_K\tau$, as in (Amblard and Coeurjolly, 2011a). Indeed, for each scale $m_k\tau$, we have, using the continuous mapping theorem:

$$\log(V_{m_k}^i) \longrightarrow 2H_i \log(m_k\tau) + \log(\mathbb{V}[X_1^i]) \quad \text{a.s..} \tag{3.188}$$

Then, we carry out a linear regression of the log-variances versus the log-time-scales. The estimation of the Hurst exponent is given by half the slope, and volatility by the exponential of half the intercept. The two estimation methods are almost equivalent.

Other estimators that are more robust when $H > 0.75$ are proposed in (Coeurjolly, 2001), a class of consistent estimators of the parameters of a fractional Brownian motion based on the asymptotic behavior of the $k-th$ absolute moment of discrete variations of

its sampled paths over a discrete grid of the interval $[0, 1]$. They derive explicit convergence rates for these types of estimators, valid through the entire range $0 < H < 1$ of the self-similarity parameter. They concluded that when the scaling coefficient of an fBm is known, and when the sample paths are discretized on the grid $\{0, 1/N, \ldots, N - 1/N\}$, they obtained estimates of the self-similarity parameter that are strongly consistent with the asymptotic rate of the order $1/\sqrt{N}\log(N)$ whatever the value of H is. In (Jean-franÇois Coeurjolly and Vidakovic, 2014), the authors discussed the estimation of a scaling parameter σ^2 when the Hurst exponent is known. To estimate σ^2, they proposed three approaches based on maximum likelihood estimation, moment-matching, and concentration inequalities, and discussed the theoretical characteristics of the estimators and optimal-filtering guidelines. They justified the improvement in the estimation of σ^2 when the Hurst parameter is known. Using the three approaches and a parametric bootstrap methodology in a simulation study, they compared the confidence intervals of σ^2 intervals of their lengths, coverage rates, and computational complexity and discussed the empirical attributes of the tested approaches. They found that the approach based on maximum likelihood estimation was optimal in terms of efficiency and accuracy, but was computationally expensive. The moment-matching approach was found to be not only comparably efficient and accurate, but also computationally fast and robust to deviations from the fractional Brownian motion model.

Estimators can be derived from the wavelet decomposition. In (Coeurjolly et al., 2013), authors study multivariate fractional Brownian motion (mfBm) viewed through the lens of the wavelet transform. They calculated the correlation structure of the wavelet transform of the mfBm. They studied the asymptotic behavior of the correlation, showing that if the analyzing wavelet has a sufficient number of null first order moments, the decomposition eliminates any possible long-range (inter)-dependence. The cross-spectral density was also considered. Its existence is proven and its evaluation is performed using a Von Bahr-Essen like representation of the function $sign(t)\left|t\right|^{\alpha}$. The behavior of the cross-spectral density of the wavelet field at zero frequency was also developed, and the results provided by the asymptotic analysis of the correlation were confirmed.

Several studies have dealt with the estimation of the Hurst exponent. The works of (Gatheral et al., 2022, Section 2) deal with the smoothness of the volatility process. Based on the assumption that log-volatility behaves essentially as a fractional Brownian motion, they estimate the Hurst exponent via a linear regression of the logarithm of the variance of the increments versus their log-time-scale. This estimation leads to a Hurst exponent of order 0.1 at any reasonable time-scale, and adopts the fractional stochastic volatility model of (Comte and Renault, 1998).

In (Chong et al., 2022a, p.12), (Chong et al., 2022b), and (Szymanski and Takabatake, 2023), based on the observation that the local behavior of stochastic volatility is much more irregular than semimartingales and resembles that of fractional Brownian motion with Hurst parameter $H < 0.5$, consistent and asymptotic estimators of H are developed based on high-frequency price observations. Moreover, minimax lower bounds for parameter estimation were established. The convergence rates of the estimator were studied, and the optimal convergence rates were determined. Results are extended to $H > 1/2$ so far. For $H > 0.75$, estimators that converge in the second order are developed.

3.6 Method for backtesting

We consider an information set $\mathfrak{I}_{t-1} = \{\mathcal{I}_{t-1}, \mathcal{I}_{t-2}, \mathcal{I}_{t-3}, \ldots, \mathcal{I}_1\}$ which consists of a sequence of indicator functions, where the hit variable $\mathcal{I}_t$ at time t is a Bernoulli variable equal to 1 if the return between times t and $t+h$ exceeds the $\mathbf{VaR}_\alpha$ predicted at time t, and 0 otherwise. More precisely, we obtain the hit variables using a rolling window approach.

For each window of size w, we predict at time t the conditional **VaR** given the history of the window $[t-w, t]$, and we compare it to the log-return realized between times t and $t+h$, thus generating a hit variable equal to either 1 or 0.

Following traditional literature on **VaR** backtesting (Christoffersen, 1998; Davis, 2016), a satisfying **VaR** must be such that $\mathbb{E}[\mathcal{I}_t|\mathfrak{I}_{t-1}] = \alpha$. Equivalently, we evaluate two criteria on the generated **VaR**s: the conditional coverage and the independence of the hit variables, namely, it has to be shown that $\{\mathcal{I}_t\} \overset{iid}{\sim} \mathcal{B}ern(\alpha)$.

In practice, the coverage test consists of estimating the conditional expectation of the sequence of hit variables that should correspond to the risk level α. This estimation was performed by empirically averaging the hit variables. The closer the empirical mean is to α, the more satisfying the conditional coverage test is.

Regarding the independence, we propose a method inspired by the extreme value theory and first put forward by (Bücher et al., 2020). Indeed, two extreme events become approximately independent if they are separated to each other by enough time. In order to measure the degree of dependence in a time series, we use the extremal index.

The extremal index θ is an indicator that quantifies the degree of dependence in a time sequence of random variables, by counting the number of clusters of variables above a predetermined threshold, namely extreme values. It is equal to one if the sequence is independent, which means that there is no cluster of extreme values. The closer the extremal index is to 0, the more numerous the clusters of extreme values are, and the larger the serial dependence in the sequence.

In the context of **VaR** backtesting, we intend to apply this method to the sequence of hit variables, considering that an extreme value corresponds to a hit equal to one. Therefore, the extremal index quantifies the presence of clusters of **VaR** violations.

An estimator of the extremal index can be provided by performing either a block or a run declustering algorithm. More precisely, given an arbitrary block size b, we partition the sequence of n successive hit variables into $k = \left\lfloor \frac{n}{b} \right\rfloor$ non-overlapping blocks (block declustering) or $n-b$ running windows (run declustering). In the block-declustering approach, each block containing nonzero hit variables represents a cluster. In the run-declustering approach, we count the number of windows without a nonzero hit, starting at the end of a cluster. With both methods, the estimator of the extremal index is then the ratio of the estimated number of clusters to the number of nonzero hit variables.

Formally, if we note $M_{i,j} = \max\{\mathcal{I}_{i+1}, \dots, \mathcal{I}_j\}$, the two estimators are defined as follows:

$$\hat{\theta}_n^B(b) = \frac{\sum_{i=1}^{k} \mathbb{1}_{\{M_{(i-1)b,ib}=1\}}}{\sum_{i=1}^{kb} \mathbb{1}_{\{\mathcal{I}_i=1\}}} \qquad \hat{\theta}_n^R(b) = \frac{\sum_{i=1}^{n-b} \mathbb{1}_{\{\mathcal{I}_i=1, M_{i,i+b}=0\}}}{\sum_{i=1}^{n-b} \mathbb{1}_{\{\mathcal{I}_i=1\}}}. \tag{3.189}$$

3.7 Simulation study

We simulate the increments of a univariate fractional Brownian motion, for a duration of two years, with a daily frequency (the day is assumed to be the smallest time-step). As discussed in Section 3.4, the simulations are performed using the Cholesky method. The trajectory is then assimilated to the trajectory of the log-price increments of a given asset.

The trajectory was split into two subsets. The first year of data is used to estimate the parameters H and σ of fBm, and the second year is used to run the predictive model.

3.7.1 In one dimension

Distribution of the estimated parameters

The distributions of the estimators of the Hurst exponent and volatility parameter are studied for different values of H and σ through histograms. The distributions of the estimated Hurst exponent (H) and volatility (σ) are shown in Figure 3.1.

The histograms representing the distributions of $\hat{H}$ and $\hat{\sigma}$ visually confirm that the estimators are unbiased for any parameter value and that the variance of the estimator of the Hurst exponent is higher when H is lower.

Convergence of the estimators

Two types of graphs of convergence are provided.

To build the first type of graphs of convergence, trajectories of fBm increments with various lengths are simulated, and the Hurst exponent is estimated for each of them. The graph displays the estimation of the Hurst exponent as a function of trajectory length.

To build the second type of graphs of convergence, we simulate a single trajectory of fBm increments and split it into several trajectories of diverse lengths. Then, the Hurst exponent is estimated for each of them. The graph displays the estimation of the Hurst exponent as a function of subset length.

Figures 3.2 and 3.3 make it possible to appreciate the accuracy of $\hat{H}$ and $\hat{\sigma}$ with respect to the length of the trajectory, using the first approach. Figures 3.4 and 3.5 answer the same question following the second approach.

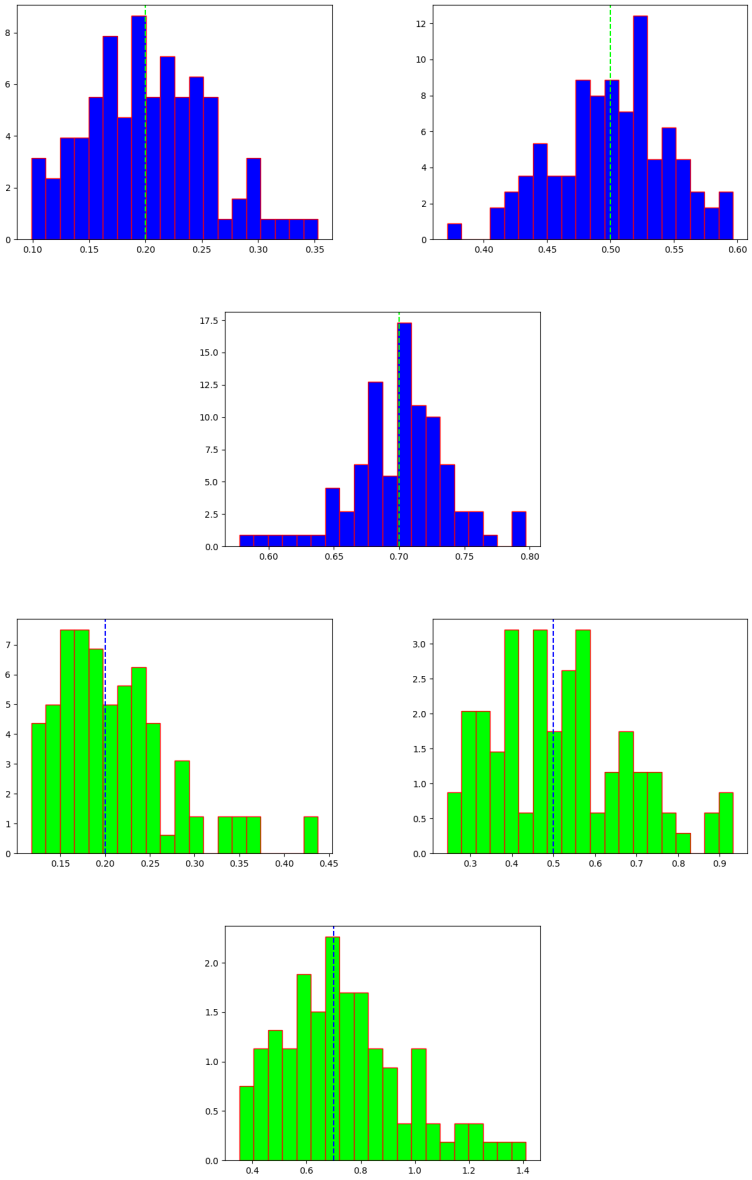


Figure 3.1: Distribution of estimated Hurst exponents (first three histograms, in blue) for various values of H (0.2, 0.5, and 0.7) and $n = 256$ observations for each of the 100 trajectories, using the method of the moments of energy levels. Distribution of estimated volatility $\hat{\sigma}$ (last three histograms, in green) for various values of σ (0.2, 0.5, and 0.7), $H = 0.2$ and $n = 256$ observations for each of the 100 trajectories.

The graphs of convergence of the estimators of H and σ show that the estimators both satisfy the convergence in probability and almost-sure convergence.

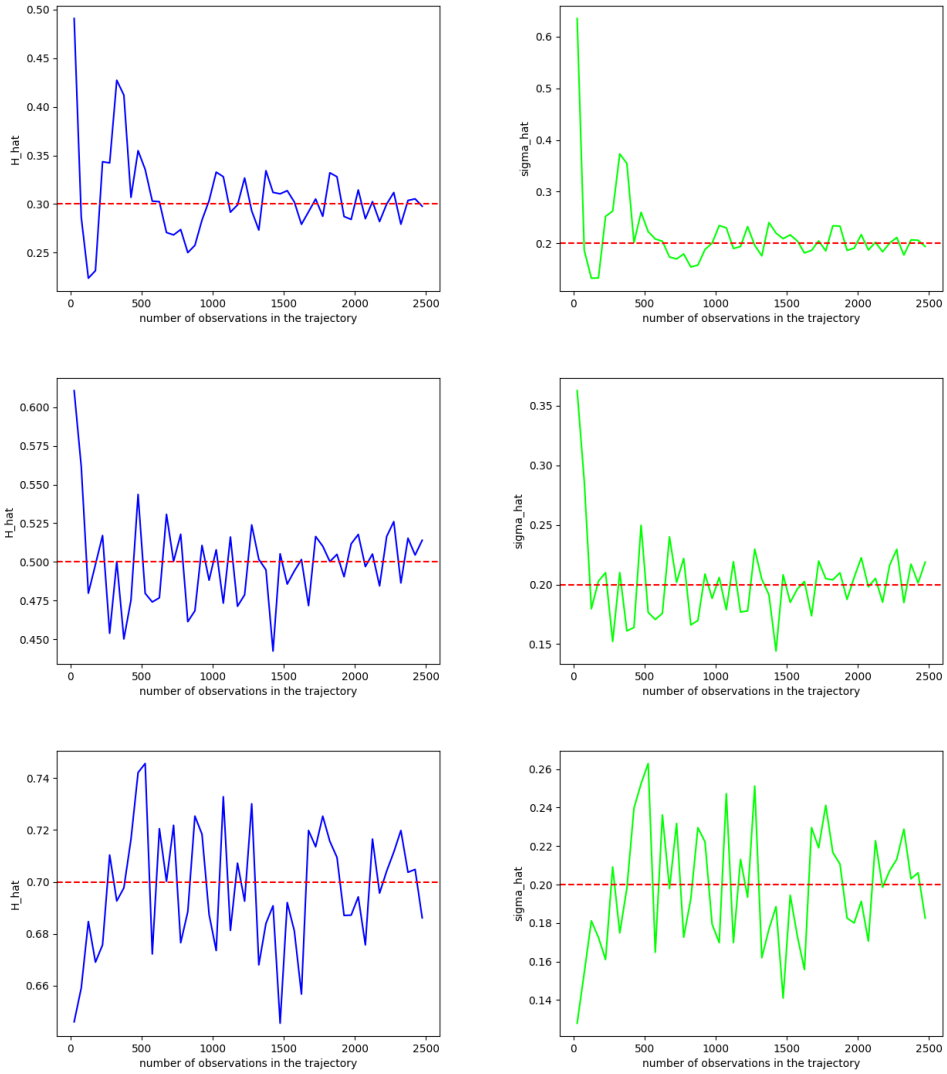


Figure 3.2: Several trajectories of different lengths (from 25 to 2500 with a step of 50), are simulated and on each of them, the parameters H and σ are estimated thanks to the method of moments of energy levels. Here, the true volatility is fixed and equal to 0.2 whereas the Hurst exponent is varying (from top to bottom : 0.3, 0.5, and 0.7).

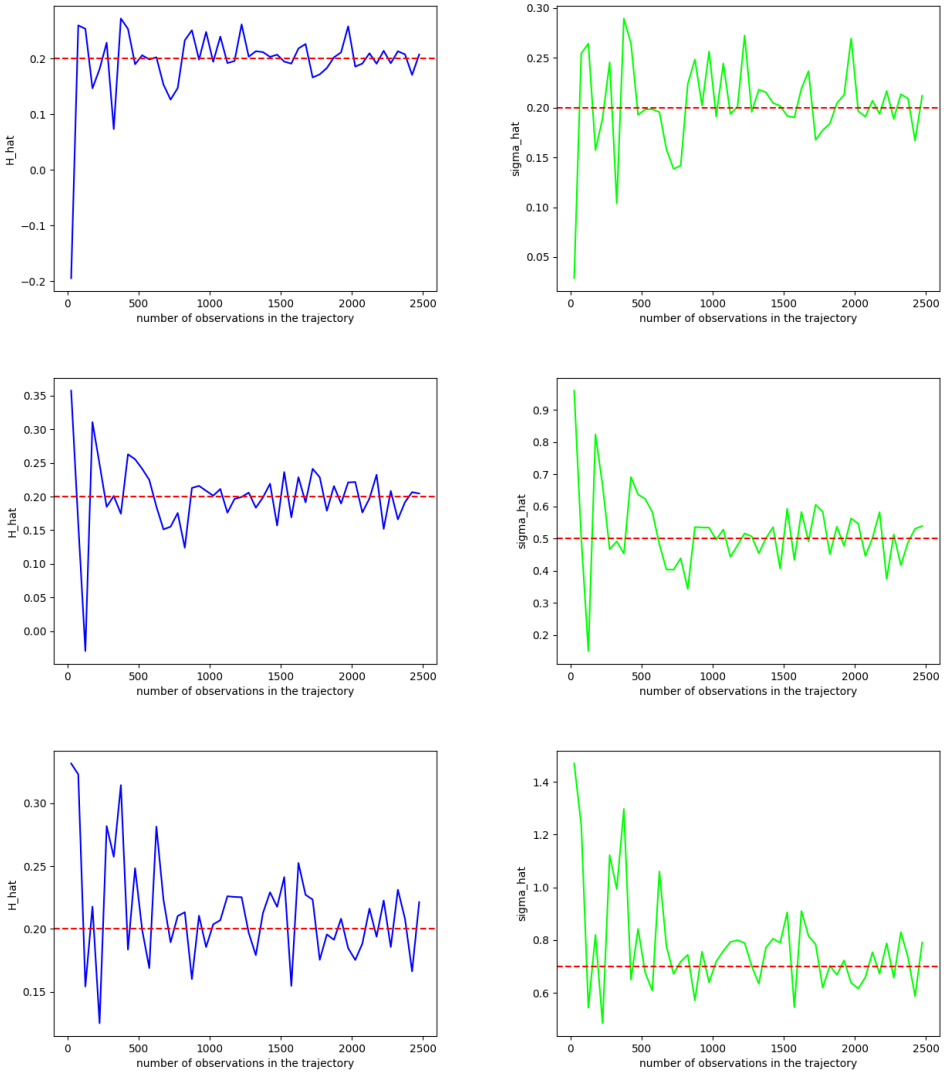


Figure 3.3: Several trajectories of different lengths (from 25 to 2500 with a step of 50), are simulated and on each of them, the parameters H and σ are estimated thanks to the method of moments of energy levels. Here, the true Hurst exponent is fixed and equal to 0.2 whereas the volatility parameter is varying (from top to bottom : 0.2, 0.5, and 0.7).

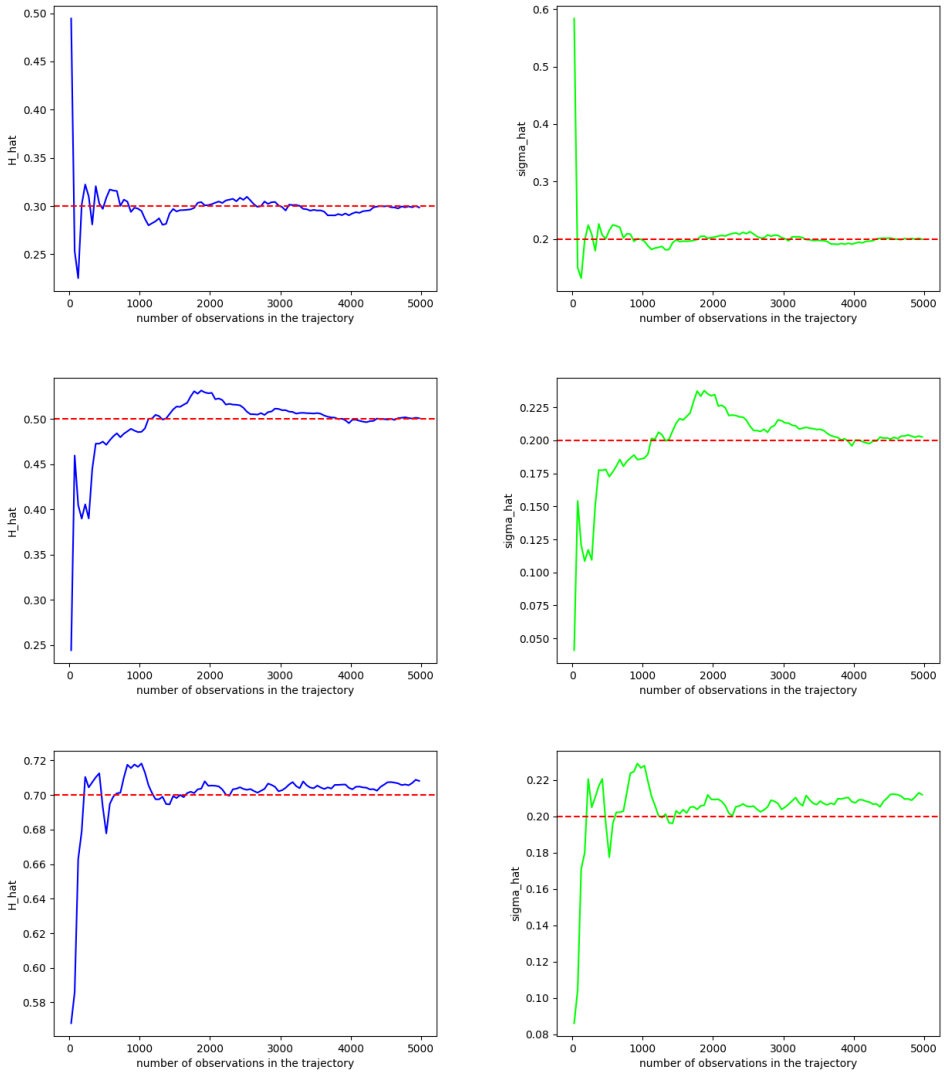


Figure 3.4: A single trajectory is simulated and subsamples of different lengths (from 25 to 5000 with a step of 50) are extracted. On each of them, the parameters H and σ are estimated thanks to the method of moments of energy levels. Here, the true volatility is fixed and equal to 0.2 whereas the Hurst exponent is varying (from top to bottom : 0.3, 0.5, and 0.7).

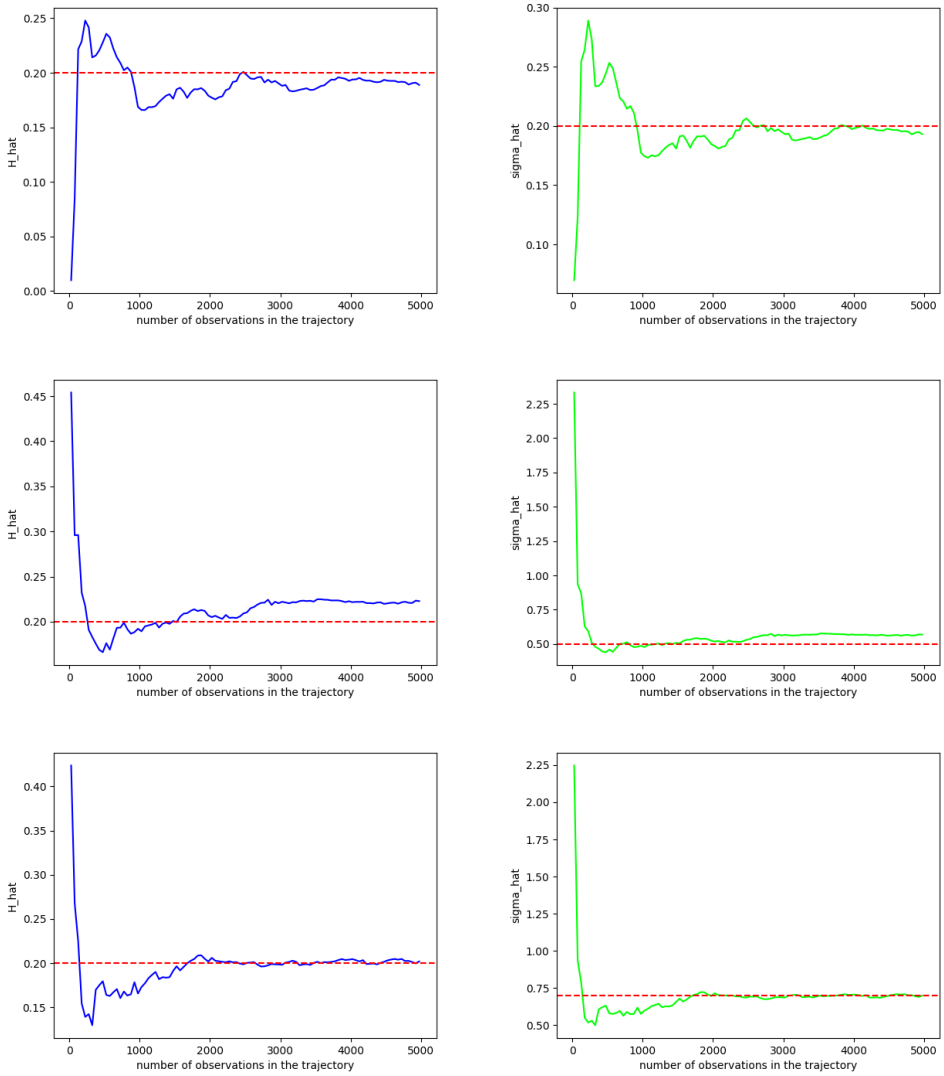


Figure 3.5: A single trajectory is simulated and subsamples of different lengths (from 25 to 5000 with a step of 50) are extracted. On each of them, the parameters H and σ are estimated thanks to the method of moments of energy levels. Here, the true Hurst exponent is fixed and equal to 0.2 whereas the volatility parameter is varying (from top to bottom : 0.2, 0.5, and 0.7).

Backtesting in univariate framework

To backtest the **VaR** model, we proceed as follows. First, we simulate a trajectory of an fBm increments of parameters H and σ, over two years (two times 256 business days) with a daily frequency. Using the parameters estimated on the first year of data, we dynamically predict, thanks to a rolling window procedure, the **VaR** at the risk level $\alpha = 1\%$ at a time horizon of $h = 1$ day after the end of each window.

We compare three **VaR** models: conditional **VaR**, Gaussian **VaR** (we force H to be equal to 0.5), and empirical **VaR**. The three **VaR** models are calculated according to three different cases: the first corresponds to the conditional **VaR** of the future increment of fBm, the second corresponds to the conditional **VaR** of the future true portfolio variation, and the third corresponds to the conditional **VaR** of the first Gaussian approximation of the future portfolio variation introduced in Section 3.3.

The best model is that for which the coverage rate $\hat{\alpha}$ is the closest to the risk level α, the extremal index is the closest to 1, and the average **VaR** is as small as possible. Indeed, the main stake is to predict the amount of money to keep aside, as accurately as possible, so that the financial institution might, in the meantime, be solvent and be able to guarantee a financial activity to generate benefits. Moreover, for a given coverage rate, we prefer a low **VaR** because it maximizes the financial institution's profit. Finally, a **VaR** with an extremal index close to 1 is encouraged because it means that the clusters of hits are independent.

We start by studying the first case, namely the conditional **VaR** of the future fBm increment. The results of backtesting are shown in Figure 3.6. In particular, the series of simulated losses is compared with the three series of **VaR**s. Because the Gaussian **VaR** is unconditional with the parameters estimated on the first half of the dataset, it remains constant in the graphs. The other two **VaR**s fluctuate and seem to reflect variations in price returns.

Beyond the visual evaluation of the three **VaR**s in Figure 3.6, we provide a more precise analysis of these risk models in Table 3.1, which focuses on the coverage of each **VaR**, and in Table 3.2, which gathers the independence statistics of the hit clusters. In each Table, for each line corresponding to a particular Hurst exponent and size of the conditioning window, we bold the most desirable values, that is, the hit probability closest to 0.01, the lowest average **VaR**, and the highest extremal index.

The best coverage rate (i.e., the closest to the target risk level) is, in most cases, the coverage rate of the **VaR** model, that is, the **VaR** based on fBm. The lowest average **VaR** is almost always the average empirical **VaR**; however, this does not mean that the empirical **VaR** is a proper **VaR** because its coverage rate is far from the target risk level. However, this does not mean that empirical **VaR** is a bad approach. Indeed, the other two models have few parameters and thus are adapted to conditioning windows with small sizes, whereas the empirical **VaR**, as a non-parametric approach, requires a much larger window to be relevant. Apart from the empirical **VaR** model, the **VaR** model is on average almost always smaller than the Gaussian **VaR**. Moreover, independence criteria of the clusters of hits is very bad for the empirical **VaR**, but very close to one for

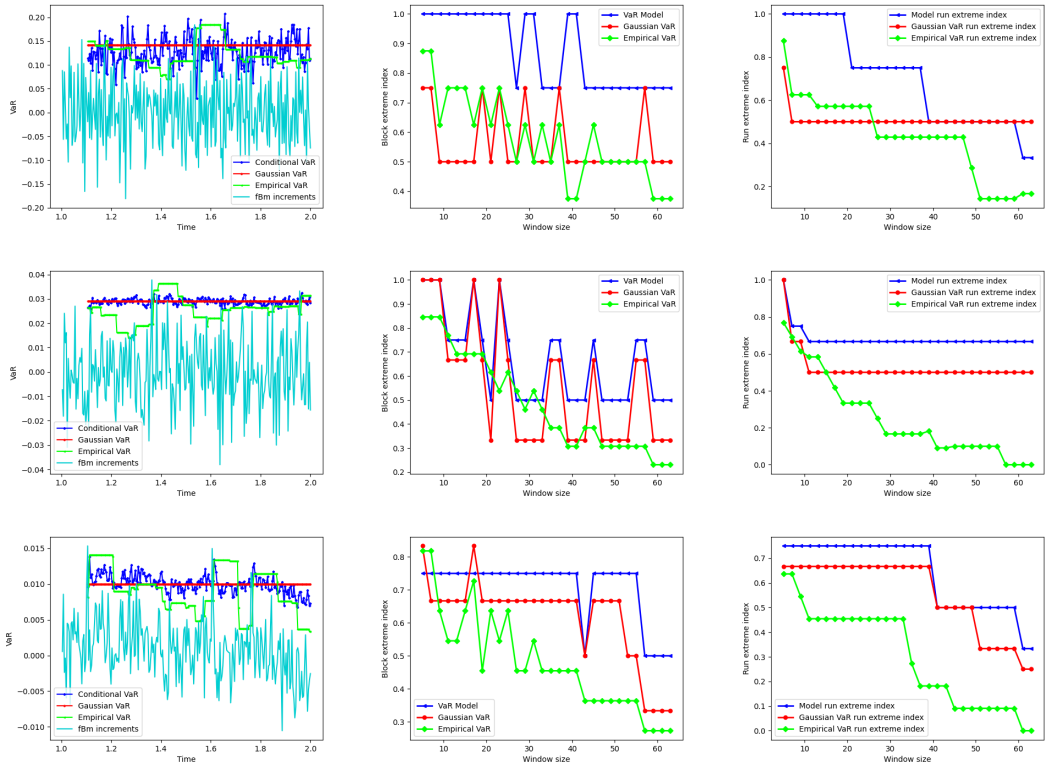


Figure 3.6: Backtesting of the conditional **VaR** of the future fBm's increment with a conditioning window of 25 days. The first column represents the trajectory of several models of **VaR** (conditional **VaR** of the future fBm's increment, Gaussian **VaR** and empirical **VaR**) in function of time, the second (respectively third) column represents the extremal index obtained by run (resp. block) declustering in function of the size of the block. The simulated trajectories have been generated with a unique $\sigma = 0.2$, and various H. The first row corresponds to $H = 0.2$, the second row corresponds to $H = 0.5$, and the third row corresponds to $H = 0.7$.

the **VaR** model or for the Gaussian **VaR**. This demonstrates the independence of the clusters of hits of the **VaR** model and Gaussian **VaR**. In conclusion, the **VaR** model seems to be the most relevant for these simulations, according to the three evaluated criteria. This is not surprising because the data generating process is consistent with the **VaR** model.

We also study the second case cited above, namely the conditional **VaR** of the future true portfolio variation. In Figure 3.7, we gather the graphs showing the portfolio variations along with the corresponding **VaR**s and extremal indexes. In this figure, the Gaussian **VaR** is no longer constant. Indeed, the future increment of the true portfolio, in the univariate framework, corresponds exactly to the increment of a log-normal random variable $S^i_{t_{N+1}} - S^i_{t_N} = S^i_{t_N}(e^{\delta_h X^i_{t_N}} - 1)$ and depends on the past, unlike the future increment of the standard Brownian motion.

H	w	$\hat{\alpha}^m$	$\hat{\alpha}^G$	$\hat{\alpha}^e$	$p_N^{m/e}$	$p_N^{m/G}$	$\overline{\mathbf{VaR}_\alpha^m}$	$\overline{\mathbf{VaR}_\alpha^G}$	$\overline{\mathbf{VaR}_\alpha^e}$
0.2	25	**0.0108**	0.0078	0.0377	1.0	0.4	0.1391	0.1536	**0.1214**
	50	**0.0102**	0.0107	0.0243	1.0	0.4	**0.1375**	0.1530	0.1397
	100	**0.0090**	0.0051	0.0198	0.5	0.6	0.1366	0.1560	**0.1330**
	200	**0.0125**	**0.0125**	0.0179	0.1	0.0	**0.1374**	0.1510	0.1490
0.5	25	**0.0104**	0.0108	0.0429	1.0	0.1	**0.0288**	0.0289	**0.0228**
	50	**0.0102**	**0.0102**	0.0257	0.5	0.0	0.0284	0.0285	**0.0247**
	100	0.0096	**0.0103**	0.0167	0.3	0.0	0.0290	0.0291	**0.0277**
	200	**0.0089**	0.0071	0.0143	0.4	0.1	0.0285	0.0284	**0.0263**
0.7	25	**0.0100**	0.0069	0.0429	1.0	0.6	0.0090	0.0095	**0.0072**
	50	**0.0092**	0.0087	0.0296	0.8	0.2	0.0087	0.0094	**0.0078**
	100	**0.0109**	0.0147	0.0301	0.9	0.8	0.0091	0.0097	**0.0084**
	200	**0.0125**	**0.0125**	**0.0125**	0.1	0.2	0.0085	0.0094	**0.0083**

Table 3.1: For the case of future fBm's increment, average estimated hit probability for the **VaR** model ($\hat{\alpha}^m$), the Gaussian **VaR** ($\hat{\alpha}^G$), and the empirical **VaR** ($\hat{\alpha}^e$), on $N = 10$ simulated time series of length 256 where each **VaR** is calculated for a risk level of $\alpha = 1\%$ and with a conditioning window of size w. We also include the probability for the model to be more accurate than the Gaussian **VaR**, $p_N^{m/G} = \frac{1}{N}\sum_{i=1}^N \mathbb{1}_{\{|\hat{\alpha}_i^m - \alpha| < |\hat{\alpha}_i^G - \alpha|\}}$, and than the empirical **VaR**, $p_N^{m/e} = \frac{1}{N}\sum_{i=1}^N \mathbb{1}_{\{|\hat{\alpha}_i^m - \alpha| < |\hat{\alpha}_i^e - \alpha|\}}$. Finally, we add the average **VaR** for each model.

Tables 3.3 and 3.4 show that, in this case, the **VaR** based on the fBm has better backtesting results than the unconditional Gaussian **VaR** and the empirical **VaR**, at least for the small conditioning windows studied.

The last case we study is the conditional **VaR** of the first Gaussian approximation of the future portfolio variation. Figure 3.8 and Tables 3.5 and 3.6 display the results. The conclusions are the same as those for the two previous cases and are thus in favor of the **VaR** based on fBm, even though the violation rates for the Gaussian **VaR** are very similar to those of the **VaR** based on fBm.

In Figure 3.8, the Gaussian **VaR** is one of the future increments of the standard Brownian motion weighted by the current asset price. The product of the Gaussian random variable with the log-normal random variable depends on the past. Therefore, this **VaR** varies over time.

In Figure 3.9 and Figure 3.10, the box plots represent the distribution of the VaR (model, Gaussian, empirical) for $H = 0.2$, $H = 0.5$, $H = 0.7$ and for four distinct window's sizes (25, 50, 100, 200). In Figure 3.9, we can observe that the empirical VaR is, on average, smaller than those of the VaR model and Gaussian VaR, except for some large window's sizes in the case $H = 0.2$. Nevertheless, in Figure 3.10, we observe that the coverage rate of the empirical **VaR** is far from the desired risk level, α. The VaR model and Gaussian VaR exhibit similar performances, and the standard deviation of the VaR model is lower than that of the Gaussian VaR for small rolling windows.

H	w	$\hat{\theta}^{m,R}$	$\hat{\theta}^{G,R}$	$\hat{\theta}^{e,R}$	$p_N^{m/e,R}$	$p_N^{m/G,R}$	$\hat{\theta}^{m,B}$	$\hat{\theta}^{G,B}$	$\hat{\theta}^{e,B}$	$p_N^{m/e,B}$	$p_N^{m/G,B}$
0.20	25	**0.78**	0.71	0.47	1.0	0.5	**0.90**	0.87	0.74	0.7	0.4
	50	0.43	**0.48**	0.03	0.7	0.4	0.57	**0.68**	0.38	0.7	0.3
	100	**0.20**	0.00	0.00	0.2	0.1	**0.45**	0.42	0.21	0.7	0.1
0.5	25	0.80	**0.83**	0.49	0.7	0.0	0.86	**0.88**	0.64	0.9	0.0
	50	0.58	**0.59**	0.07	0.9	0.0	0.60	**0.66**	0.37	0.9	0.0
	100	**0.52**	0.42	0.00	0.6	0.1	**0.65**	0.53	0.22	0.9	0.3
0.7	25	**0.73**	0.51	0.42	0.8	0.5	**0.85**	0.67	0.57	0.9	0.4
	50	0.52	**0.57**	0.06	0.7	0.2	0.57	**0.67**	0.37	0.6	0.2
	100	**0.53**	0.42	0.00	0.7	0.4	**0.69**	0.49	0.18	0.8	0.5

Table 3.2: For the case of future fBm's increment, average estimated extremal index obtained by run and block declustering, for the conditional **VaR** of the future fBm's increment ($\hat{\theta}^m$), the Gaussian **VaR** ($\hat{\theta}^G$), and the empirical **VaR** ($\hat{\theta}^e$), on $N = 10$ simulated time series of length 256 where each **VaR** is computed at the risk level of $\alpha = 1\%$ and with a conditioning window of size w. We also include the probability for the model to have a higher extremal index than empirical **VaR**, $p_N^{m/e} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{\{\hat{\theta}_i^m > \hat{\theta}_i^e\}}$, and than the Gaussian **VaR**, $p_N^{m/G} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{\{\hat{\theta}_i^m > \hat{\theta}_i^G\}}$.

H	w	$\hat{\alpha}^m$	$\hat{\alpha}^G$	$\hat{\alpha}^e$	$p_N^{m/e}$	$p_N^{m/G}$	$\overline{\mathbf{VaR}_\alpha^m}$	$\overline{\mathbf{VaR}_\alpha^G}$	$\overline{\mathbf{VaR}_\alpha^e}$
0.2	25	**0.010**	0.009	0.042	0.1	0.3	0.139	0.154	**0.130**
	50	0.013	**0.012**	0.030	0.8	0.2	0.134	0.149	**0.132**
	100	**0.008**	0.006	0.026	0.6	0.5	**0.138**	0.153	0.146
	200	**0.016**	0.021	0.027	0.5	0.2	**0.131**	0.149	0.146
0.5	25	0.016	**0.015**	0.048	0.9	0.0	0.029	0.029	**0.024**
	50	**0.008**	0.006	0.038	0.9	0.2	0.029	0.029	**0.023**
	100	**0.011**	0.013	0.021	0.6	0.1	0.029	0.029	**0.028**
	200	**0.016**	**0.016**	0.021	0.1	0.1	**0.028**	**0.028**	**0.028**
0.7	25	**0.010**	0.013	0.047	0.1	0.5	0.009	0.009	**0.008**
	50	**0.017**	0.020	0.023	0.3	0.4	**0.009**	**0.009**	0.010
	100	0.013	**0.010**	0.026	0.5	0.3	**0.009**	0.010	**0.009**
	200	0.007	**0.009**	0.021	0.4	0.0	**0.009**	**0.009**	0.011

Table 3.3: For the case of future future true portfolio variation, average estimated hit probability for the **VaR** model ($\hat{\alpha}^m$), the Gaussian **VaR** and the empirical **VaR** ($\hat{\alpha}^e$), on $N = 10$ simulated time series of length 256 where each **VaR** is calculated for a risk level of $\alpha = 1\%$ estimated using a window of size w. We also include the probability for the model to be more accurate than the Gaussian **VaR**, $p_N^{m/G} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{\{|\hat{\alpha}_i^m - \alpha| < |\hat{\alpha}_i^G - \alpha|\}}$, and than the empirical **VaR**, $p_N^{m/e} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{\{|\hat{\alpha}_i^m - \alpha| < |\hat{\alpha}_i^e - \alpha|\}}$. Finally, we add the average **VaR** for each model.

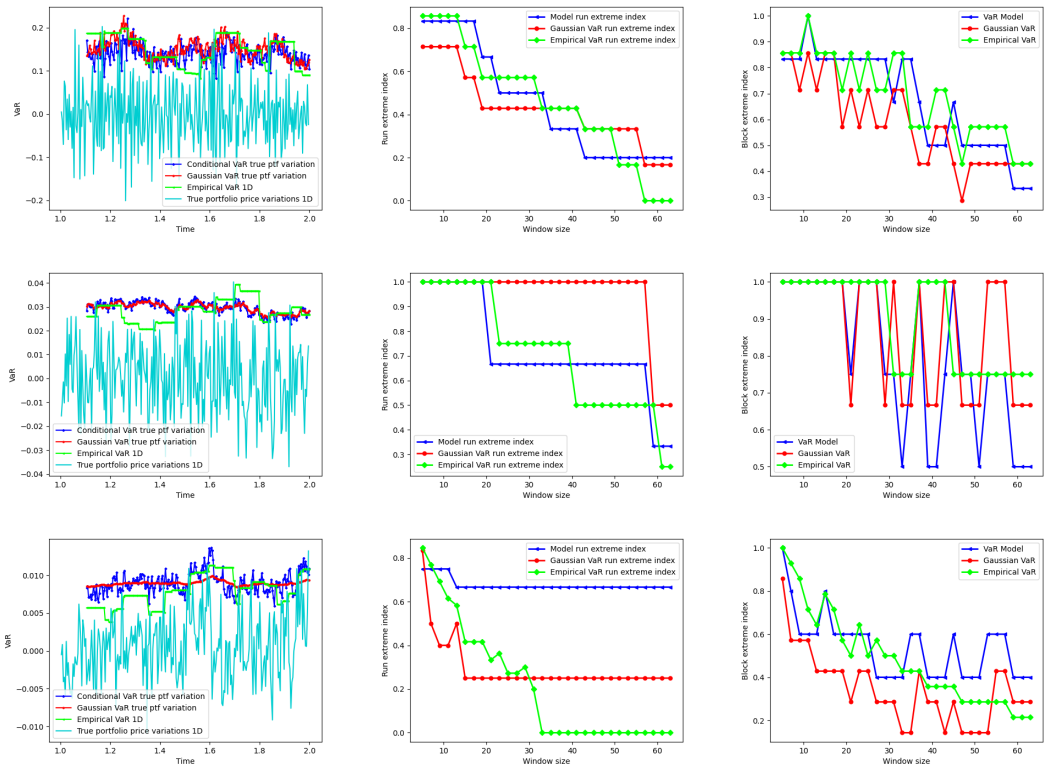


Figure 3.7: Backtesting of the conditional **VaR** of the future true portfolio variation with a conditioning window of 25 days. The first column represents the trajectory of several models of **VaR** (conditional **VaR** of the future fBm's increment, Gaussian **VaR** and empirical **VaR**) in function of time, the second (respectively third) column represents the extremal index obtained by run (resp. block) declustering in function of the size of the block. The simulated trajectories have been generated with a unique $\sigma = 0.2$, and various H. The first row corresponds to $H = 0.2$, the second row corresponds to $H = 0.5$, and the third row corresponds to $H = 0.7$.

In Figure 3.11, the estimated extremal index $\hat{\Theta}$ is represented for three VaR (model, Gaussian and empirical) and for three different Hurst exponents $H = 0.2$, $H = 0.5$ and $H = 0.7$, as a function of the window size used for backtesting. We can observe that the extremal indices of the VaR model and Gaussian VaR are very close to each other and very close to 1 for small window sizes, whereas they considerably decrease for large window sizes because there are fewer historical data. This means that the VaR model and Gaussian VaR present a small number of clusters of hits when the window size is small. In contrast, the estimated extremal index of the empirical VaR is much smaller, meaning that it presents numerous clusters of hits.

These results are supported by the following box plots.

H	w	$\hat{\theta}^{m,R}$	$\hat{\theta}^{G,R}$	$\hat{\theta}^{e,R}$	$p_N^{m/e,R}$	$p_N^{m/G,R}$	$\hat{\theta}^{m,B}$	$\hat{\theta}^{G,B}$	$\hat{\theta}^{e,B}$	$p_N^{m/e,B}$	$p_N^{m/G,B}$
0.2	25	**0.57**	0.49	0.38	0.6	0.4	**0.72**	0.66	0.60	0.6	0.2
	50	**0.63**	0.58	0.10	0.9	0.4	0.64	**0.74**	0.35	0.8	0.1
	100	0.33	**0.35**	0.00	0.4	0.2	**0.63**	0.56	0.16	1.0	0.3
0.5	25	**0.94**	0.92	0.41	1.0	0.1	**0.91**	0.89	0.64	0.9	0.1
	50	**0.78**	0.76	0.08	0.9	0.1	**0.86**	0.82	0.35	0.9	0.2
	100	**0.20**	0.20	0.00	0.3	0.0	**0.47**	**0.47**	0.21	0.8	0.0
0.7	25	**0.70**	0.50	0.43	0.8	0.6	**0.87**	0.52	0.58	0.8	0.7
	50	**0.33**	0.27	0.10	0.4	0.2	**0.43**	0.38	0.35	0.6	0.3
	100	**0.48**	0.30	0.00	0.6	0.3	**0.71**	0.45	0.18	1.0	0.6

Table 3.4: For the case of future future true portfolio variation, average estimated extremal index obtained by run and block declustering, for the conditional **VaR** of the future true portfolio's increment ($\hat{\theta}^m$), the Gaussian **VaR** ($\hat{\theta}^G$), and the empirical **VaR** ($\hat{\theta}^e$), on $N = 10$ simulated time series of length 256 where each **VaR** is computed at the risk level of $\alpha = 1\%$ and estimated using a window of size w. We also include the probability for the model to be more independent than empirical **VaR**, $p_N^{m/e} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{\{\hat{\theta}_i^m > \hat{\theta}_i^e\}}$, and than the Gaussian **VaR** $p_N^{m/G} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{\{\hat{\theta}_i^G > \hat{\theta}_i^G\}}$.

H	w	$\hat{\alpha}^m$	$\hat{\alpha}^G$	$\hat{\alpha}^e$	$p_N^{m/e}$	$p_N^{m/G}$	$\overline{\mathbf{VaR}_\alpha^m}$	$\overline{\mathbf{VaR}_\alpha^G}$	$\overline{\mathbf{VaR}_\alpha^e}$
0.2	25	**0.017**	0.017	0.047	1.0	0.5	0.138	0.156	**0.127**
	50	0.021	**0.019**	0.024	0.3	0.3	**0.136**	0.155	0.150
	100	0.015	**0.011**	0.027	0.7	0.1	**0.138**	0.160	**0.138**
	200	**0.020**	**0.020**	**0.020**	0.6	0.7	**0.135**	0.156	0.158
0.5	25	**0.010**	**0.010**	0.046	1.0	0.0	0.029	0.029	**0.023**
	50	**0.013**	**0.013**	0.040	1.0	0.0	0.029	0.029	**0.024**
	100	0.013	**0.011**	0.029	0.8	0.1	0.029	0.029	**0.026**
	200	**0.009**	0.011	0.007	0.1	0.0	**0.027**	**0.027**	0.030
0.7	25	**0.020**	**0.020**	0.045	0.9	0.7	0.009	0.009	**0.008**
	50	**0.009**	0.008	0.008	0.8	0.1	0.009	0.010	**0.008**
	100	**0.013**	**0.013**	0.018	0.4	0.6	**0.009**	0.010	**0.009**
	200	0.013	**0.009**	0.018	0.6	0.5	**0.010**	**0.010**	**0.010**

Table 3.5: For the case of the first Gaussian approximation of the future portfolio variation, average estimated hit probability for the **VaR** model ($\hat{\alpha}^m$), the Gaussian **VaR** and the empirical **VaR** ($\hat{\alpha}^e$), on $N = 10$ simulated time series of length 256 where each **VaR** is calculated for a risk level of $\alpha = 1\%$ estimated using a window of size w. We also include the probability for the model to be more accurate than the Gaussian **VaR**, $p_N^{m/G} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{\{|\hat{\alpha}_i^m - \alpha| < |\hat{\alpha}_i^G - \alpha|\}}$, and than the empirical **VaR**, $p_N^{m/e} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{\{|\hat{\alpha}_i^m - \alpha| < |\hat{\alpha}_i^e - \alpha|\}}$. Finally, we add the average **VaR** for each model.

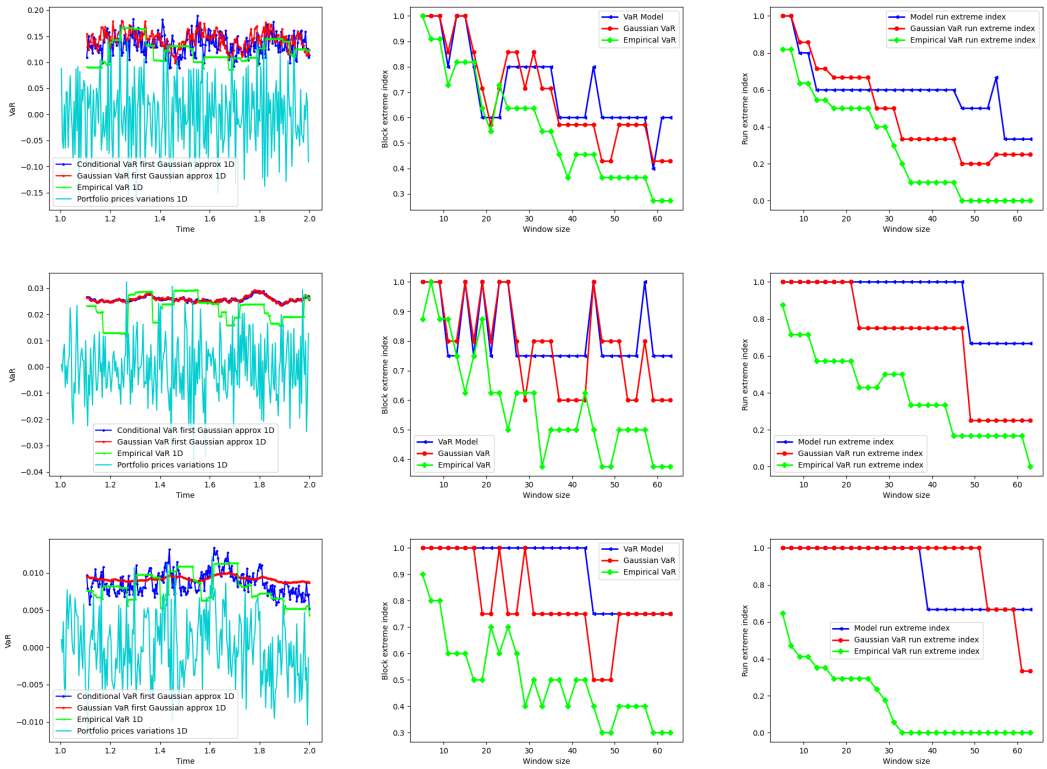


Figure 3.8: Backtesting of the conditional **VaR** of the first Gaussian approximation of the future portfolio variation with a conditioning window of 25 days. The first column represents the trajectory of several models of **VaR** (conditional **VaR** of the future fBm's increment, Gaussian **VaR** and empirical **VaR**) in function of time, the second (respectively third) column represents the extremal index obtained by run (resp. block) declustering in function of the size of the block. The simulated trajectories have been generated with a unique $\sigma = 0.2$, and various H. The first row corresponds to $H = 0.2$, the second row corresponds to $H = 0.5$, and the third row corresponds to $H = 0.7$.

In Figure 3.12 and in Figure 3.13, for each of the three VaRs (model, Gaussian and empirical) and for each of the four portfolios, the box plots respectively represent the distribution of the VaR and of the estimated coverage rate $\hat{\alpha}$, for four different window sizes (25, 50, 100, 200). In Figure 3.12, the box plots show that the empirical VaR is, on average, always lower than the VaR model and Gaussian VaR. However, the empirical VaR does not present the best performance because in Figure 3.13, we can observe that, for the first three portfolios whose Hurst exponents are homogeneous, the empirical model always has a high coverage rate far from the desired one (0.01), whereas the VaR model and the Gaussian VaR are more competitive. For the last portfolio whose Hurst exponents are heterogeneous, we can observe that the best VaR in terms of coverage rate is the Gaussian VaR because its coverage rate is the closest to 0.01, then we have the empirical VaR, whereas the VaR model reaches its limits.

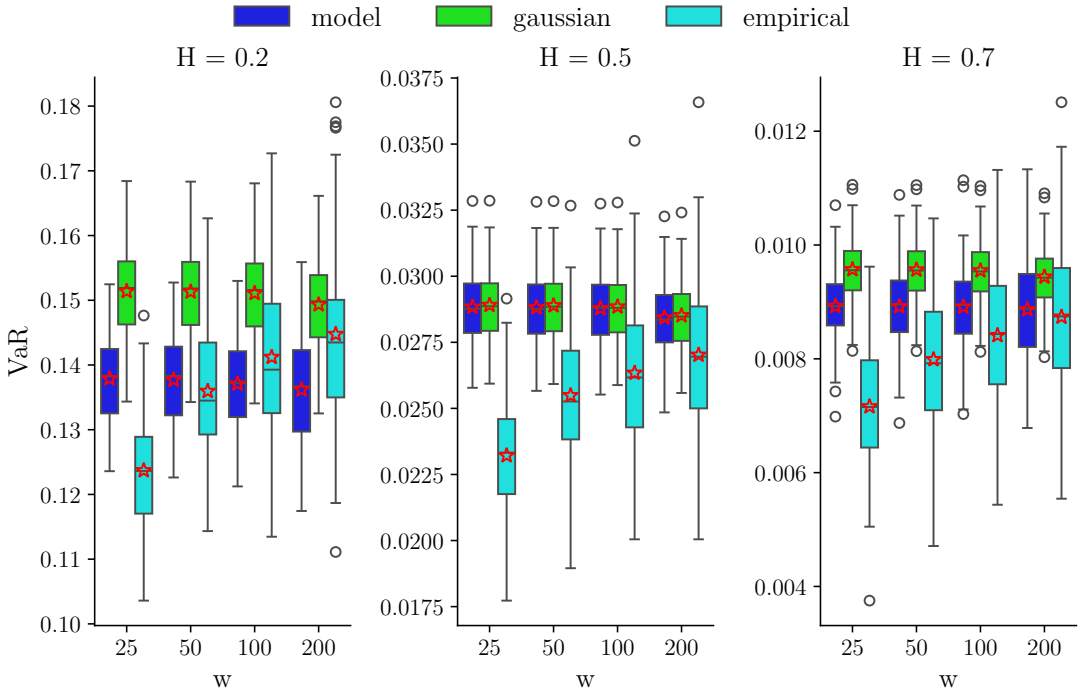


Figure 3.9: VaR distribution (100 runs)

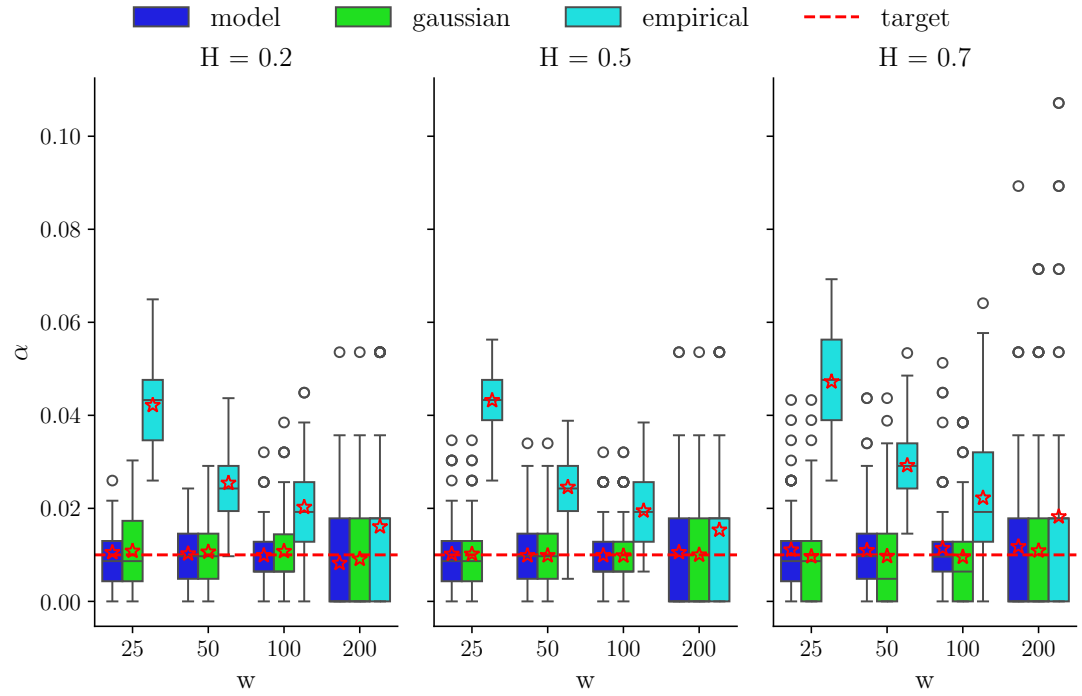


Figure 3.10: Coverage test (100 runs)

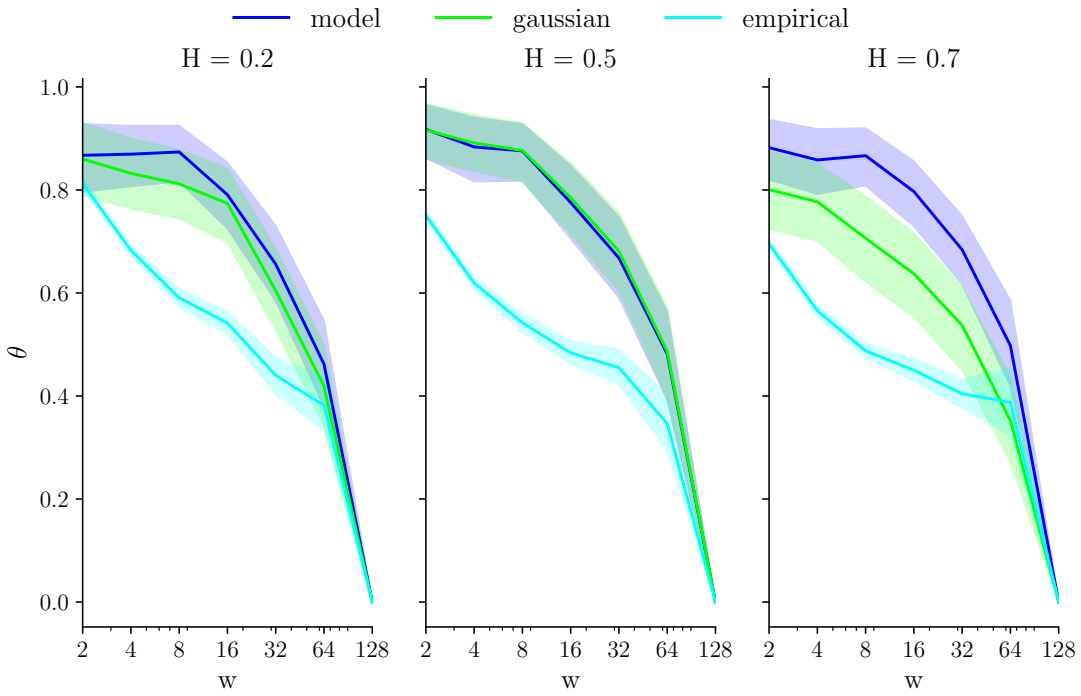


Figure 3.11: Independence test (100 runs)

H	w	$\hat{\theta}^{m,R}$	$\hat{\theta}^{G,R}$	$\hat{\theta}^{e,R}$	$p_N^{m/e,R}$	$p_N^{m/G,R}$	$\hat{\theta}^{m,B}$	$\hat{\theta}^{G,B}$	$\hat{\theta}^{e,B}$	$p_N^{m/e,B}$	$p_N^{m/G,B}$
0.2	25	**0.74**	0.71	0.51	0.8	0.2	**0.94**	0.89	0.64	1.0	0.3
	50	0.45	**0.51**	0.05	0.6	0.2	0.51	**0.65**	0.39	0.7	0.3
	100	**0.30**	0.28	0.00	0.5	0.3	**0.50**	**0.50**	0.24	0.6	0.4
0.5	25	**0.66**	**0.66**	0.41	0.6	0.0	**0.69**	**0.69**	0.60	0.7	0.0
	50	**0.57**	0.47	0.22	0.8	0.1	**0.74**	0.64	0.47	0.7	0.1
	100	**0.56**	0.41	0.00	0.7	0.2	**0.52**	0.49	0.19	0.8	0.2
0.7	25	0.73	**0.78**	0.40	0.8	0.2	**0.85**	0.74	0.61	0.9	0.4
	50	**0.63**	0.45	0.19	0.8	0.4	**0.72**	0.62	0.38	0.8	0.3
	100	0.34	**0.48**	0.00	0.6	0.2	0.55	**0.59**	0.20	0.8	0.3

Table 3.6: For the case of the first Gaussian approximation of the future portfolio variation, average estimated extremal index obtained by run and block declustering, for the conditional **VaR** of the first Gaussian approximation of the future portfolio's increment ($\hat{\theta}^m$), the Gaussian **VaR** ($\hat{\theta}^G$), and the empirical **VaR** ($\hat{\theta}^e$), on $N = 10$ simulated time series of length 256 where each **VaR** is computed at the risk level of $\alpha = 1\%$ and estimated using a window of size w. We also include the probability for the model to be more independent than empirical **VaR**, $p_N^{m/e} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{\{\hat{\theta}_i^m > \hat{\theta}_i^e\}}$, and than the Gaussian **VaR** $p_N^{m/G} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{\{\hat{\theta}_i^G > \hat{\theta}_i^G\}}$.

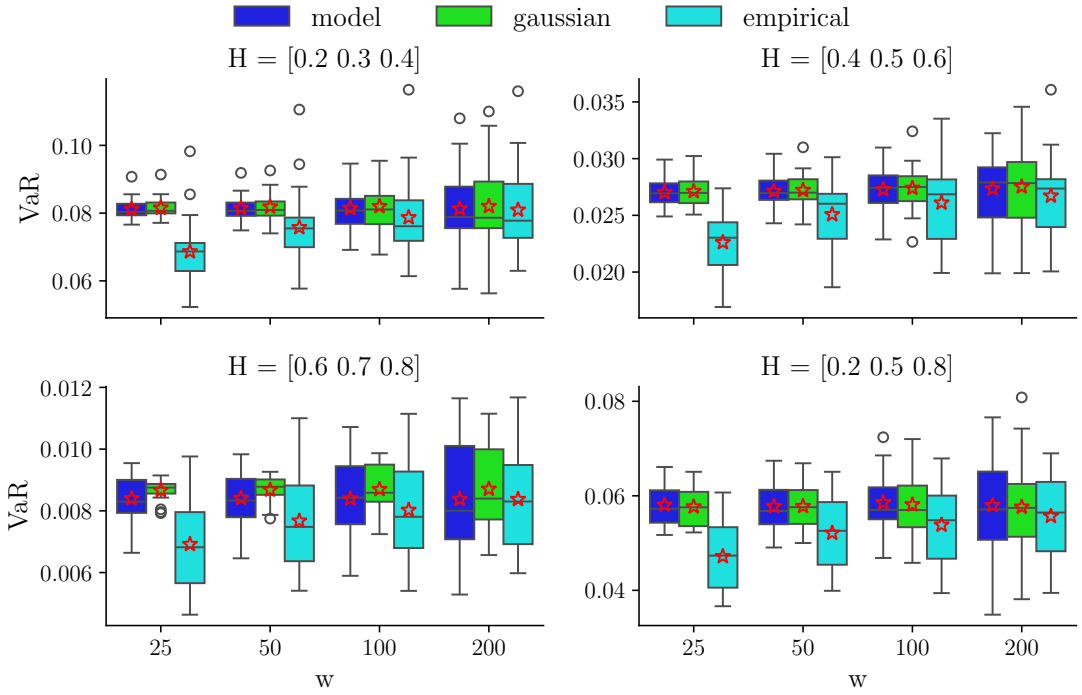


Figure 3.12: VaR distribution (20 runs)

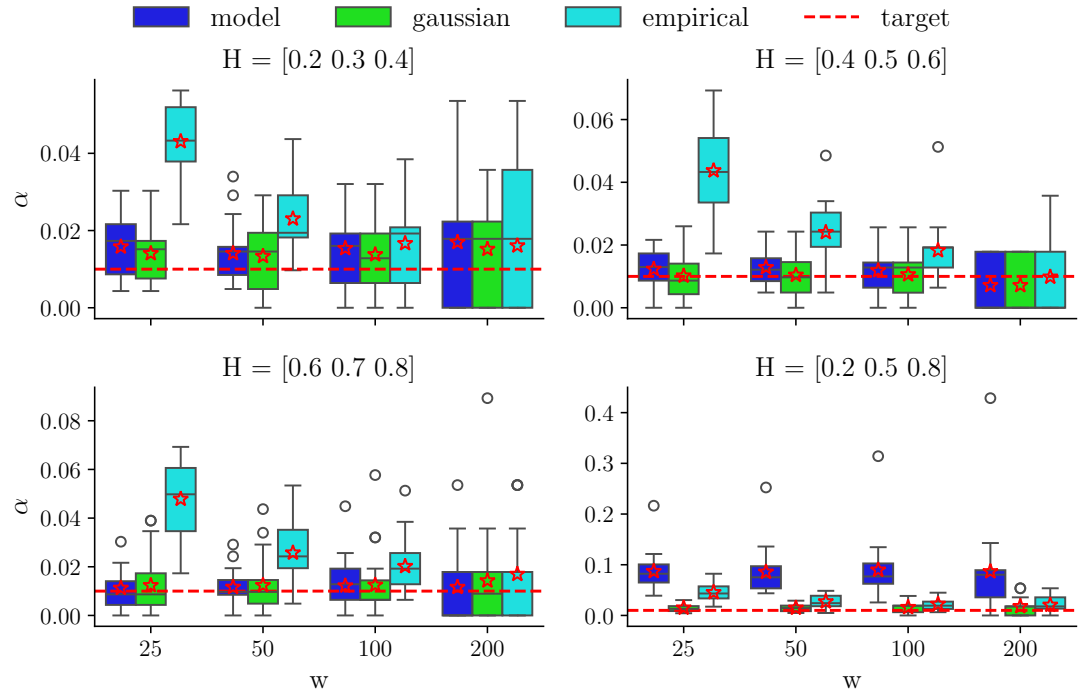


Figure 3.13: Coverage test (20 runs)

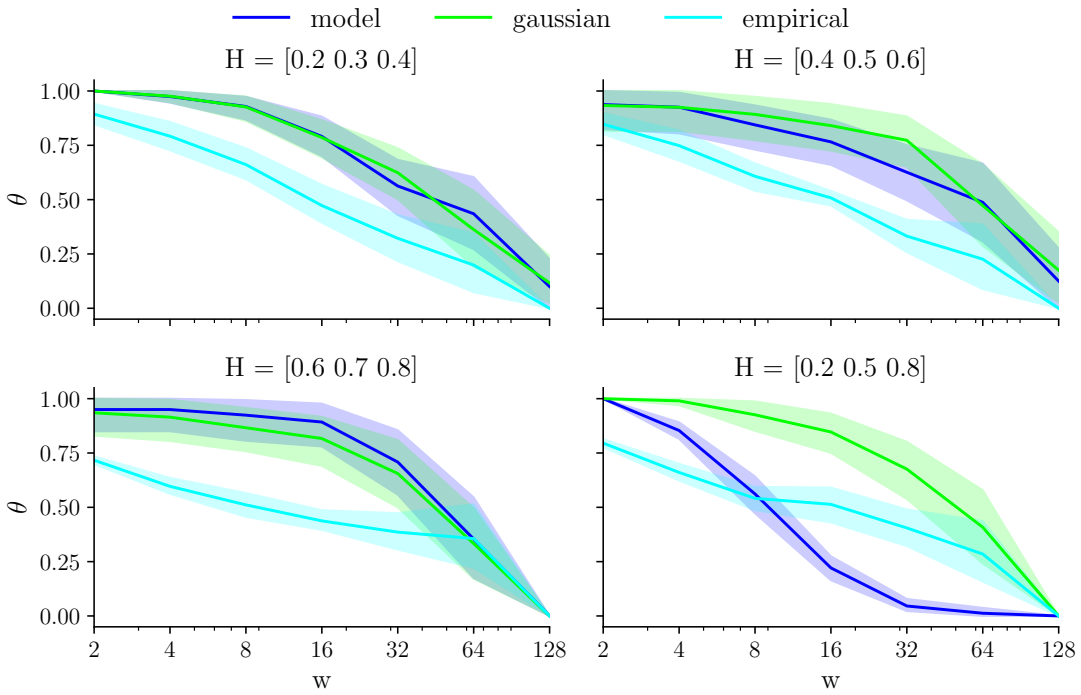


Figure 3.14: Independence test (20 runs)

In Figure 3.14, for each of the three VaRs (model, Gaussian and empirical) and for each of the four portfolios, the line plots represent the evolution of the estimated extremal index $\hat{\theta}$ as a function of the window size. We can observe that for the first three portfolios, whose Hurst exponents are homogeneous, the extremal indices are very to close to 1 for small window sizes and becomes increasingly closer to 0 when the window size increases. A correct window size to perform the independence test has to take into account the size of the available data. It neither has to be too small with respect to the size of the available data nor too large. Moreover, the VaR model is competitive with the Gaussian VaR for the first three portfolios with an extremal index close to 1 for window sizes between 2 and 16. The empirical VaR is less relevant since its extremal index is closer to 0. For the last portfolio composed of heterogeneous Hurst exponents, the VaR model badly performs whereas the Gaussian VaR presents an extremal index very close to 1 for reasonable window sizes.

3.7.2 Backtesting in multivariate framework

To backtest the **VaR** model in the multivariate framework, we proceed as follows. First, we simulate a trajectory of an mfBm increments of parameters $\boldsymbol{H} = (H_1, \ldots, H_d)$, of volatility parameters $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_d)$ and of the correlation matrix $\boldsymbol{R} = (\rho_{ij})_{(i,j)\in[\![1,d]\!]^2}$, over two years (two times 256 business days) with a daily frequency. Using the parameters estimated on the first year of data, we dynamically predict, thanks to a rolling window procedure, the **VaR** at the risk level $\alpha = 1\%$ at a time horizon $h = 1$ day after the end of each window.

As in the univariate framework, we analyze the coverage rate, average **VaR**, and extremal indices.

Three **VaR** models are compared: the conditional $\mathbf{VaR}_\alpha$ of the first Gaussian approximation of the future portfolio increment at time horizon h, the $\mathbf{VaR}_\alpha$ of the future increment at time horizon h of the Gaussian portfolio whose Hurst exponents have been forced to be equal to 0.5 for all the assets, and the empirical $\mathbf{VaR}_\alpha$ estimated on the increment's portfolio trajectory.

Four cases were distinguished. Indeed, we simulated mfBm with Hurst exponents either all below 0.5, or all above, or all close to 0.5, or spread in all possible ranges of values.

We display in Figure 3.15 and in Tables 3.7, 3.8, the corresponding backtesting results.

H	w	$\hat{\alpha}^m$	$\hat{\alpha}^G$	$\hat{\alpha}^e$	$p_N^{m/e}$	$p_N^{m/G}$	$\overline{\mathbf{VaR}_\alpha^m}$	$\overline{\mathbf{VaR}_\alpha^G}$	$\overline{\mathbf{VaR}_\alpha^e}$
0.2 - 0.3 - 0.4	25	0.013	**0.012**	0.044	1.0	0.1	0.081	0.082	**0.065**
0.4 - 0.5 - 0.6	25	0.009	**0.010**	0.043	1.0	0.7	0.027	0.027	**0.023**
0.6 - 0.7 - 0.8	25	**0.009**	0.012	0.052	1.0	0.5	0.009	0.009	**0.007**
0.2 - 0.5 - 0.8	25	0.085	**0.012**	0.047	0.1	0.0	0.058	0.057	**0.047**

Table 3.7: Average estimated hit probability for the **VaR** model ($\hat{\alpha}^m$), the Gaussian **VaR** and the empirical **VaR** ($\hat{\alpha}^e$), on $N = 10$ simulated time series of length 256 where each **VaR** is calculated for a risk level of $\alpha = 1\%$ with a conditioning window of size w. We also include the probability for the model to be more accurate than the Gaussian **VaR**, $p_N^{m/G} = \frac{1}{N}\sum_{i=1}^N \mathbb{1}_{\{|\hat{\alpha}_i^m - \alpha| < |\hat{\alpha}_i^G - \alpha|\}}$, and than the empirical **VaR**, $p_N^{m/e} = \frac{1}{N}\sum_{i=1}^N \mathbb{1}_{\{|\hat{\alpha}_i^m - \alpha| < |\hat{\alpha}_i^e - \alpha|\}}$. Finally, we add the average **VaR** for each model.

H	w	$\hat{\theta}^{m,R}$	$\hat{\theta}^{G,R}$	$\hat{\theta}^{e,R}$	$p_N^{m/G,R}$	$p_N^{m/e,R}$	$\hat{\theta}^{m,B}$	$\hat{\theta}^{G,B}$	$\hat{\theta}^{e,B}$	$p_N^{m/G,B}$	$p_N^{m/G,B}$
0.2 - 0.3 - 0.4	25	0.84	**0.87**	0.47	0.00	0.90	0.90	**0.92**	0.61	0.00	0.80
0.4 - 0.5 - 0.6	25	**0.77**	0.72	0.57	0.2	0.7	**0.88**	0.80	0.72	0.2	0.7
0.6 - 0.7 - 0.8	25	**0.72**	0.50	0.55	0.50	0.80	**0.77**	0.65	0.68	0.2	0.6
0.2 - 0.5 - 0.8	25	0.15	**0.84**	0.57	0.00	0.00	0.048	**0.97**	0.76	0.00	0.20

Table 3.8: For the case of the first Gaussian approximation of the future portfolio variation, average estimated extremal index obtained by run and block declustering, for the conditional **VaR** of the first Gaussian approximation of the future portfolio's increment ($\hat{\theta}^m$), the Gaussian **VaR** ($\hat{\theta}^G$), and the empirical **VaR** ($\hat{\theta}^e$), on $N = 10$ simulated time series of length 256 where each **VaR** is computed at the risk level of $\alpha = 1\%$ and estimated using a window of size w. We also include the probability for the model to be more independent than empirical **VaR**, $p_N^{m/e} = \frac{1}{N}\sum_{i=1}^N \mathbb{1}_{\{\hat{\theta}_i^m > \hat{\theta}_i^e\}}$, and than the Gaussian **VaR** $p_N^{m/G} = \frac{1}{N}\sum_{i=1}^N \mathbb{1}_{\{\hat{\theta}_i^G > \hat{\theta}_i^G\}}$.

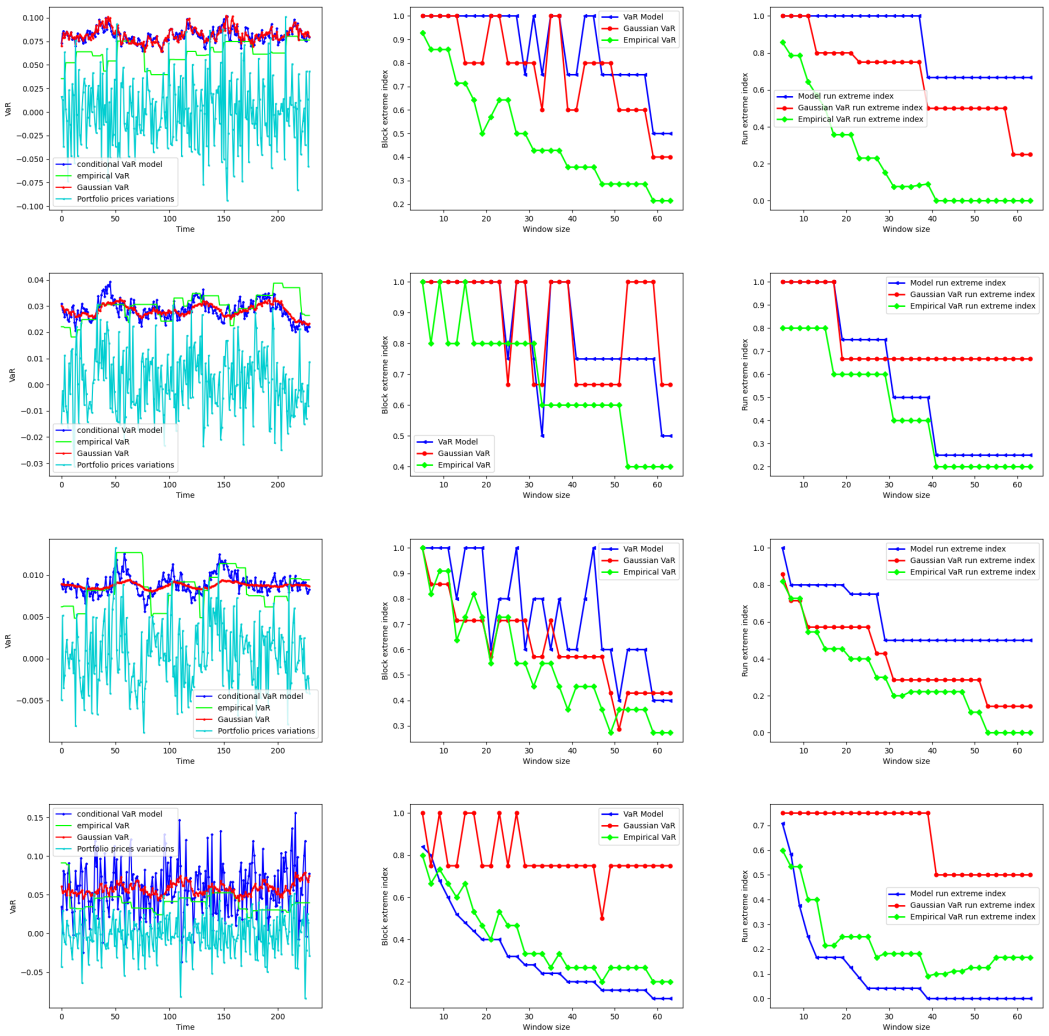


Figure 3.15: Backtesting of the conditional **VaR** in the multivariate approach. The first column represents the trajectory of several models of **VaR** (conditional **VaR**, Gaussian **VaR** and empirical **VaR**) in function of time, the second (respectively third) column represents the evolution of the extremal index obtained by run (resp. block) declustering in function of the size of the block. The simulated trajectories have been generated with a unique $\boldsymbol{\sigma} = (0.2, 0.2, 0.2)$, and $\boldsymbol{H}$ equal to $(0.2, 0.3, 0.4)$ (first row), to $(0.4, 0.5, 0.6)$ (second row), to $(0.6, 0.7, 0.8)$ (third row), and to $(0.2, 0.5, 0.8)$ (fourth row).

When the Hurst exponents of the assets contained in the portfolio are all concentrated in the same range of values $(0.25, 0.5)$, or in $(0.5, 0.75)$ or even in $(0.4, 0.6)$, the conditional **VaR** model is relevant because the coverage rate is very close to the risk level and the average **VaR** is low. However, when the Hurst exponents of the assets contained in the portfolio are very diversified and spread over different ranges of values, the conditional **VaR** model is not relevant because the coverage rate is far from the target risk level and the average **VaR** is high. The empirical **VaR** suffers from the same limitations as in the univariate case.

The papers (Chong et al., 2022a) and (Chong et al., 2022b) deal with the estimation of the Hurst exponent and studied the convergence rate of the estimators. In their study, they explained that the behavior of the estimator of the Hurst exponent depends on the range of values assumed for the true Hurst exponent. This can be explained by the convergence rate of estimator $n^{-1/(4H+2)}$. When H is very low, that is, $H \in (0, 0.25)$, thus the convergence rate is high. The fact that the estimator might converge quickly implies that it is able to properly distinguish the different values of H and provide an accurate estimation of H. For $H \in (0.25, 0.5)$, the convergence rate of the estimator is continuously decreasing. The fact that the estimator might converge slowly implies that the estimator is not able to distinguish the values of H that are close. For instance, it is not possible to distinguish between values in the range $(0.4, 0.6)$. However, it can distinguish values that are far apart such as 0.3 and 0.7. This implies that the estimation of the Hurst exponent in the range $(0.25, 0.75)$ is less accurate than for small H. Finally, when $H \in (0.75, 1)$, the quadratic variation at the first order does not converge anymore, and we develop the quadratic variation in the second order (i.e., for price increments at the second order). The estimator of H based on second-order quadratic variation converges with the same convergence rate as the other ranges of values. Thus, the estimation of H for $H \in (0.75, 1)$ is even less accurate.

3.8 Application to market data

3.8.1 Univariate framework

In this section, we compare three **VaR** models based on fBm: the conditional **VaR** of the future fBm increment, the conditional **VaR** of the future true portfolio increment, and the conditional **VaR** of the first Gaussian approximation of the future portfolio increment, with the Gaussian **VaR** based on standard Brownian motion and the empirical **VaR**. All **VaR** models are computed at the risk level $\alpha = 1\%$ and time horizon $h = 1$ day. We consider historical observations of equity indices, especially the S&P 500 index, for different periods of two years (i.e., two times 256 business days) with a daily frequency. As previously mentioned, the first year of data is used to estimate the parameters of the model, and the second year is used to perform the predictions. We consider the log-price increment dynamics that can be properly described thanks to an fBm, and that present long-range memory. Indeed, some empirical studies have especially shown the relevance of an fBm compared to non-fractional models (Garcin, 2017, 2020), (Cont, 2005), (Surgailis et al., 2008), (Lillo and Farmer, 2004).

Beyond the visual evaluation of the performances of the **VaR**s in Figure 3.16, we provide a more precise analysis of these risk models in Table 3.9, which focuses on the coverage of each **VaR**, and in Table 3.10, which gathers the independence statistics of hit clusters. In each table, for each line corresponding to a given **VaR** model, the estimation of the Hurst exponent and volatility are provided, and we put in bold the most desirable values, that is, the hit probability the closest to 0.01, the lowest average **VaR**, and the highest extremal index.

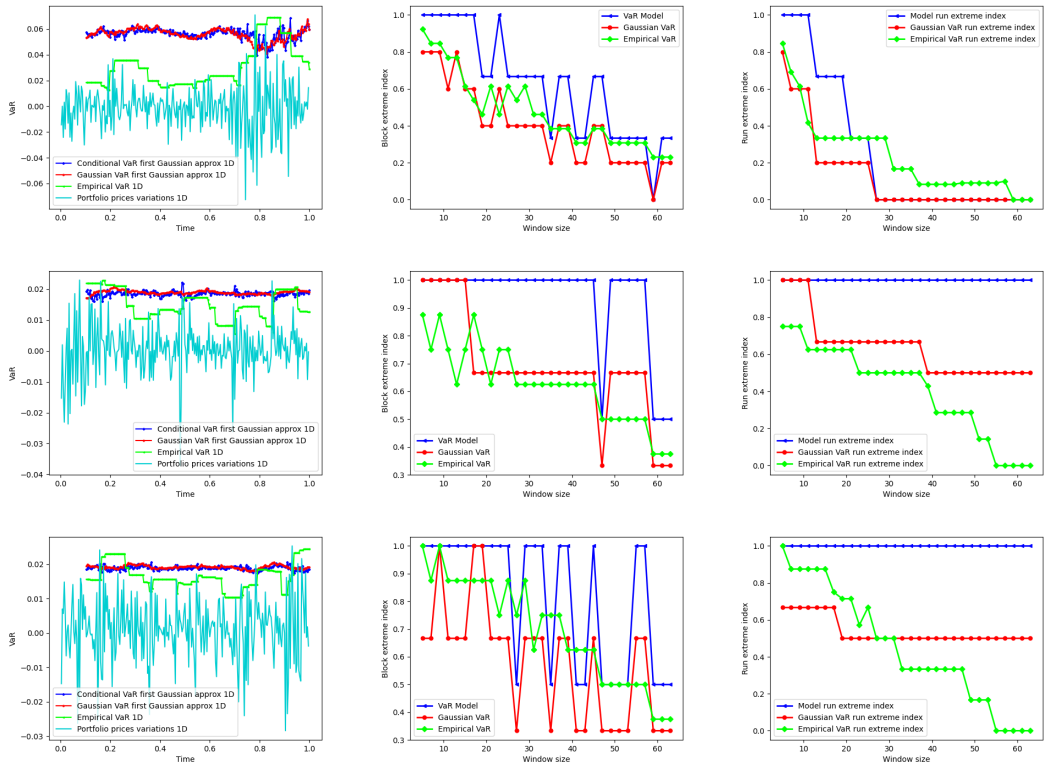


Figure 3.16: Backtesting of the conditional $\mathbf{VaR}_\alpha$ of the first Gaussian approximation of the future portfolio increment for three distinct periods $2007-2009$ (first row), $2015-2017$ (second row) and $2020-2022$ (third row). For each row, the first column represents the trajectory of several models of **VaR** (conditional **VaR**, Gaussian **VaR** and empirical **VaR**) in function of time, the second (respectively third) column represents the evolution of the extremal index obtained by block (resp. run) declustering in function of the size of the block. At each time, two years of historical data are considered. The first year of data is used to estimate the Hurst exponent and the volatility parameter, and the second year of data is used for the backtesting of the models.

Table 3.9 shows that the **VaR** model presents a coverage rate closest to the risk level. Moreover, Table 3.10, shows that the extremal index closest to one is that of the **VaR** models. However, the lowest **VaR** is the empirical **VaR**. However, the coverage rate of the empirical **VaR** is far from the risk level and the extremal index is far from one.

Period	$\hat{H}$	$\hat{\sigma}$	w	$\hat{\alpha}^m$	$\hat{\alpha}^G$	$\hat{\alpha}^e$	$\overline{\mathbf{VaR}_\alpha^m}$	$\overline{\mathbf{VaR}_\alpha^G}$	$\overline{\mathbf{VaR}_\alpha^e}$
2007-2009	0.381	0.210	25	**0.013**	0.021	0.055	0.056	0.056	**0.031**
2015-2017	0.411	0.079	25	**0.009**	0.013	0.034	0.018	0.019	**0.015**
2020-2022	0.446	0.097	25	**0.009**	0.013	0.034	0.019	0.019	**0.016**

Table 3.9: Estimated hit probability and average $\mathbf{VaR}_\alpha$ for the **VaR** model (conditional **VaR** of the first Gaussian approximation of the future portfolio's increment) ($\hat{\alpha}^m$), the Gaussian **VaR** ($\hat{\alpha}^G$), and the empirical **VaR** ($\hat{\alpha}^e$), on real data. The different **VaR**s are compared on three distinct periods $2007-2009$ (first row), $2015-2017$ (second row) and $2020-2022$. The historical data that is considered here is the $S\&P$ index. On each period, the first year is used to estimate the parameters of the model and the second yea is used to backtest the model.

Period	$\hat{H}$	$\hat{\sigma}$	w	$\hat{\theta}^{m,R}$	$\hat{\theta}^{G,R}$	$\hat{\theta}^{e,R}$	$\hat{\theta}^{m,B}$	$\hat{\theta}^{G,B}$	$\hat{\theta}^{e,B}$
2007-2009	0.381	0.210	25	**0.33**	0.2	**0.33**	**0.67**	0.4	0.62
2015-2017	0.411	0.079	25	**1.0**	0.67	0.5	**1.0**	0.67	0.75
2020-2022	0.446	0.097	25	**1.0**	0.67	0.5	**1.0**	0.67	0.88

Table 3.10: Estimated extremal index obtained by run and block declustering, for the **VaR** model ($\hat{\theta}^{m,R}$, $\hat{\theta}^{m,B}$), the Gaussian **VaR** ($\hat{\theta}^{G,R}$, $\hat{\theta}^{G,B}$) and the empirical **VaR** ($\hat{\theta}^{e,R}$, $\hat{\theta}^{e,B}$), on real data. The different **VaR**s are compared on three distinct periods $2007-2009$ (first row), $2015-2017$ (second row) and $2020-2022$. The historical data that is considered here is the $S\&P$ index. On each period, the first year is used to estimate the parameters of the model and the second yea is used to backtest the model.

This does not mean that the empirical **VaR** is not efficient but requires a longer data history. Concerning the Gaussian **VaR**, the coverage rate is better than that of the empirical **VaR**, and worse than that of the **VaR** model. Similarly, the extremal index is larger than that of the empirical **VaR** and lower than the one of the **VaR** model. Finally, the average Gaussian **VaR** is almost equal to that of the **VaR** model. Overall, the **VaR** model performs better than the empirical and Gaussian **VaR** models for the given periods.

3.8.2 Multivariate framework

In the multivariate framework, we consider several equity indices, namely the S&P 500, CAC 40 and DAX indices for different periods of two years, with a daily frequency. We create a portfolio that contains these assets and compare the **VaR** model corresponding to the conditional $\mathbf{VaR}_\alpha$ of the first Gaussian approximation of the future portfolio increment, with the Gaussian **VaR** corresponding to the **VaR** model where all the Hurst exponents have been forced to be equal to 0.5, and the empirical $\mathbf{VaR}_\alpha$. All the

$\mathbf{VaR}_\alpha$ are computed at risk level $\alpha = 1\%$ and time horizon prediction $h = 1$ day.

Beyond the visual evaluation of the three **VaR**s for different periods in Figures 3.17, we provide a more precise analysis of these risk models in Table 3.11 which focuses on the coverage of each **VaR**. In this table, for each line corresponding to a given **VaR** model, the estimation of the Hurst exponents and the volatility parameters are provided, and we bold the most desirable values, that is, the hit probability is closest to 0.01, the lowest average **VaR**, and the highest extremal index.

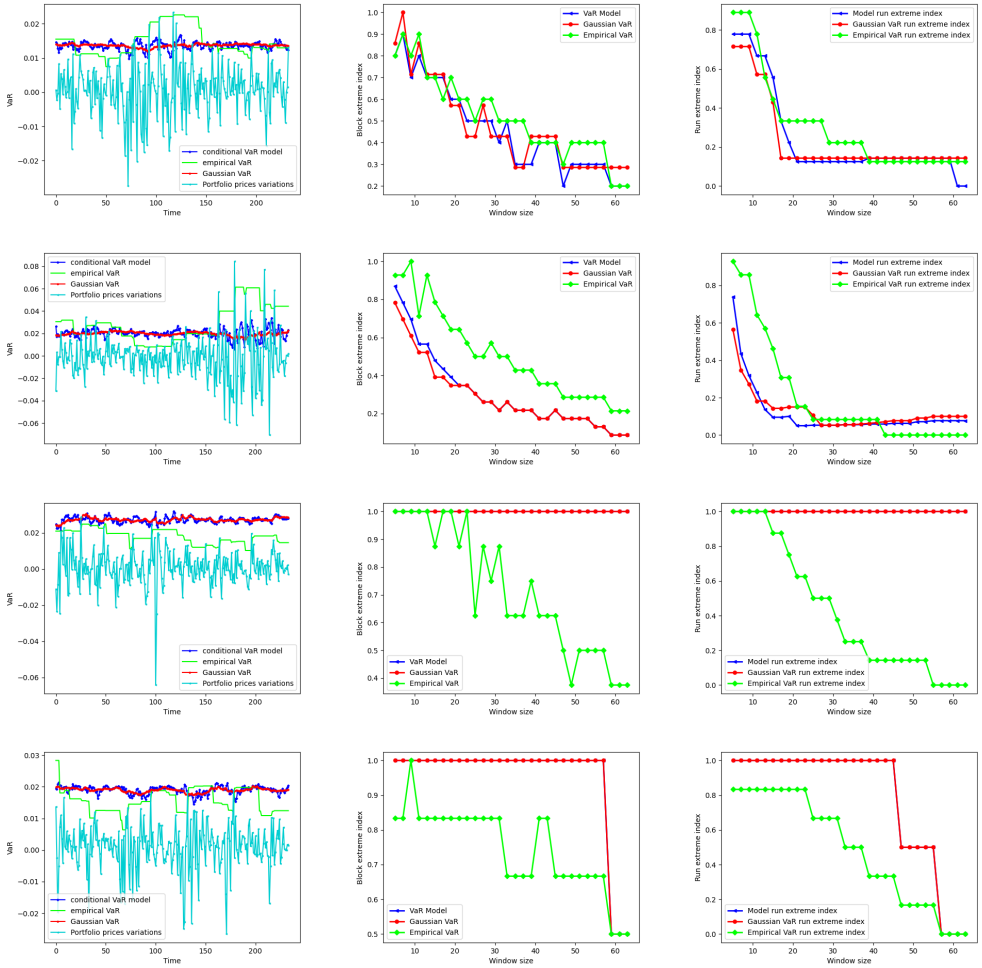


Figure 3.17: Backtesting of the conditional **VaR** for the portfolio composed of the S&P 500, CAC 40, and DAX indices on several periods. The parameters of the model are estimated on the first year of data and the backtesting is performed on the second year. Each row corresponds to a period of data: 2005−2007, 2007−2009, 2015−2017, 2018−2020 (first to fourth row). The first column represents the trajectory of several models of **VaR** (conditional **VaR**, Gaussian **VaR** and empirical **VaR**), the second (respectively third) represents the extremal index obtained by run (resp. block) declustering in function of the size of the block.

Period	$\hat{H}$	$\hat{\sigma}$	w	$\hat{\alpha}^m$	$\hat{\alpha}^G$	$\hat{\alpha}^e$	$\overline{\mathbf{VaR}_\alpha^m}$	$\overline{\mathbf{VaR}_\alpha^G}$	$\overline{\mathbf{VaR}_\alpha^e}$
2005 - 2007	0.438 - 0.448 - 0.483	0.072 - 0.083 - 0.114	25	0.043	**0.030**	0.043	**0.014**	**0.014**	0.015
2007 - 2009	0.412 - 0.428 - 0.488	0.097 - 0.117 - 0.148	25	0.098	0.098	**0.060**	**0.020**	**0.020**	0.028
2015 - 2017	0.40 - 0.472 - 0.492	0.094 - 0.20 - 0.225	25	0.004	0.004	**0.034**	0.027	0.027	**0.018**
2018 - 2020	0.490 - 0.550 - 0.513	0.164 - 0.183 - 0.166	25	**0.009**	**0.009**	0.026	0.019	0.019	**0.016**

Table 3.11: Estimated hit probability and average $\mathbf{VaR}_\alpha$ for the **VaR** model ($\hat{\alpha}^m$), the Gaussian **VaR** ($\hat{\alpha}^G$), and the empirical **VaR** ($\hat{\alpha}^e$), on real data. The historical data that is considered here is the $S\&P$ index on two years on the following periods $2005-2007$, $2007-2009$, $2015-2017$, $2018-2020$. For each period, the first year is used to estimate the parameters of the model and the second year is used to backtest the model.

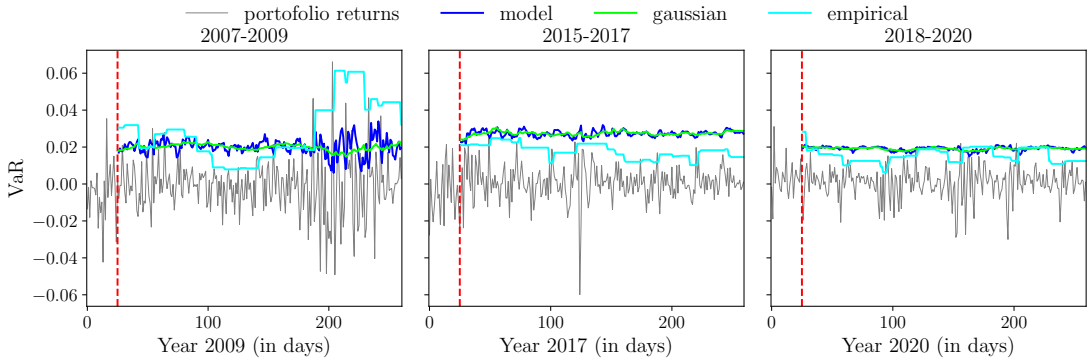


Figure 3.18: Evolution of the $\mathbf{VaR}_\alpha$ trajectories with $\alpha = 1\%$ at time and time-horizon $h = 1$ day. The portfolio is composed of the three following indices SPX 500, DAX, CAC 40 (for which the mid value is considered); for 3 periods of 2 years (1 year for parameters estimation, 1 year for prediction): 2007-2009 $\sim$ subprime crisis, 2015-2017 $\sim$ smooth period, 2018-2020 $\sim$ beginning of Covid crisis.

The **VaR** model and the Gaussian **VaR** are very close in terms of coverage rate and of average **VaR**. The best performance of the **VaR** model is on the period $2018-2020$. Moreover, the performance of the **VaR** model is not so bad on the period $2015-2017$. However, the model does not perform well at all on the periods $2007-2009$. The empirical **VaR** performs well on the periods $2007-2009$ and $2015-2017$.

In Figure 3.18, three periods are considered: 2007-2009 $\sim$ subprime crisis, 2015-2017 $\sim$ smooth period, 2018-2020 $\sim$ beginning of Covid crisis, and for each of them we represent the evolution of the portfolio returns and three $\mathbf{VaR}_\alpha$, the $\mathbf{VaR}_\alpha$ model, the Gaussian $\mathbf{VaR}_\alpha$ and the empirical $\mathbf{VaR}_\alpha$. The $\mathbf{VaR}_\alpha$ model and the Gaussian $\mathbf{VaR}_\alpha$ present good performance in terms of coverage on the period $2018-2020$.

"The essence of mathematics is not to make simple things complicated, but to make complicated things simple."

(**Stanley Gudder**, A Mathematical Journey, 1976)

Chapter 4

Mean estimation of Expected-Shortfall in heavy-tailed distributions

Contents

This chapter addresses the research questions ($RQ\#4$, ($RQ\#5$)). The objective of this chapter, is to explore robust methods for estimating the Expected-Shortfall in heavy-tailed distributions. The Expected-Shortfall is the average of losses exceeding the VaR, and an ES estimator is an estimator of the mean applied to the distribution tail beyond the VaR. Thus, we explore robust estimators of the mean as an alternative to a simple empirical mean, in heavy-tailed distributions using the toy case of the Pareto distribution. First, we recall the theory on the Expected-Shortfall and on the Pareto distribution, and we present the characteristic properties of the latter. Then, several non-asymptotic mean estimators, such as the Median-of-Means, the Trimmed-Mean, or even the Lee-Valiant estimator, are presented with their characteristics and compared to the classical empirical mean. We study their bias and provide explicit formulae when possible. Moreover, we evaluate the convergence rate of the bias. Finally, we support the theoretical analysis with experiments, and compare the performances of the different estimators.

The problem, we are interested in is the efficient numerical estimation of the Expected-Shortfall in the Pareto distribution. In Section 4.1, we provide a theoretical background on Expected-Shortfall (ES) and heavy-tailed distributions, especially the Pareto distribution. In Section 4.2, a classical asymptotic estimator, namely the empirical mean, is presented. In the other sections, the non-asymptotic estimators of the mean are presented. In Section 4.3, we study the empirical median along with its bias and convergence rate. The Median-of-Means (MoM) estimator is discussed in Section 4.4. The Section 4.5 is dedicated to the Trimmed-Mean estimator (TM). In Section 4.6, the Lee-Valiant estimator and its properties are discussed. Finally, a comparative study of the performance of the different estimators is presented in Section 4.7.

4.1 Background

In this section, we provide a theoretical background on Expected-Shortfall and heavy-tailed distributions. Special focus is placed on the Pareto distribution and its properties.

4.1.1 Expected-Shortfall

The Expected-Shortfall (**ES**), is a risk assessment measure that quantifies the amount of tail risk an investment portfolio has. It is derived by taking a weighted average of the extreme losses in the tail of the distribution of possible returns, beyond the Value-at-Risk (**VaR**), cutoff point. If an investment has shown stability over time, then **VaR** may be sufficient for risk management in a portfolio containing that investment. However, the less stable the investment, the greater the chance that **VaR** will not provide a full picture of the risks, as it is indifferent to anything beyond its own threshold. **ES** attempts to address the shortcomings of the **VaR** metric, which is a statistical technique used to measure the level of financial risk within a firm or investment portfolio over a specific time frame. While **VaR** represents a worst-case loss associated with a probability and time horizon, **ES** is the expected loss if the worst-case threshold is crossed.

The Expected-Shortfall is particularly useful in risk management and is applied in several financial risk areas. In the context of *market risk*, **ES** is used to estimate potential losses in financial assets, portfolios and positions in adverse market conditions. This helps investors and institutions understand the magnitude of potential losses beyond the **VaR** level. The Expected Shortfall is also used to assess the *credit risk* of portfolios and individual assets. For instance, in credit risk modeling, **ES** can provide insights into expected losses given a default or credit event. In addition, **ES** is applied to evaluate potential losses related to illiquid assets or positions under stressed market conditions. This helps financial institutions determine the amount of capital required to address such risks. For *operational risk* management, **ES** is used to understand the potential losses arising from operational failures beyond the **VaR** level, such as technology glitches, fraud, or human errors. Concerning *systemic risk*, **ES** can be used to assess the potential losses that a financial system or broader economy may face, in extreme market conditions. This framework, helps regulators and policymakers identify the potential vulnerabilities and causes of systemic risks. Finally, **ES** plays a crucial role in stress-testing financial models and portfolios. This allows risk managers to analyze how severe scenarios may impact the overall risk exposure and capital adequacy of financial institutions.

Overall, **ES** is a valuable metric for understanding tail risks and making informed decisions regarding risk management, capital allocation, and hedging strategies in various areas of financial risk. It provides a more comprehensive and conservative measure of risk than **VaR**.

Definition 4.1 (Expected-Shortfall (**ES**)). *Let $X \in \mathcal{L}^p(\mathcal{F})$ be a random variable representing loss*[1] *of a portfolio at some future time, with a cumulative distribution function $F_X(x) = \mathbb{P}(X \leq x); \forall x \in \mathbb{R}$. Let $\alpha \in (0,1)$ be the risk level. Then, based on (Tasche, 2002b, Prop 3.4, Eq 3.3), the Expected-Shortfall is defined as:*

$$\mathbf{ES}_\alpha(X) = \frac{1}{1-\alpha}\int_\alpha^1 \mathbf{VaR}_\beta(X)\mathrm{d}\beta \tag{4.1}$$

where $\mathbf{VaR}_\beta$ is the Value-at-Risk given by:

$$\mathbf{VaR}_\beta(X) := \inf\{x \in \mathbb{R}, F_X(x) \geq \beta\} \tag{4.2}$$

as defined in (Tasche, 2002b, Def 2.1, Eq 2.1a).

When the distribution is continuous, an equivalent definition can be given (see (Sarykalin et al., 2008, Def.2, p.273)) by:

$$\mathbf{ES}_\alpha(X) = \mathbb{E}[X|X \geq \mathbf{VaR}_\alpha(X)]. \tag{4.3}$$

The problem of estimating the Expected-Shortfall amounts to estimating the conditional expectation of a random variable based on a sample of n independent, identically distributed random draws $Y_1, \ldots, Y_n$ from the distribution of $Y \overset{d}{=} X \mid X \geq \mathbf{VaR}_\alpha(X)$ which can be approximated by the draws of X restricted to be above the empirical $\mathbf{VaR}_\alpha$. A well-known estimator of the $\mathbf{VaR}_\alpha$ is the empirical α-quantile of a sample.

Definition 4.2 (Empirical α-quantile). *Let $X_1, \ldots, X_n$ be a sample of size n, drawn from a probability distribution F and $X_1^\star, \ldots, X_n^\star$ the (increasingly) ordered sample (order statistic). Let $\alpha \in]0,1[$ be the order of the quantile. The order statistic $X^\star_{\lceil \alpha n \rceil}$ (where $\lceil \alpha n \rceil$ denotes the upper integer part of αn) is called the empirical quantile at the order α of the sample, and is denoted by $\mathbf{q}_{n,\alpha}$.*

Remark 4.3. *Let n, N be integers and $\alpha \in (0,1)$, such that $N \leq \alpha n < N+1$: if $\alpha n = N$ then $\lceil \alpha n \rceil = N$, if $\alpha n > N$ then $\lceil \alpha n \rceil = N+1$.*

Recall that the Expected-Shortfall is defined as follows:

$$\mathbf{ES}_\alpha(X) = \mathbb{E}[X|X \geq \mathbf{VaR}_\alpha] = \frac{\mathbb{E}[X\mathbb{1}_{\{X \geq \mathbf{VaR}_\alpha\}}]}{\mathbb{P}(X \geq \mathbf{VaR}_\alpha)}. \tag{4.4}$$

Let's define the empirical Expected-Shortfall.

Definition 4.4 (Empirical Expected-Shortfall). *Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sequence of Pareto i.i.d. random variables such that $\forall i \in [\![1,n]\!], X_i \sim \mathcal{P}(1,\gamma)$, with $\gamma > 2$. Let's denote by $\boldsymbol{X}^\star = (X_1^\star, \ldots, X_n^\star)$ the order statistics related to $\mathbf{X}$. The empirical α-quantile is given by $q_\alpha^n = X^\star_{\lceil n\alpha \rceil}$, and the empirical Expected-Shortfall at the risk level α is defined as follows :*

$$\mathbf{ES}_\alpha^n = \frac{\frac{1}{n}\sum_{i=1}^n X_i \mathbb{1}_{\{X_i \geq q_\alpha^n\}}}{\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{X_i \geq q_\alpha^n\}}} = \frac{1}{n - \lceil n\alpha \rceil}\sum_{i=1}^n X_i \mathbb{1}_{\{X_i \geq q_\alpha^n\}}. \tag{4.5}$$

[1] as a positive quantity, i.e. we take the convention that big losses correspond to large positive numbers

The *idealized case* assumes that the empirical α-quantile q_α^n matches the true $\mathbf{VaR}_\alpha$: although only true in the asymptotic framework but not in practice for small sample sizes, we will use this assumption in some parts of our study to simplify the analysis. In such a case, the estimation of the Expected-Shortfall corresponds to the estimation of the conditional expectation of a random variable based on a sample of n independent and identically distributed (i.i.d.) random draws, given that these random draws are larger than the true $\mathbf{VaR}_\alpha$. This assumption simplifies the expression of the estimation error. However, this assumption can be satisfied asymptotically (when the sample size is sufficiently large), but in most cases, when the sample size is far from the asymptotic case, it is not. Therefore, we are interested in the *realistic case* in which the empirical α-quantile does not match the theoretical $\mathbf{VaR}_\alpha$. In this last case, the estimation of the Expected-Shortfall corresponds to the estimation of the conditional expectation of a random variable based on a sample of n random draws, given that these random draws are larger than the empirical α-quantile. In the realistic case, the estimation of the $\mathbf{ES}_\alpha$ is more difficult because the threshold above which we want to estimate the average excess loss is an order statistics and depends on the underlying sample. This implies that the samples larger than the threshold are no longer i.i.d. and the distribution of the excess loss is unknown. Moreover, the realistic case introduces, in the estimation of the $\mathbf{ES}_\alpha$, an additional error term corresponding to the bias estimation between the true $\mathbf{VaR}_\alpha$ and the empirical α-quantile.

In all the sequels, the estimator of the $\mathbf{ES}_\alpha$ will be studied both in the *idealized case*, in which the empirical α-quantile matches the true $\mathbf{VaR}_\alpha$, and in the *realistic case* in which the two quantities differ.

4.1.2 Heavy-tailed distributions

Heavy-tailed distributions are probability distributions whose tails are not exponentially bounded, that is, they have heavier tails than the exponential distribution. It is often the right tail of the distribution that is of interest, but a distribution may have a heavy left tail, or both tails may be heavy. There are five important subclasses of heavy-tailed distributions: fat-tailed, long-tailed, and subexponential distributions.

Definition 4.5 (Heavy-tailed distribution)**.** *The distribution of a random variable X with distribution function F_X is said to have a heavy (right) tail if the moment-generating function of X, $M_X(t)$ is infinite for all $t > 0$. That means:*

$$\int_{-\infty}^{+\infty} e^{tx} dF_X(x) = +\infty; \quad \forall t > 0. \tag{4.6}$$

This is also written in terms of the tail distribution function $\bar{F}(x) = 1 - F(x) = \mathbb{P}(X > x)$:

$$\lim_{x \to +\infty} e^{tx} \mathbb{P}(X > x) = +\infty; \quad \forall t > 0. \tag{4.7}$$

(i) *The distribution of an r.v. X with distribution function F is said to have a fat tail if there exists a positive exponent γ, called the tail index, such that:*

$$\bar{F}(x) := \mathbb{P}(X > x) \sim x^{-\gamma} \text{ as } x \to +\infty. \tag{4.8}$$

"$\sim$" refers to an equivalent up to a constant. Alternatively, the tail part is proportional to the power law.

The power law class is conventionally defined by the property of the survival function, as follows. If X is a random variable belonging to the class of distributions with a power law right tail, that is:

$$\bar{F}(x) = L(x)x^{-\gamma} \tag{4.9}$$

for some $\gamma > 0$, where $L(x) > 0$ is a slowly varying function, defined as:

$$\lim_{x\to+\infty} \frac{L(kx)}{L(x)} = 1 \tag{4.10}$$

for any $k > 0$. See (Haan and Ferreira, 2006, Chapter 1) for more details.

(ii) *The distribution of an r.v. X with distribution function F is said to have a long-right tail if, for all $t > 0$,*

$$\lim_{x\to+\infty} \mathbb{P}(X > x + t|X > x) = 1 \tag{4.11}$$

or equivalently, $\bar{F}(x + t) \sim \bar{F}(x)$ as $x \to +\infty$. This has an intuitive interpretation for a right-tailed long-tailed distributed quantity that if the long-tailed quantity exceeds a certain high level, the probability approaches 1 that it will exceed any other higher level. All long-tailed distributions are heavy-tailed, but the converse is false, and heavy-tailed distributions that are not long-tailed can be constructed.

(iii) *Subexponentiality is defined in terms of the convolutions of probability distributions. For two independent, identically distributed non-negative random variables X_1, X_2 with a common distribution function F, the convolution of F with itself, written $F^{\star 2}$ and called the convolution square, is defined using the Lebesgue-Stieltjes integration by:*

$$\mathbb{P}(X_1 + X_2 \leq x) = F^{\star 2}(x) = \int_0^x F(x - y)dF(y), \tag{4.12}$$

and the n-fold convolution $F^{\star n}$ is defined inductively by the rule:

$$F^{\star n}(x) = \int_0^x F(x - y)dF^{\star n-1}(y). \tag{4.13}$$

A distribution F on the positive half-line is subexponential if:

$$\overline{F^{\star 2}}(x) \sim 2\bar{F}(x) \text{ as } x \to +\infty. \tag{4.14}$$

This implies that, for any $n \geq 1$,

$$\overline{F^{\star n}}(x) \sim n\bar{F}(x) \text{ as } x \to +\infty. \tag{4.15}$$

The probabilistic interpretation of this is that, for a sum of n independent random variables $X_1, \ldots, X_n$ with a common distribution F,

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i > x\right) \sim \mathbb{P}(\max(X_1, \ldots, X_n) > x) \text{ as } x \to +\infty. \tag{4.16}$$

This is often known as the single big jump principle or catastrophe principle. A distribution F on the entire real line is subexponential if the distribution $F\mathbb{1}_{[0,+\infty)}$ is. Here, $\mathbb{1}_{[0,+\infty)}$ is the indicator function of the positive half-line. Alternatively, an r.v. X supported on the real line is subexponential if and only if $X^+ = \max(0, X)$ is subexponential. All subexponential distributions are long-tailed; however, examples can be constructed of long-tailed distributions that are not subexponential.

Heavy-tailed distributions are useful for describing two main phenomena. On the one hand, they allow modelling the fact that some severe extreme events can occur with a non-zero probability. The further the event occurs in the distribution tail, the more severe it is. Therefore, the longer the distribution tail, the more capable it is to describe severe extreme events. On the other hand, heavy-tailed distributions enable the description of the frequency of occurrence of extreme events, that is, how frequently extreme events occur. See (Sigman, 1999) for more details on the heavy-tailed distributions.

A well-known example of a heavy-tailed distribution is the Pareto distribution, as detailed in (Arnold, 2014). There are several reasons for focusing on this distribution.

Pareto distribution plays a crucial role in extreme value theory (EVT), a branch of statistics that deals with the statistical behavior of extreme events. EVT is particularly relevant in finance, where extreme events such as market crashes, have significant implications for investors and financial institutions. In finance, data often exhibit heavy-tailed behavior, meaning that extreme events such as large price movements or financial crises, occur more severely than expected from a normal distribution. Tail risk refers to the risk of extreme events occurring in the tails of a distribution. Although rare, these events can significantly impact financial markets and portfolios. The Pareto distribution, with its ability to model right-hand heavy tails, provides a better fit to such data than traditional distributions such as the normal distribution. Because Gaussian modeling assigns small weight to distribution tails, it ignores extreme events that can lead to inaccurate **VaR** prediction. On the contrary, the Pareto distribution attributes more weight to the right-hand tail of the distribution, taking into account extreme events, which enables better **VaR** predictions, for instance when dealing with assets or portfolios that exhibit right heavy-tailed behavior.

Mathematically speaking, the Pareto distribution exhibits interesting properties. First, the distribution is relatively simple to understand and work with. It only has two parameters (scale $x_m > 0$ and shape $\gamma > 0$ parameters), making it easier to estimate and interpret compared to more complex heavy-tailed distributions such as the stable distribution. The Pareto distribution is a power-law and is well-adapted to fit data that display power-law behavior such as financial data. Indeed, the right-hand tail of the Pareto distribution follows a linear function on a log-log plot. This is supplementary evidence that extreme events occur with more severe intensity than those predicted by other heavy-tailed distributions. This is particularly useful when modeling financial phenomena that follow a power-law distribution, such as income distribution or extreme price movements. Moreover, the Pareto distribution presents some interesting properties of invariance and stability. On the one hand, the Pareto distribution is scale-invariant, which means that multiplying (or scaling) all values of any Pareto random variable by a

constant does not change the shape of the distribution, and the new scaled distribution is still a Pareto distribution with a new scale parameter. On the other hand, the Pareto distribution is stable by conditioning, which means that conditioning any Pareto distribution above a certain threshold does not change the shape of the distribution, and the conditional distribution is still a Pareto distribution with a new scale parameter. These properties are not shared by all the heavy-tailed distributions. The moments of the Pareto distribution exist only for certain ranges of shape parameter γ. Specifically, the k-th moment exists if and only if $\gamma > k$. This means that for small values of γ, moments may not exist, which is related to the heavy-tailed nature of the distribution. As the shape parameter γ decreases, the distribution tail becomes heavier and heavier, giving more weight to extreme events. This leads to finite moments for only very small orders. This is a characteristic of heavy-tailed distributions, where extreme observations have a more significant impact on higher moments than distributions with lighter tails. For instance, the Pareto distribution has an infinite mean for shape parameter $\gamma \leq 1$. For $\gamma \geq 2$, the Pareto distribution has a finite mean but an infinite variance. For $\gamma > 2$, the variance of the Pareto distribution exists and is finite. This means that if the shape parameter is greater than 2, the distribution is characterized by both a finite mean and finite variance, making it more manageable in certain statistical analyses. Finally, the probability density function and cumulative distribution function of the Pareto distribution have relatively simple analytical forms, making it easier to perform mathematical and statistical calculations compared to some other heavy-tailed distributions.

Although the Pareto distribution has applications and properties that make it useful for modeling heavy-tailed behavior, it also presents some limitations. First, the Pareto distribution is defined only for values greater than or equal to the minimum value that corresponds to the scale parameter $x_m > 0$. It cannot account for data in which the support is arbitrarily negative. Although the Pareto distribution can capture heavy-tailed behavior, it may not always provide the best fit for real-world financial data; indeed, its parametric form is imposed. In some cases, other heavy-tailed distributions like the generalized Pareto distribution (GPD) or the Student's t-distribution might offer better fit data. In general, the sensitivity of heavy-tailed distribution parameters to extreme values in the data can lead to significant distortions in parameter estimation and may result in poor model fit when extreme observations are present. Estimating the parameters of the Pareto distribution can be challenging, particularly when the shape parameter γ is small or close to one. Small sample sizes can lead to inaccurate parameter estimates and a poor model fit. Overall, since our study focuses on tail behavior and because Pareto distribution captures this main feature, we use it as a toy model. Investigations of other heavy-tailed models are postponed in future works.

Definition 4.6 (Pareto distribution $\mathcal{P}(x_m, \gamma)$)**.** *If X is a random variable following a Pareto distribution $\mathcal{P}(x_m, \gamma)$, $(x_m > 0, \gamma > 0)$, then the probability that X is larger than some number x, that is, the survival function, also called the tail function, is given by:*

$$\bar{F}_X(x) = \mathbb{P}(X > x) = \mathbb{1}_{\{x<x_m\}} + \left(\frac{x_m}{x}\right)^{\gamma} \mathbb{1}_{\{x\geq x_m\}} \tag{4.17}$$

where x_m is the (necessarily positive) minimum possible value of X, and γ is a positive parameter. The Pareto distribution is characterized by a scale parameter x_m and a shape parameter γ, which is known as the tail index.

The cumulative distribution function (c.d.f.) of a Pareto random variable with parameters x_m and γ is:

$$F_X(x) = \left(1 - \left(\frac{x_m}{x}\right)^{\gamma}\right) \mathbb{1}_{\{x \geq x_m\}}. \tag{4.18}$$

And the probability density function (p.d.f.) is given by:

$$f_X(x) = \frac{\gamma x_m^{\gamma}}{x^{\gamma+1}} \mathbb{1}_{\{x \geq x_m\}}. \tag{4.19}$$

For a graphic representation of the density of the Pareto distribution, see Figures 4.1 and 4.2.

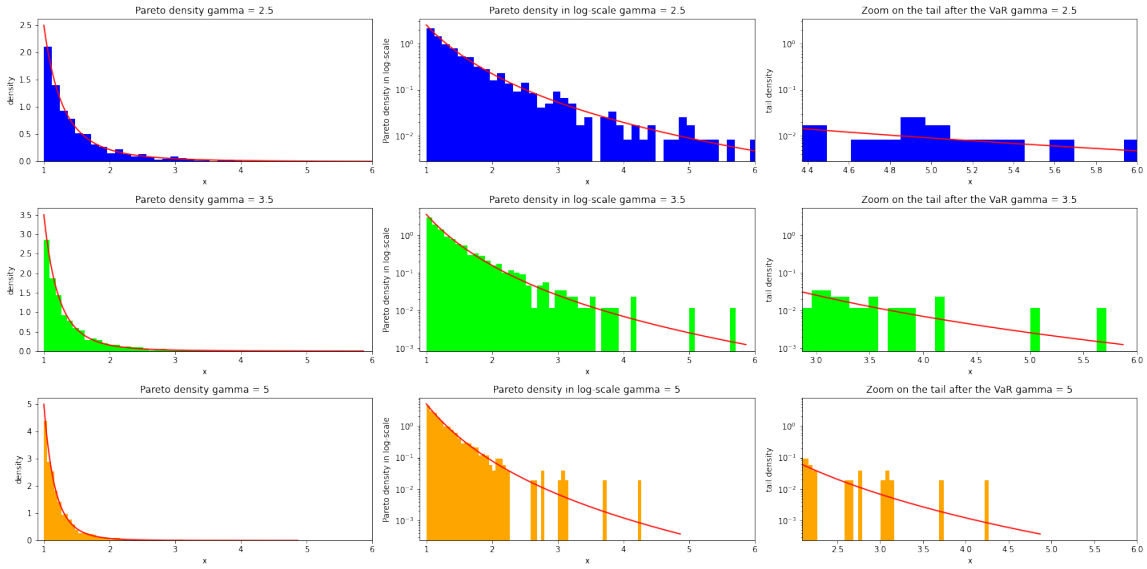


Figure 4.1: The first two columns represent the Pareto distribution for three different Pareto indices $\gamma = 2.5, 3.5, 5, x_m = 1$ from the heaviest to the thinnest tail, and histograms based on $mc = 1000$ i.i.d. samples. The third column represents the Pareto's tail distribution above the empirical α-quantile with $\alpha = 0.975$.

In Figure 4.3, we represent standardized Pareto density functions $\mathcal{P}(1, \gamma)$ for different shape parameters γ varying from 0.5 to 10. The scale parameter of the standardized Pareto distribution is equal to 1, this implies that the minimal value of the distribution support is equal to 1. If the scale parameter is modified with another positive value, then the Pareto distribution $\mathcal{P}(x_m, \gamma)$ is shifted on the left if $0 < x_m < 1$ and on the right if $x_m > 1$. Changing the scale parameter of a Pareto distribution is equivalent to proceed to a scaling operation of the initial Pareto distribution with a scaling factor equal to the ratio between the new scale parameter and the former one. The scaling property of the Pareto distribution establishes a proportionality relationship between the Pareto distributions with the same shape parameter γ but different scale parameters

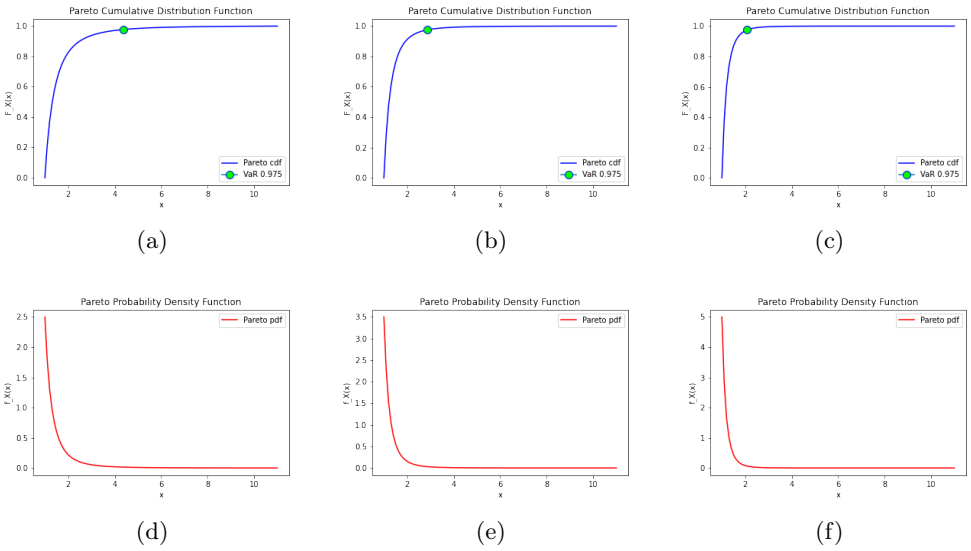


Figure 4.2: The first row represents the Pareto cumulative distribution function and the second row represents the Pareto probability distribution function for three Pareto indices: (a-d) $\gamma = 2.5$ (b-e) $\gamma = 3.5$ (c-f) $\gamma = 5$.

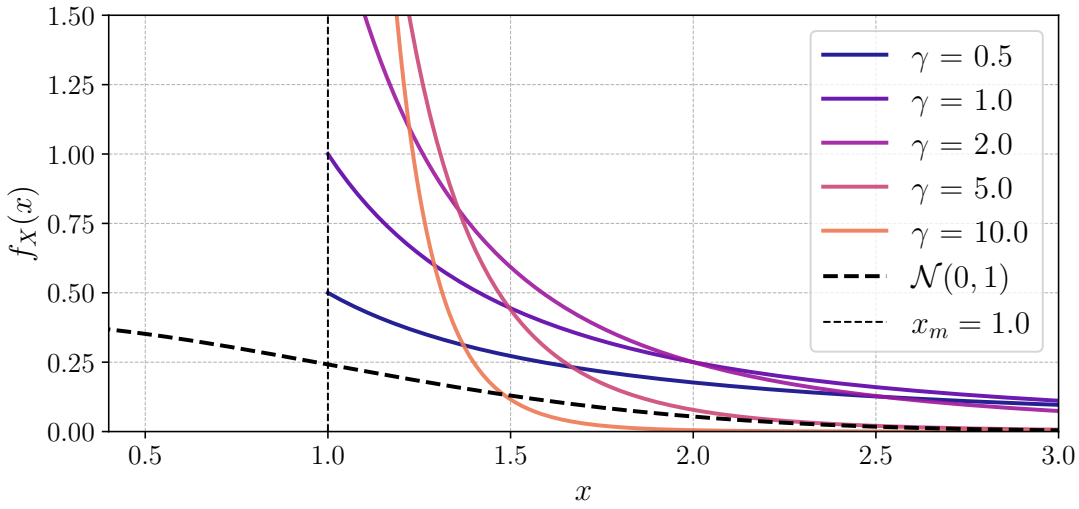


Figure 4.3: Standardized Pareto density function for different Pareto indices γ in comparison with the standard normal distribution.

x_m, with a proportionality factor equal to the ratio between the new scale parameter and the initial scale parameter. The lower the Pareto index, the thicker the tail of the distribution, whereas the lower the Pareto index, the thinner the tail of the distribution and the closer it is to the standard normal distribution.

One of the most interesting properties of Pareto distribution is its *scaling property*. Indeed, it allows establishing links between two Pareto distributions with the same shape parameter γ but different scale parameters x_m. For instance, the scaling property allows switching from the standardized Pareto distribution $\mathcal{P}(1, \gamma)$, to any non-standardized Pareto distribution $\mathcal{P}(x_m, \gamma)$, by a simple multiplication of the standardized Pareto dis-

tribution with the scaling parameter x_m of the non-standardized Pareto distribution. Conversely, the scaling property allows switching from any non-standardized Pareto distribution $\mathcal{P}(x_m, \gamma)$ to the standardized Pareto distribution $\mathcal{P}(1, \gamma)$, by a simple division of the non-standardized Pareto distribution by its scaling parameter x_m. More precisely, the scaling property establishes a proportionality relationship between the non-standardized and the standardized Pareto distributions, with a proportionality factor equal to the scale parameter x_m of the non-standardized Pareto distribution. Conversely, the scaling property allows switching from the non-standardized Pareto distribution $\mathcal{P}(x_m, \gamma)$ to the standardized Pareto distribution $\mathcal{P}(1, \gamma)$, by dividing the non-standardized Pareto distribution by its scale parameter x_m. Consequently, there exists a proportionality relationship between the standardized Pareto distribution $\mathcal{P}(1, \gamma)$ and the non-standardized Pareto distribution $\mathcal{P}(x_m, \gamma)$, with a proportionality factor equal to $\frac{1}{x_m}$. More generally, the scaling property establishes a link between any non-standardized Pareto distributions $\mathcal{P}(x_m^{(1)}, \gamma)$ and $\mathcal{P}(x_m^{(2)}, \gamma)$ with $x_m^{(1)} > 0$ and $x_m^{(2)} > 0$. Indeed, the Pareto distribution $\mathcal{P}(x_m^{(1)}, \gamma)$ is proportional to the Pareto distribution $\mathcal{P}(x_m^{(2)}, \gamma)$ with a proportionality factor equal to the ratio between the two scaling parameters $\frac{x_m^{(1)}}{x_m^{(2)}}$, and vice versa. The proportionality factor is a ratio whose numerator corresponds to the scale parameter of the non-standardized Pareto distribution that we want to reach $\mathcal{P}(x_m^{(1)}, \gamma)$ (the target distribution), and denominator corresponds to the scale parameter of the initial non-standardized Pareto distribution $\mathcal{P}(x_m^{(2)}, \gamma)$. The division of the non-standardized Pareto distribution $\mathcal{P}(x_m^{(2)}, \gamma)$ by its own scaling parameter $x_m^{(2)}$ allows the standardization of the distribution, thus reaching the standardized Pareto distribution $\mathcal{P}(1, \gamma)$. Then, the multiplication of the standardized Pareto distribution by $x_m^{(1)}$ allows reaching the desired Pareto distribution $\mathcal{P}(x_m^{(1)}, \gamma)$. Conversely, the Pareto distribution $\mathcal{P}(x_m^{(2)}, \gamma)$ is proportional to the Pareto distribution $\mathcal{P}(x_m^{(1)}, \gamma)$ with a proportionality factor equal to the ratio between the two scaling parameters $\frac{x_m^{(2)}}{x_m^{(1)}}$. The proportionality factor is a ratio whose numerator corresponds to the scale parameter of the non-standardized Pareto distribution that we want to reach $\mathcal{P}(x_m^{(2)}, \gamma)$, and denominator corresponds to the scale parameter of the initial non-standardized Pareto distribution $\mathcal{P}(x_m^{(1)}, \gamma)$. The division of the Pareto distribution $\mathcal{P}(x_m^{(1)}, \gamma)$ by its own scaling parameter $x_m^{(1)}$ allows the standardization of the distribution, thus reaching the standardized Pareto distribution $\mathcal{P}(1, \gamma)$. Then, the multiplication of the standardized Pareto distribution by $x_m^{(2)}$ allows reaching the desired Pareto distribution $\mathcal{P}(x_m^{(2)}, \gamma)$. Consequently, Pareto distribution is scale-invariant, which means that scaling a Pareto distribution using a constant parameter does not change the shape of the distribution (γ remains the same). The new scaled distribution is still a Pareto distribution with the same shape parameter γ but a new scale parameter. The new Pareto distribution is obtained by multiplying the initial Pareto distribution by the ratio between the scaling parameter of the target Pareto distribution and that of the initial Pareto distribution.

In addition to the scaling property, the Pareto distribution satisfies the *property of stability by conditioning*. This implies that the tail of any Pareto distribution above a certain positive threshold is still a Pareto distribution with the same shape parameter but a new scaling parameter. For instance, the tail of any standardized Pareto distribution $\mathcal{P}(1, \gamma)$ above a given positive threshold $s_m^{(1)} > 0$, is still a Pareto distri-

bution with the same shape parameter γ but a new scaling parameter equal to the conditioning parameter $s_m^{(1)}$: $\mathcal{P}(s_m^{(1)}, \gamma)$. Owing to the scaling property, there exists a proportionality relationship linking the non-standardized Pareto distribution $\mathcal{P}(s_m^{(1)}, \gamma)$ and the standardized Pareto distribution $\mathcal{P}(1, \gamma)$ with a proportionality factor equal to $s_m^{(1)}$. Consequently, by combining both the stability by conditioning and the scaling properties, a proportionality relationship is established between the standardized Pareto distribution conditional on its values being above the threshold $s_m^{(1)}$ and the marginal Pareto distribution $\mathcal{P}(1, \gamma)$, with a proportionality factor equal to the conditioning parameter $s_m^{(1)}$. Consequently, any standardized Pareto distribution $\mathcal{P}(1, \gamma)$ conditional on its values being above a given threshold $s_m^{(1)}$ is still a Pareto distribution with the same shape parameter γ but with a new scaling parameter equal to the conditioning parameter $s_m^{(1)}$.

More generally, the tail of any non-standardized Pareto distribution $\mathcal{P}(x_m^{(1)}, \gamma)$ above a given positive threshold $s_m^{(1)} > x_m^{(1)}$ is still a Pareto distribution with the same shape parameter γ but a new scale parameter equal to the conditioning parameter $s_m^{(1)}$: $\mathcal{P}(s_m^{(1)}, \gamma)$. Owing to the scaling property, there exists a proportionality relationship between the non-standardized Pareto distribution $\mathcal{P}(s_m^{(1)}, \gamma)$ and the standardized Pareto distribution $\mathcal{P}(1, \gamma)$, with a proportionality factor equal to $s_m^{(1)}$. Moreover, the scaling property also establishes a proportionality relationship between the two non-standardized Pareto distributions $\mathcal{P}(s_m^{(1)}, \gamma)$ and $\mathcal{P}(x_m^{(1)}, \gamma)$, with a proportionality factor equal to the ratio between the two scaling parameters $\frac{s_m^{(1)}}{x_m^{(1)}}$. Consequently, by combining the stability by conditioning and the scaling properties, a proportionality relationship is established between the non-standardized Pareto distribution $\mathcal{P}(x_m^{(1)}, \gamma)$ conditional on its values being above the threshold $s_m^{(1)}$ and the standardized Pareto distribution $\mathcal{P}(1, \gamma)$ with a proportionality factor equal to the conditioning parameter $s_m^{(1)}$. And, a proportionality relationship is also established between the non-standardized Pareto distribution $\mathcal{P}(x_m^{(1)}, \gamma)$ conditional on its values being above the threshold $s_m^{(1)}$ and the marginal distribution $\mathcal{P}(x_m^{(1)}, \gamma)$ with a proportionality factor equal to the ratio $\frac{s_m^{(1)}}{x_m^{(1)}}$ where the numerator is the conditioning parameter $s_m^{(1)}$ and denominator is the scaling parameter $x_m^{(1)}$ of the marginal distribution. Consequently, any non-standardized Pareto distribution $\mathcal{P}(x_m^{(1)}, \gamma)$ conditional on its values being above the threshold $s_m^{(1)}$ is still a Pareto distribution with the same shape parameter γ but with a new scaling parameter equal to the conditioning parameter $s_m^{(1)}$.

These properties are very convenient. Moreover, they are specific to the Pareto distribution, that is, they are not shared by all heavy-tailed distributions. They are mathematically formulated in the following theorem.

Theorem 4.7 (Pareto stability by conditioning and rescaling)**.** (i) *Let X be a standard Pareto random variable $X \sim \mathcal{P}(1, \gamma)$, $\gamma > 0$. Let $x_m > 0$ be a new scaling parameter. Let Y be a non-standard Pareto random variable such that $Y \sim \mathcal{P}(x_m, \gamma)$. Then we have:*

$$\mathbb{P}(X \leq x \mid X \geq x_m) = \mathbb{P}(x_m X \leq x) = \mathbb{P}(Y \leq x). \tag{4.20}$$

Equivalently,

$$X \mid X \geq x_m \overset{d}{=} x_m X \overset{d}{=} Y \quad \text{or} \quad \frac{Y}{x_m} \overset{d}{=} \frac{X \mid X \geq x_m}{x_m} \overset{d}{=} X. \tag{4.21}$$

In other words, conditioning a standard Pareto random variable X from a given threshold x_m, amounts to scaling the standardized Pareto distribution, that is, the marginal distribution of X, with the conditioning parameter x_m. Then, the standardized Pareto distribution conditional on its values being above a certain threshold remains a Pareto distribution with the same shape parameter γ but a new scaling parameter equal to the conditioning parameter x_m.

(ii) *Let Z_1, Z_2 be two non-standard Pareto random variables such that $Z_1 \sim \mathcal{P}(x_m^{(1)}, \gamma)$ and $Z_2 \sim \mathcal{P}(x_m^{(2)}, \gamma)$ with $x_m^{(1)} > 0, x_m^{(2)} > 0, \gamma > 0$. Then we have:*

$$Z_1 \overset{d}{=} \frac{x_m^{(1)}}{x_m^{(2)}} Z_2 \overset{d}{=} x_m^{(1)} X \quad \text{or} \quad Z_2 \overset{d}{=} \frac{x_m^{(2)}}{x_m^{(1)}} Z_1 \overset{d}{=} x_m^{(2)} X. \tag{4.22}$$

In other words, any non-standardized Pareto distribution can be expressed from any other non-standardized Pareto distribution by dividing the initial non-standardized Pareto distribution by its own scaling parameter to reach the standardized Pareto distribution $\mathcal{P}(1, \gamma)$, then by scaling it with the scaling parameter of the target distribution.

(iii) *Let $s_m^{(1)}$ be a conditioning parameter such that $s_m^{(1)} > x_m^{(1)} > 0$. Then, Equations (4.21) and (4.22) lead to:*

$$Z_1 \mid Z_1 \geq s_m^{(1)} \overset{d}{=} \frac{s_m^{(1)}}{x_m^{(1)}} Z_1 \overset{d}{=} s_m^{(1)} X. \tag{4.23}$$

In other words, conditioning any non-standardized Pareto distribution $\mathcal{P}(x_m^{(1)}, \gamma)$ from a certain threshold $s_m^{(1)}$ amounts to scaling the standardized Pareto distribution $\mathcal{P}(1, \gamma)$ with the conditioning parameter $s_m^{(1)}$, or equivalently to scaling the marginal distribution $\mathcal{P}(x_m^{(1)}, \gamma)$ with the ratio between the conditioning parameter and the scaling parameter of the marginal distribution $\frac{s_m^{(1)}}{x_m^{(1)}}$.

Proof Let $\gamma > 0$ be a shape parameter and $x_m > 0$ be a scaling parameter.

Let $X \sim \mathcal{P}(1, \gamma)$ be a standardized Pareto r.v. and $Y \sim \mathcal{P}(x_m, \gamma)$ be a non-standardized Pareto distribution. The c.d.f. of X and Y are given by $F_X(x) = \mathbb{P}(X \leq x) = \left(1 - \left(\frac{1}{x}\right)^\gamma\right) \mathbb{1}_{\{x \geq 1\}}$ and $F_Y(x) = \mathbb{P}(Y \leq x) = \left(1 - \left(\frac{x_m}{x}\right)^\gamma\right) \mathbb{1}_{\{x \geq x_m\}}$.

(i) Scaling the standard Pareto r.v. X with the parameter x_m leads to the following distribution: $F_{x_m X}(x) = \mathbb{P}\left(x_m X \leq x\right) = \mathbb{P}\left(X \leq \frac{x}{x_m}\right) = \left(1 - \left(\frac{x_m}{x}\right)^\gamma\right) \mathbb{1}_{\{x \geq x_m\}} = F_Y(x)$. Then, we can deduce the following equality in distribution: $x_m X \overset{d}{=} Y$ or equivalently $X \overset{d}{=} \frac{Y}{x_m}$.

Consequently, the standardized Pareto distribution $\mathcal{P}(1, \gamma)$ scaled with x_m is still a Pareto distribution with shape parameter γ but scale parameter x_m instead of 1: $\mathcal{P}(x_m, \gamma)$.

The non-standardized Pareto distribution $\mathcal{P}(x_m, \gamma)$ is proportional to the standardized Pareto distribution $\mathcal{P}(1, \gamma)$ with a proportionality factor equal to the scaling parameter x_m. Conversely, the standardized Pareto distribution $\mathcal{P}(1, \gamma)$ is proportional to the non-standardized Pareto distribution $\mathcal{P}(x_m, \gamma)$ with a proportionality factor equal to $\frac{1}{x_m}$.

Let us compute the conditional distribution of X given that $X \geq x_m$: for $x \geq x_m$ we have:

$$\begin{aligned}
\mathbb{P}(X \leq x \mid X \geq x_m) &= \frac{\mathbb{P}(x_m \leq X \leq x)}{\mathbb{P}(X \geq x_m)} = \frac{F_X(x) - F_X(x_m)}{1 - F_X(x_m)} \\
&= \frac{\left(1 - \left(\frac{1}{x}\right)^\gamma\right) - \left(1 - \left(\frac{1}{x_m}\right)^\gamma\right)}{1 - \left(1 - \left(\frac{1}{x_m}\right)^\gamma\right)} \\
&= \frac{\left(\frac{1}{x_m}\right)^\gamma - \left(\frac{1}{x}\right)^\gamma}{\left(\frac{1}{x_m}\right)^\gamma} = \left(1 - \left(\frac{x_m}{x}\right)^\gamma\right) \mathbb{1}_{\{x \geq x_m\}} \\
&= F_Y(x).
\end{aligned}$$

And, the scaling property states that $F_Y(x) = F_{x_m X}(x)$, then we can write:

$$\mathbb{P}(X \leq x \mid X \geq x_m) = F_{x_m X}(x).$$

Consequently, we can write: $X \mid X \geq x_m \overset{d}{=} Y \overset{d}{=} x_m X$. Therefore, conditioning a standard Pareto r.v. X above a given parameter x_m leads to get the non-standard Pareto r.v. Y with the same shape parameter γ and a scaling parameter equal to the conditioning parameter x_m. Therefore, the standardized Pareto distribution conditional on its values being above a certain threshold is still a Pareto distribution with the same shape parameter γ but a new scaling parameter equal to the conditioning parameter. Moreover, the scaling property states that there exists a proportionality relationship between the non-standardized Pareto distribution $\mathcal{P}(x_m, \gamma)$ and the standardized $\mathcal{P}(1, \gamma)$ with a proportionality factor equal to x_m. Consequently, the standardized Pareto distribution conditional on its values being above a certain threshold x_m is proportional to the standardized Pareto distribution with a proportionality factor equal to the conditioning parameter x_m. Therefore, conditioning the standardized Pareto distribution from a given threshold x_m amounts to scaling the standardized Pareto distribution with the conditioning parameter x_m. Thus, the Pareto distribution is stable by conditioning.

(ii) Let Z_1, Z_2 be two non-standard Pareto r.v. such that $Z_1 \sim \mathcal{P}(x_m^{(1)}, \gamma)$ and $Z_2 \sim \mathcal{P}(x_m^{(2)}, \gamma)$.

Equality in distribution can be obtained between Z_1 and Z_2. For this purpose, in a first time we standardize Z_2 by dividing it by its own scaling parameter $x_m^{(2)}$ so that the distribution might start from 1. Second, we rescale the standard Pareto r.v. $\frac{Z_2}{x_m^{(2)}}$ with the scaling parameter $x_m^{(1)}$ of Z_1, and conversely. Indeed, from Equation (4.21),

we have:

$$Z_1 \overset{d}{=} x_m^{(1)} X \quad \text{and} \quad X \overset{d}{=} \frac{Z_2}{x_m^{(2)}}, \tag{4.24}$$

hence:

$$Z_1 \overset{d}{=} \frac{x_m^{(1)}}{x_m^{(2)}} Z_2 \quad \text{or equivalently} \quad Z_2 \overset{d}{=} \frac{x_m^{(2)}}{x_m^{(1)}} Z_1. \tag{4.25}$$

The scaling property allows establishing equality in distribution between any non-standardized Pareto distributions. Indeed, any non-standardized Pareto distribution $\mathcal{P}(x_m^{(1)}, \gamma)$ can be expressed from any other non-standardized Pareto distribution $\mathcal{P}(x_m^{(2)}, \gamma)$ by dividing it by its own scaling parameter $x_m^{(2)}$ in order to obtain the standardized Pareto distribution and by scaling the standardized Pareto distribution of $\frac{Z_2}{x_m^{(2)}}$ with the new scaling parameter $x_m^{(1)}$. Therefore, any non-standardized Pareto distribution $\mathcal{P}(x_m^{(1)}, \gamma)$ is proportional to any other non-standardized Pareto distribution $\mathcal{P}(x_m^{(2)}, \gamma)$ with a proportionality factor equal to the ratio of the two scaling parameters $\frac{x_m^{(1)}}{x_m^{(2)}}$, and vice versa.

(iii) Let us prove that a non-standard Pareto r.v. $Z_1 \sim \mathcal{P}(x_m^{(1)}, \gamma)$ conditional on the event $\{Z_1 \geq s_m^{(1)}\}$ with $s_m^{(1)} > x_m^{(1)}$, is still a non-standard Pareto r.v. with a scaling parameter equal to $s_m^{(1)}$. Indeed,

$$\mathbb{P}\left(Z_1 \leq x \mid Z_1 \geq s_m^{(1)}\right) = \frac{\mathbb{P}\left(s_m^{(1)} \leq Z_1 \leq x\right)}{\mathbb{P}\left(Z_1 \geq s_m^{(1)}\right)} = \frac{F_{Z_1}(x) - F_{Z_1}(s_m^{(1)})}{1 - F_{Z_1}(s_m^{(1)})} \tag{4.26}$$

$$= \frac{\left(1 - \left(\frac{x_m^{(1)}}{x}\right)^\gamma\right) - \left(1 - \left(\frac{x_m^{(1)}}{s_m^{(1)}}\right)^\gamma\right)}{\left(1 - \left(\frac{x_m^{(1)}}{s_m^{(1)}}\right)^\gamma\right)} = \left(1 - \left(\frac{s_m^{(1)}}{x}\right)^\gamma\right) \mathbb{1}_{\{x \geq s_m^{(1)}\}} \tag{4.27}$$

$$= F_{s_m^{(1)} X}(x) = F_{\frac{s_m^{(1)}}{x_m^{(1)}} Z_1}(x). \tag{4.28}$$

Consequently, the non-standardized Pareto distribution with as scaling parameter $x_m^{(1)}$ conditional on event $\{Z_1 \geq s_m^{(1)}\}$ is still a Pareto distribution with a scaling parameter equal to the conditioning parameter $s_m^{(1)}$. The scaling property states that the non-standardized Pareto distribution $\mathcal{P}(s_m^{(1)}, \gamma)$ is proportional to the standardized Pareto distribution with a proportionality factor equal to $s_m^{(1)}$. This leads to the following equality in distribution:

$$Z_1 \mid Z_1 \geq s_m^{(1)} \overset{d}{=} s_m^{(1)} X. \tag{4.29}$$

To establish the equality in distribution between the conditional distribution of $Z_1 \mid Z_1 \geq s_m^{(1)}$ and the marginal distribution $\mathcal{P}(x_m^{(1)}, \gamma$, that is the distribution of Z_1, we have to proceed as follows. On the one hand, we standardize the distribution of Z_1

by its own scaling parameter $x_m^{(1)}$ so that the distribution might start from 1 instead of $x_m^{(1)}$. On the other hand, we have to scale the standard Pareto r.v. $\frac{Z_1}{x_m^{(1)}}$ with the conditioning parameter $s_m^{(1)}$ to obtain a Pareto distribution with a scaling parameter equal to $s_m^{(1)}$ that corresponds to the conditional distribution $Z_1 \mid Z_1 \geq s_m^{(1)}$ as proved in the previous point.

Indeed, because $X \overset{d}{=} \frac{Z_1}{x_m^{(1)}}$, we obtain:

$$Z_1 \mid Z_1 \geq s_m^{(1)} \overset{d}{=} \frac{s_m^{(1)}}{x_m^{(1)}} Z_1. \tag{4.30}$$

■

Furthermore, the moments of the Pareto distribution exist only for certain ranges of shape parameter γ. Specifically, the k-th moment exists if and only if $\gamma > k$. This means that for small values of γ, moments may not exist, which is related to the heavy-tailed nature of the distribution. As the shape parameter γ decreases, the distribution becomes even more heavy-tailed, giving more weight to extreme events. This leads to finite moments for only very small orders. This is a characteristic of heavy-tailed distributions, where extreme observations have a more significant impact on higher moments than distributions with lighter tails. For instance, the Pareto distribution has an infinite mean for shape parameter $\gamma \leq 1$. For $\gamma \leq 2$, the Pareto distribution has a finite mean but an infinite variance. For $\gamma > 2$, the variance of the Pareto distribution exists and is finite. This means that if the shape parameter is greater than 2, the distribution is characterized by both a finite mean and finite variance, making it more manageable in certain statistical analyses.

The scaling property of the Pareto distribution states that any non-standardized Pareto distribution $\mathcal{P}(x_m, \gamma)$ is proportional to the standardized Pareto distribution $\mathcal{P}(1, \gamma)$ with a proportionality factor equal to x_m. This implies that all the quantities (statistics) computed on the non-standardized Pareto distribution $\mathcal{P}(x_m, \gamma)$ are proportional to the ones computed on the standardized Pareto distribution $\mathcal{P}(1, \gamma)$, with a proportionality factor equal to x_m. Therefore, the calculus can be carried out on the standardized Pareto distribution, and the equivalent quantities on any non-standardized Pareto distribution $\mathcal{P}(x_m, \gamma)$, can be recovered by the multiplication of the standardized quantities with the proper scaling parameter x_m. For instance, the scaling property of the Pareto distribution allows defining any k-th moment of any non-standardized Pareto distribution $\mathcal{P}(x_m, \gamma)$ by a simple multiplication of the k-th moment of the standardized Pareto distribution $\mathcal{P}(1, \gamma)$ with the scaling parameter x_m raised to the power k.

These properties are mathematically formulated in the following lemma.

Lemma 4.8 (Moments of Pareto distribution)**.** *Let $X \sim \mathcal{P}(1, \gamma)$, $Y \sim \mathcal{P}(x_m, \gamma)$ be two Pareto random variables, with $\gamma > k$, $k \in \mathbb{N}^\star$. Subsequently, the Pareto random variables X and Y admit the finite first k moments.*

(i) *Moments of order k ($\gamma > k$):*

$$\mathbb{E}\left[X^k\right] = \frac{\gamma}{\gamma - k} \quad \textit{and} \quad \mathbb{E}\left[Y^k\right] = x_m^k \mathbb{E}\left[X^k\right]. \tag{4.31}$$

(ii) *Central moment of order k ($\gamma > k$):*

$$\mathbb{E}\left[(X - \mathbb{E}[X])^k\right] = \sum_{j=0}^{k} \binom{k}{j} \frac{\gamma^{j+1}}{(1-\gamma)^j(\gamma + j - k)} \quad \textit{and} \quad \mathbb{E}\left[(Y - \mathbb{E}[Y])^k\right] = x_m^k \mathbb{E}\left[(X - \mathbb{E}[X])^k\right]. \tag{4.32}$$

Proof In view of the scaling property of Theorem 4.7 $x_m X \overset{d}{=} Y$. This implies that $\mathbb{E}\left[Y^k\right] = \mathbb{E}\left[(x_m X)^k\right] = x_m^k \mathbb{E}\left[X^k\right]$. Thus, it is sufficient to prove the formula for X. Let $X \sim \mathcal{P}(1, \gamma)$. Recall that $f_X(x) = \frac{\gamma}{x^{\gamma+1}} \mathbb{1}_{\{x \geq 1\}}$.

(i) The first k moments of X are given by:

$$\mathbb{E}[X^k] = \gamma \int_1^{+\infty} x^{k-\gamma-1} \mathrm{d}x = \begin{cases} \frac{\gamma}{\gamma - k} & \text{if } \gamma > k \\ +\infty & \text{if } \gamma \leq k. \end{cases}$$

(ii) The first k central moments of X are defined as:

$$\mathbb{E}[(X - \mathbb{E}[X])^k] = \sum_{j=0}^{k} \binom{k}{j} \mathbb{E}[X^{k-j}(-\mathbb{E}[X])^j] = \sum_{j=0}^{k} \binom{k}{j} \frac{\gamma^{j+1}}{(1-\gamma)^j(\gamma + j - k)}.$$

■

These properties hold for all statistics computed on the Pareto distribution, particularly for the $\mathbf{VaR}_\alpha$ and the $\mathbf{ES}_\alpha$. Let $X \sim \mathcal{P}(1, \gamma)$ be a standardized Pareto random variable and $Y \sim \mathcal{P}(x_m, \gamma)$, $Z_1 \sim \mathcal{P}(x_m^{(1)}, \gamma)$ and $Z_2 \sim \mathcal{P}(x_m^{(2)}, \gamma)$ be non-standardized Pareto random variables. The scaling property states that the non-standardized Pareto distribution $\mathcal{P}(x_m, \gamma)$ is proportional to the standardized Pareto distribution $\mathcal{P}(1, \gamma)$, with a proportionality factor equal to x_m. This implies that the $\mathbf{VaR}_\alpha$ of the non-standardized Pareto distribution $\mathcal{P}(x_m, \gamma)$ is proportional to the $\mathbf{VaR}_\alpha$ of the standardized Pareto distribution with a proportionality coefficient equal to the scaling parameter x_m. Similarly, the scaling property of the Pareto distribution states that the non-standardized Pareto distribution $\mathcal{P}(x_m^{(1)}, \gamma)$ is proportional to the non-standardized Pareto distribution $\mathcal{P}(x_m^2, \gamma)$, with a proportionality factor equal to the ratio between the two scaling parameters $\frac{x_m^{(1)}}{x_m^{(2)}}$. This implies that the $\mathbf{VaR}_\alpha$ of the non-standardized Pareto distribution $\mathcal{P}(x_m^{(1)}, \gamma)$ is proportional to the $\mathbf{VaR}_\alpha$ of the non-standardized Pareto distribution $\mathcal{P}(x_m^{(2)}, \gamma)$ with a proportionality coefficient equal to $\frac{x_m^{(1)}}{x_m^{(2)}}$.

On the other hand, the property of stability by conditioning states that the standardized Pareto distribution $\mathcal{P}(1, \gamma)$ conditional on its values being above a certain threshold, is still a Pareto distribution with the same shape parameter γ but a new scale parameter equal to the conditioning parameter. Moreover, the scaling property implies that the

standardized Pareto distribution $\mathcal{P}(1,\gamma)$ conditional on its values being above a certain threshold, is proportional to the standardized Pareto distribution with a proportionality factor equal to the conditioning threshold. The Expected-Shortfall of the standardized Pareto distribution is the conditional expectation given that the standard Pareto random variable X is larger than $\mathbf{VaR}_\alpha(X)$: $\mathbf{ES}_\alpha(X) = \mathbb{E}\Big[X \mid X \geq \mathbf{VaR}_\alpha(X)\Big]$ (defined by (4.3)). Combining the stability by conditioning and the scaling properties leads to a proportionality relationship between the standardized Pareto distribution $\mathcal{P}(1,\gamma)$ conditional on its values being above $\mathbf{VaR}_\alpha(X)$, and the marginal distribution which is the standardized Pareto distribution $\mathcal{P}(1,\gamma)$, with a proportionality factor equal to the conditioning threshold $\mathbf{VaR}_\alpha(X)$. Consequently, the Expected-Shortfall of the standardized Pareto distribution $\mathbf{ES}_\alpha(X)$ is proportional to its expectation $\mathbb{E}\big[X\big]$ with a proportionality factor equal to $\mathbf{VaR}_\alpha(X)$, which is supposed to be known. Therefore, $\mathbf{ES}_\alpha(X)$ can be easily recovered owing to the simple scaling by $\mathbf{VaR}_\alpha$ of the expectation of the standardized Pareto distribution.

Similarly, the property of stability by conditioning states that the non-standardized Pareto distribution $Z_1 \sim \mathcal{P}(x_m^{(1)},\gamma)$ conditional on its values being above a certain threshold, is still a Pareto distribution with the same shape parameter γ but a new scale parameter equal to the conditioning parameter. Moreover, the scaling property implies that the non-standardized Pareto distribution $\mathcal{P}(x_m^{(1)},\gamma)$ conditional on its values being above a certain threshold, is proportional to the standardized Pareto distribution $\mathcal{P}(1,\gamma)$ with a proportionality factor equal to the conditioning parameter, and is also proportional to the marginal Pareto distribution $\mathcal{P}(x_m^{(1)},\gamma)$ with a proportionality factor equal to the ratio between the conditioning parameter and the scaling parameter $x_m^{(1)}$. Combining the stability by conditioning and the scaling properties implies that the Expected-Shortfall of the non-standardized Pareto distribution $\mathcal{P}(x_m^{(1)},\gamma)$, which is the expectation of the distribution $\mathcal{P}(x_m^{(1)},\gamma)$ conditional on its values being above $\mathbf{VaR}_\alpha(Z_1)$, is proportional to the expectation of the standardized Pareto distribution $\mathcal{P}(1,\gamma)$ with a proportionality factor equal to conditioning parameter $\mathbf{VaR}_\alpha(Z_1)$, and is proportional to the expectation of the marginal Pareto distribution $\mathcal{P}(x_m^{(1)},\gamma)$ with a proportionality factor equal to the ratio $\frac{\mathbf{VaR}_\alpha(Z_1)}{x_m^{(1)}}$. Therefore, the Expected-Shortfall of any non-standardized Pareto distribution $\mathcal{P}(x_m^{(1)},\gamma)$ can be recovered by a simple scaling of the expectation of the standardized Pareto distribution $\mathcal{P}(1,\gamma)$ with the conditioning threshold $\mathbf{VaR}_\alpha(Z_1)$, and by a simple scaling of the expectation of the marginal Pareto distribution $\mathcal{P}(x_m^{(1)},\gamma)$ with the ratio $\frac{\mathbf{VaR}_\alpha(Z_1)}{x_m^{(1)}}$.

All these properties are mathematically formulated in the following proposition.

Proposition 4.9 (Pareto $\mathbf{VaR}_\alpha$ and $\mathbf{ES}_\alpha$)**.** *If $X \sim \mathcal{P}(1,\gamma)$ and $Y \sim \mathcal{P}(x_m,\gamma)$, $x_m > 0, \gamma > 0$, then their respective $\mathbf{VaR}_\alpha$ are defined as follows:*

$$\begin{cases} \mathbf{VaR}_\alpha(X) &= F_X^{-1}(\alpha) = (1-\alpha)^{-\frac{1}{\gamma}}, \\ \mathbf{VaR}_\alpha(Y) &= F_Y^{-1}(\alpha) = x_m \mathbf{VaR}_\alpha(X), \end{cases} \quad \alpha \in (0,1), \tag{4.33}$$

and the respective $\mathbf{ES}_\alpha$ *are given by:*

$$\begin{cases} \mathbf{ES}_\alpha(X) = \mathbf{VaR}_\alpha(X)\mathbb{E}\left[X\right] = \frac{\gamma}{\gamma-1}(1-\alpha)^{-\frac{1}{\gamma}}, \\ \mathbf{ES}_\alpha(Y) = x_m \mathbf{ES}_\alpha(X), \end{cases} \qquad \alpha \in (0,1). \tag{4.34}$$

Proof Let $X \sim \mathcal{P}(1,\gamma)$ and $Y \sim \mathcal{P}(x_m,\gamma)$.

(i) Recall from Definition 4.1, Equation (4.2), that $\mathbf{VaR}_\alpha(X) := \inf\{x \in \mathbb{R}, F_X(x) \geq \alpha\}$ and $F_X(x) = \left(1 - \left(\frac{1}{x}\right)^\gamma\right) \mathbb{1}_{\{x \geq 1\}}$. Because the distribution of interest is the Pareto distribution, we work with a continuous distribution. We solve the following equation:

$$F_X(x) = \alpha \quad \text{i.e.} \quad 1 - \left(\frac{1}{x}\right)^\gamma = \alpha, \quad x \geq 1. \tag{4.35}$$

This leads to:

$$x = (1-\alpha)^{-\frac{1}{\gamma}}. \tag{4.36}$$

Consequently,

$$\mathbf{VaR}_\alpha(X) = (1-\alpha)^{-\frac{1}{\gamma}}. \tag{4.37}$$

(ii) From the scaling property we obtain $Y \overset{d}{=} x_m X$, which leads to:

$$F_Y(x) = F_{x_m X}(x) = F_X\left(\frac{x}{x_m}\right). \tag{4.38}$$

This implies that solving the following equation:

$$F_Y(x) = \alpha \quad \text{amounts to} \quad F_X\left(\frac{x}{x_m}\right) = \alpha. \tag{4.39}$$

From the previous case, we fall onto:

$$\frac{x}{x_m} = \mathbf{VaR}_\alpha(X) \quad \text{i.e.} \quad x = x_m \mathbf{VaR}_\alpha(X). \tag{4.40}$$

Consequently, the scaling property of the Pareto distribution establishes a proportionality relationship between the $\mathbf{VaR}_\alpha(Y)$ of the non-standardized Pareto distribution $\mathcal{P}(x_m,\gamma)$ and the $\mathbf{VaR}_\alpha(X)$ of the standardized Pareto distribution $\mathcal{P}(1,\gamma)$ with a proportionality factor equal to the scaling parameter x_m.

(iii) Recall from Definition 4.1, Equation (4.3) that: $\mathbf{ES}_\alpha(X) = \mathbb{E}\left[X \mid X \geq \mathbf{VaR}_\alpha(X)\right]$. From the stability by conditioning and the scaling properties of the Pareto distribution stated in Theorem 4.7, Equation (4.21), the following equality in distribution holds:

$$X \mid X \geq \mathbf{VaR}_\alpha(X) \overset{d}{=} \mathbf{VaR}_\alpha(X) X. \tag{4.41}$$

This leads to:

$$\mathbb{E}\left[X \mid X \geq \mathbf{VaR}_\alpha(X)\right] = \mathbf{VaR}_\alpha(X)\mathbb{E}\left[X\right]. \tag{4.42}$$

Consequently:

$$\mathbf{ES}_\alpha(X) = \mathbf{VaR}_\alpha(X)\mathbb{E}\left[X\right]. \tag{4.43}$$

From Lemma 4.8, Equation (4.31), we have $\mathbb{E}\left[X\right] = \frac{\gamma}{\gamma-1}$, and from Proposition 4.9, Equation (4.33) we have $\mathbf{VaR}_\alpha(X) = (1-\alpha)^{-\frac{1}{\gamma}}$, thus we obtain:

$$\mathbf{ES}_\alpha(X) = \frac{\gamma}{\gamma-1}(1-\alpha)^{-\frac{1}{\gamma}}. \tag{4.44}$$

(iv) Recall that $Y \sim \mathcal{P}(x_m, \gamma)$ and $\mathbf{ES}_\alpha(Y) = \mathbb{E}\left[Y \mid Y \geq \mathbf{VaR}_\alpha(Y)\right]$. From the stability by conditioning and scaling properties stated in Theorem 4.7, Equation (4.23), we can write the following equality in distribution:

$$Y \mid Y \geq \mathbf{VaR}_\alpha(Y) \overset{d}{=} \frac{\mathbf{VaR}_\alpha(Y)}{x_m} Y \overset{d}{=} \mathbf{VaR}_\alpha(Y)X. \tag{4.45}$$

This implies that:

$$\mathbb{E}\left[Y \mid Y \geq \mathbf{VaR}_\alpha(Y)\right] = \mathbf{VaR}_\alpha(Y)\mathbb{E}\left[X\right] = x_m\mathbf{VaR}_\alpha(X)\mathbb{E}\left[X\right]. \tag{4.46}$$

Therefore:

$$\mathbf{ES}_\alpha(Y) = \mathbf{VaR}_\alpha(Y)\mathbb{E}\left[X\right] = x_m\mathbf{VaR}_\alpha(X)\mathbb{E}\left[X\right]. \tag{4.47}$$

Since $\mathbf{VaR}_\alpha(X) = (1-\alpha)^{-\frac{1}{\gamma}}$ and $\mathbb{E}\left[X\right] = \frac{\gamma}{\gamma-1}$, then we finally get:

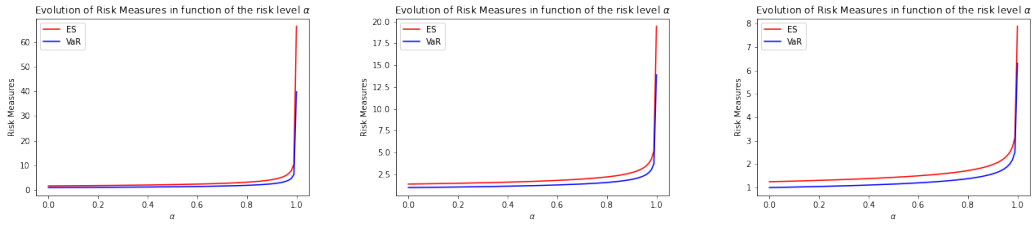
$$\mathbf{ES}_\alpha(Y) = x_m\frac{\gamma}{\gamma-1}(1-\alpha)^{-\frac{1}{\gamma}}. \tag{4.48}$$

■

As previously mentioned, owing to the scaling property of the Pareto distribution, all the statistics of any non-standardized Pareto distribution are proportional to the corresponding statistics of the standardized Pareto distribution, with a proportionality factor equal to the scaling parameter of the target Pareto distribution. Consequently, in the sequel, all the statistics will be computed using the standardized Pareto distribution. A simple scaling of these standardized statistics by the proper scaling parameter of the target Pareto distribution will allow for the recovery of the corresponding statistics of the desired non-standardized target Pareto distribution.

As previously explained, owing to the property of stability by conditioning, the Expected-Shortfall of the standardized Pareto distribution is proportional to its expectation, with a proportionality factor equal to the Value-at-Risk of the standardized Pareto distribution $\mathbf{VaR}_\alpha$, which is supposed to be known. Therefore, the main step in computing the Expected-Shortfall of the standardized Pareto distribution, is to compute its expectation. The Expected-Shortfall of the standardized Pareto distribution can be recovered thanks to a simple scaling of this expectation by the $\mathbf{VaR}_\alpha$, which is supposed to be known.

In addition, in Figure 4.4, we compare the evolution of the $\mathbf{VaR}_\alpha$ and $\mathbf{ES}_\alpha$ as functions of the risk level α.



(a) (b) (c)

Figure 4.4: $\mathbf{VaR}_\alpha$ and $\mathbf{ES}_\alpha$ as functions of α. (a) $\gamma = 2.5$ (b) $\gamma = 3.5$ (c) $\gamma = 5$.

4.1.3 Issues

The presence of extreme values with low probability in a heavy-tailed distribution, can influence the estimator (empirical mean) of the expectation and lead to poor performance. In this context, the goal is to determine an estimator of the expectation that can be more resistant to extreme values. For that purpose, different mean estimators will be studied based on their concentration and fluctuation properties. An overview of this study is shown in Figure 4.5.

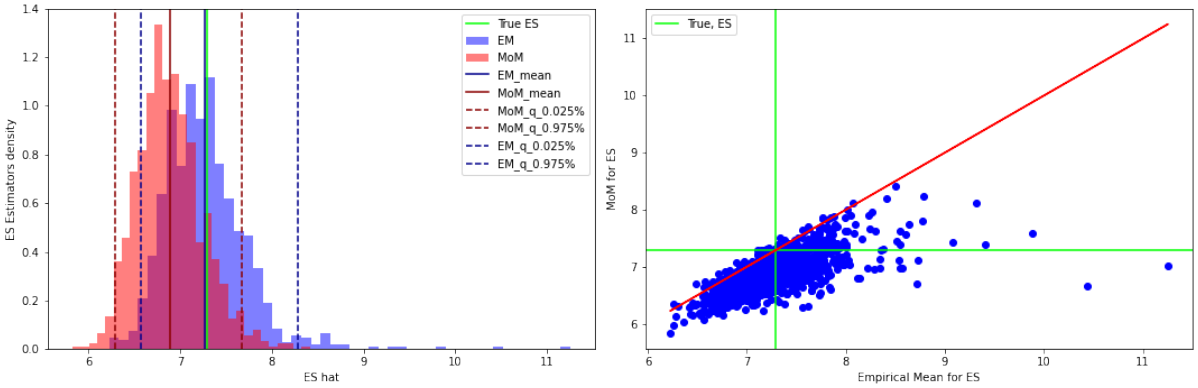


Figure 4.5: Left: Distribution of the Empirical Mean estimator versus distribution of the Median-of-Means estimator for the Expected-Shortfall. Right: Scatterplot of the Median-of-Means estimator versus the Empirical-Mean estimator for the Expected-Shortfall.

Definition 4.10 (Robust statistics)**.** *Robust statistics are statistics with good performance for data drawn from a wide range of probability distributions, especially for distributions that are far from being normal. The goal is to produce statistical methods that are not unduly affected by outliers, or to provide methods with good performance when there are small departures from a parametric distribution.*

(i) *Strictly speaking, a robust statistic is resistant to errors in the results, produced by deviations from assumptions (e.g., of normality). This means that if the assumptions are only approximately met, the robust estimator will still have a reasonable efficiency, and reasonably small bias, as well as being asymptotically unbiased (i.e., having a bias tending towards* 0 *as the sample size tends towards infinity).*

(ii) *Usually, the most important case is distributional robustness - robustness to breaking assumptions about the underlying distribution of the data. Classical statistical procedures are typically sensitive to longtailedness (e.g., when the distribution of the data has longer tails than the assumed normal distribution). This implies that they will be strongly affected by the presence of outliers in the data, and the estimates they produce may be heavily distorted if there are extreme outliers in the data, compared to what they would be if the outliers were not included in the data.*

(iii) *In contrast, more robust estimators that are not very sensitive to distributional distortions, such as longtailedness, are resistant to the presence of outliers. Thus, in the context of robust statistics, distributionally robust and outlier-resistant are synonymous. Some experts prefer the term resistant statistics for distributional robustness, and reserve robustness for non-distributional robustness, such as, robustness to violation of assumptions about the probability model or estimator, but this is a minority usage. Plain robustness to mean distributional robustness is common. When considering how robust an estimator is to the presence of outliers, it is useful to test what happens when an extreme outlier is added to the dataset and to test what happens when an extreme outlier replaces one of the existing data points, and then to consider the effect of multiple additions or replacements.*

Ideally, an estimator is expected to satisfy the following constraints: be robust to heavy tailed-distribution, that is, to the presence of extreme values, and reach a high-level of accuracy with a high-level of confidence. In other words, we are interested in assessing, for any sample size n, and confidence parameter $\delta \in (0, 1)$, the smallest possible value $\epsilon = \epsilon(n, \delta)$ such that:

$$\mathbb{P}(\left|\hat{\mu}_n - \mu\right| > \epsilon) \leq \delta \tag{4.49}$$

where $\hat{\mu}_n$ is a mean-estimator computed on a sample of size n and μ is the true mean. It should be noted that this criterion is non-asymptotic. These constraints are satisfied by sub-Gaussian estimators.

A random variable is considered sub-Gaussian if its tail behavior is no worse than that of a Gaussian random variable. In other words, it does not have heavy tails and decays reasonably quickly, similar to a Gaussian distribution.

An estimator is said to be sub-Gaussian if the distribution of its deviation is sub-Gaussian. Sub-Gaussian deviations are desirable because they allow for better control and analysis of the estimation performance. Indeed, sub-Gaussian estimators enjoy concentration inequalities such as Hoeffding's inequality and Bernstein's inequality. These inequalities provide bounds on the estimator's deviation from its expected value with high probability. Consequently, sub-Gaussian estimators tend to be more stable and predictable.

Moreover, sub-Gaussian estimators are more robust to outliers and extreme observations in data. They are less affected by extreme values, which makes them suitable for noisy or imperfect datasets.

Finally, sub-Gaussian estimators often provide tighter error bounds, allowing for a more precise quantification of estimation accuracy.

Mathematically speaking, a sub-Gaussian estimator is defined as follows.

Definition 4.11 (Sub-Gaussian estimator)**.** *A mean estimator $\hat{\mu}_n$ is said to be $L-$sub-Gaussian if the distribution in which the random variables are drawn admits a finite second order moment σ^2, and if there exists a constant $L > 0$ such that for any sample size n with probability at least $1 - \delta$:*

$$\left|\hat{\mu}_n - \mu\right| \leq \frac{L\sigma\sqrt{\log(2/\delta)}}{\sqrt{n}}. \tag{4.50}$$

or equivalently,

$$\mathbb{P}\left(\left|\hat{\mu}_n - \mu\right| \leq \epsilon\right) \geq 1 - 2e^{-\frac{n\epsilon^2}{L^2\sigma^2}} \tag{4.51}$$

where ϵ quantifies the deviation.

Despite the advantages of sub-Gaussian estimators, they also present some limitations. Indeed, proving that an estimator is sub-Gaussian is not so obvious because it relies on restrictive assumptions on the models, or it requires some tips to develop a new estimator such as Median-of-Means (see later in this chapter). Moreover, some sub-Gaussian estimators might have a higher computational complexity than simpler estimators. This can be a drawback when dealing with large datasets or in real-time applications.

In the following section, we will study different types of mean estimators with the aim of estimating the Expected-Shortfall in the Pareto distribution. Because the stability by conditioning and the scaling properties of the Pareto distribution establish a proportionality relationship between the Expected-Shortfall and the expectation of the marginal distribution with a proportionality factor equal to the α-quantile, we will be interested in the two following cases. The first case consists of estimating the expectation of a random variable on the support of the entire distribution. The second corresponds to the estimation of the expectation of a random variable above a certain threshold, the α-quantile. In the second case, for each estimator, two configurations will be studied. The first configuration corresponds to the idealized case, in which the empirical α-quantile match the true $\mathbf{VaR}_\alpha$. This case is reached only in the asymptotic framework. The second case is the realistic case and the most common one, in which the empirical α-quantile differs from the true $\mathbf{VaR}_\alpha$. In the last case, the estimation of the expectation above the given threshold is more challenging because the threshold above which we consider the excess loss is an order statistics and depends on the underlying sample. This implies that the samples larger than the empirical α-quantile are no longer i.i.d. and the distribution of the excess loss is unkown. Moreover, the realistic case introduces, in the estimation of the $\mathbf{ES}_\alpha$, an additional error term corresponding to the bias estimation between the true $\mathbf{VaR}_\alpha$ and the empirical α-quantile.

4.2 Empirical Mean (EM) for heavy-tailed distributions

4.2.1 Presentation and properties of the Empirical Mean

Usually, an obvious choice for the mean estimator is the empirical mean, defined by:

$$\bar{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{4.52}$$

with a convergence speed of $\frac{1}{\sqrt{n}}$. However, this estimator does not satisfy the aforementioned constraints.

First, the empirical mean is not robust. Such an estimator is influenced by the presence of extreme values in the distribution, which leads to poor performance.

Then, the deviation of the empirical mean estimator with respect to the expectation of the distribution can reach a Gaussian regime, but the amount of data required to reach this regime is so large that it can be qualified as an asymptotic regime. This is what the Fuk-Nagaev inequality shows. Indeed, the Fuk-Nagaev inequality is a useful concentration inequality for sums of independent heavy-tailed random variables with only a limited number of finite moments, as it provides an upper bound on the tail probability of the sum of random variables, based on the variance proxies of the individual random variables. This inequality is especially powerful when dealing with large sums of heavy-tailed random variables, as it allows us to control the tail behavior of the sum and understand its deviation from its expected value with high probability.

From (Rio, 2017) the Fuk-Nagaev inequality is stated as follows.

Lemma 4.12 (Fuk-Nagaev inequality)**.** *If $X_1, \dots, X_n$ are n i.i.d. random variables following the same distribution as an r.v. X that admits a finite moment of order $\kappa > 2$ (i.e., $\mathbb{E}[|X|^\kappa] < +\infty$). Then, the Fuk-Nagaev inequality states that:*

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}\left[X\right] \geq \frac{\lambda}{\sqrt{n}}\right) \leq \left(\frac{\kappa+2}{\kappa}\right)^{\kappa} \frac{\mathbb{E}[(X - \mathbb{E}\left[X\right])_+^\kappa]}{n^{\frac{\kappa}{2}-1}\lambda^\kappa} + e^{-\frac{2\lambda^2}{(\kappa+2)^2 e^\kappa \mathbb{V}[X]}} \tag{4.53}$$

where $\lambda > 0$ is the deviation related to the sum of the recentered random variables and $\mathbb{V}[X]$ represents the variance of X.

To reach a sub-Gaussian regime, we can conserve only the second term on the right-hand side. So that the first term might disappear, either the deviation has to be huge what we do not want, or the amount of data has to be very large. The amount of data required to make the first term disappear has to be so high that we can say that the Gaussian regime is *only* asymptotic; consequently, the empirical mean (under heavy-tail assumptions on X) can have large deviations for a small sample size.

The fact that the Gaussian regime is asymptotic is also aligned with the Central Limit Theorem (CLT).

Lemma 4.13 (Central Limit Theorem (CLT))**.** *If $X_1, \dots, X_n$ are n i.i.d. random variables of mean μ and finite variance σ^2. Then, the mean squared error of $\bar{\mu}_n$ is equal to $\frac{\sigma^2}{n}$. CLT guarantees that the estimator asymptotically has Gaussian tails.*

$$\lim_{n\to+\infty} \mathbb{P}\left(\left|\bar{\mu}_n - \mu\right| > \frac{\sigma\Phi^{-1}(1-\delta/2)}{\sqrt{n}}\right) = \delta, \tag{4.54}$$

where $\Phi(x) = \mathbb{P}(G \leq x)$ is the c.d.f. of the standard normal variable G.

For any $x > 0$, the exponential Markov inequality yields $1 - \Phi(x) = \mathbb{P}(e^{xX} \geq e^{x^2}) \leq \mathbb{E}[e^{xX}]e^{-x^2} = e^{-\frac{x^2}{2}}$. This implies that $\Phi^{-1}(1-\delta/2) \leq \sqrt{2\log(2/\delta)}$, and the CLT asserts that:

$$\lim_{n\to+\infty} \mathbb{P}\left(\left|\bar{\mu}_n - \mu\right| > \frac{\sigma\sqrt{2\log(2/\delta)}}{\sqrt{n}}\right) \leq \delta. \tag{4.55}$$

In other words, it exhibits sub-Gaussian performance but in the *asymptotic case*. Is it possible to outperform the sub-Gaussian performance? (Lugosi and Mendelson, 2019, Theorem 1) states that for any mean estimator $\bar{\mu}_n$ there always exists a distribution (with mean μ and variance $\sigma^2 > 0$) such that:

$$\mathbb{P}\left(\left|\bar{\mu}_n - \mu\right| > \sigma\sqrt{\frac{\log(1/\delta)}{n}}\right) \geq \delta \tag{4.56}$$

provided that $\delta \in (2e^{-n/4}, 1/2)$ and $n > 5$. However, from Definition 4.11, to reach sub-Gaussian performance, we should have:

$$\mathbb{P}\left(\left|\bar{\mu}_n - \mu\right| > \sigma\sqrt{\frac{\log(1/\delta)}{n}}\right) \leq \delta. \tag{4.57}$$

This implies that the sub-Gaussian bound (up to constants) is the best bound that can be reached.

Now, let us assume that there exists an estimator that satisfies the above non-asymptotic inequality for all n and for all δ, up to adjusting the constant in front of σ and inside the log. Specifically, for $n = 1$, we obtain:

$$\mathbb{P}\left(\left|X - \mu\right| > L\sigma\sqrt{\log(1/\delta)}\right) \leq \delta, \qquad \forall\delta \in (0,1), \tag{4.58}$$

for some L independent of δ. This implies, from Definition 4.11, that:

$$\mathbb{P}\left(\left|X - \mu\right| > \epsilon\right) \leq e^{-\left(\frac{\epsilon}{L\sigma}\right)^2}. \tag{4.59}$$

This implies that X is sub-Gaussian and has light tails. The deviation bound is sub-Gaussian and the distribution of the samples cannot be heavy-tailed. This implies that an estimator built on a sample that follows a Pareto distribution cannot have a deviation bound that is sub-Gaussian.

Consequently, alternatives to the sample mean must be found to estimate the expectation of heavy-tailed distributions.

Our goal is to study the bias between the empirical $\mathbf{ES}_\alpha$ when the latter is estimated by the empirical mean estimator applied to the standardized Pareto distribution tail and the true $\mathbf{ES}_\alpha$. This bias is studied in two frameworks: in the *idealized case* where the empirical α-quantile matches the true $\mathbf{VaR}_\alpha$ and in the *realistic case* where the empirical α-quantile does not match the true $\mathbf{VaR}_\alpha$. To this end, we first study the bias between the empirical mean estimator applied to the entire standardized Pareto distribution $\mathcal{P}(1,\gamma)$ and the expectation.

4.2.2 Bias between the Empirical Mean estimator and the expectation of the standardized Pareto distribution

Let $X_1,\dots,X_n$ be n independent and identically distributed (i.i.d.) random variables following the distribution of $X\mathcal{P}(1,\gamma)$. The expectation of X is denoted by $\mu=\mathbb{E}[X]$.

The bias between the Empirical Mean and the expectation of the distribution is defined as follows:

$$B_\mu(\bar{\mu}_n)=\mathbb{E}[\mu_n]-\mu=\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i]-\mu. \tag{4.60}$$

Since $(X_i)_{i\in[\![1,n]\!]}$ are i.i.d., then $\forall i\in[\![1,n]\!],\mathbb{E}[X_i]=\mu$. And we get:

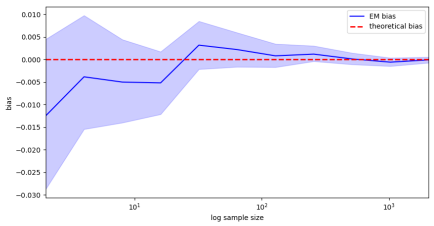
$$B_\mu(\bar{\mu}_n)=0. \tag{4.61}$$

Consequently, the Empirical Mean is an unbiased estimator of the expectation of the distribution, as shown in Figure 4.6.

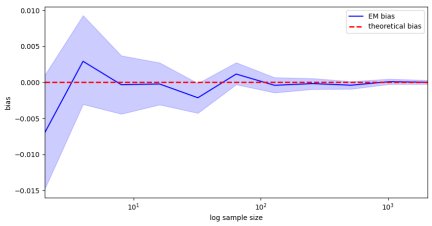
From Central Limit Theorem (CLT), the empirical mean of a standardized Pareto distribution $\mathcal{P}(1,\gamma)$ asymptotically converges to the expectation of the given distribution, with a convergence rate equal to $\frac{1}{\sqrt{n}}$. Then, the bias goes to 0 with a convergence rate equal to $\frac{1}{\sqrt{n}}$.

Owing to the stability by conditioning and scaling properties of the Pareto distribution, the bias between the empirical mean and the expectation of any non-standardized Pareto distribution $\mathcal{P}(x_m,\gamma)$ is proportional to the bias between the empirical mean and the expectation of the standardized Pareto distribution $\mathcal{P}(1,\gamma)$ with a proportionality factor equal to the scaling parameter x_m. Because x_m does not depend on the underlying sample, the convergence speed of the bias between the empirical mean and the expectation of the non-standardized Pareto distribution $\mathcal{P}(x_m,\gamma)$ is the same as the convergence speed of the bias between the empirical mean and the expectation of the standardized Pareto distribution $\mathcal{P}(1,\gamma)$.

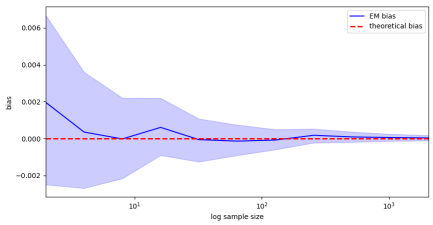
Now that we have studied the bias between the empirical mean estimator applied to the entire standardized Pareto distribution $\mathcal{P}(1,\gamma)$ and the expectation, we are interested in studying the bias between the empirical $\mathbf{ES}_\alpha$ and the true $\mathbf{ES}_\alpha$ in two frameworks: the *idealized case* where the empirical α-quantile matches the true $\mathbf{VaR}_\alpha$ and the *realistic case* where the empirical α-quantile does not match the true $\mathbf{VaR}_\alpha$.



(a)



(b)



(c)

Figure 4.6: Evolution of the bias between the empirical mean estimator and the expectation of the whole Pareto distribution, for three different shape parameters versus the sample size, with 95% CI, over 10^4 repetitions, (a) $\mathcal{P}(1, 2.5)$, (b) $\mathcal{P}(1, 3.5)$, (c) $\mathcal{P}(1, 5)$.

4.2.3 Bias between the empirical mean estimator applied to the standardized Pareto distribution tail above the true $\mathbf{VaR}_\alpha$ and the true ES (Idealized case)

Let us recall that the true $\mathbf{ES}_\alpha$ is defined as follows:

$$\mathbf{ES}_\alpha(X) = \mathbb{E}[X|X \geq \mathbf{VaR}_\alpha] = \mathbb{E}[X\mathbb{1}_{\{X\geq \mathbf{VaR}_\alpha\}}]/\mathbb{P}(X \geq \mathbf{VaR}_\alpha) = \mathbb{E}\left[X\mathbb{1}_{\{X\geq \mathbf{VaR}_\alpha\}}\right]/(1-\alpha). \tag{4.62}$$

In the idealized case where the empirical α-quantile matches the true $\mathbf{VaR}_\alpha$, the estimation of the $\mathbf{ES}_\alpha$ is given by the empirical mean estimator applied to the distribution tail above the true $\mathbf{VaR}_\alpha$, as follows:

$$\mathbf{ES}_\alpha(X) \approx \frac{1}{n(1-\alpha)} \sum_{i=1}^{n} X_i \mathbb{1}_{\{X_i \geq \mathbf{VaR}_\alpha\}} =: \bar{\mu}_{n_{|>\mathbf{VaR}_\alpha}}. \tag{4.63}$$

The bias is defined by:

$$B_{\mathbf{ES}_\alpha}(\bar{\mu}_{n_{|>\mathbf{VaR}_\alpha}}) = \mathbb{E}\left[\bar{\mu}_{n_{|>\mathbf{VaR}_\alpha}}\right] - \mathbf{ES}_\alpha(X). \tag{4.64}$$

On the one hand,

$$\mathbb{E}\left[\bar{\mu}_{n_{|>\mathbf{VaR}_\alpha}}\right] = \frac{1}{n(1-\alpha)} \sum_{i=1}^{n} \mathbb{E}\left[X_i \mathbb{1}_{\{X_i \geq \mathbf{VaR}_\alpha\}}\right] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[X_i | X_i \geq \mathbf{VaR}_\alpha\right]. \tag{4.65}$$

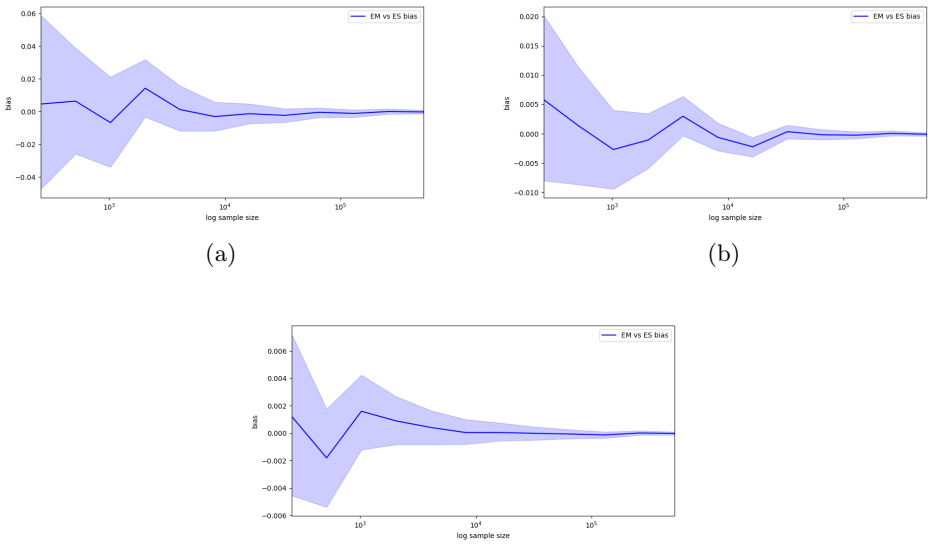
But the random variables X_i are i.i.d., then $\mathbb{E}\left[X_i|X_i \geq \mathbf{VaR}_\alpha\right] = \mathbb{E}\left[X|X \geq \mathbf{VaR}_\alpha\right]$ and we get:

$$\mathbb{E}\left[\bar{\mu}_{n_{|>\mathbf{VaR}_\alpha}}\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[X|X \geq \mathbf{VaR}_\alpha\right] = \mathbb{E}\left[X|X \geq \mathbf{VaR}_\alpha\right] = \mathbf{ES}_\alpha(X). \tag{4.66}$$

Consequently, $B_{\mathbf{ES}_\alpha}(\bar{\mu}_{n_{|>\mathbf{VaR}_\alpha}}) = 0$.

The empirical mean applied to the standardized Pareto distribution conditional on its values being above the true $\mathbf{VaR}_\alpha$ is an unbiased estimator of the true $\mathbf{ES}_\alpha$, as shown in Figure 4.7.

Indeed, in the idealized case, the empirical α-quantile matches the true $\mathbf{VaR}_\alpha$, which is supposed to be known. In this case, the estimator of the $\mathbf{ES}_\alpha$ in the standardized Pareto distribution $\mathcal{P}(1,\gamma)$ is the empirical average of the standardized Pareto distribution conditional on its values being above the true $\mathbf{VaR}_\alpha$. The conditioning threshold is independent on the underlying sample. This implies that the samples larger than the true $\mathbf{VaR}_\alpha$ are independent and identically distributed (i.i.d.) and the stability by conditioning and scaling properties of the Pareto distribution, as stated in Theorem 4.7, are valid. The stability by conditioning property implies that the standardized Pareto distribution $\mathcal{P}(1,\gamma)$ conditional on its values being above $\mathbf{VaR}_\alpha$ is still a Pareto distribution, with the same shape parameter γ, but a new scaling parameter equal to the conditioning parameter $\mathbf{VaR}_\alpha$. The scaling property states that this non-standardized Pareto distribution $\mathcal{P}(\mathbf{VaR}_\alpha,\gamma)$ is proportional to the marginal Pareto distribution $\mathcal{P}(1,\gamma)$, with a proportionality factor equal to the conditioning parameter $\mathbf{VaR}_\alpha$. Therefore, the standardized Pareto distribution conditional on its values being above $\mathbf{VaR}_\alpha$ is proportional to the standardized Pareto distribution with a proportionality factor equal to $\mathbf{VaR}_\alpha$. This implies that the empirical mean estimator applied to the standardized Pareto distribution conditional on its values being above $\mathbf{VaR}_\alpha$, is proportional to the empirical mean estimator applied to the entire standardized Pareto distribution, with a proportionality factor equal to $\mathbf{VaR}_\alpha$. In the same way, the $\mathbf{ES}_\alpha$ is proportional to the expectation of the standardized Pareto distribution, with a proportionality factor equal to $\mathbf{VaR}_\alpha$. Thus, the bias between the empirical mean estimator applied to the standardized Pareto distribution conditional on its values being above $\mathbf{VaR}_\alpha$ and the $\mathbf{ES}_\alpha$ is proportional to the bias between the empirical mean estimator applied to the entire standardized Pareto distribution and the expectation, with a proportionality factor equal to $\mathbf{VaR}_\alpha$. Moreover, because the proportionality factor (conditioning threshold) is independent on the underlying sample, the convergence rate of the bias between the empirical mean estimator applied to the standardized Pareto distribution conditional on its values being above $\mathbf{VaR}_\alpha$ and the true $\mathbf{ES}_\alpha$ is the same as that of the bias between the empirical mean estimator applied to the entire standardized Pareto distribution and the expectation.



(a) (b)

(c)

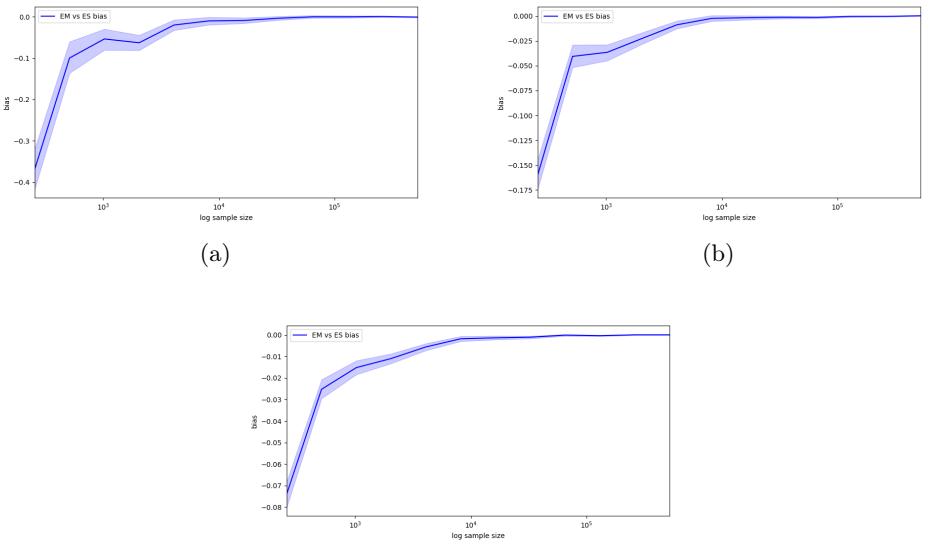
Figure 4.7: Evolution of the bias between the empirical mean estimator applied to the Pareto distribution tail above the true $\mathbf{VaR}_\alpha$ and the Expected-Shortfall, for three different shape parameters versus the sample size, with 95% CI, over 10^4 repetitions, (a) $\mathcal{P}(1, 2.5)$, (b) $\mathcal{P}(1, 3.5)$, (c) $\mathcal{P}(1, 5)$.

4.2.4 Bias between the empirical mean estimator applied to the standardized Pareto distribution tail above the empirical α-quantile and the true ES (Realistic case)

In the realistic case, the empirical α-quantile does not match the true $\mathbf{VaR}_\alpha$. In this case, the estimator of the $\mathbf{ES}_\alpha$ in the standardized Pareto distribution $\mathcal{P}(1, \gamma)$ is the empirical average of the standardized Pareto distribution conditional on its values being above the empirical α-quantile. The conditioning threshold is an order statistics and depends on the underlying sample. This implies that the samples larger than the empirical α-quantile are no longer independent and identically distributed (i.i.d.). Therefore, the distribution of the samples larger than the empirical α-quantile is not necessarily a Pareto distribution. The stability by conditioning and scaling properties of the Pareto distribution, as stated in Theorem 4.7, are no longer valid. The distribution of the samples larger than the empirical α-quantile is unknown and it is more challenging to establish an analytic closed-form formula for the bias between the empirical mean estimator applied to the standardized Pareto distribution tail above the empirical α-quantile and the true $\mathbf{ES}_\alpha$. For this reason, we provide some empirical study in order to gain insight into the bias behavior and its convergence speed as the sample size goes to infinity. See Figure 4.8.

From Equation (4.5), we recall that the empirical Expected-Shortfall is defined as:

$$\bar{\mu}_{n_{|>\mathbf{q}_{n,\alpha}}} = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i \mathbb{1}_{\{X_i \geq q_\alpha^n\}}}{\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i \geq q_\alpha^n\}}} = \frac{\sum_{i=1}^{n} X_i \mathbb{1}_{\{X_i \geq q_\alpha^n\}}}{n - \lceil n\alpha \rceil}. \tag{4.67}$$



(a) (b)

(c)

Figure 4.8: Evolution of the bias between the empirical mean estimator applied to the Pareto distribution tail above the empirical α-quantile and the Expected-Shortfall, for three different shape parameters versus the sample size, with 95% CI, over 10^4 repetitions, (a) $\mathcal{P}(1, 2.5)$, (b) $\mathcal{P}(1, 3.5)$, (c) $\mathcal{P}(1, 5)$.

4.3 Empirical median for heavy-tailed distribution

4.3.1 Presentation and properties of the Median

The median is the value separating the higher half from the lower half of a data sample, population, or probability distribution. For a dataset, it may be thought of as "the middle" value. The basic feature of the median in describing the data compared to the mean (often simply described as the "average") is that it is not skewed by a small proportion of extremely large or small values, and therefore provides a better representation of the distribution centrality. Therefore, the median is of key importance in robust statistics.

Formally, the median of a population is any value such that at least half of the population is less than or equal to the proposed median and at least half is greater than or equal to the proposed median. Medians may not be unique. If each set contains more than half of the population, then some of the population is exactly equal to the unique median.

The median is well-defined for any ordered (one-dimensional) data, and is independent of any distance metric. Thus, the median can be applied to classes that are ranked but not numerical, although the result might be halfway between classes if there is an even number of cases.

On the other hand, a geometric median is defined in any number of dimensions. A related concept, in which the outcome is forced to correspond to a member of the sample, is medoid.

The median is a special case of other ways of summarizing the typical values associated with a statistical distribution: the 2^{nd} quartile, 5^{th} decile and 50^{th} percentile.

The median of a sequence is defined as follows. It consists of ordering the values into a non-decreasing sequence and choosing the value that is at the center of this ordered sequence. For an ordered sequence composed of n elements, where n is odd, the median belongs to the sequence and corresponds to term of order $\frac{n+1}{2}$. For an ordered sequence composed of n elements, where n is even, the median does not belong to the sequence and corresponds to the arithmetical mean of the terms of orders $\frac{n}{2}$ and $\frac{n}{2}+1$. The complexity of the algorithm that allows computing the median corresponds to the complexity of the sorting algorithm used, that is, at best $O(n\log(n))$.

Mathematically speaking, the Median is defined as follows.

Definition 4.14 (Median of probabilistic distributions)**.** *Let X be a real random variable with as cumulative distribution function $F_X(x) = \mathbb{P}(X \leq x)$ and survival function $\bar{F}_X(x) = 1 - F_X(x)$. Subsequently, the median M of the distribution satisfies the following equality:*

$$M = \inf\left\{x \in \mathbb{R} : F_X(x) \geq \frac{1}{2} \text{ and } \bar{F}_X(x) \geq \frac{1}{2}\right\}. \tag{4.68}$$

Definition 4.15 (Empirical Median)**.** *The median of n real numbers $x_1, \ldots, x_n \in \mathbb{R}$ is defined as $\hat{M}_n = x_i$ where x_i is such that i satisfies:*

$$i = \min\left\{j \in [\![1, n]\!] : \left|\{k \in [\![1, n]\!] : x_k \leq x_j\}\right| \geq \frac{n}{2} \quad \text{and} \quad \left|\{k \in [\![1, n]\!] : x_k \geq x_j\}\right| \geq \frac{n}{2}\right\}. \tag{4.69}$$

According to the above definition, whatever the parity of n, that is, whether n is even or odd, then the median belongs to the sequence. When n is odd, the median corresponds to the term of order $\frac{n+1}{2}$ and each subset is composed of $\frac{n}{2}+1$ samples including the median. When n is even, then both terms of order $\frac{n}{2}$ and $\frac{n}{2}+1$ satisfy the above definition, but by convention, the term with the smallest index, that is, the term of order $\frac{n}{2}$, is retained for the median. In this case, the lower subset contains $\frac{n}{2}$ samples, whereas the upper subset contains $\frac{n}{2}+1$, including the median.

Another definition of the median, in which the median belongs to the sequence if n is odd, and does not belong to the sequence if n is even, is given as follows:

$$x_i^\star = \begin{cases} \frac{1}{2}\left(x^\star_{\frac{n}{2}} + x^\star_{\frac{n}{2}+1}\right) & \text{if } n \in 2\mathbb{N} \\ x^\star_{\frac{n+1}{2}} & \text{if } n \notin 2\mathbb{N}. \end{cases} \tag{4.70}$$

where $x_1^\star, \ldots, x_n^\star$ is the increasing reordering of sequence $x_1, \ldots, x_n$. With this last definition, when n is odd, the median corresponds to the term of the sequence of order $\frac{n+1}{2}$. In such a case, both subsets contain $\frac{n}{2}+1$ samples, including the median. When n is even, the median does not belong to the sequence and corresponds to the arithmetic mean of the terms of orders $\frac{n}{2}$ and $\frac{n}{2}+1$. In this case, both subsets contain $\frac{n}{2}$ samples, excluding the median, which does not belong to the sequence.

There are many definitions for the empirical median. We have provided two of these. The first is the one used in (Lugosi and Mendelson, 2019), but in the sequel, we will work with the second formulation.

From (Reiss, 2012, Lemma 3.1.1. and p.85), a deviation inequality of the empirical median with respect to the theoretical median is derived.

Theorem 4.16 (Deviation inequality for order statistics)**.** *Let $U_1, \ldots, U_n$ be n i.i.d. random variables following a uniform distribution on $(0,1)$ and $U_1^\star, \ldots, U_n^\star$ be the related sequence of ordered statistics. Let $X_1, \ldots, X_n$ be n i.i.d. random variables with a common distribution function F_X, and $X_1^\star, \ldots, X_n^\star$ be the related sequence of ordered statistics. Then, the k^{th} ordered statistics $X_k^\star$ satisfies:*

$$\mathbb{P}\left(\sqrt{n}\frac{g(\alpha)}{\sigma}(X_k^\star - F_X^{-1}(\alpha)) \begin{array}{l}\leq -\epsilon \\ \geq \epsilon\end{array}\right) \leq \mathbb{P}\left(\frac{\sqrt{n}}{\sigma}(U_k^\star - \alpha) \begin{array}{l}\leq h(-\epsilon) \\ \geq h(\epsilon)\end{array}\right), \quad \forall \epsilon > 0, \tag{4.71}$$

where $g(\alpha)$ is a non-negative constant, $h(x) = \frac{\sqrt{n}}{\sigma}\left(F_X\left(F_X^{-1}(\alpha) + \frac{x\sigma}{g(\alpha)\sqrt{n}}\right) - \alpha\right)$, and:

$$\mathbb{P}\left(\frac{\sqrt{n}}{\sigma}(U_k^\star - \alpha) \begin{array}{l}\leq h(-\epsilon) \\ \geq h(\epsilon)\end{array}\right) \leq e^{-\frac{(h(\mp\epsilon))^2}{3\left(1+\frac{|h(\mp\epsilon)|}{\sigma\sqrt{n}}\right)}} \tag{4.72}$$

with $\alpha = \frac{k}{n+1}$ and $\sigma^2 = \alpha(1-\alpha)$. In the specific case of the empirical median, $k = \left\lceil \frac{n}{2} \right\rceil$.

Another key result on the order statistics is the following theorem, explained in detail in (Reiss, 2012, Thm 1.3.2. p.21).

Theorem 4.17 (Density of order statistics)**.** *Let $X_k^\star$ be the k-th order statistics of n i.i.d. random variables of common distribution function F_X and density f_X. Then the density of $X_k^\star$ is given by:*

$$f_{X_k^\star}(x) = \frac{n!}{(k-1)!(n-k)!} f_X(x)(F_X(x))^{k-1}(1-F_X(x))^{n-k}. \tag{4.73}$$

Theorem 4.18 (Expectation of empirical median when n is odd)**.** (i) *Let $X_1, \ldots, X_{2k+1}$ be $2k+1$ i.i.d. random variables that follow the same distribution as $X \sim \mathcal{P}(1, \gamma)$. Let $X_1^\star, \ldots, X_{2k+1}^\star$ be the related sequence of ordered statistics. The density of the empirical median $X_{k+1}^\star$ is given by:*

$$f_{\hat{M}_{2k+1}^{(1)}}(x) = \frac{(2k+1)!}{k!k!}\left(1 - \frac{1}{x^\gamma}\right)^k \frac{1}{x^{k\gamma}}\left(\frac{\gamma}{x^{\gamma+1}}\right)\mathbb{1}_{\{x \geq 1\}}. \tag{4.74}$$

The expectation of the empirical median is defined as follows:

$$\mathbb{E}[\hat{M}_{2k+1}^{(1)}] = \frac{(2k+1)!}{k!k!}\mathfrak{B}\left(k+1, k+1-\frac{1}{\gamma}\right) \tag{4.75}$$

where $\mathfrak{B}(x,y) = \int_0^1 (1-t)^{x-1}t^{y-1}\mathrm{d}t = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$ *is the Beta function, and* $\Gamma(z) = \int_0^{+\infty} t^{z-1}e^{-t}\mathrm{d}t$ *is the Gamma function.*

(ii) *Let* $Y_1, \ldots, Y_{2k+1}$ *be* $2k+1$ *i.i.d. random variables that follow the same distribution as* $Y \sim \mathcal{P}(x_m, \gamma)$*. Then, from the scaling property of the Pareto distribution (Theorem 4.7), we obtain the following proportionality relationship between the distributions* $\mathcal{P}(x_m, \gamma)$ *and* $\mathcal{P}(1, \gamma)$*:*

$$(Y_1, \ldots, Y_{2k+1}) \overset{d}{=} (x_m X_1, \ldots, x_m X_{2k+1}). \tag{4.76}$$

In particular, a proportionality relationship of factor x_m *links the empirical median of the non-standardized Pareto distribution* $\mathcal{P}(x_m, \gamma)$ *and that of the standardized Pareto distribution* $\mathcal{P}(1, \gamma)$ *such that* $\hat{M}_{2k+1}^{(x_m)} \overset{d}{=} x_m \hat{M}_{2k+1}^{(1)}$*, as well as their expectations:*

$$\mathbb{E}[\hat{M}_{2k+1}^{(x_m)}] = x_m \mathbb{E}[\hat{M}_{2k+1}^{(1)}]. \tag{4.77}$$

Remark 4.19. *When* $\gamma \to +\infty$ *(the distribution tails are lighter and lighter),* X *converges in distribution to* 1*. In this asymptotic case, we expect the median expectation to converge to 1, which is consistent with*

$$\lim_{\gamma \to +\infty} \mathbb{E}\left[\hat{M}_{2k+1}\right] = \frac{(2k+1)!}{k!k!} \frac{\Gamma(k+1)\Gamma\left(k+1\right)}{\Gamma\left(2k+2\right)} = 1$$

using that $\Gamma(m+1) = m!$ *for integers* m*.*

Proof (i) Let $X_1, \ldots, X_{2k+1}$ be $2k+1$ i.i.d. random variables that follow the same distribution as $X \sim \mathcal{P}(1, \gamma)$. Let $X_1^\star, \ldots, X_{2k+1}^\star$ be the related sequence of ordered statistics.

From Equation (4.73), the density of the empirical median of a sample of i.i.d. Pareto r.v. is defined as follows:

$$f_{\hat{M}_{2k+1}^{(1)}}(x) = \frac{(2k+1)!}{k!k!} \left(1 - \left(\frac{1}{x}\right)^\gamma\right)^k \left(\frac{1}{x}\right)^{k\gamma} \left(\frac{\gamma}{x^{\gamma+1}}\right) \mathbb{1}_{\{x \geq 1\}}.$$

Then we get: $\mathbb{E}\left[\hat{M}_{2k+1}^{(1)}\right] = \int_{\mathbb{R}} x f_{\hat{M}_{2k+1}^{(1)}}(x)\mathrm{d}x = \int_1^{+\infty} x \frac{(2k+1)!}{k!k!} \left(1 - \left(\frac{1}{x}\right)^\gamma\right)^k \left(\frac{1}{x}\right)^{k\gamma} \left(\frac{\gamma}{x^{\gamma+1}}\right) \mathrm{d}x.$

By the following substitution, $y = \left(\frac{1}{x}\right)^\gamma$, $\mathrm{d}y = -\gamma \frac{1}{x^{\gamma+1}}\mathrm{d}x$, when $x = 1$ then $y = 1$ and when $x \to +\infty$ then $y \to 0$, we obtain:

$$\mathbb{E}\left[\hat{M}_{2k+1}^{(1)}\right] = \frac{(2k+1)!}{k!k!} \int_0^1 (1-y)^k y^{k-\frac{1}{\gamma}} \mathrm{d}y = \frac{(2k+1)!}{k!k!} \mathfrak{B}\left(k+1, k+1-\frac{1}{\gamma}\right).$$

(ii) The second point comes from the scaling property of the Pareto distribution which establishes a proportionality relationship between the non-standardized Pareto distribution $\mathcal{P}(x_m, \gamma)$ and the standardized Pareto distribution $\mathcal{P}(1, \gamma)$ with a proportionality factor equal to the scaling parameter x_m. This implies that the order statistics of the

non-standardized Pareto distribution $\mathcal{P}(x_m, \gamma)$ are proportional to the order statistics of the standardized Pareto distribution $\mathcal{P}(1, \gamma)$ with a proportionality factor equal to x_m. The k-th order moment of the non-standardized Pareto distribution $\mathcal{P}(x_m, \gamma)$ is thus proportional to the k-th order moment of the standardized Pareto distribution $\mathcal{P}(1, \gamma)$ with a proportionality factor equal to x_m^k. With $k = 1$, we find the expected result. ■

Remark 4.20. *Theorem 4.18 can be easily adjusted when the number of data n is not odd.*

Theorem 4.21 (Expectation of empirical median when n is even)**.** (i) *Let $X_1, \ldots, X_{2k}$ be $2k$ independent and identically distributed (i.i.d.) random variables following the same distribution as $X \sim \mathcal{P}(1, \gamma)$. Let $X_1^\star, \ldots, X_{2k}^\star$ be the related sequence of ordered statistics. We define*

$$\hat{M}_{2k} := \frac{1}{2}(X_k^\star + X_{k+1}^\star). \tag{4.78}$$

The expectation of the empirical median is given by:

$$\mathbb{E}\left[\hat{M}_{2k}\right] = \frac{1}{2}\left[\frac{(2k)!}{(k-1)!k!}\mathfrak{B}\left(k, k+1-\frac{1}{\gamma}\right) + \frac{(2k)!}{k!(k-1)!}\mathfrak{B}\left(k+1, k-\frac{1}{\gamma}\right)\right] \tag{4.79}$$

(ii) *Let $Y_1, \ldots, Y_{2k}$ be $2k$ i.i.d. random variables that follow the same distribution as $Y \sim \mathcal{P}(x_m, \gamma)$. Then, from the scaling property of the Pareto distribution (Theorem 4.7), we obtain the following proportionality relationship between the distributions $\mathcal{P}(x_m, \gamma)$ and $\mathcal{P}(1, \gamma)$:*

$$(Y_1, \ldots, Y_{2k+1}) \overset{d}{=} (x_m X_1, \ldots, x_m X_{2k+1}). \tag{4.80}$$

In particular, a proportionality relationship of factor x_m links the empirical median of the non-standardized Pareto distribution $\mathcal{P}(x_m, \gamma)$ and that of the standardized Pareto distribution $\mathcal{P}(1, \gamma)$ such that $\hat{M}_{2k}^{(x_m)} \overset{d}{=} x_m \hat{M}_{2k}^{(1)}$, as well as their expectations:

$$\mathbb{E}[\hat{M}_{2k}^{(x_m)}] = x_m \mathbb{E}[\hat{M}_{2k}^{(1)}]. \tag{4.81}$$

Proof (i) From Equation (4.73), we have the expression for the density of $X_k^\star$ which yields

$$\mathbb{E}\left[X_k^\star\right] = \int_{\mathbb{R}} x f_{X_k^\star}(x) \mathrm{d}x \tag{4.82}$$

$$= \frac{(2k)!}{(k-1)!k!}\int_1^{+\infty} x\left(1-\frac{1}{x^\gamma}\right)^{k-1}\frac{1}{x^{\gamma k}}\frac{\gamma}{x^{\gamma+1}}\mathrm{d}x. \tag{4.83}$$

By the following substitution, $y = \frac{1}{x^\gamma}$, $\mathrm{d}y = -\gamma\frac{1}{x^{\gamma+1}}\mathrm{d}x$, $y = 1$ when $x = 1$, $y \to 0$ when $x \to +\infty$, we get:

$$\mathbb{E}\left[X_k^\star\right] = \frac{(2k)!}{(k-1)!k!}\int_0^1 (1-y)^{k-1} y^{k-\frac{1}{\gamma}}\mathrm{d}y = \frac{(2k)!}{(k-1)!k!}\mathfrak{B}\left(k, k+1-\frac{1}{\gamma}\right). \tag{4.84}$$

On the other hand, let's compute $\mathbb{E}\left[X^\star_{k+1}\right]$:

$$\mathbb{E}\left[X^\star_{k+1}\right] = \int_{\mathbb{R}} x f_{X^\star_{k+1}}(x)\mathrm{d}x \tag{4.85}$$

$$= \frac{(2k)!}{k!(k-1)!}\int_1^{+\infty} x\left(1-\frac{1}{x^\gamma}\right)^k \frac{1}{x^{\gamma(k-1)}}\frac{\gamma}{x^{\gamma+1}}\mathrm{d}x. \tag{4.86}$$

Using the same substitution as previously, we fall onto:

$$\mathbb{E}\left[X^\star_k\right] = \frac{(2k)!}{k!(k-1)!}\int_0^1 (1-y)^k y^{k-1-\frac{1}{\gamma}}\mathrm{d}y = \frac{(2k)!}{k!(k-1)!}\mathfrak{B}\left(k+1, k-\frac{1}{\gamma}\right). \tag{4.87}$$

(ii) Same proof as in the odd case. ■

4.3.2 Bias of the Median in the specific case of Pareto distribution

In this section, we are interested in computing the bias between the empirical and the theoretical medians for a Pareto distribution. The scaling property of the Pareto distribution establishes a proportionality relationship between any non-standardized Pareto distribution $\mathcal{P}(x_m, \gamma)$ and the standardized Pareto distribution $\mathcal{P}(1, \gamma)$ with a proportionality factor equal to x_m. This proportionality relationship of factor x_m is still true between the statistics (and their expectation) of the non-standardized and standardized Pareto distributions. Therefore, the bias between the empirical and the theoretical medians of the non-standardized Pareto $\mathcal{P}(x_m, \gamma)$ distribution is proportional to the bias of the standardized Pareto distribution $\mathcal{P}(1, \gamma)$ with a proportionality factor equal to x_m. This implies that it suffices to compute the bias in the case of a standardized Pareto distribution, and then multiply the result by the scaling parameter x_m. Thus, in the following analysis, we use the unitary scale $x_m = 1$.

(i) **First, let us assume that the sample size n is odd i.e. $\exists k \in \mathbb{N}$ s.t. $n = 2k+1$.** According to (4.75) and (4.33), the bias between the empirical and theoretical medians is given by:

$$B_M(\hat{M}_{2k+1}) = \mathbb{E}\left[\hat{M}_{2k+1}\right] - M = \frac{(2k+1)!}{k!k!}\frac{\Gamma(k+1)\Gamma\left(k+1-\frac{1}{\gamma}\right)}{\Gamma\left(2k+2-\frac{1}{\gamma}\right)} - \sqrt[\gamma]{2}. \tag{4.88}$$

Our purpose is to prove the following asymptotic result for the bias of the median.

Theorem 4.22. *As $n \to +\infty$,*

$$\mathbb{E}\left[\hat{M}_n\right] = \sqrt[\gamma]{2}\left(1 + \frac{1}{2\gamma n} + \frac{1}{2\gamma^2 n} + o\left(\frac{1}{n}\right)\right). \tag{4.89}$$

The smaller γ, the heavier the tails, and the bigger the bias. The empirical median systematically has a (asymptotically) positive bias.

Proof

(i) **We start by considering sequences of odd integers n, i.e. $n = 2k + 1$.**

The equality $\Gamma(k+1) = k!$ combined with (4.75) implies that:

$$\mathbb{E}\left[\hat{M}_{2k+1}\right] = \frac{(2k+1)!}{k!} \frac{\Gamma\left(k+1-\frac{1}{\gamma}\right)}{\Gamma\left(2k+2-\frac{1}{\gamma}\right)} = \frac{\Gamma(2k+2)}{\Gamma(k+1)} \frac{\Gamma\left(k+1-\frac{1}{\gamma}\right)}{\Gamma\left(2k+2-\frac{1}{\gamma}\right)}. \tag{4.90}$$

From (Xu et al., 2016, Eq. 1.2.), the extended Stirling formula gives the expansion of the Gamma function as follows:

$$\Gamma(z+1) = \sqrt{2\pi z}\left(\frac{z}{e}\right)^z \left(1 + \frac{1}{12z} + o\left(\frac{1}{z}\right)\right), \quad z \to +\infty. \tag{4.91}$$

Let's write the expansion of each factor in (4.90):

$$\Gamma(2k+2) = \sqrt{2\pi(2k+1)}\left(\frac{2k+1}{e}\right)^{2k+1} \left(1 + \frac{1}{12(2k+1)} + o\left(\frac{1}{k}\right)\right), \tag{4.92}$$

$$\Gamma(k+1) = \sqrt{2\pi k}\left(\frac{k}{e}\right)^{k} \left(1 + \frac{1}{12k} + o\left(\frac{1}{k}\right)\right), \tag{4.93}$$

$$\Gamma\left(k+1-\frac{1}{\gamma}\right) = \sqrt{2\pi\left(k-\frac{1}{\gamma}\right)}\left(\frac{k-\frac{1}{\gamma}}{e}\right)^{k-\frac{1}{\gamma}} \left(1 + \frac{1}{12\left(k-\frac{1}{\gamma}\right)} + o\left(\frac{1}{k}\right)\right), \tag{4.94}$$

$$\Gamma\left(2k+2-\frac{1}{\gamma}\right) = \sqrt{2\pi\left(2k+1-\frac{1}{\gamma}\right)}\left(\frac{2k+1-\frac{1}{\gamma}}{e}\right)^{2k+1-\frac{1}{\gamma}} \left(1 + \frac{1}{12\left(2k+1-\frac{1}{\gamma}\right)} + o\left(\frac{1}{k}\right)\right). \tag{4.95}$$

We rewrite the expectation of the median estimator (4.90) in the following form:

$$\mathbb{E}\left[\hat{M}_{2k+1}\right] = A \times B \times C \tag{4.96}$$

where:

$$A := \frac{\sqrt{2\pi(2k+1)}}{\sqrt{2\pi k}} \frac{\sqrt{2\pi\left(k-\frac{1}{\gamma}\right)}}{\sqrt{2\pi\left(2k+1-\frac{1}{\gamma}\right)}} = \frac{\sqrt{2k+1}}{\sqrt{k}} \frac{\sqrt{k-\frac{1}{\gamma}}}{\sqrt{2k+1-\frac{1}{\gamma}}}, \tag{4.97}$$

$$B := \left(\frac{2k+1}{e}\right)^{2k+1} \left(\frac{e}{k}\right)^{k} \left(\frac{k-\frac{1}{\gamma}}{e}\right)^{k-\frac{1}{\gamma}} \left(\frac{e}{2k+1-\frac{1}{\gamma}}\right)^{2k+1-\frac{1}{\gamma}} \tag{4.98}$$

$$= \left(\frac{2k+1-\frac{1}{\gamma}}{2k+1}\right)^{-(2k+1)} \left(\frac{k-\frac{1}{\gamma}}{k}\right)^{k} \left(\frac{2k+1-\frac{1}{\gamma}}{k-\frac{1}{\gamma}}\right)^{\frac{1}{\gamma}}, \tag{4.99}$$

$$C := \frac{\left(1+\frac{1}{12(2k+1)}+o\left(\frac{1}{k}\right)\right)\left(1+\frac{1}{12\left(k-\frac{1}{\gamma}\right)}+o\left(\frac{1}{k}\right)\right)}{\left(1+\frac{1}{12k}+o\left(\frac{1}{k}\right)\right)\left(1+\frac{1}{12\left(2k+1-\frac{1}{\gamma}\right)}+o\left(\frac{1}{k}\right)\right)}. \tag{4.100}$$

We simplify each of the above expressions. Standard successive first-order Taylor expansions give

$$A = \sqrt{\frac{1-\frac{1}{k\gamma}}{1-\frac{1}{(2k+1)\gamma}}} \tag{4.101}$$

$$= \sqrt{\left(1-\frac{1}{k\gamma}\right)\left(1+\frac{1}{(2k+1)\gamma}+o\left(\frac{1}{k}\right)\right)} \tag{4.102}$$

$$= \sqrt{1-\frac{1}{k\gamma}+\frac{1}{(2k+1)\gamma}+o\left(\frac{1}{k}\right)} \tag{4.103}$$

$$= \sqrt{1-\frac{1}{2k\gamma}+o\left(\frac{1}{k}\right)} \tag{4.104}$$

$$= 1-\frac{1}{4\gamma k}+o\left(\frac{1}{k}\right). \tag{4.105}$$

Now, let us expand the expression of B. Notice that $\frac{2k+1-\frac{1}{\gamma}}{k-\frac{1}{\gamma}} = 2\left(1+\frac{1+\frac{1}{\gamma}}{2(k-\frac{1}{\gamma})}\right)$, then

$$B = \left(1-\frac{1}{(2k+1)\gamma}\right)^{-(2k+1)} \left(1-\frac{1}{k\gamma}\right)^{k} \left(2\left(1+\frac{1+\frac{1}{\gamma}}{2(k-\frac{1}{\gamma})}\right)\right)^{\frac{1}{\gamma}} \tag{4.106}$$

$$= 2^{\frac{1}{\gamma}} e^{-(2k+1)\log\left(1-\frac{1}{(2k+1)\gamma}\right)+k\log\left(1-\frac{1}{k\gamma}\right)+\frac{1}{\gamma}\log\left(1+\frac{1+\frac{1}{\gamma}}{2(k-\frac{1}{\gamma})}\right)}. \tag{4.107}$$

Proceed to a second order Taylor expansion to the second order of the first two logarithmic terms, and to the first order of the last logarithmic term:

$$\log\left(1-\frac{1}{(2k+1)\gamma}\right)=-\frac{1}{(2k+1)\gamma}-\frac{1}{2(2k+1)^2\gamma^2}+o\left(\frac{1}{k^2}\right), \tag{4.108}$$

$$\log\left(1-\frac{1}{k\gamma}\right)=-\frac{1}{k\gamma}-\frac{1}{2k^2\gamma^2}+o\left(\frac{1}{k^2}\right), \tag{4.109}$$

$$\log\left(1+\frac{1+\frac{1}{\gamma}}{2(k-\frac{1}{\gamma})}\right)=\frac{1+\frac{1}{\gamma}}{2(k-\frac{1}{\gamma})}+o\left(\frac{1}{k}\right). \tag{4.110}$$

Substituting these expansions into (4.107) and proceeding to the standard Taylor expansions of the exponential function yields

$$B=2^{\frac{1}{\gamma}}e^{\frac{1}{\gamma}+\frac{1}{2(2k+1)\gamma^2}-\frac{1}{\gamma}-\frac{1}{2k\gamma^2}+\frac{1+\frac{1}{\gamma}}{2\gamma(k-\frac{1}{\gamma})}+o\left(\frac{1}{k}\right)} \tag{4.111}$$

$$=2^{\frac{1}{\gamma}}\left(1+\frac{1}{2(2k+1)\gamma^2}-\frac{1}{2k\gamma^2}+\frac{1+\frac{1}{\gamma}}{2\gamma(k-\frac{1}{\gamma})}+o\left(\frac{1}{k}\right)\right) \tag{4.112}$$

$$=2^{\frac{1}{\gamma}}\left(1+\frac{1}{4\gamma^2k}-\frac{1}{2\gamma^2k}+\frac{1+\frac{1}{\gamma}}{2\gamma k}+o\left(\frac{1}{k}\right)\right) \tag{4.113}$$

$$=2^{\frac{1}{\gamma}}\left(1+\frac{1}{4\gamma^2k}+\frac{1}{2\gamma k}+o\left(\frac{1}{k}\right)\right). \tag{4.114}$$

Now, let us focus on C. Simple first-order expansion of $1/(1+x)$ for $x\to 0$ gives

$$C=\left(1+\frac{1}{12(2k+1)}+o\left(\frac{1}{k}\right)\right)\left(1+\frac{1}{12(k-\frac{1}{\gamma})}+o\left(\frac{1}{k}\right)\right) \tag{4.115}$$

$$\times\left(1-\frac{1}{12k}+o\left(\frac{1}{k}\right)\right)\left(1-\frac{1}{12(2k+1-\frac{1}{\gamma})}+o\left(\frac{1}{k}\right)\right) \tag{4.116}$$

$$=1+\frac{1}{24k}+\frac{1}{12k}-\frac{1}{12k}-\frac{1}{24k}+o\left(\frac{1}{k}\right)=1+o\left(\frac{1}{k}\right). \tag{4.117}$$

Gathering A, B and C into (4.90), we obtain:

$$\mathbb{E}\left[\hat{M}_{2k+1}\right]=A\times B\times C=\left(1-\frac{1}{4\gamma k}+o\left(\frac{1}{k}\right)\right)2^{\frac{1}{\gamma}}\left(1+\frac{1}{4\gamma^2k}+\frac{1}{2\gamma k}+o\left(\frac{1}{k}\right)\right)\left(1+o\left(\frac{1}{k}\right)\right). \tag{4.118}$$

Rearranging different terms achieves the proof of Theorem 4.22 along sequences of odd integers n.

(ii) **Let us assume that n is even i.e. $\exists k \in \mathbb{N}$ s.t. $n = 2k$.** From Theorem 4.21, we obtain

$$\mathbb{E}\left[\hat{M}_{2k}\right] = \frac{1}{2}\left[\frac{(2k)!}{(k-1)!k!}B\left(k, k+1-\frac{1}{\gamma}\right) + \frac{(2k)!}{k!(k-1)!}B\left(k+1, k-\frac{1}{\gamma}\right)\right] \tag{4.119}$$

$$= \frac{1}{2}\left[\frac{\Gamma(2k+1)}{\Gamma(k)\Gamma(k+1)}\frac{\Gamma(k)\Gamma\left(k+1-\frac{1}{\gamma}\right)}{\Gamma\left(2k+1-\frac{1}{\gamma}\right)} + \frac{\Gamma(2k+1)}{\Gamma(k+1)\Gamma(k)}\frac{\Gamma\left(k+1\right)\Gamma\left(k-\frac{1}{\gamma}\right)}{\Gamma\left(2k+1-\frac{1}{\gamma}\right)}\right] \tag{4.120}$$

$$= \frac{1}{2}\left[\frac{\Gamma\left(k+1-\frac{1}{\gamma}\right)}{\Gamma(k+1)}\frac{\Gamma(2k+1)}{\Gamma\left(2k+1-\frac{1}{\gamma}\right)} + \frac{\Gamma\left(k-\frac{1}{\gamma}\right)}{\Gamma\left(k\right)}\frac{\Gamma(2k+1)}{\Gamma\left(2k+1-\frac{1}{\gamma}\right)}\right] \tag{4.121}$$

$$= \frac{\Gamma\left(k+1-\frac{1}{\gamma}\right)}{\Gamma(k+1)}\frac{\Gamma(2k+1)}{\Gamma\left(2k+1-\frac{1}{\gamma}\right)}\frac{1}{2}\left[1+\frac{k}{k-\frac{1}{\gamma}}\right] =: A\cdot\frac{1}{2}\cdot B \tag{4.122}$$

using that $\Gamma(x+1) = x\Gamma(x)$. To study the asymptotic behavior of $\mathbb{E}\left[\hat{M}_{2k}\right]$, we use a Taylor expansion at the first order of the Gamma function, (see (4.91)):

$$\Gamma(k+1) = \sqrt{2\pi k}\left(\frac{k}{e}\right)^k\left(1+\frac{1}{12k}+o\left(\frac{1}{k}\right)\right), \tag{4.123}$$

$$\Gamma(2k+1) = \sqrt{4\pi k}\left(\frac{2k}{e}\right)^{2k}\left(1+\frac{1}{24k}+o\left(\frac{1}{k}\right)\right), \tag{4.124}$$

$$\Gamma\left(k+1-\frac{1}{\gamma}\right) = \sqrt{2\pi\left(k-\frac{1}{\gamma}\right)}\left(\frac{k-\frac{1}{\gamma}}{e}\right)^{k-\frac{1}{\gamma}}\left(1+\frac{1}{12\left(k-\frac{1}{\gamma}\right)}+o\left(\frac{1}{k}\right)\right), \tag{4.125}$$

$$\Gamma\left(2k+1-\frac{1}{\gamma}\right) = \sqrt{2\pi\left(2k-\frac{1}{\gamma}\right)}\left(\frac{2k-\frac{1}{\gamma}}{e}\right)^{2k-\frac{1}{\gamma}}\left(1+\frac{1}{12\left(2k-\frac{1}{\gamma}\right)}+o\left(\frac{1}{k}\right)\right). \tag{4.126}$$

Using the same method as in the odd case, we obtain the following results. Write

$$A = \frac{\Gamma\left(k+1-\frac{1}{\gamma}\right)}{\Gamma(k+1)}\frac{\Gamma(2k+1)}{\Gamma\left(2k+1-\frac{1}{\gamma}\right)} = A_1\times A_2\times A_3 \tag{4.127}$$

where:

$$A_1 = \frac{\sqrt{2k}}{\sqrt{2k-\frac{1}{\gamma}}}\frac{\sqrt{k-\frac{1}{\gamma}}}{\sqrt{k}} = 1 - \frac{1}{4\gamma k} + o\left(\frac{1}{k}\right), \tag{4.128}$$

$$A_2 = \left(\frac{2k}{e}\right)^{2k}\left(\frac{2k-\frac{1}{\gamma}}{e}\right)^{-\left(2k-\frac{1}{\gamma}\right)}\left(\frac{k-\frac{1}{\gamma}}{e}\right)^{k-\frac{1}{\gamma}}\left(\frac{k}{e}\right)^{-k} = 2^{\frac{1}{\gamma}}\left(1+\frac{1}{4\gamma^2 k}+o\left(\frac{1}{k}\right)\right), \tag{4.129}$$

$$A_3 = \frac{\left(1+\frac{1}{24k}+o\left(\frac{1}{k}\right)\right)\left(1+\frac{1}{12\left(k-\frac{1}{\gamma}\right)+o\left(\frac{1}{k}\right)}\right)}{\left(1+\frac{1}{12\left(2k-\frac{1}{\gamma}\right)}+o\left(\frac{1}{k}\right)\right)\left(1+\frac{1}{12k}+o\left(\frac{1}{k}\right)\right)} = 1+o\left(\frac{1}{k}\right). \tag{4.130}$$

Therefore:

$$A = 2^{\frac{1}{\gamma}}\left(1+\frac{1}{4\gamma^2 k}-\frac{1}{4\gamma k}+o\left(\frac{1}{k}\right)\right) \tag{4.131}$$

On the other hand:

$$B = 1+\frac{k}{k-\frac{1}{\gamma}} = 2+\frac{1}{k\gamma}+o\left(\frac{1}{k}\right). \tag{4.132}$$

Finally, we get:

$$\mathbb{E}\left[\hat{M}_{2k}\right] = A\cdot\frac{1}{2}\cdot B = 2^{\frac{1}{\gamma}}\left(1+\frac{1}{4\gamma^2 k}-\frac{1}{4\gamma k}+o\left(\frac{1}{k}\right)\right)\left(1+\frac{1}{2k\gamma}+o\left(\frac{1}{k}\right)\right) \tag{4.133}$$

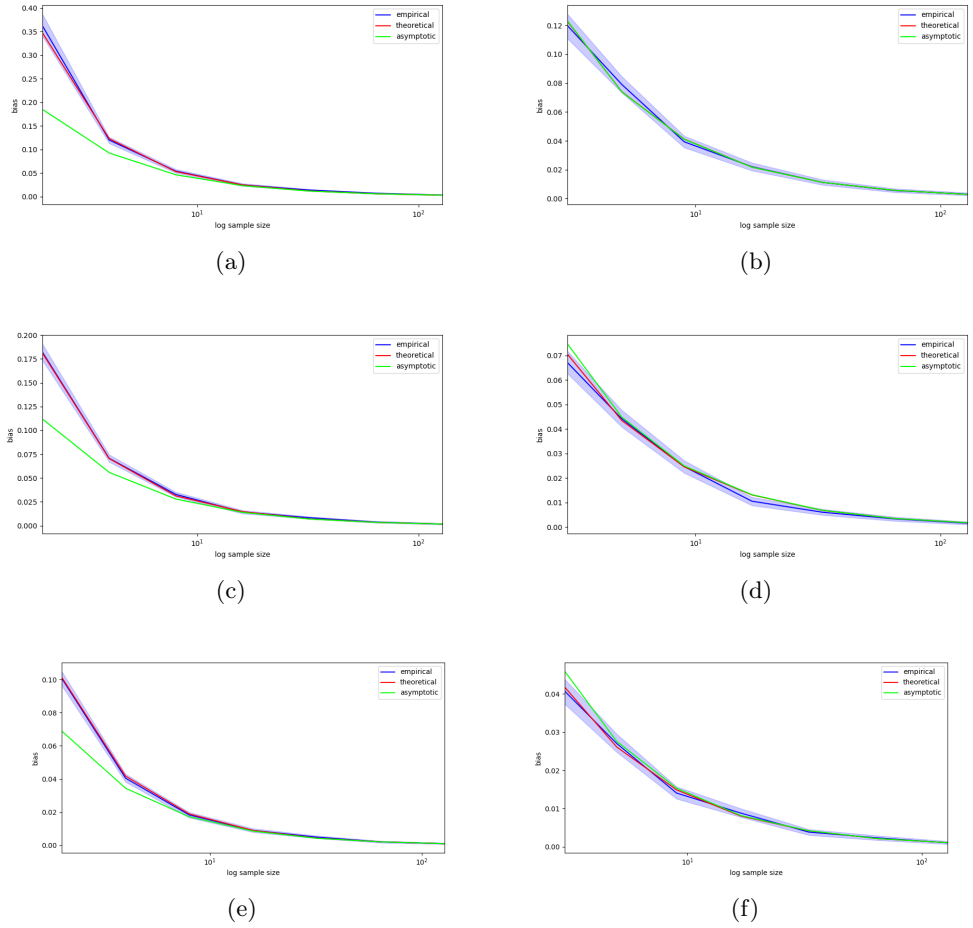
$$= 2^{\frac{1}{\gamma}}\left(1+\frac{1}{4\gamma^2 k}+\frac{1}{4\gamma k}+o\left(\frac{1}{k}\right)\right). \tag{4.134}$$

This completes the proof of the expansion of Theorem 4.22 when k is even. ■

In the graphs below, we display the evolution of the bias between the empirical and theoretical medians of the standardized Pareto distribution as a function of the log-sample size, both when the sample size is even and odd. See Figure 4.9.

Remark 4.23. *The logarithmic scale is used only in the abscissa to better visualize the decrease of the bias.*

In addition to the evolution of the bias of the empirical median, we are interested in the speed of convergence of this bias. Since the bias is a power function of the sample size, the logarithm of the bias as a function of the logarithm of the sample size is a linear



(a) (b) (c) (d) (e) (f)

Figure 4.9: Evolution of the bias between the empirical and theoretical medians of the Pareto distribution $\mathcal{P}(1, 2.5)$ ((a) - (b)), $\mathcal{P}(1, 3.5)$ ((c) - (d)) and $\mathcal{P}(1, 5)$ ((e) - (f)) with 95% CI over 10^4 repetitions versus the log sample size: for even sample sizes ((a) - (c) - (e)), for odd sample sizes ((b) - (d) - (f)).

function whose slope is the speed of convergence. Therefore, we study the log-log plot of the bias versus the sample size.

For both γ values, the slopes of the log-log plots are approximately -1; therefore, the speed of convergence of the bias of the empirical median is in $\frac{1}{n}$ where n is the sample size, as shown in Figure 4.10.

Moreover, this result can be extended to any non-standardized Pareto distribution $\mathcal{P}(x_m, \gamma)$ with a scaling parameter equal to x_m instead of 1. Indeed, owing to the scaling property of the Pareto distribution, the bias between the empirical and true medians of the non-standardized Pareto distribution $\mathcal{P}(x_m, \gamma)$ is proportional to the bias between the empirical and true medians of the standardized Pareto distribution with a proportionality factor equal to x_m.

Therefore, the bias between the empirical and true medians of the non-standardized Pareto distribution $\mathcal{P}(x_m, \gamma)$ corresponds to the bias between the empirical median and the true median of the standardized Pareto distribution $\mathcal{P}(1, \gamma)$, scaled by the

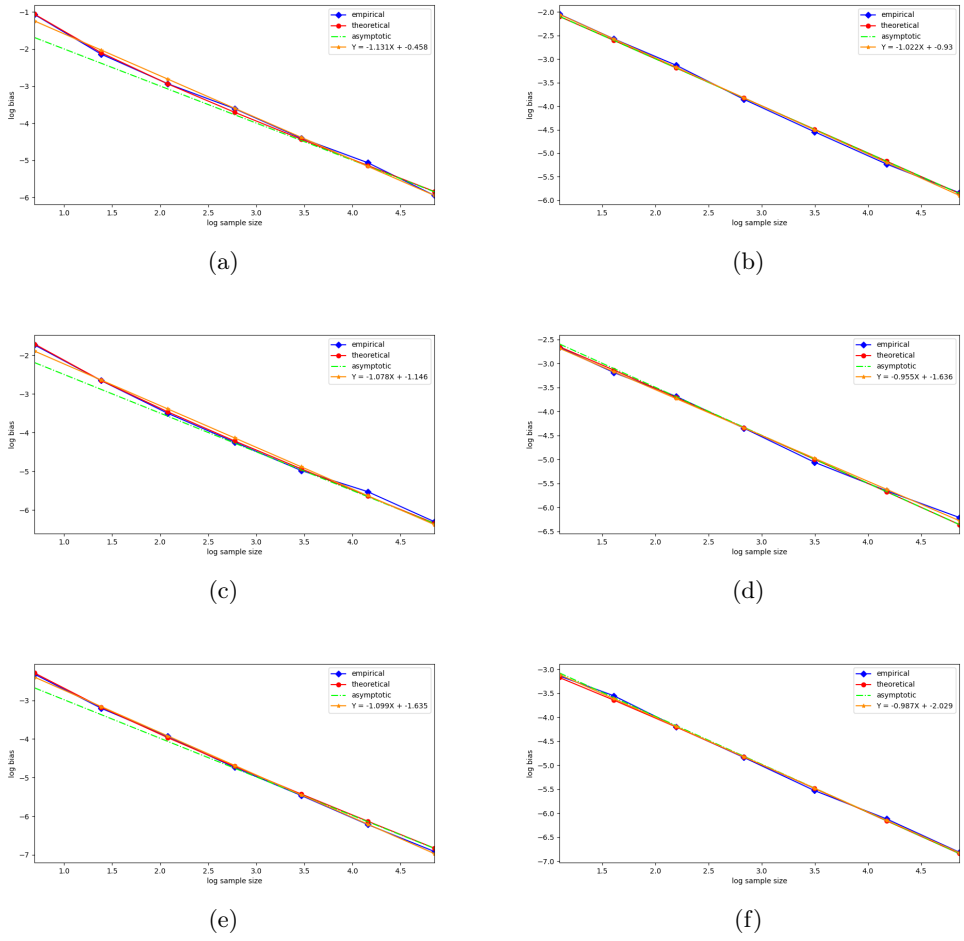


Figure 4.10: Evolution of the log bias between the empirical and theoretical medians of the Pareto distribution for $\mathcal{P}(1, 2.5)$ ((a) - (b)), $\mathcal{P}(1, 3.5)$ ((c) - (d)), and $\mathcal{P}(1, 2.5)$ ((e) - (f)), with 95% CI over 10^4 repetitions versus the log sample size: for even sample sizes ((a) - (c) - (e)), for odd sample sizes ((b) - (d) - (f)).

parameter x_m. Because x_m is independent of the sample, the speed of convergence of the bias between the empirical and true medians of the non-standardized Pareto distribution $\mathcal{P}(x_m, \gamma)$ is the same as the speed of convergence between the empirical and true medians of the Pareto distribution $\mathcal{P}(1, \gamma)$.

The empirical median estimator is a centrality estimator of a distribution, but is not an estimator of the expectation of the distribution. Thus, the empirical median of the distribution tail above the $\mathbf{VaR}_\alpha$ is not a proper estimator of the Expected-Shortfall. Therefore, this study will not be extended to the tail of the Pareto distribution. However, the study of the empirical median estimator has been carried out with a view to better understand the properties of a robust mean estimator, the Median-of-Means.

4.4 Median-of-Means (MoM) for heavy-tailed distribution

4.4.1 Presentation of the estimator

The MoM estimator combines the empirical mean and median estimators. The Median-of-Means estimator requires partitioning the data into k groups of roughly equal size, computing the empirical mean in each group, and taking the median of the sequence composed by the k empirical means.

Definition 4.24 (Median-of-Means (**MoM**))**.** *Let $X_1, \dots, X_n$ be n independent, identically, distributed random draws from the distribution of X. Let m, k be two positive integers. We assume that n is a multiple of k such that $n = mk$. The empirical mean of each block is defined as follows:*

$$\forall j \in [\![1, k]\!], \quad \bar{\mu}_{B_j} = \frac{1}{\left|B_j\right|} \sum_{i \in B_j} X_i. \tag{4.135}$$

The **MoM** *estimator is then defined by $\widehat{MoM}_n = M(\bar{\mu}_1, \dots, \bar{\mu}_k)$.*

Why can the MoM estimator be considered as a good estimator of the expectation ? (i) The first step in the construction of the MoM estimator consists of partitioning the set composed of n i.i.d. random draws into k disjoint blocks, and computing the empirical mean $\bar{\mu}_j$ of each of the k blocks. For the sake of simplicity, let us assume that the number of random draws n is a multiple of the number of blocks k such that $n = mk$, with m, k, n being three positive integers. For each of the k blocks, the empirical mean is an unbiased estimator of the expectation of the distribution. Indeed, because the random draws X_i are i.i.d. with mean $\mu = \mathbb{E}[X]$, then $\mathbb{E}[\bar{\mu}_j] = \mu$. Moreover, because the random draws are i.i.d. with variance $\mathbb{V}[X_i] = \sigma^2$ then, the standard deviation of the empirical means $\bar{\mu}_j$ is given by $\frac{\sigma}{\sqrt{m}}$. This means that, in each of the k blocks, the empirical mean does not deviate from the expectation of the distribution by more than a few units of $\frac{\sigma}{\sqrt{m}}$.

(ii) The second step consists of taking the median of the sequence composed of the k block-wise empirical means. Why can we say that the median of the block-wise means is a good estimator of the expectation of the distribution ?

Actually the empirical median does not deviate from the empirical mean by more than the empirical standard deviation, this is the following statement.

Lemma 4.25. *Let X be a scalar random variable, with standard deviation σ and median M. Then, the following inequality holds:*

$$\left|\mathbb{E}\left[X\right] - M\right| \leq \sigma. \tag{4.136}$$

Assuming that σ is sufficiently small, we can say that the median is close enough to the expectation of the distribution to be a good estimator.

Proof The triangular inequality leads to:

$$\left|\mathbb{E}\left[X\right] - M\right| \leq \mathbb{E}\left[\left|X - M\right|\right]. \tag{4.137}$$

Moreover, since the median minimizes the L_1-error, then we get:

$$\mathbb{E}\left[\left|X - M\right|\right] \leq \mathbb{E}\left[\left|X - \mathbb{E}\left[X\right]\right|\right]. \tag{4.138}$$

The Cauchy-Schwarz inequality implies that:

$$\mathbb{E}\left[\left|X - \mathbb{E}\left[X\right]\right|\right] \leq \sqrt{\mathbb{E}\left[\left|X - \mathbb{E}\left[X\right]\right|^2\right]}. \tag{4.139}$$

Consequently, we have:

$$\left|\mathbb{E}\left[X\right] - M\right| \leq \sqrt{\mathbb{E}\left[\left|X - \mathbb{E}\left[X\right]\right|^2\right]}. \tag{4.140}$$

■

In our case, if $\mathbf{X} = (X_1, \ldots, X_n)$ is a sample composed of n independent and identically distributed random variables, then $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ is the related empirical mean. Let us denote by $B_j, \forall j \in [\![1, k]\!]$ the k blocks composed of m random variables. The block-wise empirical mean $\bar{\mu}_{B_j} = \frac{1}{\left|B_j\right|}\sum_{i \in B_j} X_i$ is an unbiased estimator of the expectation μ of the distribution. Moreover, because the initial random draws are i.i.d. and the blocks are disjoint, the block-wise means are i.i.d. random variables. The empirical mean of the block-wise means is equal to the empirical mean of the initial sample $\hat{\mu}_n$ and the empirical variance is denoted by $\hat{\sigma}_n^2 = \frac{1}{k}\sum_{j=1}^{k}(\bar{\mu}_{B_j} - \hat{\mu}_n)^2$. Then, from Equation (4.136), the following concentration inequality is obtained:

$$\left|\hat{\mu}_n - \widehat{\mathbf{MoM}}_n\right| \leq \hat{\sigma}_n. \tag{4.141}$$

If the empirical standard deviation of the sample is small, we can conclude that the Median-of-Means estimator and empirical mean estimator are close to each other, and thus they should both estimate the true expectation. Nonetheless, in general, the fluctuations of the MoM estimator are smaller than those of empirical mean estimator, as we will see later.

Advantages of the MoM estimator This estimator allows for the combination of two centrality estimators: the empirical mean and empirical median. The empirical mean is not robust to extreme values and is significantly influenced by their presence. Therefore, the k block-wise empirical means consider the extreme values; they are unbiased with respect to the expectation of the sequence, but they have a large standard deviation. The empirical median is robust to extreme values because it depends only on the central values of the sequence, and the standard deviation of the empirical median

is expected to be much lower than that of the empirical mean; however, the median may exhibit a bias with respect to the expectation of the sequence. This implies that the final estimator can be sensitive to the data while not being too influenced by the extreme values of the distribution. Moreover, because the second layer of construction of the estimator relies on taking the empirical median of the sequence of block-wise empirical means, a bias is introduced with respect to the expectation of the sequence.

4.4.2 Concentration inequalities

As previously mentioned, the MoM estimator is sufficiently concentrated around the true expectation of the distribution to provide a good estimation.

Some results on the probability of deviation of the estimator provide insights into its performance. From (Lugosi and Mendelson, 2019, Theorem 2, p.7), the following result provides a good idea of the sub-Gaussian performance of the MoM estimator.

Theorem 4.26. *Let $X_1, \ldots, X_n$ be independent and, identically distributed random variables with mean μ and variance σ^2. Let m, k be positive integers, and assume that $n = mk$. Then, the Median-of-Means estimator $\widehat{MoM}_n$ with k blocks satisfies:*

$$\mathbb{P}\left(\left|\widehat{MoM}_n - \mu\right| > \sigma\sqrt{4/m}\right) \leq e^{-k/8}. \tag{4.142}$$

In particular, for any $\delta \in (0,1)$, if $k = \lceil 8\log(1/\delta)\rceil$, then, with probability at least $1-\delta$,

$$\left|\widehat{MoM}_n - \mu\right| \leq \sigma\sqrt{\frac{32\log(1/\delta)}{n}}. \tag{4.143}$$

Another result with more constraints (third absolute moment condition) is stated in (Lugosi and Mendelson, 2019, Theorem 4, p.9), as follows.

Theorem 4.27. *Let $X_1, \ldots, X_n$ be n independent and, identically distributed random variables with mean μ, variance σ^2, and third central moment $\rho = \mathbb{E}[|X-\mu|^3]$. Let m, k be positive integers, and assume that $n = mk$. Assume that:*

$$\sqrt{\frac{\log(2/\delta)}{2k}} + \frac{\rho}{2\sigma^3\sqrt{m}} \leq \frac{1}{4}. \tag{4.144}$$

Then, the **MoM** *estimator with k blocks satisfies, with a probability of at least $1-\delta$,*

$$\left|\hat{\mu}_n - \mu\right| \leq \frac{1}{c}\left(\sigma\sqrt{\frac{\log(2/\delta)}{n}} + \frac{\rho k}{2\sigma^2 n}\right) \tag{4.145}$$

where $c = \phi\left(\Phi^{-1}\left(\frac{3}{4}\right)\right)$ denotes a constant. Here, ϕ and Φ refer to the standard normal density and distribution function, respectively.

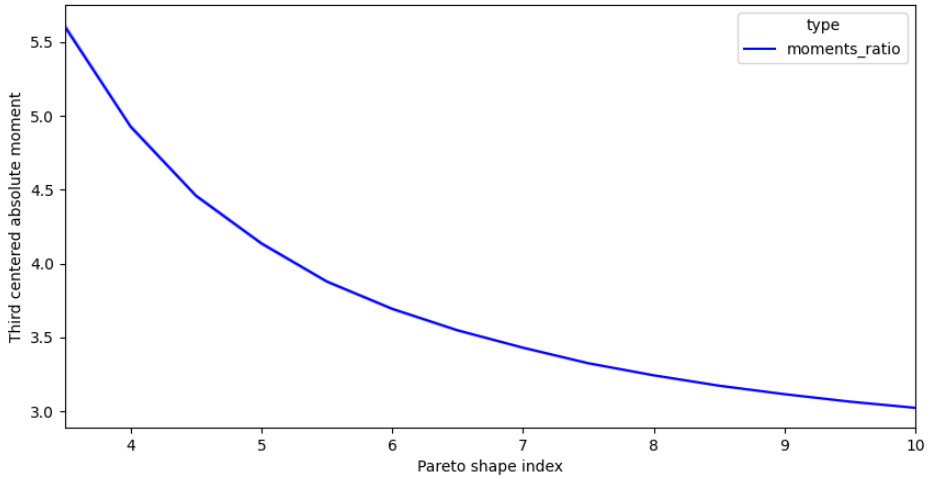


Figure 4.11: Evolution of the ratio of the absolute third centered absolute moment over the variance raised to the power $\frac{3}{2}$: $\frac{\rho}{\sigma^3}$, as a function of the shape parameter γ of the Pareto distribution. Graphs produced by Monte-Carlo simulation with 1000 random draws and 10^5 repetitions, in the Pareto distribution of scaling parameter $x_m = 1$ and with shape parameter γ varying from 3.5 to 10.5.

Interpretation: The ratio between the absolute third central moment ρ and the second central moment σ^2 raised to the power $\frac{3}{2}$ decreases as a function of γ and its limit is equal to 2, as displayed in Figure 4.11. This implies that the larger γ is, the smaller the ratio between the central moments, and the smaller the required size of the blocks. This can be explained by the fact that the larger the Pareto index γ is, the thinner the tails, that is, the fewer the extreme values, then, a small block size is sufficient to provide a good estimation.

4.4.3 Parameterizations of the MoM estimator

With Theorem 4.26, the chosen parameterization is the following $k = \lceil 8 \log(1/\delta) \rceil$ and $m = \lfloor n/k \rfloor$. The advantage of this theorem is that, even with a high level of confidence, it does not require too much data, therefore, it provides non-asymptotic results.

With Theorem 4.27, two different parameterizations are studied.

(i) The first parameterization assumes that:

$$\sqrt{\frac{\log(2/\delta)}{2k}} \leq \frac{1}{8} \quad \text{and} \quad \frac{\rho}{2\sigma^3\sqrt{m}} \leq \frac{1}{8}. \tag{4.146}$$

Under these assumptions we get:

$$k \geq 32 \log(2/\delta) \quad \text{and} \quad m \geq 16 \frac{\rho^2}{\sigma^6}. \tag{4.147}$$

This leads to $n \geq 512 \frac{\rho^2}{\sigma^6} \log(2/\delta)$.

Assuming that $\delta = 0.05$ and $\gamma = 2.5$, then $\frac{\rho}{\sigma^3} \simeq 7.551$, $\frac{\rho^2}{\sigma^6} \simeq 57.018$ and finally, we get $n \geq 107690$ which is too large if we want to work in a non-asymptotic framework.

(ii) The second parameterization assumes that:

$$\sqrt{\frac{\log(2/\delta)}{2k}} \leq \frac{1}{16} \quad \text{and} \quad \frac{\rho}{2\sigma^3\sqrt{m}} \leq \frac{3}{16}. \tag{4.148}$$

Under these assumptions we get:

$$k \geq 128 \log(2/\delta) \quad \text{and} \quad m \geq \frac{64}{9}\frac{\rho^2}{\sigma^6}. \tag{4.149}$$

This results in $n \geq 342\frac{\rho^2}{\sigma^6}\log(2/\delta)$. Assuming that $\delta = 0.05$ and $\gamma = 2.5$, then $\frac{\rho}{\sigma^3} \simeq 7.551$, $\frac{\rho^2}{\sigma^6} \simeq 57.018$ and finally, we get $n \geq 71934$ which is too large if we want to work in a non-asymptotic framework.

Theorem 4.27 requires the existence of a finite absolute third centered moment. This implies that, for small Pareto shape parameters, that is, for a heavy-tailed distribution, and for a high level of confidence, a very large amount of data is required. Such a parameterization can be similar to the asymptotic case.

In the sequel, only the parameterization of Theorem 4.26 will be retained to carry out the experiments, because the parameterization of Theorem 4.27 requires an extremely large number of samples with respect to the low performance improvement.

Recall that the goal of the survey is to study the bias between the empirical $\mathbf{ES}_\alpha$, when the latter is estimated by the MoM estimator applied to the standardized Pareto distribution tail, and the true $\mathbf{ES}_\alpha$ in two frameworks: the *idealized case* where the empirical α-quantile matches the true $\mathbf{VaR}_\alpha$ and the *realistic case* where the empirical α-quantile does not match the true $\mathbf{VaR}_\alpha$. To this end, we first study the bias between the **MoM** estimator applied to the entire standardized Pareto distribution $\mathcal{P}(1,\gamma)$ and the expectation.

4.4.4 Bias between the Median-of-Means (MoM) estimator applied to the entire standardized Pareto distribution and the expectation

Let $X_1, \ldots, X_n$ be n i.i.d. random variables following the standardized Pareto distribution $\mathcal{P}(1,\gamma)$, that is, $\forall i \in [\![1,n]\!]$, $X_i \overset{d}{=} X$ with $X \sim \mathcal{P}(1,\gamma)$. The expectation of X is denoted by $\mu = \mathbb{E}[X]$. The n random variables $X_1, \ldots, X_n$ are divided into k blocks of size $m = \lfloor n/k \rfloor$. The empirical mean $\bar{\mu}_k$ is computed for each of the k blocks. The MoM estimator then corresponds to the empirical median of the k block-wise empirical means: $\widehat{MoM}_n = \hat{M}(\bar{\mu}_1, \ldots, \bar{\mu}_k)$.

The bias between the **MoM** estimator and the expectation of the distribution is given by the following formula:

$$B_\mu(\widehat{MoM}_n) = \mathbb{E}\left[\widehat{MoM}_n\right] - \mu. \tag{4.150}$$

The stake is to compute the expectation of the $\widehat{\mathbf{MoM}}_n$ estimator.

Sketch of the analysis

Computation of the bias of the Median-of-Means is a delicate task. The goal of this section is to detail the approach used to establish an analytic closed-form formula for the bias of the Median-of-Means estimator, and to point out the difficulties encountered. First, an overview of the chosen approach is provided. Recall that the **MoM** estimator corresponds to the empirical median of the k block-wise empirical means computed on the sample composed of the n independently and identically distributed standardized Pareto random variables $X_1, \dots, X_n$:

$$\widehat{\mathbf{MoM}}_n = \hat{M}(\bar{\mu}_1, \dots, \bar{\mu}_k) \quad \text{with} \quad \forall j \in [\![1, k]\!], \quad \bar{\mu}_j = \frac{1}{\left|B_j\right|} \sum_{i \in B_j} X_i \quad \text{and } \hat{M} \text{ is the empirical median.} \tag{4.151}$$

The bias of the **MoM** estimator is defined as the difference between the expectation of the estimator and the expectation of the standardized Pareto distribution $\mathcal{P}(1, \gamma)$. Thus, the challenge is to compute the expectation of the **MoM** estimator. The natural way to reach this goal is to determine the density of the estimator and compute the first-order moment with respect to this density. The **MoM** estimator is the order statistic at the level 50% of the sequence of block-wise empirical means. As mentioned in Reiss 2012 (Reiss, 2012, p.21, Eq.1.3.5.), its density is given by:

$$f_{\widehat{\mathbf{MoM}}_n}(x) = \frac{k!}{(j-1)!(k-j)!} f_{\bar{\mu}_k}(x)(F_{\bar{\mu}_k}(x))^{j-1}(1 - F_{\bar{\mu}_k}(x))^{k-j}. \tag{4.152}$$

where $j = \frac{k+1}{2}$ if k is odd, $j = \left\lfloor \frac{k}{2} \right\rfloor$ if k is even, $f_{\bar{\mu}_k}$ and $F_{\bar{\mu}_k}$ are the probability density function and cumulative distribution function of the block-wise empirical means, respectively.

The issue is then to determine the density and cumulative distribution function of the block-wise empirical means. However, the probability density function of the block-wise empirical means is the inverse Fourier transform of the characteristic function of the block-wise empirical means:

$$f_{\bar{\mu}_k}(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \phi_{\bar{\mu}_j}(t) \mathrm{d}t \tag{4.153}$$

where $\phi_{\bar{\mu}_j}(t)$ is the characteristic function of the block-wise empirical means

$$\phi_{\bar{\mu}_k}(t) = \mathbb{E}[e^{it\bar{\mu}_k}], \tag{4.154}$$

which is supposed to be integrable for Equation (4.153) to be valid. A good reference is Durett 2019 (Durrett, 2019, p.111, Thm.3.1.1.).

Now, let us enter in more details in each step of the approach.

Expectation of the Median-of-Means estimator

The expectation of the **MoM** estimator is then defined as follows:

$$\mathbb{E}[\widehat{\mathbf{MoM}}_n] = \int_{-\infty}^{+\infty} x f_{\widehat{\mathbf{MoM}}_n}(x) \mathrm{d}x. \tag{4.155}$$

where $f_{\widehat{\mathbf{MoM}}_n}$ is the density of the **MoM** estimator.

Knowledge of the estimator density is useful for computing the expectation of the MoM estimator. Nevertheless, the density of the **MoM** estimator is more complicated than it sounds. The **MoM** estimator relies on two estimation layers. The first layer corresponds to the computation of the k block-wise empirical means. The second layer consists of taking the empirical median of the k block-wise empirical means. The median density can be determined from Equation (4.152). However, it relies on the density and cumulative distribution function of the underlying sample, on which the empirical mean is based. In our case, the underlying sample refers to the sequence of the k block-wise empirical means. The distribution of the k block-wise empirical means is not simple to determine.

However, two important aspects must be considered. On the one hand, the k block-wise empirical means are i.i.d. random variables since they are computed on disjoint blocks. This implies that determining the distribution of one block-wise empirical mean is sufficient. On the other hand, each of the k block-wise means is computed on m i.i.d. standardized Pareto random variables. For the sum of i.i.d. random variables, it is natural to compute the characteristic function, which leads to the next step.

Characteristic function of the Median-of-Means estimator

In this part, we are interested in computing the characteristic function of the k block-wise means $\bar{\mu}_j,\ \forall j \in [\![1, k]\!]$:

$$\bar{\mu}_j = \frac{1}{\left|B_j\right|} \sum_{l \in B_j} X_l := \frac{S_m}{m} \quad \text{with} \quad X_l \overset{i.i.d.}{\sim} \mathcal{P}(1, \gamma). \tag{4.156}$$

The characteristic function related to $\bar{\mu}_j$ is given by:

$$\phi_{\bar{\mu}_j}(t) = \mathbb{E}\left[e^{it\frac{S_m}{m}}\right]. \tag{4.157}$$

Because random variables X_l are identically and independently distributed, the following is obtained:

$$\phi_{\bar{\mu}_j}(t) = \mathbb{E}\left[\prod_{l \in B_j} e^{i\frac{t}{m}X_l}\right] = \prod_{l \in B_j} \mathbb{E}\left[e^{i\frac{t}{m}X_l}\right] \tag{4.158}$$

$$= \prod_{l \in B_j} \phi_{X_l}\left(\frac{t}{m}\right) \tag{4.159}$$

$$= \left(\phi_X\left(\frac{t}{m}\right)\right)^m \quad \text{with} \quad X \in \mathcal{P}(1, \gamma). \tag{4.160}$$

The formula of $\phi_X\left(\frac{t}{m}\right)$ is expressed in the following theorem.

Theorem 4.28 (Standard Pareto characteristic function)**.** *Let $X \sim \mathcal{P}(1,\gamma)$ with $\gamma > 1$. The characteristic function of X, given by $\phi_X(t) = \mathbb{E}\left[e^{itX}\right] = \int_1^{+\infty} \frac{\gamma}{x^{\gamma+1}} \mathrm{d}x$ is a solution of the Ordinary Differential Equation (ODE):*

$$\frac{\partial}{\partial t} J_\gamma(t) - \frac{\gamma}{t} J_\gamma(t) = -\frac{\gamma}{t} e^{it}, \tag{4.161}$$

and takes the following form:

$$\forall t \in \mathbb{R}, \phi_X(t) = \gamma(-it)^\gamma \Gamma(-\gamma, -it). \tag{4.162}$$

where $\Gamma(z) = \int_0^{+\infty} u^{z-1} e^{-u} \mathrm{d}u$, and satisfies the following relationship $\Gamma(z+1) = z\Gamma(z)$.

Proof ▷ Let us prove Equation (4.161), that is, let us establish the ODE for the Pareto characteristic function. Let $X \sim \mathcal{P}(1,\gamma)$. Let $\phi_X(t) = \mathbb{E}[e^{itX}] = \int_1^{+\infty} \frac{\gamma}{x^{\gamma+1}} e^{itx} \mathrm{d}x$ be the characteristic function of the standardized Pareto distribution. In the following, we consider $\phi_X(t) := J_\gamma(t)$.

Let us proceed to an integration by part taking: $u(x) = \frac{\gamma}{x^{\gamma+1}}$, $u'(x) = -\frac{\gamma(\gamma+1)}{x^{\gamma+2}}$, $v'(x) = e^{itx}$ and $v(x) = \frac{e^{itx}}{it}$: $J_\gamma(t) = \left[\frac{\gamma}{x^{\gamma+1}} \frac{e^{itx}}{it}\right]_{x=1}^{x\to+\infty} + \int_1^{+\infty} \frac{\gamma(\gamma+1)}{x^{\gamma+2}} \frac{e^{itx}}{it} \mathrm{d}x$.

However, the modulus of the complex number e^{itx} is bounded to 1. Consequently, $\lim_{x\to+\infty} \frac{\gamma}{x^{\gamma+1}} \frac{e^{itx}}{it} = 0$. Then we fall onto: $J_\gamma(t) = -\frac{\gamma e^{it}}{it} + \frac{\gamma}{it} \int_1^{+\infty} \frac{\gamma+1}{x^{\gamma+2}} e^{itx} \mathrm{d}x = -\frac{\gamma e^{it}}{it} + \frac{\gamma}{it} J_{\gamma+1}(t)$

Moreover, $J_\gamma(t) = \int_1^{+\infty} \frac{\gamma}{x^{\gamma+1}} e^{itx} \mathrm{d}x$, let us differentiate once with respect to t: $\frac{\partial}{\partial t} J_\gamma(t) = i\frac{\gamma}{\gamma-1} J_{\gamma-1}(t)$. To justify the differentiation, we used $\gamma > 1$.

Then we get: $\frac{\partial}{\partial t} J_{\gamma+1}(t) = i\frac{\gamma+1}{\gamma} J_\gamma(t)$, and thus: $J_\gamma(t) = \frac{\gamma}{i(\gamma+1)} \frac{\partial}{\partial t} J_{\gamma+1}(t)$.

Therefore, we have, $J_\gamma(t) = -\frac{\gamma}{it} e^{it} + \frac{\gamma}{it} J_{\gamma+1}(t)$ and $J_\gamma(t) = \frac{\gamma}{i(\gamma+1)} \frac{\partial}{\partial t} J_{\gamma+1}(t)$.

Then we obtain: $\frac{\gamma}{i(\gamma+1)} \frac{\partial}{\partial t} J_{\gamma+1}(t) = -\frac{\gamma}{it} e^{it} + \frac{\gamma}{it} J_{\gamma+1}(t)$ and $\frac{\partial}{\partial t} J_{\gamma+1}(t) = -\frac{\gamma+1}{t} e^{it} + \frac{\gamma+1}{t} J_{\gamma+1}(t)$.

Finally, we fall onto: $\frac{\partial}{\partial t} J_\gamma(t) = -\frac{\gamma}{t} e^{it} + \frac{\gamma}{t} J_\gamma(t)$.

▷ Let us now prove Equation (4.162), that is, let us determine the form of the standardized Pareto characteristic function. ODE (4.161) is difficult to solve. Therefore, another method is used to determine the explicit form of the characteristic function. Finally, a test is performed to check whether the founded function satisfies the ODE well. The method is explained in detail in the following proof.

Let $X \sim \mathcal{P}(1,\gamma)$. To determine the characteristic function of X we use complex analysis.

We define beforehand the framework of the study, which will allow us to use the theorems of complex analysis.

Let H be an open, star-connected subset of $\mathbb{C}$. This property is required for using the theorems of complex analysis. We choose H open because we assume that the Gamma function admits a holomorphic extension, and a holomorphic function is defined on an open subset.

The goal is to find an explicit form for the characteristic function of X defined by $\phi_X(t) = \mathbb{E}[e^{itx}] = \int_1^{+\infty} \frac{\gamma}{x^{\gamma+1}} e^{itx} \mathrm{d}x$.

(i) First, we generalize ϕ_X. For this purpose, we fix t and take:

$$h(s) = \int_1^{+\infty} \frac{\gamma}{x^{\gamma+1}} e^{sx} \mathrm{d}x. \tag{4.163}$$

(ii) Let us prove that h is well-defined on an open star-connected subset H of $\mathbb{C}$.

Let $H = \left\{ s \in \mathbb{C}^\star \text{s.t. } Re(s) < 0 \right\}$ be a convex part of $\mathbb{C}$. To prove that h is well-defined on H, we have to prove that the integral is converging for $s \in \mathbb{C}^\star$ with $Re(s) < 0$.

$$\left| \frac{\gamma}{x^{\gamma+1}} e^{sx} \right| = \gamma \left| \frac{e^{sx}}{x^{\gamma+1}} \right| = \gamma \left| \frac{e^{(Re(s)+iIm(s))x}}{x^{\gamma+1}} \right| = \gamma \left| \frac{e^{Re(s)x} e^{iIm(s)x}}{x^{\gamma+1}} \right| \tag{4.164}$$

However, the modulus of the complex exponential $e^{iIm(s)x}$ is bounded to 1. Moreover, $Re(s) < 0$ then $\left| e^{Re(s)x} \right| \leq 1$ and we get:

$$\left| \frac{\gamma}{x^{\gamma+1}} e^{sx} \right| \leq \left| \frac{\gamma}{x^{\gamma+1}} \right|. \tag{4.165}$$

Let us consider $g(x) = \frac{\gamma}{x^{\gamma+1}}$. Since $\gamma > 0$ then $\int_1^{+\infty} g(x) \mathrm{d}x$ is converging. Consequently, h is well-defined on H.

(iii) Now, let us prove that h is holomorphic on H. Let us take $f(s,x) = \gamma x^{-(\gamma+1)} e^{sx}$ as the integrand of the function h. Function $f(s,\cdot)$ is holomorphic using standard operations on classic standard functions.

First, $f(s,\cdot)$ is holomorphic on H. Then, $f(s,\cdot)$ is measurable with respect to the Borelian sigma-algebra, by the product of continuous measurable functions. Finally, $\left| f(s,x) \right| \leq g(x)$ with $g(x) \in L^1(\mathbb{R})$.

The integrand satisfies all the conditions of the theorem of holomorphism under the integral, then h is holomorphic on H.

(iv) In a first time, we consider $s \in \mathbb{R}^\star_-$ ($\mathbb{R}^\star_-$ is a subset of $\mathbb{C}^\star$ with Re(s) < 0.)

If $s \in \mathbb{R}^\star_-$, the following substitution: $u = -xs$, $\mathrm{d}u = -s\mathrm{d}x$, when $x = 1 \Rightarrow u = -x$ and when $x \to +\infty \Rightarrow u \to +\infty$, leads to:

$$h(s) = \gamma(-s)^\gamma \int_{-s}^{+\infty} u^{-\gamma-1} e^{-u} \mathrm{d}u = \gamma(-s)^\gamma \Gamma(-\gamma, -s). \tag{4.166}$$

Consequently, $h(s)$ and $\tilde{h}(s) = \gamma(-s)^\gamma \Gamma(-\gamma, -s)$ match on $\mathbb{R}^\star_-$ that is a non-isolated part of H.

(v) The isolated-zero principle is used to prove that $h(s) = \tilde{h}(s)$ on the entire subset H.

We proved that h is holomorphic on H based on the theorem of holomorphism under the integral. Moreover, function $\tilde{h}$ is holomorphic on H by standard operations on the holomorphic functions on H. The function $h - \tilde{h}$ is worth 0 on $\mathbb{R}^\star_-$ since these two

functions match on $\mathbb{R}_-^\star$ that is a subset of H. However, $h-\tilde{h}$ is holomorphic because it is zero everywhere or on an isolated subset of H. But $\mathbb{R}_-^\star$ is a non-isolated part of H then $h-\tilde{h}$ is zero everywhere on H. Therefore, h and $\tilde{h}$ match on the whole subset H. Consequently, $\forall s \in H$, $h(s) = \gamma(-s)^\gamma \Gamma(-\gamma, -s)$.

(vi) Now, we want to evaluate function h at the complex point it whose real part is worth zero. But $0 \notin H$, then we have to go to the limit. We must prove that we can go to the limit.

First, we have to prove that $h(it)$ is still well-defined, that is, the integral defined for $s = it$ is still converging:

$$h(it) = \int_1^{+\infty} \frac{\gamma}{x^{\gamma+1}} e^{itx} \mathrm{d}x. \tag{4.167}$$

Since $\left|e^{itx}\right| \leq 1$, then $\left|\frac{\gamma}{x^{\gamma+1}} e^{itx}\right| \leq \left|\frac{\gamma}{x^{\gamma+1}}\right|$.

But $g(x) = \frac{\gamma}{x^{\gamma+1}} \in L^1(\mathbb{R})$ then $\int_1^{+\infty} g(x)\mathrm{d}x$ converges. Therefore, $\int_1^{+\infty} \frac{\gamma}{x^{\gamma+1}} e^{itx}\mathrm{d}x$ converges. Consequently, $h(it)$ is well-defined.

Now, let us prove that we can go to the limit in 0 by lower values. We want to prove that $h(it) = \lim_{r\to 0^-} h(r+it) \in H$.

Let us take $f(r+it, x) = \gamma x^{-(\gamma+1)} e^{(r+it)x}$. The limit exists at fixed t. Indeed, at fixed t, for $r < 0$, we have $\left|e^{itx}\right| \leq 1$ and $e^{rx} \leq 1$ then, we still have:

$$\left|f(r+it, x)\right| = \left|\gamma x^{-(\gamma+1)} e^{(r+it)x}\right| \leq \left|\frac{\gamma}{x^{\gamma+1}}\right|; \quad \forall r, \forall x. \tag{4.168}$$

Then, $\int_1^{+\infty} f(r+it, x)\mathrm{d}x$ is well-defined.

The upper bound $g(x) = \frac{\gamma}{x^{\gamma+1}}$ is a function that does not depend on r and that is integrable with respect to x.

By dominated convergence theorem, we can swap limit and integral.

$$\int_1^{+\infty} \lim_{r\to 0^-} f(r+it, x)\mathrm{d}x = \lim_{r\to 0^-} \int_1^{+\infty} f(r+it, x)\mathrm{d}x = \int_1^{+\infty} \gamma x^{-(\gamma+1)} e^{itx}\mathrm{d}x = h(it). \tag{4.169}$$

Therefore, $h(it) = \lim_{r\to 0^-} h(r+it) = \lim_{r\to 0^-} \tilde{h}(r+it)$ because $r+it \in H$ and h and $\tilde{h}$ match on H. But the function $\tilde{h}$ is holomorphic by standard operations on holomorphic functions on $\mathbb{C} \times \mathbb{C} \setminus \{(-k, 0), \forall k \in \mathbb{N}\}$ since the incomplete Γ function is not defined when γ is a negative integer. Consequently, h and $\tilde{h}$ match on $\mathbb{C}\times\mathbb{C}\setminus\{(-k, 0), \forall k \in \mathbb{N}\}$.

Finally, we conclude that $\phi_X(t) = \tilde{h}(it) = \gamma(-it)^\gamma \Gamma(-\gamma, -it)$.

▷ Now, let's check that $\phi_X(t) = \gamma(-it)^\gamma \Gamma(-\gamma, -it)$ satisfies the ODE.

Let us consider the holomorphic function $\phi_X(t) = \gamma(-it)^\gamma \Gamma(-\gamma, -it)$.

(i) Let us rewrite the function ϕ owing to the complex logarithm. $\forall z \in \mathbb{C}$, the complex logarithm is defined up to a multiple of 2π by $\mathrm{Log}(z) = \log(|z|) + i\arg(z)$ where $|z|$ refers to the modulus of z and $\arg(z)$ refers to the argument of z.

Then ϕ_X can be rewritten as follows: $\phi_X(t) = \gamma e^{\gamma \text{Log}(-it)}\Gamma(-\gamma, -it)$ where $\text{Log}(-it) = \log(|-it|) + i\arg(-it) = \log(|t|) - i\frac{\pi}{2}$.

Moreover, let us rewrite the incomplete Gamma function based on the Gamma function. First, we recall some important points. So far, we have called the incomplete Gamma function the upper incomplete Gamma function defined as: $\Gamma(s,x) = \int_x^{+\infty} u^{s-1}e^{-u}\mathrm{d}u$. The lower incomplete Gamma function is defined by: $\bar{\gamma}(s,x) = \int_0^x u^{s-1}e^{-u}\mathrm{d}u$.

The Gamma function is defined as: $\Gamma(s) = \Gamma(s,0) = \int_0^{+\infty} u^{s-1}e^{-u}\mathrm{d}u$. The relationship between the Gamma function and incomplete Gamma functions is $\Gamma(-\gamma, -it) = \Gamma(-\Gamma, 0) - \bar{\gamma}(-\gamma, -it)$.

Then, we fall onto $\phi_X(t) = e^{\gamma\log(|t|) - \gamma i\frac{\pi}{2}}\Big(\Gamma(-\gamma,0) - \bar{\gamma}(-\gamma,-it)\Big)$.

(ii) Let us compute the first derivative of ϕ_X with respect to t as follows:

$$\phi'_X(t) = \gamma\frac{\partial}{\partial t}\left(e^{\gamma\log(|t|) - \gamma i\frac{\pi}{2}}\right)\Big(\Gamma(-\gamma,0) - \bar{\gamma}(-\gamma,-it)\Big) + \gamma e^{\gamma\log(|t|) - \gamma i\frac{\pi}{2}}\frac{\partial}{\partial t}\Big(\Gamma(-\gamma,0) - \bar{\gamma}(-\gamma,-it)\Big).$$

On the one hand, $\frac{\partial}{\partial t}\left(e^{\gamma\log(|t|) - \gamma i\frac{\pi}{2}}\right) = \frac{\gamma}{|t|}(-it)^{\gamma}$.

On the other hand, $\frac{\partial}{\partial t}\Big(\Gamma(-\gamma,0) - \bar{\gamma}(-\gamma,-it)\Big) = -\frac{\partial}{\partial t}\bar{\gamma}(-\gamma,-it) = ie^{-(\gamma+1)\log(|t|)}e^{(\gamma+1)i\frac{\pi}{2}}e^{it}$.

Plugging in the formula of ϕ'_X we get: $\phi'_X(t) = \frac{\gamma}{|t|}\Big(\gamma(-it)^{\gamma}\Gamma(-\gamma,-it)\Big) - \frac{\gamma}{|t|}e^{it} = \frac{\gamma}{|t|}\phi_X(t) - \frac{\gamma}{|t|}e^{it}$.

Therefore, we fall onto:

$$\phi'_X(t) - \frac{\gamma}{|t|}\phi_X(t) = -\frac{\gamma}{|t|}e^{it}. \tag{4.170}$$

Consequently, $\phi_X(t) = \gamma(-it)^{\gamma}\Gamma(-\gamma,-it)$ satisfies the ODE $\frac{\partial}{\partial t}J_\gamma(t) - \frac{\gamma}{t}J_\gamma(t) = -\frac{\gamma}{t}e^{it}$.
■

From Equation (4.162), the characteristic function of the block-wise mean $\bar{\mu}_j$ can be defined as follows:

$$\phi_{\bar{\mu}_j}(t) = \gamma^m\left(-i\frac{t}{m}\right)^{m\gamma}\left(\Gamma\left(-\gamma, -i\frac{t}{m}\right)\right)^m. \tag{4.171}$$

Inverse Fourier transform of the characteristic function

Let us recall that the initial motivation is to determine the density and cumulative distribution of the block-wise mean. In this part, the objective is to derive from the previously established characteristic function, the density function.

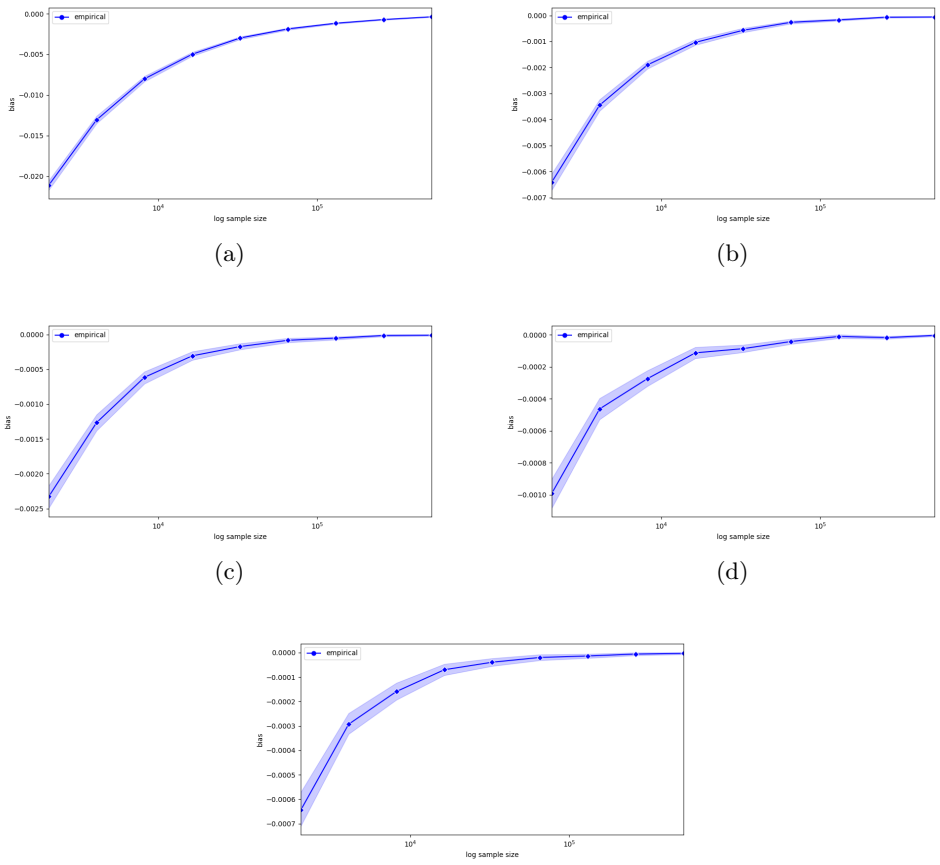
Recall that the density function is the inverse Fourier transform of the characteristic function. Let us denote by $F\circ\phi_{\bar{\mu}_j}(x) = \int_{\mathbb{R}}\frac{1}{2\pi}e^{-itx}\phi_{\bar{\mu}_j}(t)\mathrm{d}t$ the inverse Fourier transform of the characteristic function of the block-wise mean $\bar{\mu}_j$. Then, the density function of

the block-wise mean can be recovered using the inverse Fourier transform, as follows:

$$f_{\bar{\mu}_j}(x) = \int_{\mathbb{R}} \frac{1}{2\pi} e^{-itx} \phi_{\bar{\mu}_j}(t) \mathrm{d}t. \tag{4.172}$$

However, because the characteristic function of the block-wise mean $\bar{\mu}_j$ depends on the Gamma function, which is already an integral, it is very complicated to determine an explicit analytic form for the density of the block-wise mean. In this framework, it is difficult to obtain an analytic form of the bias between the **MoM** estimator applied to the standardized Pareto distribution and its expectation. This is still a work in progress. We provide numerical experiments to have an insight about the convergence rate of the bias of the MoM estimator applied to the standardized Pareto distribution.

The evolution of the bias between the **MoM** estimator applied to the entire Pareto distribution and its expectation is shown in Figure 4.12.



(a) (b) (c) (d) (e)

Figure 4.12: Evolution of the bias between the **MoM** estimator and the expectation of the entire standardized Pareto distribution $\mathcal{P}(1, \gamma)$ for five different shape parameters, versus the log sample size, with 95% CI, over 10^4 repetitions, (a) $\mathcal{P}(1, 2.5)$, (b) $\mathcal{P}(1, 3.5)$, (c) $\mathcal{P}(1, 5)$, (d) $\mathcal{P}(1, 7.5)$, (e) $\mathcal{P}(1, 10)$.

Remark 4.29. *On the above graphs, the logarithmic scale is only used in the abscissa to better visualize the evolution of the bias.*

We are interested in the convergence speed of the bias between the **MoM** estimator applied to the entire standardized Pareto distribution $\mathcal{P}(1,\gamma)$ and the expectation. Because this bias is presumably a power function of the sample size, the logarithm of the bias is a expected to be a linear function of the logarithm of the sample size. Therefore, we study the log-log plot of bias versus sample size.

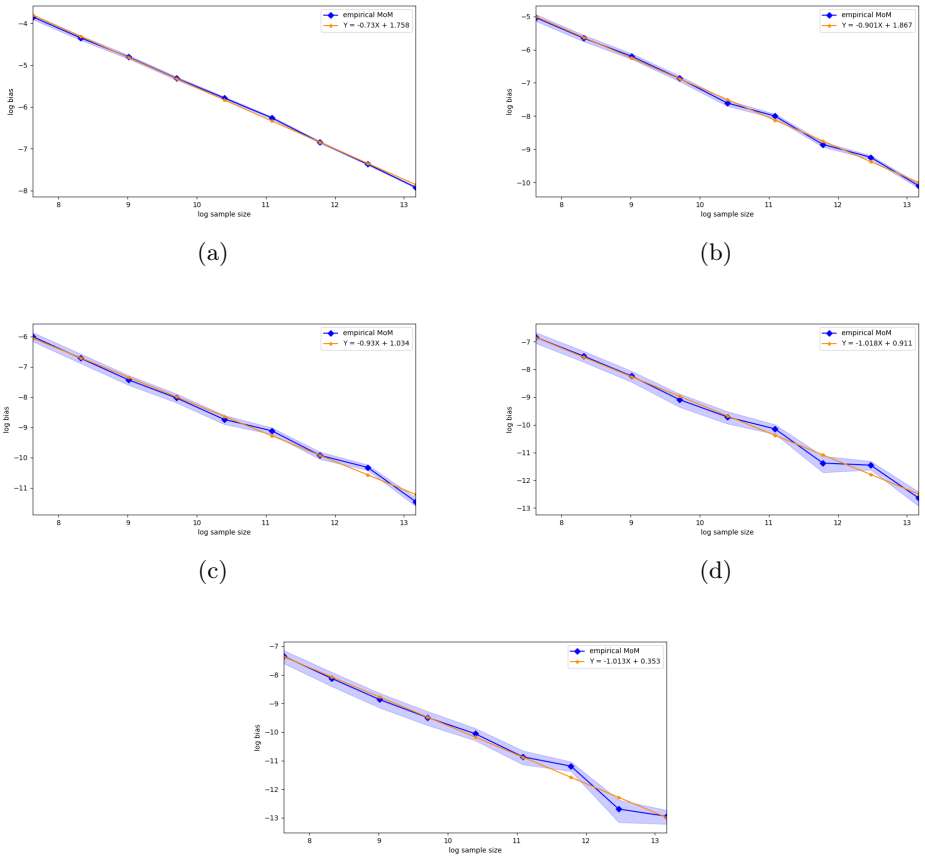


Figure 4.13: Evolution of the log of the bias between the **MoM** estimator and the expectation of the entire Pareto distribution for five different shape parameters versus the log sample size, with 95% CI, over 10^5 repetitions, (a) $\mathcal{P}(1,2.5)$, (b) $\mathcal{P}(1,3.5)$, (c) $\mathcal{P}(1,5)$, (d) $\mathcal{P}(1,7.5)$, (e) $\mathcal{P}(1,10)$.

The convergence rate of the bias between the **MoM** estimator and the expectation of the entire standardized Pareto distribution $\mathcal{P}(1,\gamma)$ seems to be dependent on the Pareto index γ. Indeed, the convergence rate varies between $\frac{1}{n}$ and $\frac{1}{n^{0.7}}$ and becomes closer to $\frac{1}{n}$ when the distribution tail becomes increasingly thinner, that is, when γ is increasing. See Figure 4.13.

Now that we have studied the bias between the MoM estimator applied to the entire standardized Pareto distribution and the expectation, we focus on the bias between the **MoM** estimator applied to the entire standardized Pareto distribution tail and the $\mathbf{ES}_\alpha$ in two different frameworks. The first framework refers to the idealized case where the empirical α-quantile matches the theoretical $\mathbf{VaR}_\alpha$. The second framework

corresponds to the realistic case where the empirical α-quantile does not match the theoretical $\mathbf{VaR}_\alpha$.

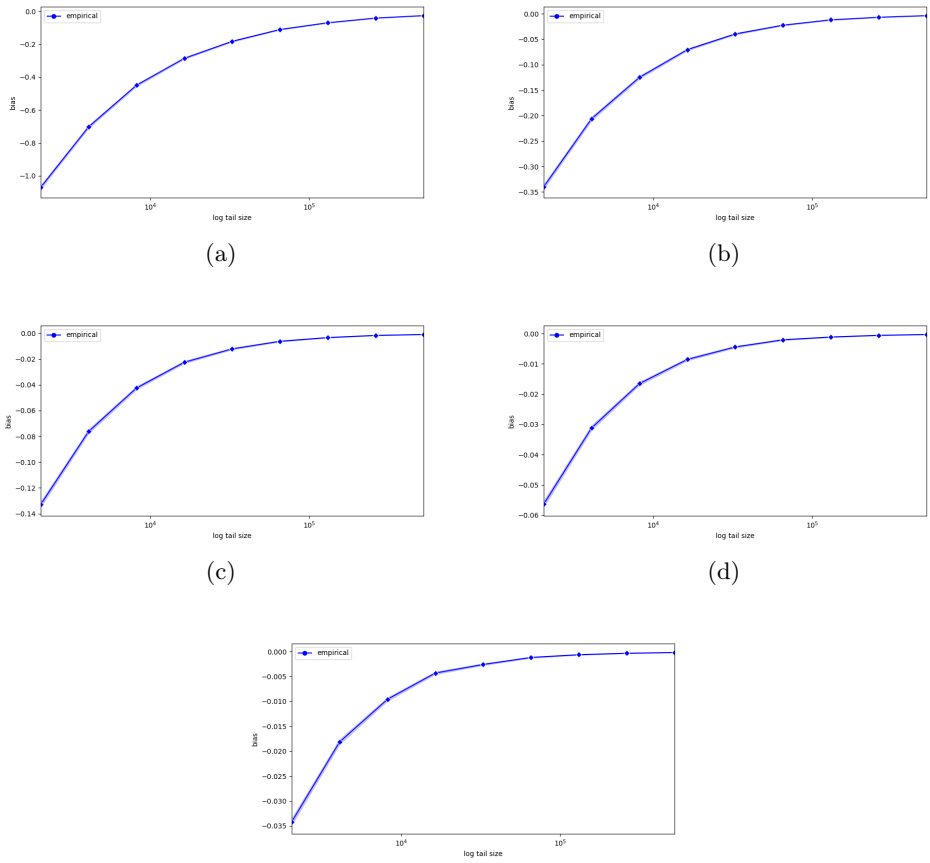
4.4.5 Bias between the MoM estimator applied to the standardized Pareto distribution tail above the true $\mathbf{VaR}_\alpha$ and the true ES (Idealized case)

In the idealized case, the empirical α-quantile matches the true $\mathbf{VaR}_\alpha$, which is supposed to be known. In this case, the estimator of the $\mathbf{ES}_\alpha$ in the standardized Pareto distribution $\mathcal{P}(1,\gamma)$, is the empirical average of the standardized Pareto distribution conditional on its values being above the true $\mathbf{VaR}_\alpha$. The conditioning threshold is independent on the underlying sample. This implies that the samples larger than the true $\mathbf{VaR}_\alpha$ are independent and identically distributed (i.i.d.) and the stability by conditioning and scaling properties of the Pareto distribution, as stated in Theorem 4.7, are valid. The stability by conditioning property implies that the standardized Pareto distribution $\mathcal{P}(1,\gamma)$ conditional on its values being above the true $\mathbf{VaR}_\alpha$ is still a Pareto distribution, with the same shape parameter γ, but a new scaling parameter equal to the conditioning parameter $\mathbf{VaR}_\alpha$. The scaling property states that this non-standardized Pareto distribution $\mathcal{P}(\mathbf{VaR}_\alpha,\gamma)$ is proportional to the marginal Pareto distribution $\mathcal{P}(1,\gamma)$, with a proportionality factor equal to the conditioning parameter $\mathbf{VaR}_\alpha$. Therefore, the standardized Pareto distribution conditional on its values being above the true $\mathbf{VaR}_\alpha$ is proportional to the standardized Pareto distribution with a proportionality factor equal to $\mathbf{VaR}_\alpha$. This implies that the **MoM** estimator applied to the standardized Pareto distribution conditional on its values being above the true $\mathbf{VaR}_\alpha$, is proportional to the **MoM** estimator applied to the entire standardized Pareto distribution, with a proportionality factor equal to $\mathbf{VaR}_\alpha$. Similarly, the $\mathbf{ES}_\alpha$ is proportional to the expectation of the standardized Pareto distribution, with a proportionality factor equal to $\mathbf{VaR}_\alpha$. Thus, the bias between the **MoM** estimator applied to the standardized Pareto distribution conditional on its values being above the true $\mathbf{VaR}_\alpha$ and the $\mathbf{ES}_\alpha$ is proportional to the bias between the **MoM** estimator applied to the entire standardized Pareto distribution and the expectation, with a proportionality factor equal to $\mathbf{VaR}_\alpha$.

$$B_{\mathbf{ES}_\alpha}(\widehat{\mathbf{MoM}}_n) = \mathbf{VaR}_\alpha \times B_\mu(\widehat{\mathbf{MoM}}_n). \tag{4.173}$$

Because $B_\mu(\widehat{\mathbf{MoM}}_n)$ converges asymptotically to 0, the bias $B_{\mathbf{ES}_\alpha}(\widehat{\mathbf{MoM}}_n)$ converges asymptotically to 0 too.

In the idealized case, we approximately recover the convergence rate of the bias between the MoM estimator applied to the entire standardized Pareto distribution and the expectation, between $\frac{1}{n}$ and $\frac{1}{n^{0.7}}$. This can be explained as follows. Owing to the stability by conditioning and scaling properties of the Pareto distribution, the bias between the MoM estimator applied to the standardized Pareto distribution tail above the true $\mathbf{VaR}_\alpha$ and $\mathbf{ES}_\alpha$ is proportional to the bias between the MoM estimator applied to the entire standardized Pareto distribution and the expectation, with a proportionality factor equal to $\mathbf{VaR}_\alpha$. Because the proportionality factor $\mathbf{VaR}_\alpha$ is independent on the



(a) (b) (c) (d) (e)

Figure 4.14: Evolution of the bias between the **MoM** estimator applied to the standardized Pareto distribution tail above the true $\mathbf{VaR}_\alpha$ and the $\mathbf{ES}_\alpha$, for five different shape parameters versus the sample size, with 95% CI, over 10^4 repetitions, (a) $\mathcal{P}(1, 2.5)$, (b) $\mathcal{P}(1, 3.5)$, (c) $\mathcal{P}(1, 5)$, (d) $\mathcal{P}(1, 7.5)$, (e) $\mathcal{P}(1, 10)$.

underlying sample, the convergence rate of the bias between the MoM estimator applied to the standardized Pareto distribution tail above $\mathbf{VaR}_\alpha$ and the true $\mathbf{ES}_\alpha$ is the same as that of the bias between the MoM estimator applied to the entire standardized Pareto distribution and the expectation. See Figure 4.15.

4.4.6 Bias between the MoM estimator applied to the standardized Pareto distribution tail above the empirical α-quantile and the true ES (Realistic case)

In the realistic case, the empirical α-quantile does not match the true $\mathbf{VaR}_\alpha$. In this case, the estimator of the $\mathbf{ES}_\alpha$ in the standardized Pareto distribution $\mathcal{P}(1, \gamma)$ is the empirical average of the standardized Pareto distribution conditional on its values being above the empirical α-quantile. The conditioning threshold is an order statistics and depends on the underlying sample. This implies that the samples larger than the empirical α-quantile are no longer independent and identically distributed (i.i.d.). The

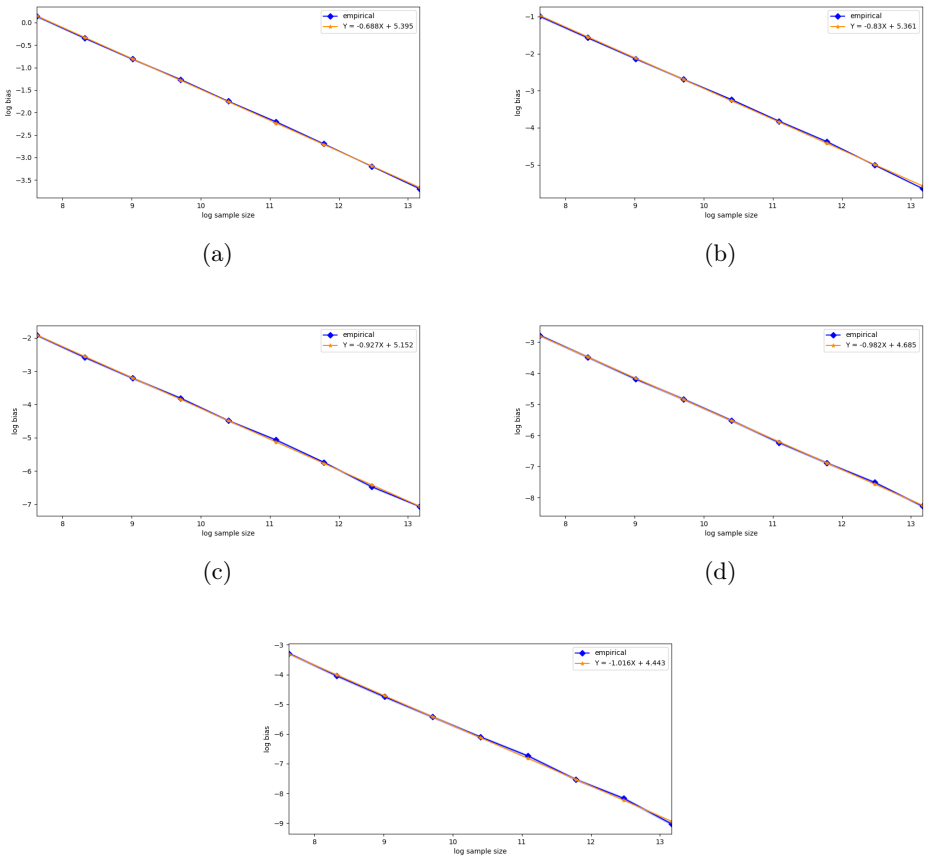
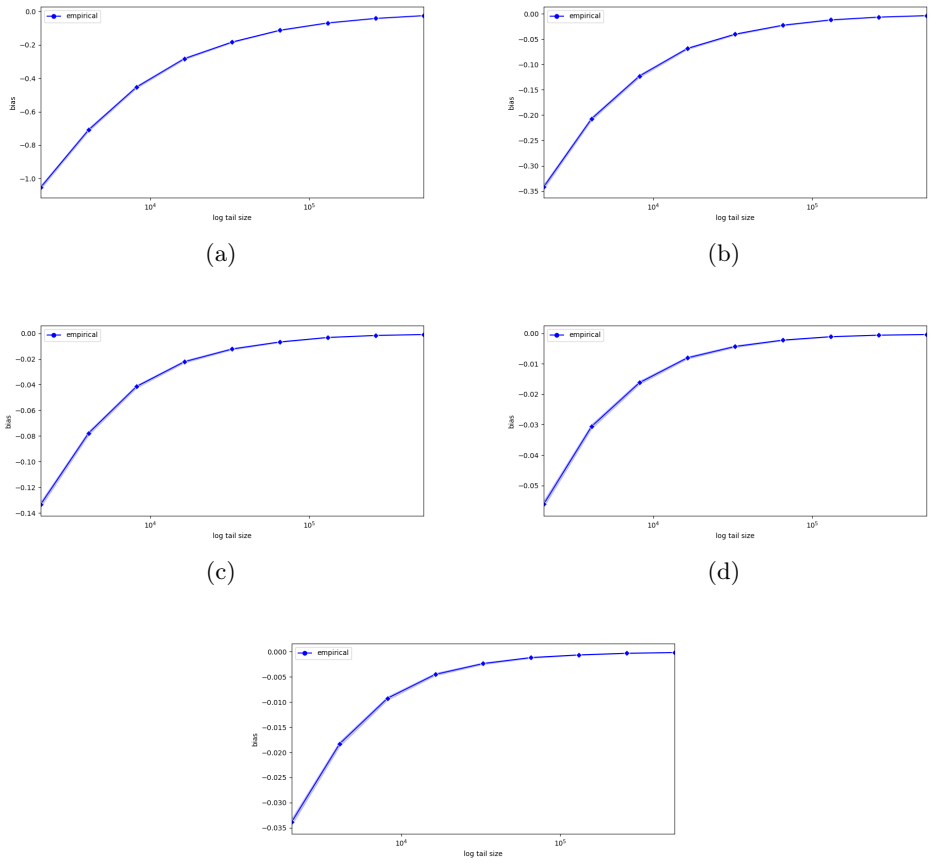


Figure 4.15: Log bias between the **MoM** estimator applied to the standardized Pareto distribution tail above the true $\mathbf{VaR}_\alpha$ and the Expected-Shortfall, for five shape parameters, versus the log sample size, with 95% CI, over 10^4 repetitions, (a) $\mathcal{P}(1, 2.5)$, (b) $\mathcal{P}(1, 3.5)$, (c) $\mathcal{P}(1, 5)$, (d) $\mathcal{P}(1, 7.5)$, (e) $\mathcal{P}(1, 10)$.

stability by conditioning and scaling properties of the Pareto distribution, as stated in Theorem 4.7, are no longer valid. Therefore, the distribution of the samples larger than the empirical α-quantile is not necessarily a Pareto distribution. The distribution of the samples larger than the empirical α-quantile is unknown and it is more challenging to establish an analytic closed-form formula for the bias between the **MoM** estimator applied to the standardized Pareto distribution tail above this threshold and the $\mathbf{ES}_\alpha$. For this reason, we only provide simulations. See Figures 4.16 and 4.17.

It is expected that the bias between the **MoM** estimator applied to the standardized Pareto distribution tail above the empirical α-quantile and the $\mathbf{ES}_\alpha$, be a power function of the sample size. Therefore, it is expected that the logarithm of the bias be a linear function of the logarithm of the sample size. For this reason, to have an insight into the convergence speed of the bias between the **MoM** applied to the standardized Pareto distribution tail above the empirical α-quantile and the $\mathbf{ES}_\alpha$, we study the log-log plot of the bias versus the sample size. See Figure 4.17.



(a) (b) (c) (d) (e)

Figure 4.16: Evolution of the bias between the **MoM** estimator applied to the standardized Pareto distribution tail above the empirical α-quantile and the $\mathbf{ES}_\alpha$, for five different shape parameters versus the log sample size, with 95% CI, over 10^5 repetitions, (a) $\mathcal{P}(1, 2.5)$, (b) $\mathcal{P}(1, 3.5)$, (c) $\mathcal{P}(1, 5)$, (d) $\mathcal{P}(1, 7.5)$, (e) $\mathcal{P}(1, 10)$.

The convergence rate of the bias between the **MoM** estimator applied to the standardized Pareto distribution tail above the empirical α-quantile and the true $\mathbf{ES}_\alpha$, seems to be the same as the convergence rate in the idealized case, and thus the same as the convergence rate of the bias between the **MoM** estimator applied to the entire standardized Pareto distribution and the expectation. This suggests that it would be sufficient to study the bias between the **MoM** estimator applied to the entire standardized Pareto distribution and the expectation. Deriving a closed-form formula for this bias is still a work in progress.

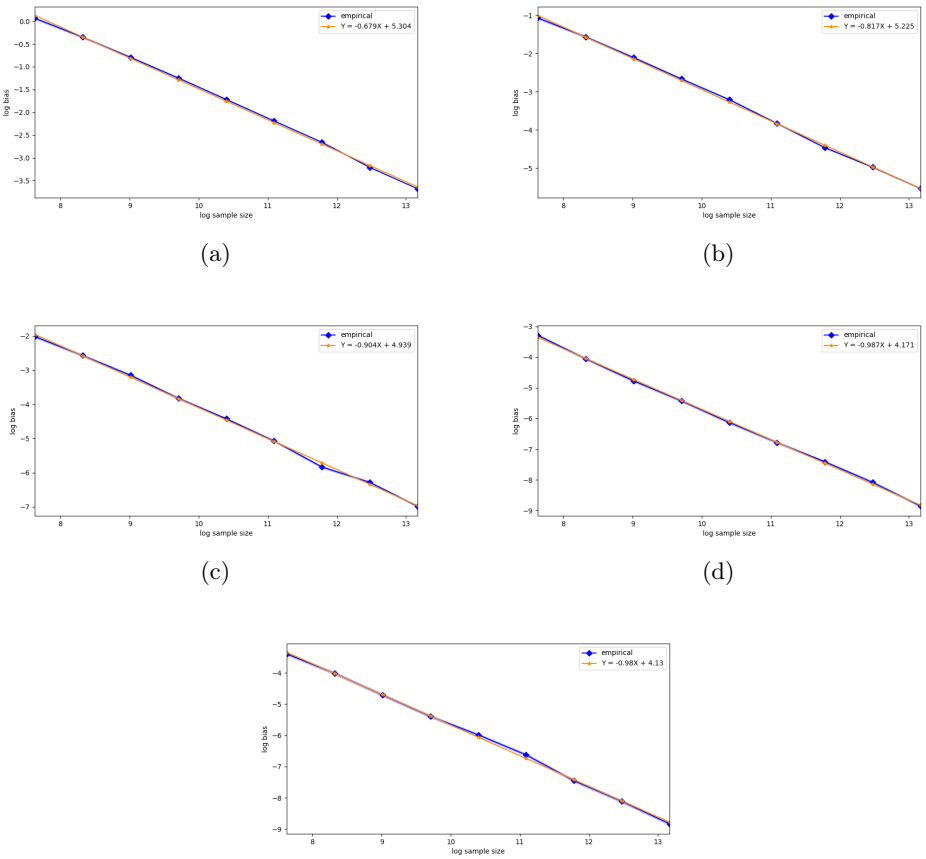


Figure 4.17: Log bias between the **MoM** estimator applied to the standardized Pareto distribution tail above the empirical α-quantile and the $\mathbf{ES}_\alpha$, for five different shape parameters versus the log sample size, with 95% CI, over 10^5 repetitions, (a) $\mathcal{P}(1, 2.5)$, (b) $\mathcal{P}(1, 3.5)$, (c) $\mathcal{P}(1, 5)$, (d) $\mathcal{P}(1, 7.5)$, (e) $\mathcal{P}(1, 10)$

4.5 Trimmed-Mean (TM) Estimator for heavy-tailed distribution

4.5.1 Presentation of the estimator

The Trimmed-Mean estimator is based on the following principle. The most natural attempt to improve the performance of the empirical mean is to remove possible outliers by truncating the sample. The Trimmed-Mean estimator is defined by removing a fraction of the sample, representing the ϵn largest and smallest points for some parameter $\epsilon \in (0, 1)$, and then averaging over the rest.

The estimator splits data into two equal parts. One half is used to determine the correct truncation level. The points from the other half are averaged, except for the data points that fall outside the truncation region, which are thresholded. For convenience of notation, we assume that the data are composed of $2n$ independent copies

of random variable $X\mathcal{P}(1,\gamma)$, denoted by $X_1,\dots,X_n,Y_1,\dots,Y_n$. We respectively denote by $X_1^\star,\dots,X_n^\star$ and $Y_1^\star,\dots,Y_n^\star$ the order statistics sequences related to samples $X_1,\dots,X_n =: \mathbf{X}_n$ and $Y_1,\dots,Y_n =: \mathbf{Y}_n$.

Define the truncation function:

$$\phi_{\mathbf{Y}_n}(x) = \begin{cases} Y^\star_{\lceil(1-\epsilon)n\rceil} & \text{if } x > Y^\star_{\lceil(1-\epsilon)n\rceil}, \\ x & \text{if } x \in \left[Y^\star_{\lceil\epsilon n\rceil}, Y^\star_{\lceil(1-\epsilon)n\rceil}\right], \\ Y^\star_{\lceil\epsilon n\rceil} & \text{if } x < Y^\star_{\lceil\epsilon n\rceil}. \end{cases} \tag{4.174}$$

With this notation in place, the definition of the estimator is as follows:

(i) Given a confidence level $\delta \geq 8e^{-3n/16}$, we set:

$$\epsilon_n = \frac{16\log(8/\delta)}{3n}. \tag{4.175}$$

(ii) Set:

$$\widehat{TM}_{2n} = \frac{1}{n}\sum_{i=1}^{n} \phi_{\mathbf{Y}_n}(X_i). \tag{4.176}$$

The drawback of this estimator is that, for a high confidence threshold, the amount of data required to obtain a reasonable truncation threshold is very large.

4.5.2 Concentration inequality

From (Lugosi and Mendelson, 2019, p.14), a concentration inequality of the Trimmed-Mean estimator is provided to quantify the performance of the estimator.

Theorem 4.30. *Let $X_1,\dots,X_n,Y_1,\dots,Y_n$, be independent and identically distributed (i.i.d.) random variables with mean μ and variance σ^2. Let $\delta \in (0,1)$ be such that $n > \frac{16\log(8/\delta)}{3}$. Then, with a probability of at least $1-\delta$:*

$$\left|\widehat{TM}_{2n} - \mu\right| \leq 9\sigma\sqrt{\frac{\log(8/\delta)}{n}}. \tag{4.177}$$

Recall that the goal of the survey is to study the bias between the empirical $\mathbf{ES}_\alpha$, when the latter is estimated by the Trimmed-Mean estimator applied to the standardized Pareto distribution tail, and the true $\mathbf{ES}_\alpha$ in two frameworks: the *idealized case* where the empirical α-quantile matches the true $\mathbf{VaR}_\alpha$ and the *realistic case* where the empirical α-quantile does not match the true $\mathbf{VaR}_\alpha$. To this end, we first study the bias between the Trimmed-Mean estimator applied to the entire standardized Pareto distribution $\mathcal{P}(1,\gamma)$ and the expectation.

4.5.3 Bias of the Trimmed-Mean applied to the entire standardized Pareto distribution and the expectation

Let $X_1, \dots, X_n, Y_1, \dots, Y_n$ be $2n$ independent copies of the random variable X that follows a standardized Pareto distribution $\mathcal{P}(1, \gamma)$. The expectation of the distribution is denoted as $\mu = \mathbb{E}[X] = \frac{\gamma}{\gamma - 1}$. Truncation on both sides of the distribution is relevant when both tails of the distribution are unbounded and contain extreme values. However, the Pareto distribution presents the particularity that its left tail is bounded, whereas its right tail is not bounded. This implies that to eliminate the outliers, it is not required to threshold the left tail of the distribution, but only the right tail.

In this specific case, the truncation function is given by:

$$\phi_{\mathbf{Y}_n}(x) = \begin{cases} Y^\star_{\lceil (1-\epsilon)n \rceil} & \text{if } x > Y^\star_{\lceil (1-\epsilon)n \rceil}, \\ x & \text{if } x \in \left[1, Y^\star_{\lceil (1-\epsilon)n \rceil}\right]. \end{cases} \tag{4.178}$$

We then define the Trimmed-Mean estimator as in Equation (4.176) with the same parameter ϵ_n as given in Equation (4.175).

Moreover, it is important to note that the truncation thresholds are fixed on a sample that is independent of the sample on which the estimator is built.

The bias between the Trimmed-Mean estimator and the expectation of the standardized Pareto distribution is defined by:

$$B_\mu(\widehat{TM}_{2n}) = \mathbb{E}\left[\widehat{TM}_{2n}\right] - \mu. \tag{4.179}$$

Proposition 4.31 (Bias of the Trimmed-Mean estimator in the standardized Pareto distribution)**.** *Let $X_1, \dots, X_n, Y_1, \dots, Y_n$ and $X_1^\star, \dots, X_n^\star, Y_1^\star, \dots, Y_n^\star$, be respectively $2n$ independent copies of the random variable X that follows a standardized Pareto distribution $\mathcal{P}(1, \gamma)$, and the related order statistics. Let $\epsilon_n \in (0, 1)$ be the truncation threshold that satisfies Equation (4.175). The bias between the Trimmed-Mean estimator applied to the entire standardized Pareto distribution $\mathcal{P}(1, \gamma)$ and the expectation is then given as follows:*

$$B_\mu(\widehat{TM}_{2n}) = -\frac{1}{\gamma - 1}\mathbb{E}\left[\left(Y^\star_{\lceil (1-\epsilon)n \rceil}\right)^{1-\gamma}\right] = -\frac{1}{\gamma - 1}\kappa_{\epsilon,n}\mathfrak{B}\left(n - \left\lceil (1-\epsilon)n \right\rceil + 2 - \frac{1}{\gamma}, \left\lceil (1-\epsilon)n \right\rceil\right). \tag{4.180}$$

where $\kappa_{\epsilon,n} = \frac{n!}{\left(\left\lceil (1-\epsilon)n \right\rceil - 1\right)!\left(n - \left\lceil (1-\epsilon)n \right\rceil\right)!}$ *and* $\mathfrak{B}(x, y) = \int_0^1 u^{x-1}(1-u)^{y-1}\mathrm{d}u = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$.

Proof To compute the expectation of the Trimmed-Mean estimator, we have to consider the conditioning with respect to the sequence of random variables $Y_1, \ldots, Y_n$ on which the truncation thresholds are fixed. The tower property states that:

$$\mathbb{E}\left[\widehat{TM}_{2n}\right] = \mathbb{E}\left[\mathbb{E}\left[\widehat{TM}_{2n}|\mathbf{Y}_n\right]\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\mathbb{E}\left[\phi_{\mathbf{Y}_n}(X_i)|\mathbf{Y}_n\right]\right] \tag{4.181}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\mathbb{E}\left[Y^{\star}_{\lceil(1-\epsilon)n\rceil}\mathbb{1}_{\{X_i>Y^{\star}_{\lceil(1-\epsilon)n\rceil}\}} + X_i\mathbb{1}_{\{1\leq X_i\leq Y^{\star}_{\lceil(1-\epsilon)n\rceil}\}}|\mathbf{Y}_n\right]\right]. \tag{4.182}$$

Since the random draws X_i are i.i.d., we get:

$$\mathbb{E}\left[\widehat{TM}_{2n}\right] = \mathbb{E}\left[\mathbb{E}\left[Y^{\star}_{\lceil(1-\epsilon)n\rceil}\mathbb{1}_{\{X_1>Y^{\star}_{\lceil(1-\epsilon)n\rceil}\}} + X_1\mathbb{1}_{\{1\leq X_1\leq Y^{\star}_{\lceil(1-\epsilon)n\rceil}\}}|\mathbf{Y}_n\right]\right]. \tag{4.183}$$

First, we focus on the internal expectation. The conditioning allows fixing, in a first time, the threshold $Y^{\star}_{\lceil(1-\epsilon)n\rceil}$ that is a random variable which is $\mathbf{Y}_n$-measurable. Under conditioning, we are then able to compute the expectation with respect to $X_1 \sim \mathcal{P}(1,\gamma)$: the first term writes

$$\mathbb{E}\left[Y^{\star}_{\lceil(1-\epsilon)n\rceil}\mathbb{1}_{\{X_1>Y^{\star}_{\lceil(1-\epsilon)n\rceil}\}}|\mathbf{Y}_n\right] = Y^{\star}_{\lceil(1-\epsilon)n\rceil}\int_{Y^{\star}_{\lceil(1-\epsilon)n\rceil}}^{+\infty} f_{X_1}(x)\mathrm{d}x = \left(Y^{\star}_{\lceil(1-\epsilon)n\rceil}\right)^{1-\gamma}. \tag{4.184}$$

The second term writes:

$$\mathbb{E}\left[X_1\mathbb{1}_{\{1\leq X_1\leq Y^{\star}_{\lceil(1-\epsilon)n\rceil}\}}|\mathbf{Y}_n\right] = \int_1^{Y^{\star}_{\lceil(1-\epsilon)n\rceil}} x\frac{\gamma}{x^{\gamma+1}}\mathrm{d}x = \frac{\gamma}{\gamma-1} - \left(Y^{\star}_{\lceil(1-\epsilon)n\rceil}\right)^{1-\gamma}\frac{\gamma}{\gamma-1}. \tag{4.185}$$

Second, let us take the external expectation with respect to $\mathbf{Y}_n$:

$$\mathbb{E}\left[\widehat{TM}_{2n}\right] = \mathbb{E}\left[\left(Y^{\star}_{\lceil(1-\epsilon)n\rceil}\right)^{1-\gamma} + \frac{\gamma}{\gamma-1} - \left(Y^{\star}_{\lceil(1-\epsilon)n\rceil}\right)^{1-\gamma}\frac{\gamma}{\gamma-1}\right] \tag{4.186}$$

$$= \mu - \frac{1}{\gamma-1}\mathbb{E}\left[\left(Y^{\star}_{\lceil(1-\epsilon)n\rceil}\right)^{1-\gamma}\right]. \tag{4.187}$$

We have proved that the bias of the Trimmed-Mean estimator, given by Equation (4.185), is equal to:

$$B_{\mu}(\widehat{TM}_{2n}) = -\frac{1}{\gamma-1}\mathbb{E}\left[\left(Y^{\star}_{\lceil(1-\epsilon)n\rceil}\right)^{1-\gamma}\right]. \tag{4.188}$$

This bias is negative, which is consistent with the fact that the right tails are truncated.

From Equation (4.152), and taking $\kappa_{\epsilon,n} = \frac{n!}{(\lceil (1-\epsilon)n \rceil - 1)!(n - \lceil (1-\epsilon)n \rceil)!}$, the density of $Y^{\star}_{\lceil (1-\epsilon)n \rceil}$ is given by:

$$f_{Y^{\star}_{\lceil (1-\epsilon)n \rceil}}(x) = \kappa_{\epsilon,n} \left(\frac{\gamma}{x^{\gamma+1}} \right) \left(1 - \left(\frac{1}{x} \right)^{\gamma} \right)^{\lceil (1-\epsilon)n \rceil - 1} \left(\left(\frac{1}{x} \right)^{\gamma} \right)^{n - \lceil (1-\epsilon)n \rceil} \mathbb{1}_{\{x \geq 1\}}. \tag{4.189}$$

Recall that ε depends on n in the definition of the Trimmed-Mean estimator; however, for the sake of simplicity, we simply write ε instead of ε_n. Then

$$\mathbb{E}\left[\left(Y^{\star}_{\lceil (1-\epsilon)n \rceil} \right)^{1-\gamma} \right] = \int_1^{+\infty} x^{1-\gamma} f_{Y^{\star}_{\lceil (1-\epsilon)n \rceil}}(x) \mathrm{d}x \tag{4.190}$$

$$= \kappa_{\epsilon,n} \int_1^{+\infty} x^{1-\gamma} \left(\frac{\gamma}{x^{\gamma+1}} \right) \left(1 - \left(\frac{1}{x} \right)^{\gamma} \right)^{\lceil (1-\epsilon)n \rceil - 1} \left(\left(\frac{1}{x} \right)^{\gamma} \right)^{n - \lceil (1-\epsilon)n \rceil} \mathrm{d}x. \tag{4.191}$$

Using the following substitution $y = x^{-\gamma}$, $\mathrm{d}y = -\gamma x^{-\gamma-1}\mathrm{d}x$, we get:

$$\mathbb{E}\left[\left(Y^{\star}_{\lceil (1-\epsilon)n \rceil} \right)^{1-\gamma} \right] = \kappa_{\epsilon,n} \int_0^1 y^{n - \lceil (1-\epsilon)n \rceil + 1 - \frac{1}{\gamma}} (1-y)^{\lceil (1-\epsilon)n \rceil - 1} \mathrm{d}y \tag{4.192}$$

$$= \kappa_{\epsilon,n} \mathfrak{B}\left(n - \left\lceil (1-\epsilon)n \right\rceil + 2 - \frac{1}{\gamma}, \left\lceil (1-\epsilon)n \right\rceil \right). \tag{4.193}$$

Consequently,

$$B_{\mu}(\widehat{TM}_{2n}) = -\frac{1}{\gamma - 1} \kappa_{\epsilon,n} \mathfrak{B}\left(n - \left\lceil (1-\epsilon)n \right\rceil + 2 - \frac{1}{\gamma}, \left\lceil (1-\epsilon)n \right\rceil \right). \tag{4.194}$$

■

Theorem 4.32 (Asymptotic bias of the Trimmed-Mean estimator in the standardized Pareto distribution)**.** *As $n \to +\infty$,*

$$B_{\mu}(\widehat{TM}_{2n}) = -\frac{1}{\gamma - 1} \frac{\Gamma\left(\lfloor C_{\delta} \rfloor + 2 - \frac{1}{\gamma} \right)}{\Gamma(\lfloor C_{\delta} \rfloor + 1)} n^{\frac{1}{\gamma} - 1} + o\left(n^{\frac{1}{\gamma} - 1} \right). \tag{4.195}$$

where for all $x \in \mathbb{R}$, $\Gamma(x) = \int_0^{+\infty} u^{x-1} e^{-u} \mathrm{d}u$ is the Gamma function.

Proof First, let us simplify the expression $\left\lceil (1-\epsilon)n \right\rceil$.

Lemma 4.33 (Integer part)**.** *Let $x \in \mathbb{R}_+$. Let $\lfloor x \rfloor$ and $\lceil x \rceil$ be respectively the lower and upper integer parts of x, defined as the largest integer below x and smallest integer above x.*

Subsequently, for any $n \in \mathbb{N}$ and $c \in [0, +\infty)$ the upper integer part of $n - c$ satisfies the following rule:

$$\lceil n - c \rceil = n - \lfloor c \rfloor. \tag{4.196}$$

The proof is easy and left to the reader.

Recall that ϵ is defined by $\epsilon = \frac{16\log(8/\delta)}{3n}$. At fixed δ, ϵ can be rewritten as:

$$\epsilon = \frac{C_\delta}{n} \quad \text{with} \quad C_\delta = \frac{16\log(8/\delta)}{3}. \tag{4.197}$$

For $\delta = 0.05$, $C_\delta \simeq 27.068$. Then

$$\left\lceil (1-\epsilon)n \right\rceil = \lceil n - \epsilon n \rceil = \left\lceil n - C_\delta \right\rceil = n - \left\lfloor C_\delta \right\rfloor. \tag{4.198}$$

Then the bias can be rewritten as follows:

$$B_\mu(\widehat{TM}_{2n}) = -\frac{1}{\gamma - 1}\kappa_{\epsilon,n} B\left(\left\lfloor C_\delta \right\rfloor + 2 - \frac{1}{\gamma}, n - \left\lfloor C_\delta \right\rfloor\right) \quad \text{with} \quad \kappa_{\epsilon,n} = \frac{n!}{(n - \left\lfloor C_\delta \right\rfloor - 1)!(\left\lfloor C_\delta \right\rfloor)!}. \tag{4.199}$$

For any positive integer m, recall that $\Gamma(m+1) = m!$. The bias formula can then be simplified as follows:

$$B_\mu(\widehat{TM}_{2n}) = -\frac{1}{\gamma - 1}\frac{n!}{(n - \left\lfloor C_\delta \right\rfloor - 1)!(\left\lfloor C_\delta \right\rfloor)!}\frac{\Gamma\left(\left\lfloor C_\delta \right\rfloor + 2 - \frac{1}{\gamma}\right)\Gamma(n - \left\lfloor C_\delta \right\rfloor)}{\Gamma\left(n + 2 - \frac{1}{\gamma}\right)} \tag{4.200}$$

$$= -\frac{1}{\gamma - 1}\frac{\Gamma\left(\left\lfloor C_\delta \right\rfloor + 2 - \frac{1}{\gamma}\right)}{\Gamma(\left\lfloor C_\delta \right\rfloor + 1)}\frac{\Gamma(n+1)}{\Gamma\left(n + 2 - \frac{1}{\gamma}\right)}. \tag{4.201}$$

Similarly to the proof of Theorem **??**, we use a Taylor expansion of the Gamma function to estimate the bias of the Trimmed-Mean estimator, when n goes to infinity.

From Stirling's expansion for the Gamma function given in Equation (4.91), we can write the following equalities. Let us write the expansion of each factor in Equation (4.201):

$$\Gamma(n+1) = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \left(1 + \frac{1}{12n} + o\left(\frac{1}{n}\right)\right), \tag{4.202}$$

$$\Gamma\left(n + 2 - \frac{1}{\gamma}\right) = \sqrt{2\pi\left(n + 1 - \frac{1}{\gamma}\right)}\left(\frac{n + 1 - \frac{1}{\gamma}}{e}\right)^{n+1-\frac{1}{\gamma}} \left(1 + \frac{1}{12\left(n + 1 - \frac{1}{\gamma}\right)} + o\left(\frac{1}{n}\right)\right). \tag{4.203}$$

The ratio between the two Gamma functions can be written under following form:

$$\frac{\Gamma(n+1)}{\Gamma\left(n + 2 - \frac{1}{\gamma}\right)} = A \times B \times C, \tag{4.204}$$

where:

$$A = \frac{\sqrt{2\pi n}}{\sqrt{2\pi\left(n+1-\frac{1}{\gamma}\right)}} = \frac{\sqrt{n}}{\sqrt{n+1-\frac{1}{\gamma}}}, \tag{4.205}$$

$$B = \left(\frac{n}{e}\right)^n \left(\frac{e}{n+1-\frac{1}{\gamma}}\right)^{n+1-\frac{1}{\gamma}}, \tag{4.206}$$

$$C = \frac{\left(1+\frac{1}{12n}+o\left(\frac{1}{n}\right)\right)}{\left(1+\frac{1}{12\left(n+1-\frac{1}{\gamma}\right)}+o\left(\frac{1}{n}\right)\right)}. \tag{4.207}$$

Let us simplify each of the above expressions. Regarding A, we have

$$A = \sqrt{\frac{1}{1+\frac{1}{n}\left(1-\frac{1}{\gamma}\right)}} \tag{4.208}$$

$$= \sqrt{1-\frac{1}{n}\left(1-\frac{1}{\gamma}\right)+o\left(\frac{1}{n}\right)} \tag{4.209}$$

$$= 1-\frac{1}{2n}\left(1-\frac{1}{\gamma}\right)+o\left(\frac{1}{n}\right). \tag{4.210}$$

Notice that B can be rewritten as follows:

$$B = \left(\frac{n}{e}\right)^n \left(\frac{e}{n+1-\frac{1}{\gamma}}\right)^{n+1-\frac{1}{\gamma}} \tag{4.211}$$

$$= e^{1-\frac{1}{\gamma}} \left(\frac{n+1-\frac{1}{\gamma}}{n}\right)^{-n} \left(n+1-\frac{1}{\gamma}\right)^{\frac{1}{\gamma}-1} \tag{4.212}$$

$$= e^{1-\frac{1}{\gamma}-n\log\left(\frac{n+1-\frac{1}{\gamma}}{n}\right)+\left(\frac{1}{\gamma}-1\right)\log\left(n+1-\frac{1}{\gamma}\right)} \tag{4.213}$$

$$= e^{1-\frac{1}{\gamma}-n\log\left(1+\frac{1}{n}\left(1-\frac{1}{\gamma}\right)\right)+\left(\frac{1}{\gamma}-1\right)\log(n)+\left(\frac{1}{\gamma}-1\right)\log\left(1+\frac{1}{n}\left(1-\frac{1}{\gamma}\right)\right)}. \tag{4.214}$$

Providing a Taylor expansion at the first and second order of $\log\left(1+\frac{1}{n}\left(1-\frac{1}{\gamma}\right)\right)$ we get:

$$B = e^{1-\frac{1}{\gamma}-\left(1-\frac{1}{\gamma}\right)+\frac{1}{2n}\left(1-\frac{1}{\gamma}\right)^2+\left(\frac{1}{\gamma}-1\right)\log(n)-\frac{1}{n}\left(1-\frac{1}{\gamma}\right)^2+o\left(\frac{1}{n}\right)} \tag{4.215}$$

$$= n^{\frac{1}{\gamma}-1} e^{-\frac{1}{2n}\left(1-\frac{1}{\gamma}\right)^2+o\left(\frac{1}{n}\right)} \tag{4.216}$$

$$= n^{\frac{1}{\gamma}-1}\left(1-\frac{1}{2n}\left(1-\frac{1}{\gamma}\right)^2+o\left(\frac{1}{n}\right)\right). \tag{4.217}$$

Finally, let us provide a Taylor expansion of C.

$$C = \frac{\left(1 + \frac{1}{12n} + o\left(\frac{1}{n}\right)\right)}{\left(1 + \frac{1}{12\left(n+1-\frac{1}{\gamma}\right)} + o\left(\frac{1}{n}\right)\right)} \tag{4.218}$$

$$= \left(1 + \frac{1}{12n} + o\left(\frac{1}{n}\right)\right)\left(1 - \frac{1}{12\left(n+1-\frac{1}{\gamma}\right)} + o\left(\frac{1}{n}\right)\right) \tag{4.219}$$

$$= \left(1 + \frac{1}{12n} + o\left(\frac{1}{n}\right)\right)\left(1 - \frac{1}{12n} + o\left(\frac{1}{n}\right)\right) \tag{4.220}$$

$$= 1 + \frac{1}{12n} - \frac{1}{12n} + o\left(\frac{1}{n}\right) = 1 + o\left(\frac{1}{n}\right). \tag{4.221}$$

Gathering the expressions of A, B, C we obtain:

$$\frac{\Gamma(n+1)}{\Gamma\left(n+2-\frac{1}{\gamma}\right)} = A \times B \times C \tag{4.222}$$

$$= \left(1 - \frac{1}{2n}\left(1 - \frac{1}{\gamma}\right) + o\left(\frac{1}{n}\right)\right) n^{\frac{1}{\gamma}-1} \left(1 - \frac{1}{2n}\left(1 - \frac{1}{\gamma}\right)^2 + o\left(\frac{1}{n}\right)\right)\left(1 + o\left(\frac{1}{n}\right)\right) \tag{4.223}$$

$$= n^{\frac{1}{\gamma}-1}\left(1 - \frac{1}{2n}\left(2 - \frac{1}{\gamma}\right)\left(1 - \frac{1}{\gamma}\right) + o\left(\frac{1}{n}\right)\right). \tag{4.224}$$

But we want at least that the first moment of the Pareto distribution might be finite. Therefore, we take $\gamma > 1$, what implies that $\frac{1}{\gamma} - 1 < 0$. Then we get:

$$\frac{\Gamma(n+1)}{\Gamma\left(n+2-\frac{1}{\gamma}\right)} = n^{\frac{1}{\gamma}-1} + o\left(n^{\frac{1}{\gamma}-1}\right) \tag{4.225}$$

and finally, we obtain,

$$B_\mu(\widehat{TM}_{2n}) = -\frac{1}{\gamma - 1}\frac{\Gamma\left(\lfloor C_\delta \rfloor + 2 - \frac{1}{\gamma}\right)}{\Gamma(\lfloor C_\delta \rfloor + 1)} n^{\frac{1}{\gamma}-1} + o\left(n^{\frac{1}{\gamma}-1}\right). \tag{4.226}$$

■

A closed-form formula for the bias between the $\widehat{\mathbf{RTM}}_n$ estimator applied to the entire standardized Pareto distribution $\mathcal{P}(1,\gamma)$ and the expectation has been provided. Before providing results about the bias between the $\widehat{\mathbf{RTM}}_n$ estimator applied to the standardized Pareto distribution tail and the $\mathbf{ES}_\alpha$, we provide some numerical experiments to support the theory.

We are interested in studying the evolution of the bias between the Right-Trimmed-Mean (RTM) estimator applied to the entire standardized Pareto distribution $\mathcal{P}(1,\gamma)$ and the expectation of the distribution. The graph representing the evolution of this bias, for several Pareto indices, in function of the log-sample size is provided in Figure 4.18.

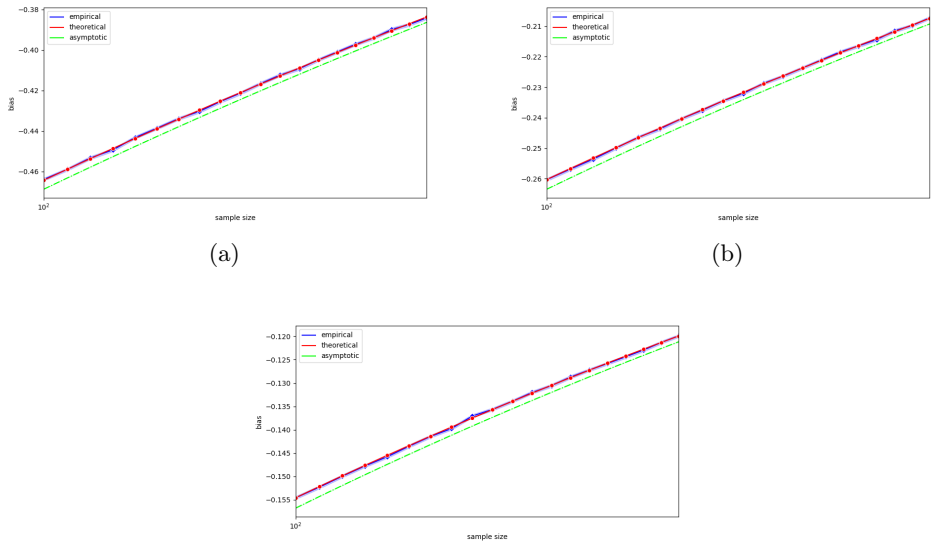
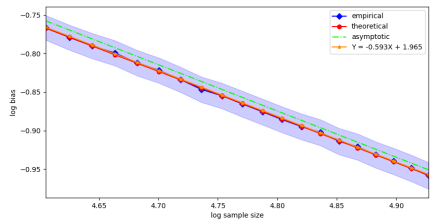


Figure 4.18: Evolution of the bias between the RTM estimator applied to the entire standardized Pareto distribution $\mathcal{P}(1,\gamma)$ and its expectation for three different shape parameters versus the lo sample size, with 95% CI, over 10^4 repetitions, (a) $\mathcal{P}(1,2.5)$, (b) $\mathcal{P}(1,3.5)$, (c) $\mathcal{P}(1,5)$.

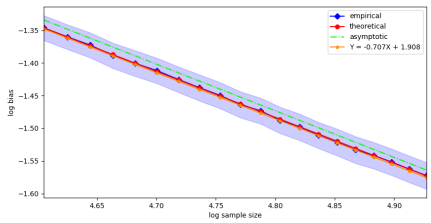
As previously mentioned, we are also interested in determining the convergence speed of the bias between the Right-Trimmed-Mean (RTM) estimator applied to the entire standardized Pareto distribution $\mathcal{P}(1,\gamma)$ and the expectation. Because the bias is a power function of the sample size, the logarithm of the bias is a linear function of the logarithm of the sample size with the convergence speed as the slope. For this reason, we display the log-log plot of the bias versus the sample size. See Figure 4.19.

For three standardized Pareto distributions $\mathcal{P}(1,2.5)$, $\mathcal{P}(1,3.5)$, $\mathcal{P}(1,5)$, we obtain convergence speeds equal to $\frac{1}{n^{0.593}}$, $\frac{1}{n^{0.707}}$ and $\frac{1}{n^{0.789}}$ respectively. Therefore, the convergence speed varies as a function of the Pareto index γ and satisfies the rule $n^{\frac{1}{\gamma}-1}$, multiplied by the proper constant. Consequently, the numerical study confirms the theoretical results.

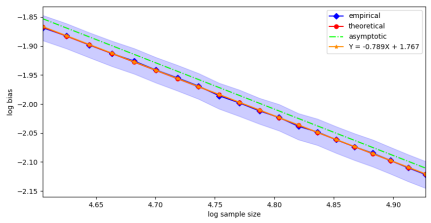
Moreover, these results can be extended to any non-standardized Pareto distribution $\mathcal{P}(x_m,\gamma)$, with $x_m > 0$. Indeed, the scaling property of the Pareto distribution states that any non-standardized Pareto distribution $\mathcal{P}(x_m,\gamma)$ is proportional to the standardized Pareto distribution $\mathcal{P}(1,\gamma)$ with a proportionality factor equal to the new scaling parameter x_m. This implies that the $\widehat{\mathbf{RTM}}_n$ estimator applied to the non-standardized Pareto distribution $\mathcal{P}(x_m,\gamma)$ is proportional to the $\widehat{\mathbf{RTM}}_n$ estimator applied to the standardized Pareto distribution $\mathcal{P}(1,\gamma)$ with a proportionality factor equal to x_m.



(a)



(b)



(c)

Figure 4.19: Logarithm of the bias between the $\widehat{\mathbf{RTM}}_n$ estimator applied to the standardized Pareto distribution $\mathcal{P}(1,\gamma)$ and its expectation, for three different shape parameters, versus the log sample size, with 95% CI, over 10^4 repetitions, (a) $\mathcal{P}(1,2.5)$, (b) $\mathcal{P}(1,3.5)$, (c) $\mathcal{P}(1,5)$.

Similarly, the expectation of the non-standardized Pareto distribution $\mathcal{P}(x_m,\gamma)$ is proportional to the expectation of the standardized Pareto distribution $\mathcal{P}(1,\gamma)$ with a proportionality factor equal to x_m. Consequently, the bias between the $\widehat{\mathbf{RTM}}_n$ estimator applied to the non-standardized Pareto distribution $\mathcal{P}(x_m,\gamma)$ and the expectation is proportional to the bias between the $\widehat{\mathbf{RTM}}_n$ estimator applied to the standardized Pareto distribution $\mathcal{P}(1,\gamma)$ and the expectation, with a proportionality factor equal to x_m. Because x_m is independent of the sample, the convergence speed of the bias between the $\widehat{\mathbf{RTM}}_n$ estimator and the expectation of the non-standardized Pareto distribution $\mathcal{P}(x_m,\gamma)$ is the same as the convergence speed of the bias between the $\widehat{\mathbf{RTM}}_n$ estimator and the expectation of the standardized Pareto distribution $\mathcal{P}(1,\gamma)$.

Now that we have studied the evolution of the bias between the $\widehat{\mathbf{RTM}}_n$ estimator and the expectation of the entire standardized Pareto distribution $\mathcal{P}(1,\gamma)$, as well as its convergence speed, we are interested in studying the evolution of the bias between the $\widehat{\mathbf{RTM}}_n$ estimator applied to the standardized Pareto distribution tail and the $\mathbf{ES}_\alpha$ in two cases: an *idealized case* in which the conditioning threshold corresponds to the theoretical $\mathbf{VaR}_\alpha$, and a *realistic case* in which the conditioning threshold corresponds to the empirical α-quantile.

4.5.4 Bias between the Trimmed-Mean estimator applied to the standardized Pareto distribution tail above the true $\mathbf{VaR}_\alpha$ and the true ES (Idealized case)

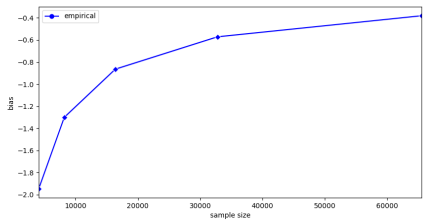
In the idealized case, the empirical α-quantile matches the true $\mathbf{VaR}_\alpha$, which is supposed to be known. In this case, the estimator of the $\mathbf{ES}_\alpha$ in the standardized Pareto distribution $\mathcal{P}(1,\gamma)$ is the empirical average of the standardized Pareto distribution conditional on its values being above the true $\mathbf{VaR}_\alpha$. The conditioning threshold is independent of the underlying sample and the stability by conditioning and the scaling properties of the Pareto distribution, as stated in Theorem 4.7, are valid. The stability by conditioning property implies that the standardized Pareto distribution $\mathcal{P}(1,\gamma)$ conditional on its values being above $\mathbf{VaR}_\alpha$ is still a Pareto distribution, with the same shape parameter γ, but a new scaling parameter equal to the conditioning parameter $\mathbf{VaR}_\alpha$. The scaling property states that this non-standardized Pareto distribution $\mathcal{P}(\mathbf{VaR}_\alpha,\gamma)$ is proportional to the marginal Pareto distribution $\mathcal{P}(1,\gamma)$, with a proportionality factor equal to the conditioning parameter $\mathbf{VaR}_\alpha$. Therefore, the standardized Pareto distribution conditional on its values being above $\mathbf{VaR}_\alpha$ is proportional to the standardized Pareto distribution with a proportionality factor equal to $\mathbf{VaR}_\alpha$. This implies that the $\widehat{\mathbf{RTM}}_n$ estimator applied to the standardized Pareto distribution conditional on its values being above $\mathbf{VaR}_\alpha$, is proportional to the $\widehat{\mathbf{RTM}}_n$ estimator applied to the entire standardized Pareto distribution, with a proportionality factor equal to $\mathbf{VaR}_\alpha$. Similarly, the $\mathbf{ES}_\alpha$ is proportional to the expectation of the standardized Pareto distribution, with a proportionality factor equal to $\mathbf{VaR}_\alpha$. Thus, the bias between the $\widehat{\mathbf{RTM}}_n$ estimator applied to the standardized Pareto distribution conditional on its values being above $\mathbf{VaR}_\alpha$ and the $\mathbf{ES}_\alpha$ is proportional to the bias between the $\widehat{\mathbf{RTM}}_n$ estimator applied to the entire standardized Pareto distribution and the expectation, with a proportionality factor equal to $\mathbf{VaR}_\alpha$. Because the conditioning threshold is independent on the underlying sample, the convergence speed of the bias between the $\widehat{\mathbf{RTM}}_n$ estimator applied to the standardized Pareto distribution conditional on its values being above $\mathbf{VaR}_\alpha$ and the true $\mathbf{ES}_\alpha$ is the same as that of the bias between the $\widehat{\mathbf{RTM}}_n$ estimator applied to the entire standardized Pareto distribution and the expectation.

$$B_{\mathbf{ES}}(\widehat{\mathbf{RTM}}_n) = \mathbf{VaR}_\alpha \times B_\mu(\widehat{\mathbf{RTM}}_n) \tag{4.227}$$

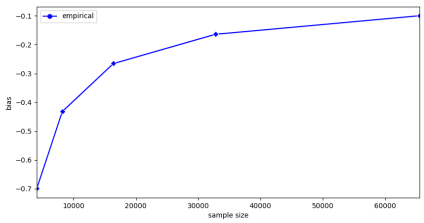
$$= \frac{1}{1-\gamma}(1-\alpha)^{-\frac{1}{\gamma}}\kappa_{\epsilon,n}B\left(n-\lfloor(1-\epsilon)n\rfloor+2-\frac{1}{\gamma},\lfloor(1-\epsilon)n\rfloor\right). \tag{4.228}$$

The graph of the evolution of the bias between the $\widehat{\mathbf{RTM}}_n$ estimator applied to the standardized Pareto distribution tail above the true $\mathbf{VaR}_\alpha$ and the true $\mathbf{ES}_\alpha$ is provided in Figure 4.20.

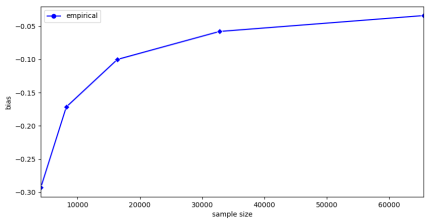
We are interested in the convergence speed of the bias between the Right-Trimmed-Mean estimator applied to the standardized Pareto distribution tail above the true $\mathbf{VaR}_\alpha$ and the $\mathbf{ES}_\alpha$. Because the bias is a power function of the sample size, then the logarithm of the bias is a linear function of the logarithm of the sample size with the



(a)



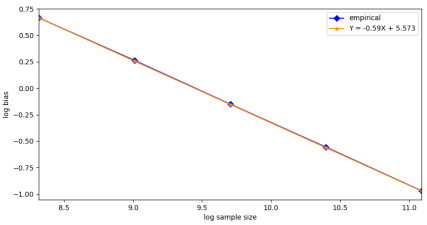
(b)



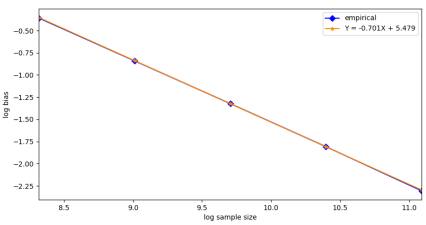
(c)

Figure 4.20: Evolution of the bias between the $\widehat{\mathbf{RTM}}_n$ estimator applied to the standardized Pareto distribution tail above the true $\mathbf{VaR}_\alpha$ and the $\mathbf{ES}_\alpha$, for three different shape parameters, versus the sample size, with 95% CI, over 10^4 repetitions, (a) $\mathcal{P}(1, 2.5)$, (b) $\mathcal{P}(1, 3.5)$, (c) $\mathcal{P}(1, 5)$.

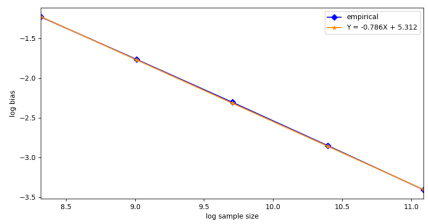
convergence speed as the slope. Therefore, we display the log-log plot of the bias versus the sample size. See Figure 4.21.



(a)



(b)



(c)

Figure 4.21: Evolution of the logarithm of the bias between the $\widehat{\mathbf{RTM}}_n$ estimator applied to the Pareto distribution tail above the true $\mathbf{VaR}_\alpha$ and the $\mathbf{ES}_\alpha$, for three different shape parameters, versus the log-sample size, with 95% CI, over 10^4 repetitions, (a) $\mathcal{P}(1, 2.5)$, (b) $\mathcal{P}(1, 3.5)$, (c) $\mathcal{P}(1, 5)$.

Because $\mathbf{VaR}_\alpha$ is supposed to be known and does not depend on the underlying sample, the convergence speed of the bias between the $\widehat{\mathbf{RTM}}_n$ estimator applied to the standardized Pareto distribution tail above $\mathbf{VaR}_\alpha$ and the true $\mathbf{ES}_\alpha$ is the same as the convergence speed of the bias between the $\widehat{\mathbf{RTM}}_n$ estimator applied to the entire Pareto distribution and the expectation, that is, of order $n^{\frac{1}{\gamma}-1}$.

4.5.5 Bias between the Trimmed-Mean estimator applied to the standardized Pareto distribution tail above the empirical α-quantile and the true ES (Realistic case)

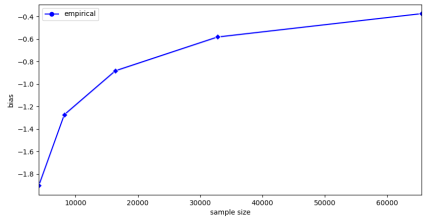
In the realistic case, the estimator of the $\mathbf{ES}_\alpha$ in the standardized Pareto distribution $\mathcal{P}(1,\gamma)$ is the empirical average of the standardized Pareto distribution conditional on its values being above the empirical α-quantile. In this case, the conditioning threshold is an order statistic and depends on the underlying sample. This implies that the samples larger than the empirical α-quantile are no longer independent and identically distributed (i.i.d.). The stability by conditioning and scaling properties, as stated in Theorem 4.7, are no longer valid. Therefore, the distribution of the samples larger than the empirical α-quantile is not necessarily a Pareto distribution. Consequently, the distribution of the samples larger than the empirical α-quantile is unknown, and it is more difficult to establish an analytic closed-form formula for the bias between the $\widehat{\mathbf{RTM}}_n$ estimator applied to the standardized Pareto distribution tail above the empirical α-quantile and the $\mathbf{ES}_\alpha$. For this reason, we provide some experimental study to give an insight about the convergence speed of the bias between the $\widehat{\mathbf{RTM}}_n$ applied to the Pareto distribution tail above the empirical α-quantile and the true $\mathbf{ES}_\alpha$. See Figure 4.22.

It is expected that the bias between the $\widehat{\mathbf{RTM}}_n$ estimator applied to the standardized Pareto distribution tail above the empirical α-quantile and the $\mathbf{ES}_\alpha$, be a power function of the sample size. Therefore, it is expected that the logarithm of the bias be a linear function of the logarithm of the sample size. For this reason, to have an insight into the convergence speed of the bias between the $\widehat{\mathbf{RTM}}_n$ applied to the standardized Pareto distribution tail above the empirical α-quantile and the $\mathbf{ES}_\alpha$, we study the log-log plot of the bias versus the sample size. See Figure 4.23.

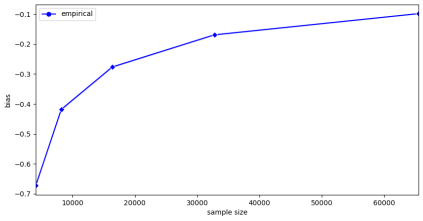
The slopes of the log-log plots are between -0.6 and -0.75, this means that the convergence speed of the bias between the $\widehat{\mathbf{RTM}}_n$ applied to the standardized Pareto distribution tail above the empirical α-quantile and the Expected-Shortfall is larger than $\frac{1}{n^{0.8}}$ and lower than $\frac{1}{n^{0.6}}$. The convergence speed of the bias between the $\widehat{\mathbf{RTM}}_n$ applied to the standardized Pareto distribution tail above the empirical α-quantile and the Expected-Shortfall could match the theoretical convergence speed $n^{\frac{1}{\gamma}-1}$.

4.6 Lee-Valiant (LV) Estimator

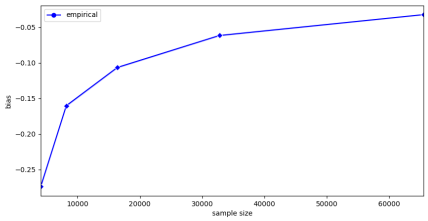
The Lee-Valiant estimator is an improved version of the **MoM** estimator, which includes a correction term.



(a)

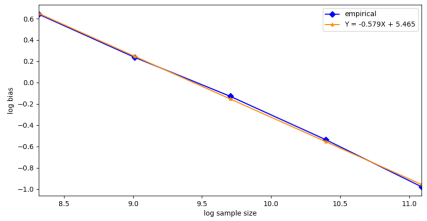


(b)

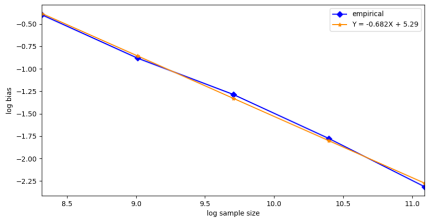


(c)

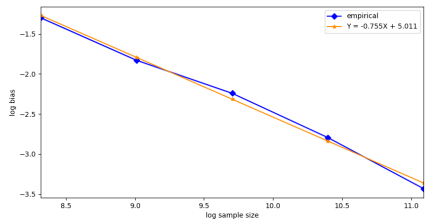
Figure 4.22: Evolution of the bias between the $\widehat{\mathbf{RTM}}_n$ estimator applied to the Pareto distribution tail above the empirical α-quantile and the $\mathbf{ES}_\alpha$, for three different shape parameters, versus the sample size, with 95% CI, over 10^4 repetitions, (a) $\mathcal{P}(1, 2.5)$, (b) $\mathcal{P}(1, 3.5)$, (c) $\mathcal{P}(1, 5)$.



(a)



(b)



(c)

Figure 4.23: Log bias between the $\widehat{\mathbf{RTM}}_n$ estimator applied to the Pareto distribution tail above the empirical α-quantile and the $\mathbf{ES}_\alpha$, for three different shape parameters, versus the log sample size, with 95% CI, over 10^4 repetitions, (a) $\mathcal{P}(1, 2.5)$, (b) $\mathcal{P}(1, 3.5)$, (c) $\mathcal{P}(1, 5)$.

Definition 4.34 (Lee-Valiant estimator)**.** *For a given δ, define the Median-of-Mean estimator $\widehat{\mathbf{MoM}}_n = \mathbf{MoM}(X_1, \dots, X_n)$, computed on $k = log\left(\frac{1}{\delta}\right) \leq n$ blocks with $\delta \geq e^{-n}$ and k an integer. The Lee-Valiant estimator is then defined as*

$$\widehat{LV}_n = \widehat{\mathbf{MoM}}_n + \frac{1}{n}\sum_{i=1}^{n}(X_i - \widehat{\mathbf{MoM}}_n)(1 - \min(\alpha(X_i - \widehat{\mathbf{MoM}}_n)^2, 1)) \tag{4.229}$$

where the parameter α is the solution of the monotonic, piecewise-linear equation

$$\sum_{i=1}^{n}\min(\alpha(X_i - \widehat{\mathbf{MoM}}_n)^2, 1) = \frac{1}{3}\log\left(\frac{1}{\delta}\right). \tag{4.230}$$

The zero of the equation:

$$\sum_{i=1}^{n}\min(\alpha(X_i - \widehat{\mathbf{MoM}}_n)^2, 1) - \frac{1}{3}\log\left(\frac{1}{\delta}\right) = 0 \tag{4.231}$$

can be solved using a dichotomy algorithm.

Concentration inequality

From (Gobet et al., 2022, p.14, Thm. 2.5), a deviation results is given as follows.

Theorem 4.35. *Let $X_1, \dots, X_n$ be independent and identically distributed (i.i.d.) random variables with mean μ and variance σ^2. Let $\delta \in (0, 1)$ be such that $\delta \geq e^{-n}$ and assume that $k = \log\left(\frac{1}{\delta}\right)$ is an integer. Then, Lee-Valiant estimator $\widehat{LV}_n$ satisfies:*

$$\mathbb{P}\left(\left|\widehat{LV}_n - \mu\right| < \sigma(1 + o(1))\sqrt{\frac{2\log(1/\delta)}{n}}\right) \geq 1 - \delta \tag{4.232}$$

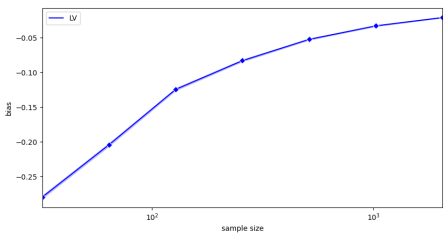
where $o(1)$ term goes to zero when $\left(\delta, \frac{\log(1/\delta)}{n}\right) \to (0, 0)$.

See (Lee and Valiant, 2022) for more details about the Lee-Valiant estimator.

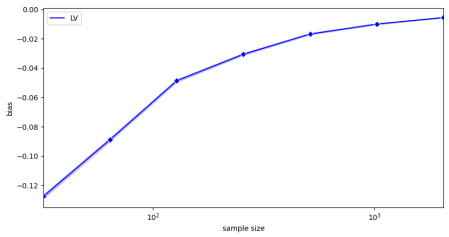
The goal of the survey is to study the bias between the empirical $\mathbf{ES}_\alpha$, when the latter is estimated by the Lee-Valiant estimator applied to the standardized Pareto distribution tail, and the true $\mathbf{ES}_\alpha$ in two frameworks: the *idealized case* where the empirical α-quantile matches the true $\mathbf{VaR}_\alpha$ and the *realistic case* where the empirical α-quantile does not match the true $\mathbf{VaR}_\alpha$. To this end, we first study the bias between the Lee-Valiant estimator applied to the entire standardized Pareto distribution $\mathcal{P}(1, \gamma)$ and the expectation.

Bias between the Lee-Valiant (LV) estimator applied to the entire standardized Pareto distribution and the expectation

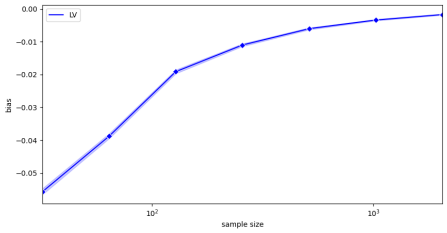
The Lee-Valiant estimator can be seen as the corrected **MoM** estimator. Because the analytic form for the bias of the **MoM** estimator is difficult to determine, the difficulty persists for the bias of the $\widehat{\mathbf{LV}}_n$ estimator. For this reason, we provide experimental results to gain insight into the bias and its convergence speed. See Figure 4.24.



(a)



(b)



(c)

Figure 4.24: Evolution of the bias between the LV estimator applied to the entire standardized Pareto distribution $\mathcal{P}(1,\gamma)$ and its expectation, for three different shape parameters versus the sample size, with 95% CI, over 10^4 repetitions, (a) $\mathcal{P}(1,2.5)$, (b) $\mathcal{P}(1,3.5)$, (c) $\mathcal{P}(1,5)$.

Moreover, we are interested in the convergence speed of the bias between the $\widehat{\mathbf{LV}}_n$ estimator applied to the entire standardized Pareto distribution $\mathcal{P}(1,\gamma)$ and the expectation. The bias seems to move towards a power function of the sample size. The logarithm of a power function is a linear function of the sample size, with the convergence speed of the bias as the slope. Therefore, we display the log-log plot of the bias versus the sample size. See Figure 4.25.

For three standardized Pareto distributions $\mathcal{P}(1,2.5)$, $\mathcal{P}(1,3.5)$, $\mathcal{P}(1,5)$, the slopes of the log-log plots are between -0.9 and -0.65. This implies that the convergence speed of the bias between the $\widehat{\mathbf{LV}}_n$ estimator applied to the entire standardized Pareto distribution $\mathcal{P}(1,\gamma)$ and the expectation varies as a function of the Pareto index γ. For three standardized Pareto distributions $\mathcal{P}(1,2.5)$, $\mathcal{P}(1,3.5)$, and $\mathcal{P}(1,5)$, the convergence speed of the bias is between $\frac{1}{n^{0.9}}$ and $\frac{1}{n^{0.6}}$, that is larger than $\frac{1}{n}$ and lower than $\frac{1}{\sqrt{n}}$. Moreover, these results can be extended to any non-standardized Pareto distribution $\mathcal{P}(x_m,\gamma)$, with $x_m > 0$. Indeed, the scaling property of the Pareto distribution states that any non-standardized Pareto distribution $\mathcal{P}(x_m,\gamma)$ is proportional to the standardized Pareto distribution $\mathcal{P}(1,\gamma)$, with a proportionality factor equal to the new scaling parameter x_m. This implies that the $\widehat{\mathbf{LV}}_n$ estimator applied to the non-standardized Pareto distribution $\mathcal{P}(x_m,\gamma)$ is proportional to the $\widehat{\mathbf{LV}}_n$ estimator applied to the standardized Pareto distribution $\mathcal{P}(1,\gamma)$, with a proportionality factor equal to x_m. Similarly, the expectation of the non-standardized Pareto distribution $\mathcal{P}(x_m,\gamma)$ is proportional to the expectation of the standardized Pareto distribution $\mathcal{P}(1,\gamma)$, with a proportionality factor equal to x_m. Consequently, the bias between the $\widehat{\mathbf{LV}}_n$ estimator applied to the non-standardized Pareto distribution $\mathcal{P}(x_m,\gamma)$ and the expectation is

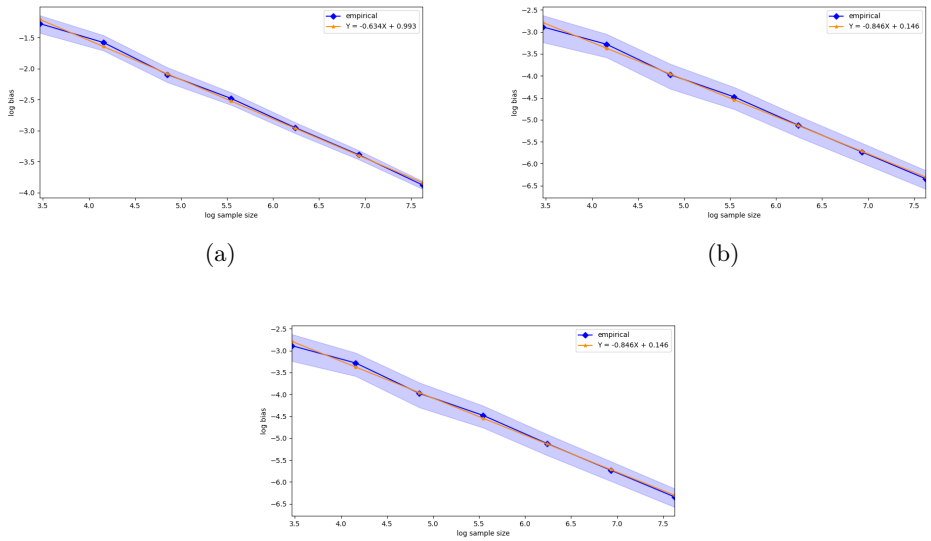


Figure 4.25: Log bias between the $\widehat{\mathbf{LV}}_n$ estimator applied to the entire standardized Pareto distribution $\mathcal{P}(1,\gamma)$ and its expectation, for three different shape parameters, versus the log sample size, with 95% CI, over 10^4 repetitions, (a) $\mathcal{P}(1,2.5)$, (b) $\mathcal{P}(1,3.5)$, (c) $\mathcal{P}(1,5)$.

proportional to the bias between the $\widehat{\mathbf{LV}}_n$ estimator applied to the standardized Pareto distribution $\mathcal{P}(1,\gamma)$ and the expectation, with a proportionality factor equal to x_m. Because x_m is independent of the sample, the convergence speed of the bias between the $\widehat{\mathbf{LV}}_n$ estimator and the expectation of the non-standardized Pareto distribution $\mathcal{P}(x_m,\gamma)$ is the same as the convergence speed of the bias between the $\widehat{\mathbf{LV}}_n$ estimator and the expectation of the standardized Pareto distribution $\mathcal{P}(1,\gamma)$.

Now that we have studied the evolution of the bias between the $\widehat{\mathbf{LV}}_n$ estimator and the expectation of the entire standardized Pareto distribution $\mathcal{P}(1,\gamma)$, as well as its convergence speed, we are interested in studying the evolution of the bias between the $\widehat{\mathbf{LV}}_n$ estimator applied to the standardized Pareto distribution tail and the $\mathbf{ES}_\alpha$ in two cases: an *idealized case* in which the conditioning threshold corresponds to the theoretical $\mathbf{VaR}_\alpha$, and a *realistic case* in which the conditioning threshold corresponds to the empirical α-quantile.

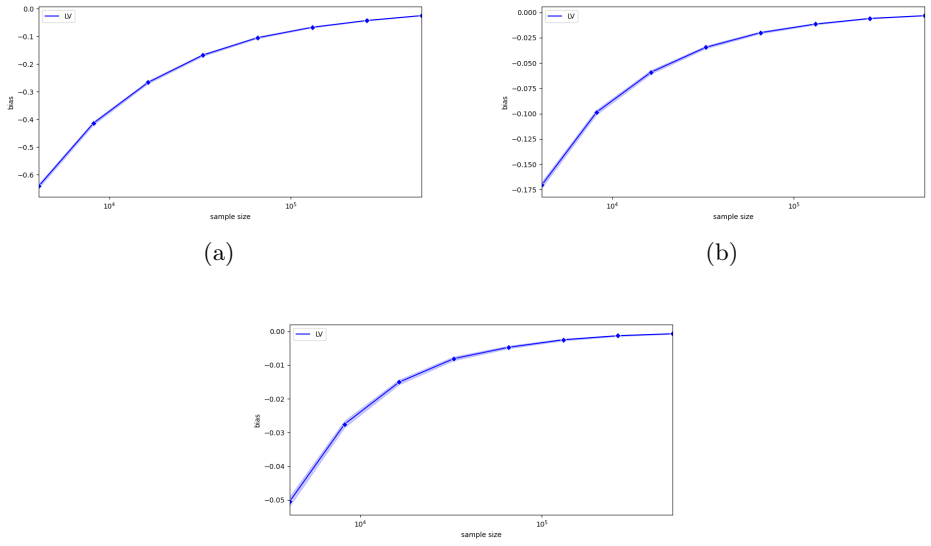
Bias between the Lee-Valiant (LV) estimator applied to the standardized Pareto distribution tail above the true $\mathbf{VaR}_\alpha$ and the $\mathbf{ES}_\alpha$ (Idealized case)

In the idealized case, the empirical α-quantile matches the true $\mathbf{VaR}_\alpha$, which is known. In this case, the estimator of the $\mathbf{ES}_\alpha$ in the standardized Pareto distribution $\mathcal{P}(1,\gamma)$ is the empirical average of the standardized Pareto distribution conditional on its values being above the true $\mathbf{VaR}_\alpha$. The conditioning threshold is independent on the underlying sample. This implies that the samples larger than the true $\mathbf{VaR}_\alpha$ are independent and identically distributed (i.i.d.) and the stability by conditioning and scaling

properties, as stated in Theorem 4.7, are still valid. The stability by conditioning property implies that the standardized Pareto distribution $\mathcal{P}(1,\gamma)$ conditional on its values being above $\mathbf{VaR}_\alpha$ is still a Pareto distribution, with the same shape parameter γ, but a new scaling parameter equal to the conditioning parameter $\mathbf{VaR}_\alpha$. The scaling property states that this non-standardized Pareto distribution $\mathcal{P}(\mathbf{VaR}_\alpha,\gamma)$ is proportional to the marginal Pareto distribution $\mathcal{P}(1,\gamma)$, with a proportionality factor equal to the conditioning parameter $\mathbf{VaR}_\alpha$. Therefore, the standardized Pareto distribution conditional on its values being above $\mathbf{VaR}_\alpha$ is proportional to the standardized Pareto distribution, with a proportionality factor equal to $\mathbf{VaR}_\alpha$. This implies that the $\widehat{\mathbf{LV}}_n$ estimator applied to the standardized Pareto distribution conditional on its values being above $\mathbf{VaR}_\alpha$, is proportional to the $\widehat{\mathbf{LV}}_n$ estimator applied to the entire standardized Pareto distribution, with a proportionality factor equal to $\mathbf{VaR}_\alpha$. Similarly, the $\mathbf{ES}_\alpha$ is proportional to the expectation of the standardized Pareto distribution, with a proportionality factor equal to $\mathbf{VaR}_\alpha$. Thus, the bias between the $\widehat{\mathbf{LV}}_n$ estimator applied to the standardized Pareto distribution conditional on its values being above $\mathbf{VaR}_\alpha$ and the $\mathbf{ES}_\alpha$ is proportional to the bias between the $\widehat{\mathbf{LV}}_n$ estimator applied to the entire standardized Pareto distribution and the expectation, with a proportionality factor equal to $\mathbf{VaR}_\alpha$.

$$B_{\mathbf{ES}}(\widehat{\mathbf{LV}}_n) = \mathbf{VaR}_\alpha \times B_\mu(\widehat{\mathbf{LV}}_n). \tag{4.233}$$

The graph of the evolution of the bias between the $\widehat{\mathbf{LV}}_n$ estimator applied to the standardized Pareto distribution tail above the true $\mathbf{VaR}_\alpha$ and the true $\mathbf{ES}_\alpha$ is provided in Figure 4.26.



(a) (b)

(c)

Figure 4.26: Evolution of the bias between the $\widehat{\mathbf{LV}}_n$ estimator applied to the standardized Pareto distribution tail above the true $\mathbf{VaR}_\alpha$ and the $\mathbf{ES}_\alpha$, for three different shape parameters, versus the sample size, with 95% CI, over 1000 repetitions, (a) $\mathcal{P}(1,2.5)$, (b) $\mathcal{P}(1,3.5)$, (c) $\mathcal{P}(1,5)$.

We are interested in the convergence speed of the bias between the Lee-Valiant estimator applied to the standardized Pareto distribution tail above the true $\mathbf{VaR}_\alpha$ and the $\mathbf{ES}_\alpha$. Because the bias is intended to be a power function of the sample size, then the logarithm of the bias is a linear function of the logarithm of the sample size with the convergence speed as the slope. Therefore, we display the log-log plot of the bias versus the sample size. See Figure 4.27.

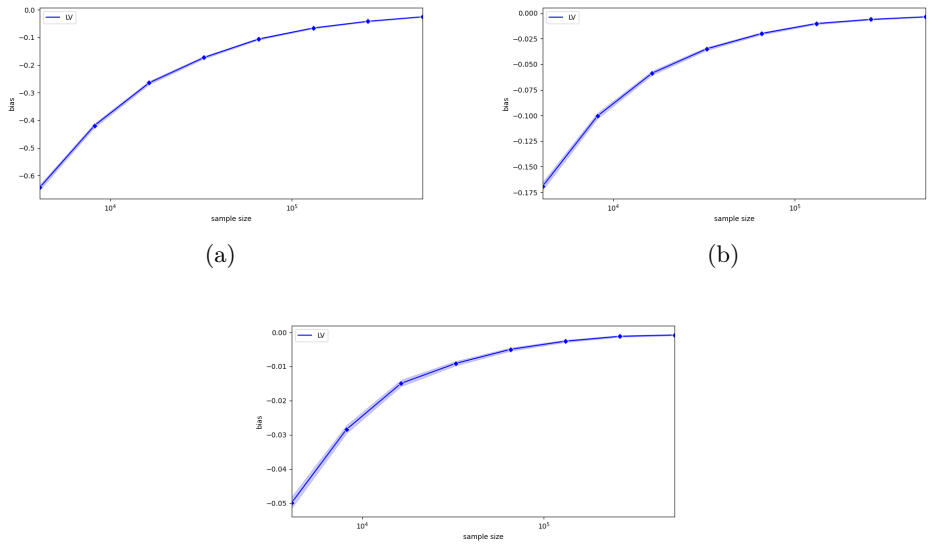


Figure 4.27: Evolution of the log bias between the $\widehat{\mathbf{LV}}_n$ estimator applied to the Pareto distribution's tail above the true $\mathbf{VaR}_\alpha$ and the $\mathbf{ES}_\alpha$, for three different shape parameters, versus the log sample size, with 95% CI, over 1000 repetitions, (a) $\mathcal{P}(1, 2.5)$, (b) $\mathcal{P}(1, 3.5)$, (c) $\mathcal{P}(1, 5)$.

Because $\mathbf{VaR}_\alpha$ is supposed to be known and does not depend on the underlying sample, the convergence speed of the bias between the $\widehat{\mathbf{LV}}_n$ estimator applied to the Pareto distribution tail above $\mathbf{VaR}_\alpha$ and the $\mathbf{ES}_\alpha$ is the same as the convergence speed of the bias between the $\widehat{\mathbf{LV}}_n$ estimator applied to the entire Pareto distribution and the expectation, that is, between $\frac{1}{n^{0.9}}$ and $\frac{1}{n^{0.6}}$.

Bias between the Lee-Valiant (LV) estimator applied to the standardized Pareto distribution tail above the empirical α-quantile and the $\mathbf{ES}_\alpha$ (Realistic case)

In the realistic case, the empirical α-quantile does not match the true $\mathbf{VaR}_\alpha$. In this case, the estimator of the $\mathbf{ES}_\alpha$ in the standardized Pareto distribution $\mathcal{P}(1, \gamma)$ is the empirical average of the standardized Pareto distribution conditional on its values being above the empirical α-quantile. The conditioning threshold is an order statistic and depends on the underlying sample. This implies that the samples larger than the empirical α-quantile are no longer independent and identically distributed (i.i.d.). The stability by conditioning and scaling properties of the Pareto distribution, as stated in

Theorem 4.7, are no longer valid. Therefore, the distribution of the samples larger than the empirical α-quantile is not necessarily a Pareto distribution. Thus, the distribution of the samples larger than the empirical α-quantile is unknown and it is difficult to establish an analytic closed-form formula for the bias between the $\widehat{\mathbf{LV}}_n$ estimator applied to the standardized Pareto distribution tail above this empirical α-quantile and the true $\mathbf{ES}_\alpha$. For this reason, we provide some experimental study to give an insight about the convergence speed of the bias between the $\widehat{\mathbf{LV}}_n$ applied to the Pareto distribution tail above the empirical α-quantile and the Expected-Shortfall. See Figure 4.28.

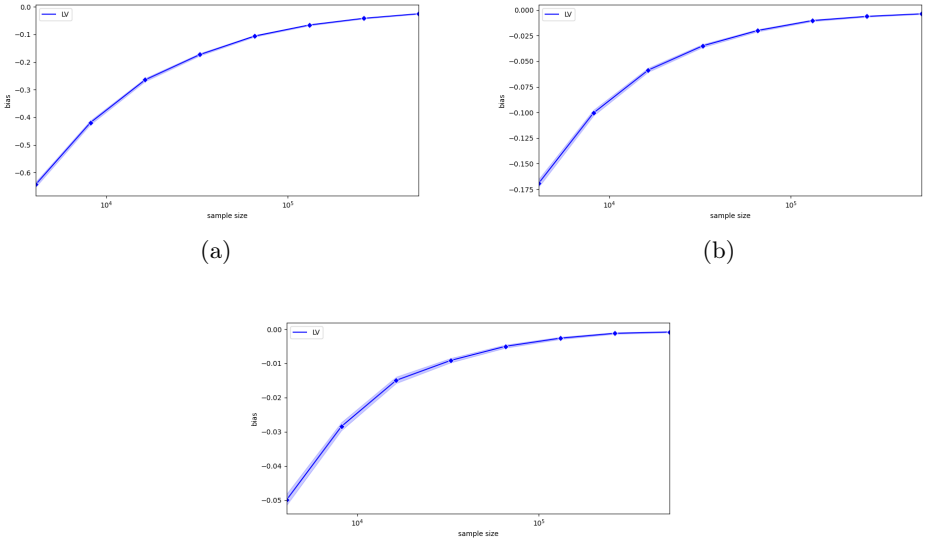
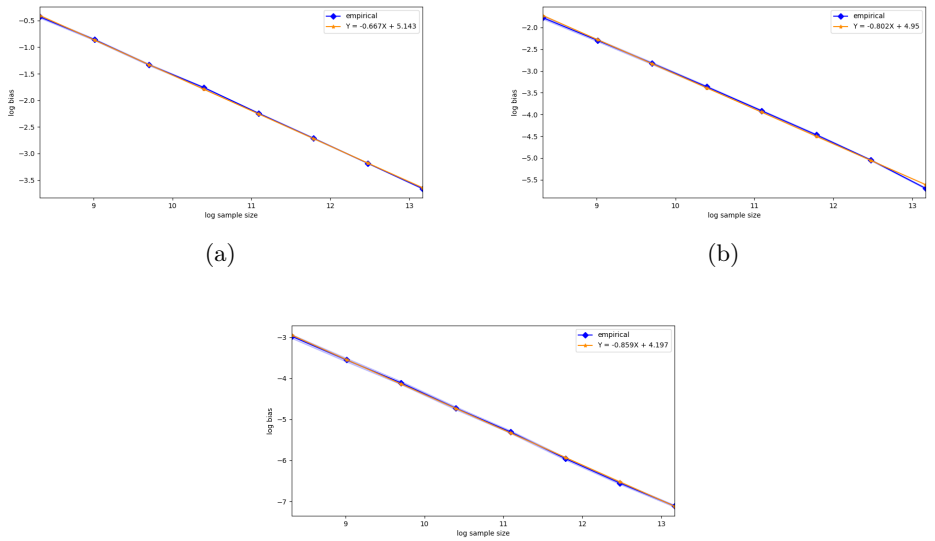


Figure 4.28: Evolution of the bias between the $\widehat{\mathbf{LV}}_n$ estimator applied to the standardized Pareto distribution tail above the empirical α-quantile and the $\mathbf{ES}_\alpha$, for three different shape parameters, versus the sample size, with 95% CI, over 1000 repetitions, (a) $\mathcal{P}(1, 2.5)$, (b) $\mathcal{P}(1, 3.5)$, (c) $\mathcal{P}(1, 5)$.

It is expected that the bias between the $\widehat{\mathbf{LV}}_n$ estimator applied to the standardized Pareto distribution tail above the empirical α-quantile and the $\mathbf{ES}_\alpha$, be a power function of the sample size. Therefore, it is expected that the logarithm of the bias be a linear function of the logarithm of the sample size. For this reason, to have an insight into the convergence speed of the bias between the $\widehat{\mathbf{LV}}_n$ applied to the standardized Pareto distribution tail above the empirical α-quantile and the $\mathbf{ES}_\alpha$, we study the log-log plot of the bias versus the sample size. See Figure 4.29.

The slopes of the log-log plots are between -0.9 and -0.7, this means that the convergence speed of the bias between the $\widehat{\mathbf{LV}}_n$ applied to the standardized Pareto distribution tail above the empirical α-quantile and the true $\mathbf{ES}_\alpha$ is larger than $\frac{1}{n^{0.9}}$ and lower than $\frac{1}{n^{0.6}}$.



(a) (b) (c)

Figure 4.29: Evolution of the log bias between the $\widehat{\mathbf{LV}}_n$ estimator applied to the standardized Pareto distribution tail above the empirical α-quantile and the $\mathbf{ES}_\alpha$, for three different shape parameters, versus the log sample size, with 95% CI, over 1000 repetitions, (a) $\mathcal{P}(1, 2.5)$, (b) $\mathcal{P}(1, 3.5)$, (c) $\mathcal{P}(1, 5)$.

4.7 Experimental results comparing all estimators

This section is dedicated to the presentation of experiments that illustrate the theoretical results. The distributions of the different estimators of the Expected-Shortfall are studied and compared. For this purpose, histograms, scatterplots and tables are presented.

Histograms: A histogram is a graphical representation of data points organized into user-specified ranges. Similar in appearance to a bar graph, the histogram condenses a data series into an easily interpreted visual by taking many data points and grouping them into logical ranges or bins. In our specific case, histograms represent the Monte-Carlo distribution of the estimators of the $\mathbf{ES}_\alpha$ and highlight the bias between the empirical mean of the Monte-Carlo distribution of these estimators and the Expected-Shortfall at the risk level α.

Scatterplots: A scatterplot, is a type of plot or mathematical diagram using Cartesian coordinates to display values for typically two variables for a set of data. If the points are coded (color/shape/size), one additional variable can be displayed. The data are displayed as a collection of points, each having the value of one variable determining the position on the horizontal axis and the value of the other variable determining the position on the vertical axis.

In our specific case, these scatterplots display the distribution of one estimator versus another estimator. This allows a comparison of the performance of one estimator with that of the other. If the samples line up along the diagonal both estimators provide the same performance. The samples that are above the diagonal testify to better performance for the estimator represented in ordinates when they are located in the lower left quarter, and to a worse performance of the estimator represented in ordinates when they are located in the upper right quarter. The samples that are below the diagonal testify to better performance for the estimator represented in abscissa when they are located in the lower left quarter, and to a worse performance of the estimator represented in abscissa when they are located in the upper right quarter.

Tables: For each estimator, we summarize descriptive statistics in tables. We present for different tail sizes, the confidence interval, the length of the confidence interval, the empirical mean and the empirical standard deviation.

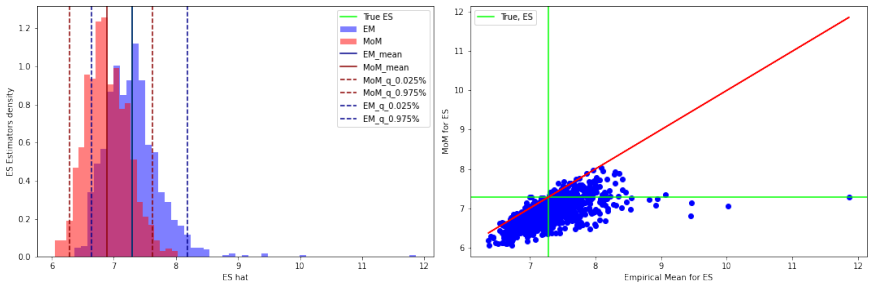
Notations: We denote the number of random draws in the Monte-Carlo trajectory by N, the risk level at which the empirical α-quantile is computed by α, the tail size above the empirical α-quantile by n, the shape parameter (Pareto index) by γ, the scaling parameter of the Pareto distribution by x_m, the confidence threshold by δ, and the truncation threshold for the Trimmed-Mean estimator by ϵ.

First, a comparison of each estimator is carried out with respect to the more common one, the empirical mean. Then, the comparison is carried out between the two best estimators.

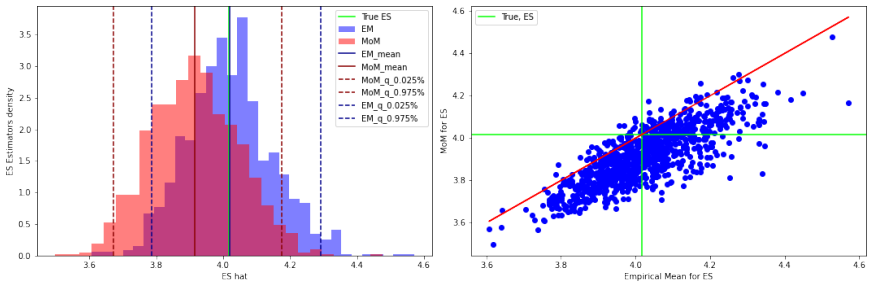
4.7.1 Median-of-Means (MoM) with parametrization of Theorem 4.26 versus Empirical Mean (EM) for ES

In this section, the distribution of the $\widehat{\mathbf{MoM}}_n$ estimator for the **ES** is compared with the distribution of the empirical mean estimator for the **ES** for three shape values of the Pareto distribution $\gamma = 2.5$ (the heavier tailed-distribution), $\gamma = 3.5$ and $\gamma = 5$ (the thinner tailed-distribution). In both cases, the experiments are carried out on the distribution tail above the empirical α-quantile of a standardized Pareto distribution $\mathcal{P}(1, \gamma)$. Moreover, the estimators are computed at a confidence level of 95%, on a number of draws equal to 10000. The risk level α is fixed at 97.5% which leads to a tail size of 250. Finally, the Monte-Carlo distribution of the estimators is built from 1000 trajectories. See Figure 4.30.

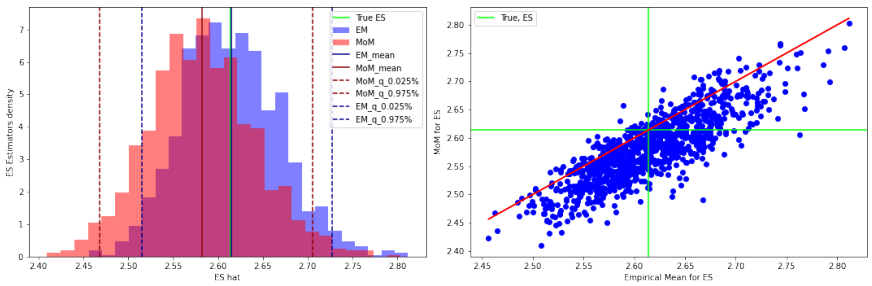
The distribution of the empirical mean for the **ES** is unbiased, whereas the distribution of the **MoM** estimator for the **ES** presents a non-negligible bias. However, the distribution of the empirical mean for the **ES** is heavy-tailed and presents a higher variance than the distribution of the **MoM** estimator for the **ES**. This bias decreases when the tail of the distribution becomes increasingly thinner, that is, when the shape parameter increases.



(a)



(b)



(c)

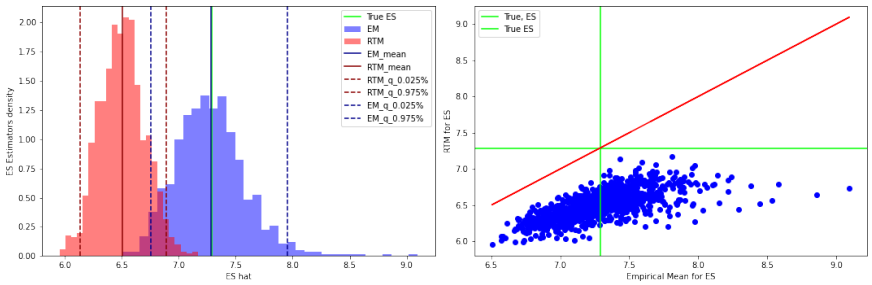
Figure 4.30: Monte-Carlo distributions of the **MoM** estimator for the **ES** versus the empirical mean estimator for the **ES**, for three standardized Pareto distributions $\mathcal{P}(1,\gamma)$ of different shape parameters: (a) $\gamma = 2.5$ (b) $\gamma = 3.5$ (c) $\gamma = 5$.

Moreover, the scatterplots show that the empirical mean estimator provides a better estimation of the ES than the **MoM** estimator for small values but is less accurate for large values.

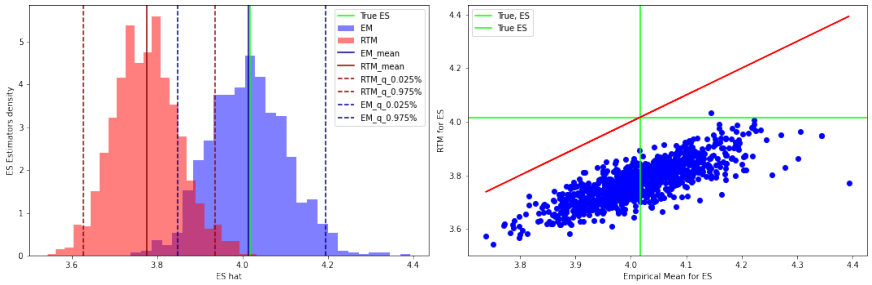
4.7.2 Right-Trimmed-Mean (RTM) versus Empirical Mean (EM) for ES

In this section, the distribution of the RTM estimator for the **ES** is compared with the empirical mean estimator for the **ES**, for three shape values of the standardized Pareto distribution, $\gamma = 2.5$, $\gamma = 3.5$ and $\gamma = 5$. In both cases, the experiments are conducted on the tail distribution above the empirical α-quantile of the standardized Pareto distribution $\mathcal{P}(1,\gamma)$. The RTM estimator for the **ES** is built only on one half of the distribution tail because the other half of the tail is used to estimate the upper trunca-

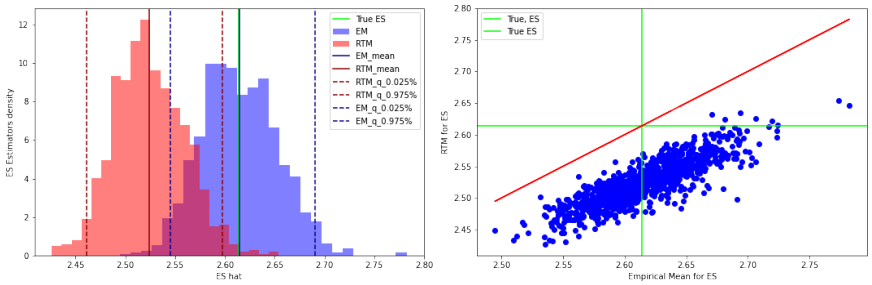
tion threshold. Therefore, the number of draws in the entire trajectories is doubled so that the number of samples on which the RTM estimator is built remains the same as previously mentioned. Therefore, the estimators are still computed at a confidence level of 95% but on a number of draws equal to 20000. The risk level α is still fixed at 97.5%, which leads to a tail size of 500. Both training and test sets are composed of 250 draws. The truncation percentage is fixed at 10%. Finally, the Monte-Carlo distribution of the estimators is built from 1000 trajectories. See Figure 4.31.



(a)



(b)



(c)

Figure 4.31: Monte-Carlo distributions of the RTM estimator for the **ES** versus the empirical mean estimator for the **ES**, for three standardized Pareto distributions $\mathcal{P}(1, \gamma)$ of different shape parameters: (a) $\gamma = 2.5$ (b) $\gamma = 3.5$ (c) $\gamma = 5$.

As previously mentioned, the distribution of the empirical mean estimator for the **ES** is unbiased, whereas the distribution of the RTM estimator for the **ES** exhibits an important bias that decreases when the tail of the Pareto distribution becomes increasingly thinner, that is, when the shape parameter increases. Moreover, the distribution of the empirical mean for the **ES** is heavy-tailed and presents a large variance whereas the distribution of the RTM estimator for the **ES** is thin-tailed; therefore, its variance is

lower.

The scatterplots allow us to conclude about the accuracy of the estimators. In both cases, the estimation of the **ES** provided by the empirical mean estimator is better than that provided by the RTM estimator.

4.7.3 Lee-Valiant (LV) versus Empirical Mean (EM) for ES

In this section, the distribution of the LV estimator for the **ES** is compared with the empirical mean estimator for the **ES**, for three shape values of the Pareto distribution, $\gamma = 2.5$, $\gamma = 3.5$ and $\gamma = 5$. In both cases, the experiments are carried out on the distribution tail above the empirical α-quantile of the standardized Pareto distribution $\mathcal{P}(1, \gamma)$. Moreover, the estimators are computed at a confidence level of 95%, on a number of draws equal to 1000. The risk level α is fixed at 97.5% which leads to a tail size of 250. Finally, the Monte-Carlo distribution of the estimators is built from 1000 trajectories. See Figure 4.32.

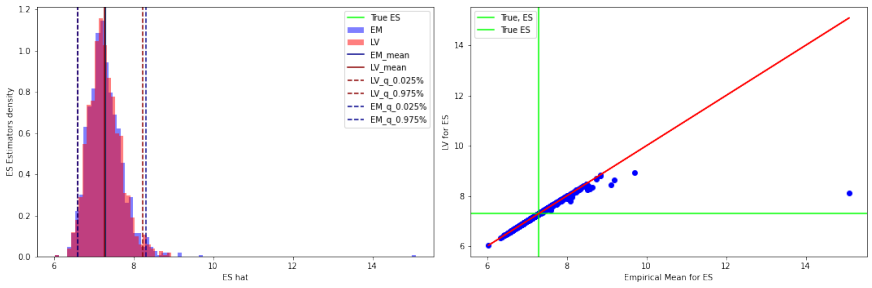
The distributions of both the empirical mean and of the LV estimators for the **ES** are unbiased. Moreover, the distribution of the empirical mean for the **ES** is heavy-tailed and exhibits a large variance, whereas the distribution of the LV estimator for the **ES** is thin-tailed. This means that the variance of the empirical mean estimator for the **ES** is higher than that of the LV estimator for the **ES**. Finally, the scatterplots show that both the performance of the empirical mean estimator for the **ES** and the performance of the LV estimator for the **ES** are the same, in most cases. They exhibit good performance for the values at the heart of the distribution but poor performances for very low or very high values (extreme values).

4.7.4 Lee-Valiant (LV) versus Median-of-Means (MoM) for ES

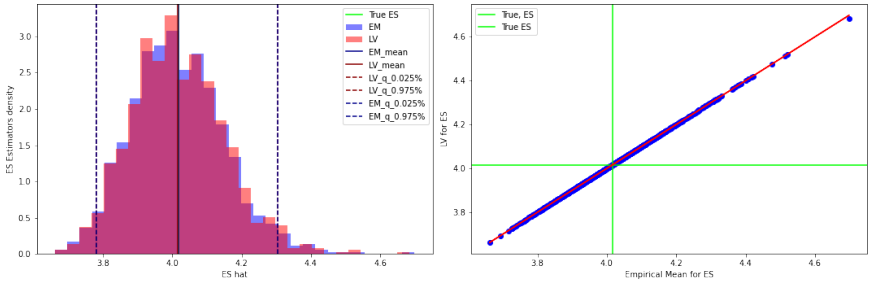
In this section, the distribution of the LV estimator for the **ES** is compared with the **MoM** estimator for the **ES**, for three shape values of the standardized Pareto distribution, $\gamma = 2.5$, $\gamma = 3.5$ and $\gamma = 5$. In both cases, the experiments are conducted on the tail distribution above the empirical α-quantile of the standardized Pareto distribution $\mathcal{P}(1, \gamma)$. Moreover, the estimators are computed at a confidence level of 95%, on a number of draws equal to 1000. The risk level α is fixed at 97.5% which leads to a tail size of 250. Finally, the Monte-Carlo distribution of the estimators is built from 1000 trajectories.

As previously mentioned, the distribution of the **MoM** estimator for the **ES** is biased, whereas the distribution of the LV estimator for the **ES** is not. The tail of the LV estimator is thinner than that of the **MoM** estimator, which means that the variance of the LV estimator is lower than that of the **MoM** estimator.

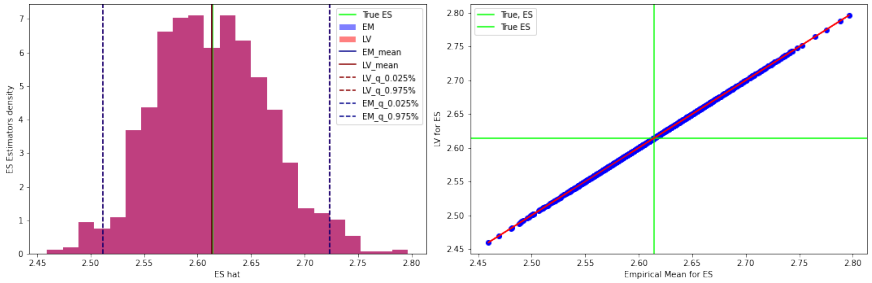
The scatterplots show that the estimations of the **ES** provided by the **MoM** estimator and those provided by the LV estimator are significantly different from each other. In most cases, the LV estimator provides an estimation of the **ES** that is more accurate



(a)



(b)

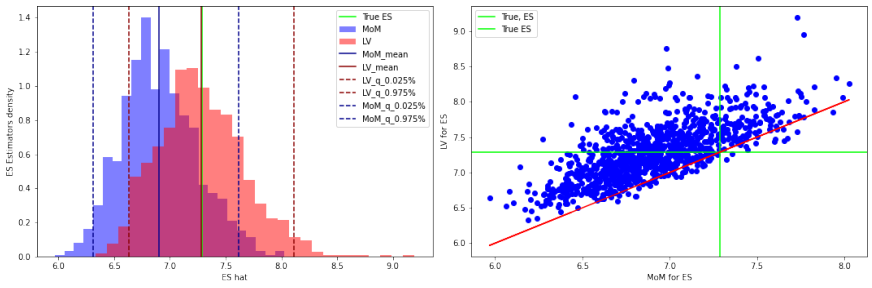


(c)

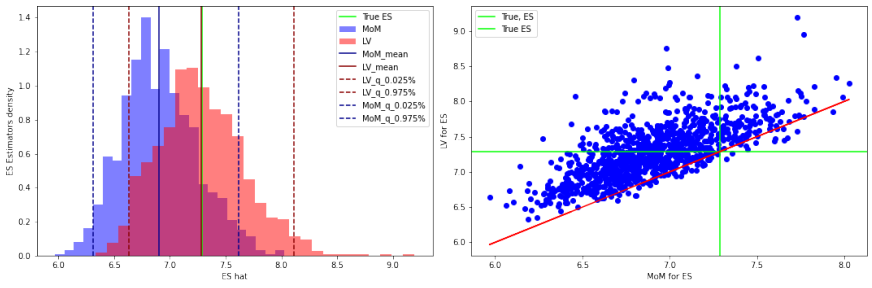
Figure 4.32: Monte-Carlo distributions of the LV estimator for the **ES** versus the empirical mean estimator for the **ES**, for three standardized Pareto distribution $\mathcal{P}(1,\gamma)$ of different shape parameters: (a) $\gamma = 2.5$ (b) $\gamma = 3.5$ (c) $\gamma = 5$.

than that provided by the **MoM** estimator. The LV estimator can be seen as an improved version of the **MoM** estimator.

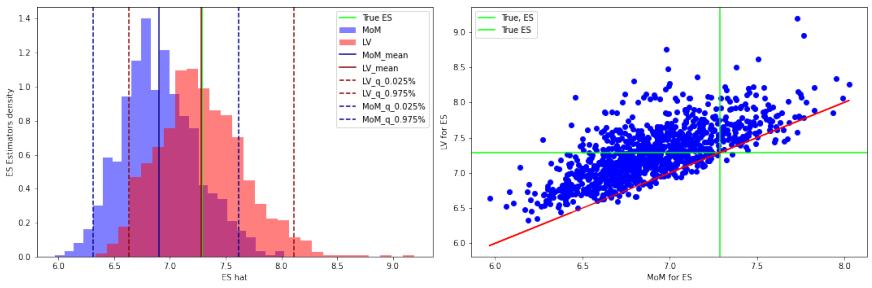
In Figure 4.34, violin plots represent for each shape parameter γ the distribution of four estimators: empirical mean (EM), **MoM**, LV and RTM. These violin plots confirm the conclusions drawn from the histograms. Indeed, the LV estimator has the same distribution as the empirical mean (EM), and these two estimators are unbiased. The **MoM** estimator has a similar distribution, but presents a bias, and the RTM estimator has a similar distribution but presents a large bias.



(a)



(b)



(c)

Figure 4.33: Monte-Carlo distributions of the LV estimator for the **ES** versus the **MoM** estimator for the **ES**, for three standardized Pareto distributions $\mathcal{P}(1,\gamma)$ of different shape parameters: (a) $\gamma = 2.5$ (b) $\gamma = 3.5$ (c) $\gamma = 5$.

4.7.5 Summary of the results

Tables summarizing the descriptive statistics and the bias of the estimators (EM, MoM, RTM, and LV) of the $\mathbf{ES}_\alpha$, with $\alpha = 0.975$, for several standardized Pareto distributions with different Pareto indices ($\mathcal{P}(1,2.5)$, $\mathcal{P}(1,3.5)$, and $\mathcal{P}(1,5)$), with various tail size, are provided. For $\mathcal{P}(1,2.5)$, see Table 4.1, for $\mathcal{P}(1,3.5)$, see Tables 4.2, and for $\mathcal{P}(1,5)$ see Tables 4.3. Moreover, the statistics are computed from 1000 repetitions of the experiments, and confidence intervals are created at the level 95%.

In these tables, the blue color is dedicated to the target value, that is, the value of the true $\mathbf{ES}_\alpha$, the red color is used to the estimator that exhibits the worst performance, and the green color highlights the estimator that exhibits the best performance.

(i) The above tables show that the $\mathbf{ES}_\alpha$ decreases when the Pareto index decreases, that is, when the Pareto distribution tail becomes increasingly thinner.

True ES = 7.289 - $\alpha = 0.975$ - $\delta = 0.05$ - $\gamma = 2.5$ - $x_m = 1$ - $n_{runs} = 1000$						
N, n	Estimator	CI ($q_{2.5\%} - q_{97.5\%}$)	$\hat{\mathbb{E}}$	$\hat{Std}$	$q_{97.5\%} - q_{2.5\%}$	Bias
$N = 25000, n = 626$	EM	6.824 – 7.782	7.288	0.287	0.958	−0.001383
	MoM - Thm2 (k = 24, m = 26, r = 1)	6.640 – 7.535	7.074	0.231	0.894	−0.215191
	RTM ($\epsilon = 0.447$)	6.277 – 6.975	6.628	0.177	0.698	−0.661202
	LV	6.823 – 7.773	7.279	0.248	0.950	−0.009763
$N = 50000, n = 1251$	EM	6.923 – 7.656	7.277	0.190	0.733	−0.011991
	MoM - Thm2 (k = 24, m = 52, r = 2)	6.780 – 7.511	7.149	0.185	0.730	−0.140317
	RTM ($\epsilon = 0.291$)	6.553 – 7.110	6.831	0.145	0.557	−0.457658
	LV	6.922 – 7.645	7.274	0.183	0.723	−0.015122
$N = 100000, n = 2501$	EM	7.039 – 7.574	7.288	0.142	0.535	−0.000797
	MoM - Thm2 (k = 24, m = 104, r = 4)	6.959 – 7.495	7.208	0.137	0.537	−0.080829
	RTM ($\epsilon = 0.217$)	6.765 – 7.234	6.988	0.120	0.469	−0.300604
	LV	7.038 – 7.571	7.286	0.137	0.533	−0.003205

Table 4.1: Descriptive statistics of the estimators (EM, MoM, RTM, LV) of the $\mathbf{ES}_\alpha$, with $\alpha = 0.975$, are computed on the standardized Pareto distribution $\mathcal{P}(1, 2.5)$. Statistics of these estimators are computed from 1000 repetitions with a confidence level equal to 95%. These experiments are carried out four three distinct sizes of the distribution tail. (N is the trajectory size and n is the tail size)

True ES = 4.017 - $\alpha = 0.975$ - $\delta = 0.05$ - $\gamma = 3.5$ - $x_m = 1$ - $n_{runs} = 1000$						
N, n	Estimator	CI ($q_{2.5\%} - q_{97.5\%}$)	$\hat{\mathbb{E}}$	$\hat{Std}$	$q_{97.5\%} - q_{2.5\%}$	Bias
$N = 25000, n = 626$	EM	3.864 – 4.180	4.0127	0.082	0.316	−0.003943
	MoM - Thm2 (k = 24, m = 26, r = 1)	3.806 – 4.152	3.965	0.088	0.346	−0.051934
	RTM ($\epsilon = 0.447$)	3.685 – 3.974	3.822	0.071	0.289	−0.194164
	LV	3.864 – 4.177	4.012	0.081	0.313	0.004163
$N = 50000, n = 1251$	EM	3.906 – 4.140	4.016	0.060	0.234	−0.000273
	MoM - Thm2 (k = 24, m = 52, r = 2)	3.876 – 4.115	3.989	0.063	0.239	−0.027715
	RTM ($\epsilon = 0.291$)	3.789 – 4.001	3.892	0.055	0.212	−0.124526
	LV	3.906 – 4.138	4.016	0.060	0.232	−0.000412
$N = 100000, n = 2501$	EM	3.939 – 4.097	4.017	0.041	0.159	0.000690
	MoM - Thm2 (k = 24, m = 104, r = 4)	3.916 – 4.088	4.003	0.045	0.171	−0.013946
	RTM ($\epsilon = 0.217$)	3.861 – 4.025	3.942	0.041	0.164	−0.074801
	LV	3.939 – 4.097	4.017	0.041	0.159	0.000620

Table 4.2: Descriptive statistics of the estimators (EM, MoM, RTM, LV) of the $\mathbf{ES}_\alpha$, with $\alpha = 0.975$, are computed on the standardized Pareto distribution $\mathcal{P}(1, 3.5)$. Statistics of these estimators are computed from 1000 repetitions with a confidence level equal to 95%. These experiments are carried out four three distinct sizes of the distribution tail. (N is the trajectory size and n is the tail's size)

True ES = 2.614 - $\alpha = 0.975$ - $\delta = 0.05$ - $\gamma = 5$ - $x_m = 1$ - $n_{runs} = 1000$						
N, n	Estimator	CI ($q_{2.5\%} - q_{97.5\%}$)	$\hat{\mathbb{E}}$	$\hat{Std}$	$q_{97.5\%} - q_{2.5\%}$	Bias
$N = 25000, n = 626$	EM	2.552 – 2.683	2.615	0.0337	0.132	0.000458
	MoM - Thm2 (k = 24, m = 26, r = 1)	2.530 – 2.673	2.599	0.037	0.143	−0.014891
	RTM ($\epsilon = 0.447$)	2.486 – 2.612	2.547	0.032	0.126	−0.067281
	LV	2.552 – 2.683	2.615	0.034	0.132	0.000429
$N = 50000, n = 1251$	EM	2.570 – 2.662	2.615	0.0237	0.092	0.000508
	MoM - Thm2 (k = 24, m = 52, r = 2)	2.560 – 2.662	2.607	0.026	0.102	−0.007571
	RTM ($\epsilon = 0.291$)	2.524 – 2.623	2.571	0.025	0.099	−0.042827
	LV	2.570 – 2.662	2.615	0.024	0.092	0.000494
$N = 100000, n = 2501$	EM	2.582 – 2.650	2.614	0.017	0.068	0.000047
	MoM - Thm2 (k = 24, m = 104, r = 4)	2.575 – 2.650	2.610	0.019	0.076	−0.003929
	RTM ($\epsilon = 0.217$)	2.553 – 2.627	2.589	0.019	0.074	−0.025168
	LV	2.582 – 2.650	2.614	0.017	0.067	0.000040

Table 4.3: Descriptive statistics of the estimators (EM, MoM, RTM, LV) of the $\mathbf{ES}_\alpha$, with $\alpha = 0.975$, are computed on the standardized Pareto distribution $\mathcal{P}(1, 5)$. Statistics of these estimators are computed from 1000 repetitions with a confidence level equal to 95%. These experiments are carried out for three distinct sizes of the distribution tail. (N is the trajectory size and n is the tail's size)

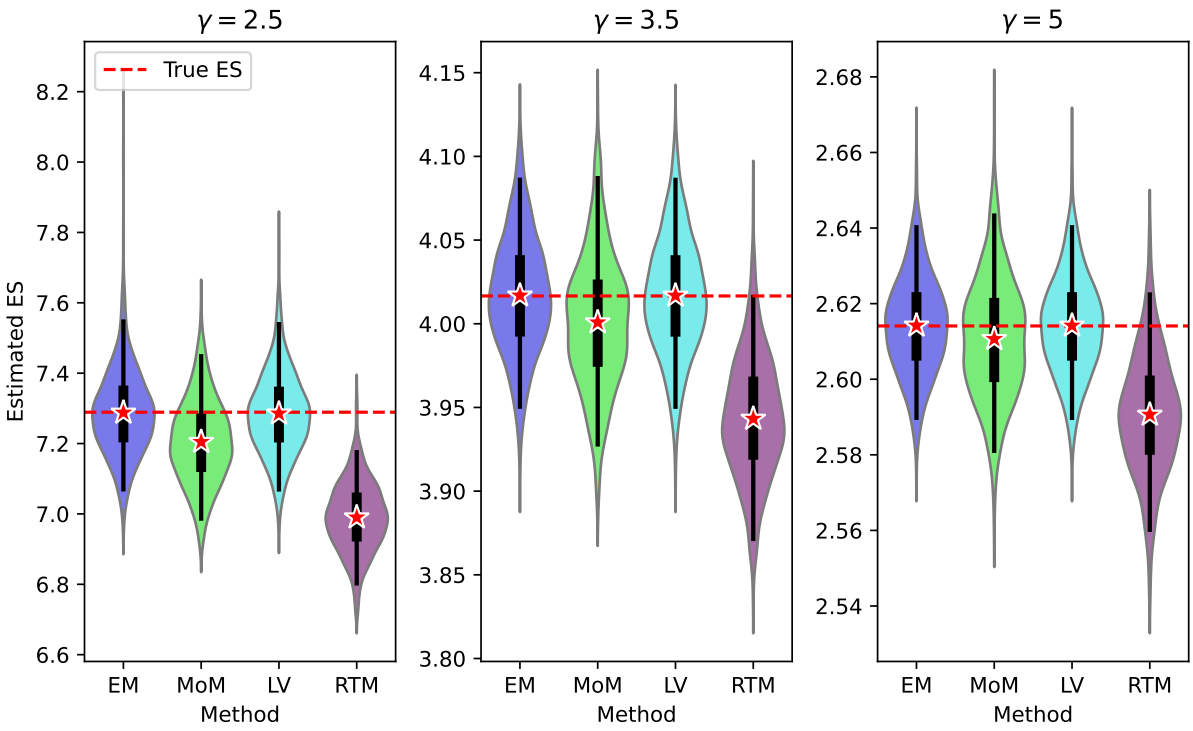


Figure 4.34: Estimators distribution (1 000 runs), whiskers: $2.5 - 97.5$ perc.

(ii) The estimator of the $\mathbf{ES}_\alpha$ with the lowest standard deviation is the RTM, however this estimator is significantly biased.

(iii) Two estimators are very interesting: the empirical mean estimator in the asymptotic case and the LV estimator which demonstrates very good performance both in the non-asymptotic and in the asymptotic cases. The empirical mean estimator has a lower bias than the LV estimator but its standard deviation is higher than that of the LV estimator. The estimator that presents the most relevant performance in the non-asymptotic framework with a good trade-of between the bias and the standard deviation is LV estimator. Indeed, the LV estimator has a low bias even if it is a bit higher than that of empirical mean, and its standard deviation is the lowest.

Proof of the bias between the empirical ES and the true ES in the realistic case (when the empirical α-quantile does not match the true $\mathbf{VaR}_\alpha$)

Proof Let us recall the formula of the empirical Expected-Shortfall, $\mathbf{ES}_\alpha^n$:

$$\mathbf{ES}_\alpha^n = \frac{1}{n - \lceil \alpha n \rceil} \sum_{i=1}^{n} X_i \mathbb{1}_{\{X_i \geq q_\alpha^n\}}. \tag{4.234}$$

In the sake of simplicity, let us denote by $r = \lceil \alpha n \rceil$ the rank of the empirical α-quantile. Then, the empirical **ES** corresponds to the average of the samples larger than the order statistics of order r and can be rewritten as $\mathbf{ES}_r^n = \frac{1}{n-r} \sum_{i=1}^{n} X_i \mathbb{1}_{\{X_i \geq X_r^\star\}}$.

The goal is to compute the bias between the empirical **ES** and the theoretical **ES** for a Pareto distribution.

$$B_{\mathbf{ES}_r}[\mathbf{ES}_r^n] = \mathbb{E}\left[\mathbf{ES}_r^n\right] - \mathbf{ES}_r. \tag{4.235}$$

where:

$$\mathbb{E}\left[\mathbf{ES}_r^n\right] = \frac{1}{n-r} \sum_{i=1}^{n} \mathbb{E}\left[X_i \mathbb{1}_{\{X_i \geq X_r^\star\}}\right] = \frac{n}{n-r} \mathbb{E}\left[X_1 \mathbb{1}_{\{X_1 \geq X_r^\star\}}\right] \tag{4.236}$$

because the sample is i.i.d..

The stake is to compute the expectation $\mathbb{E}\left[X_1 \mathbb{1}_{\{X_1 \geq X_r^\star\}}\right]$. For that purpose, the knowledge of the joint distribution of $X_i, X_r^\star$ is useful. The proof is organized into two parts. In a first time, the proof is carried out for the uniform distribution on $[0,1]$, in order to explicitly compute the joint distribution of $X_i, X_r^\star$. Second, a direct application to the standardized Pareto distribution is provided owing to a variable change, from the uniform distribution.

(i) **Random variable X_i follows a uniform distribution on $[0,1]$.**

For all $(u,v) \in [0,1]^2$,

$$\mathbb{P}(X_r^\star \leq u, X_i \leq v) = \mathbb{E}\left[\mathbb{1}_{\{X_r^\star \leq u\} \cap \{X_i \leq v\}\}}\right]. \tag{4.237}$$

The tower property leads to:

$$\mathbb{P}(X_r^\star \leq u, X_i \leq v) = \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{\{X_r^\star \leq u\} \cap \{X_i \leq v\}\}} \mid X_i\right]\right]. \tag{4.238}$$

Thanks to the measurability, we get:

$$\mathbb{P}(X_r^\star \leq u, X_i \leq v) = \mathbb{E}\left[\mathbb{1}_{\{X_i \leq v\}} \mathbb{E}\left[\mathbb{1}_{\{X_r^\star \leq u\}} \mid X_i\right]\right] = \mathbb{E}\left[\mathbb{1}_{\{X_i \leq v\}} \mathbb{P}(X_r^\star \leq u \mid X_i)\right]. \tag{4.239}$$

Let us now focus on the conditional probability $\mathbb{P}(X_r^\star \leq u \mid X_i)$.

From (Reiss, 2012, p.13, formula (1.1.7)), since there are at least r samples lower than the order statistics of order r, then the event $\{X_r^\star \leq u\}$ can be rewritten as follows:

$$\{X_r^\star \leq u\} = \left\{ \sum_{j=1}^{n} \mathbb{1}_{\{X_j \leq u\}} \geq r \right\}. \tag{4.240}$$

Then, the probability of the event $\{X_r^\star \leq u\}$ becomes:

$$\mathbb{P}(X_r^\star \leq u \mid X_i) = \mathbb{P}\left(\sum_{j=1}^{n} \mathbb{1}_{\{X_j \leq u\}} \mid X_i \right) \tag{4.241}$$

$$= \mathbb{1}_{\{X_i > u\}} \mathbb{P}\left(\sum_{j=1, j\neq i}^{n} \mathbb{1}_{\{X_j \leq u\}} \geq r \right) + \mathbb{1}_{\{X_i \leq u\}} \mathbb{P}\left(\sum_{j=1, j\neq i}^{n} \mathbb{1}_{\{X_j \leq u\}} \geq r-1 \right). \tag{4.242}$$

Because the r.v. $\mathbb{1}_{\{X_j \leq u\}}$ are Bernoulli, the sum of these r.v. is Binomial and, we can write on the one hand, $\mathbb{P}\left(\sum_{j=1, j\neq i}^{n} \mathbb{1}_{\{X_j \leq u\}} \geq r\right) = \sum_{k=r}^{n-1} \binom{n-1}{k} u^k (1-u)^{n-1-k}$, and on the other hand, $\mathbb{P}\left(\sum_{j=1, j\neq i}^{n} \mathbb{1}_{\{X_j \leq u\}} \geq r-1\right) = \sum_{k=r-1}^{n-1} \binom{n-1}{k} u^k (1-u)^{n-1-k}$. Then, we get:

$$\mathbb{P}(X_r^\star \leq u, X_i \leq v) = \mathbb{E}\left[\mathbb{1}_{X_i \in (u,v]} \sum_{k=r}^{n-1} \binom{n-1}{k} u^k (1-u)^{n-1-k} + \mathbb{1}_{\{X_i \leq v, X_i \leq u\}} \sum_{k=r-1}^{n-1} \binom{n-1}{k} u^k (1-u)^{n-1-k} \right] \tag{4.243}$$

$$= \mathbb{E}\left[\mathbb{1}_{\{X_i \leq v\}} \sum_{k=r}^{n-1} \binom{n-1}{k} u^k (1-u)^{n-1-k} + \mathbb{1}_{\{X_i \leq v, X_i \leq u\}} \binom{n-1}{r-1} u^{r-1} (1-u)^{n-r} \right] \tag{4.244}$$

$$= \mathbb{P}(X_i \leq v) \sum_{k=r}^{n-1} \binom{n-1}{k} u^k (1-u)^{n-1-k} + \mathbb{P}(X_i \leq \min(u,v)) \binom{n-1}{r-1} u^{r-1} (1-u)^{n-r}. \tag{4.245}$$

Under the uniform distribution, we get:

$$\mathbb{P}(X_r^\star \leq u, X_i \leq v) = v \sum_{k=r}^{n-1} \binom{n-1}{k} u^k (1-u)^{n-1-k} + \min(u,v) \binom{n-1}{r-1} u^{r-1} (1-u)^{n-r}. \tag{4.246}$$

(ii) **The random variable Y_i follows a standardized Pareto distribution $\mathcal{P}(1,\gamma)$.**

Let Y_i be a standardized Pareto random variable such that $Y_i \sim \mathcal{P}(1,\gamma)$ with $\gamma > 2$, with F_{Y_i}, as cumulative distribution function. We call the generalized inverse of F_{Y_i}, the function $F_{Y_i}^{-1}$ defined for all $y \in]0,1]$ by:

$$F_{Y_i}^{-1}(y) = \inf\left\{ x \in \mathbb{R} \mid F_{Y_i}(x) \geq y \right\} = \mathbf{VaR}(y) = (1-y)^{-\frac{1}{\gamma}}. \tag{4.247}$$

If X_i is a uniform random variable such that $X_i \sim \mathcal{U}([0,1])$, then $F_{Y_i}^{-1}(X_i)$ follows the same distribution as Y_i, that is the standardized Pareto distribution $\mathcal{P}(1,\gamma)$. Moreover, because F_{Y_i} is continuous on $\mathbb{R}$ and $f_{Y_i} > 0$, then $F_{Y_i}(Y_i)$ follows the uniform distribution $\mathcal{U}([0,1])$.

Therefore:

$$Y_i = F_{Y_i}^{-1}(X_i) = \mathbf{VaR}(X_i) = (1 - X_i)^{-\frac{1}{\gamma}} \sim \mathcal{P}(1,\gamma). \tag{4.248}$$

Proceeding to the following quantile transformation $\mathbf{VaR}_\alpha(X_i) = Y_i$ in Equation (4.246), we obtain:

$$\mathbb{P}(X_r^\star \leq u, X_i \leq v) = \mathbb{P}(Y_r^\star \leq \mathbf{VaR}_\alpha(u), Y_i \leq \mathbf{VaR}_\alpha(v)). \tag{4.249}$$

We would continue with the deriving of the joint cumulative distribution function to compute the average of the empirical $\mathbf{ES}_\alpha$, and by analyzing the asymptotic bias in n. Currently, there is no simple formula that would allow us to carry out this analysis. This is still a work in progress. ■

"Mathematics is the most beautiful and most powerful creation of the human spirit."

(**Stefan Banach**, International Congress of Mathematicians held in Zurich, Switzerland, 1932.)

Etat de l'art et résumé des contributions

Motivés par les différentes questions de recherche (RQ) énoncées dans la Section 4.1.3 du Chapter 1, nous fournissons maintenant un aperçu détaillé des contributions de cette thèse où chaque chapitre est dédié à l'une des directions de recherche.

4.8 Contributions

4.8.1 Chapter 2 - Introduction aux processus Gaussiens self-similaires et stationnaires

Etat de l'art

En finance, comprendre les effets temporels dans les processus est un problème majeur. Certaines propriétés bien connues appelées stationnarité et auto-similarité sont liées à ces effets temporels. Plus précisément, la propriété de stationnarité est associée à la translation temporelle aussi connue sous le nom de changement d'origine temporelle et la propriété d'auto-similarité est liée au changement d'échelle de temps.

La propriété de stationnarité d'un processus est intéressante car elle énonce que le processus des incréments (c'est-à-dire le processus translaté dans le temps) et le processus initial ont les mêmes caractéristiques. En d'autres termes, les caractéristiques d'un processus stationnaire sont invariantes par translation, ou par changement d'origine temporelle, en temps et en espace. La propriété d'autosimilarité suppose que le processus changé d'échelle temporelle et le processus initial conservent les mêmes caractéristiques par un redimensionnement spatial approprié. La propriété d'autosimilarité d'un processus établit une relation de proportionnalité spatiale entre les caractéristiques d'un processus évalués à deux échelles de temps distinctes λt et t avec $\lambda > 0$, et le coefficient de proportionnalité spatiale dépend de l'échelle temporelle λ. De plus, un processus autosimilaire présente une dépendance à long ou à court terme. Selon les hypothèses faites sur le processus, les propriétés de stationnarité et d'autosimilarité peuvent concerner les caractéristiques $\mathbb{L}^2$ du processus, sa distribution (au sens faible), ou sa trajectoire (au sens strict).

Mandelbrot et Van Ness ont été les premiers à introduire un processus Gaussien autosimilaire avec des incréments stationnaires, le célèbre mouvement Brownien fractionnaire en 1968 (Mandelbrot and Van Ness, 1968a). Ils fournissent une représentation intégrale de tels processus. Dans (Taqqu, 1978), une représentation indexée dans le temps pour une séquence de processus auto-similaires dont les moments fini-dimensionnels ont été spécifiés est fournie. La représentation des processus stationnaires via l'équation de Langevin et des processus auto-similaires a été étudiée dans (Viitasaari, 2016). Les propriétés de stationnarité et d'auto-similarité ne sont pas spécifiques au cadre Gaussien, mais sont souvent exploitées dans le cadre Gaussien du fait de ses propriétés pratiques

et font l'objet de nombreuses études. Les processus auto-similaires sont étudiés en détail dans (Das and Pan, 2011, Chap. 3), (Embrechts and Maejima, 2000), (Embrechts, 2009), (Samorodnitsky, 2006), (Chaumont, 2006). Dans (Lamperti, 1962), les auteurs étudient les propriétés des processus semi-stables, y compris les processus auto-similaires. Les processus stationnaires et auto-similaires sont utilisés dans divers domaines, comme expliqué dans (Pardo, 2007). En effet, ils peuvent être utilisés pour modéliser de nombreux phénomènes aléatoires qui présentent un changement d'échelle espace-temps et qui peuvent être observés en finance, physique, biologie et dans d'autres domaines. Par exemple, les fragments stellaires, la croissance et la généalogie des populations, la tarification des options en finance, divers domaines du traitement d'images, la climatologie, les sciences de l'environnement ne sont que quelques-uns des domaines dans lesquels les processus auto-similaires sont utilisés. Les processus auto-similaires apparaissent dans diverses parties de la théorie des probabilités, tels que les processus de Lévy, les processus de branchement, la physique statistique, la théorie de la fragmentation, la théorie de la coalescence, les champs aléatoires. Quelques exemples bien connus sont : le processus de Lévy stable, le mouvement Brownien fractionnaire, la diffusion de branchement de Feller, les processus de Bessel, la fragmentation auto-similaire (Bertoin, 2002). Les processus auto-similaires sont également utilisés en télécommunications et en traitement du signal, comme évoqué dans (Sheluhin et al., 2007). La construction de processus auto-similaires est expliquée dans (Fan et al., 2015). De tels processus sont également très intéressants lors de l'ajout d'hypothèses supplémentaires telles que l'indépendance de leurs incréments (Sato, 1991). En supposant la propriété de stationnarité en plus de l'auto-similarité des processus, certains travaux montrent que ces processus peuvent prendre une forme spécifique. En effet, dans (Samorodnitsky et al., 1996, Chap. 7) ou dans (Barndorff-Nielsen and Pérez-Abreu, 1999), les auteurs montrent qu'il existe un unique processus Gaussien auto-similaire avec des incréments stationnaires et que ce processus correspond au mouvement Brownien fractionnaire. Les processus auto-similaires avec des incréments stationnaires sont également étudiés dans la théorie du chaos dans (Maejima and Tudor, 2012). Les auteurs prouvent que les processus Gaussiens vivent dans le premier chaos de Wiener et que ces derniers sont exprimés sous forme d'intégrales simples, avec une intégrande déterministe, par rapport au processus de Wiener. Ensuite, dans le premier chaos de Wiener, le seul processus Gaussien auto-similaire avec des incréments stationnaires est le mouvement Brownien fractionnaire. De plus, les auteurs prouvent que les éléments du deuxième chaos de Wiener sont des intégrales stochastiques doublement itérées par rapport au processus de Wiener. Dans le deuxième chaos de Wiener, les processus auto-similaires avec des incréments stationnaires ne sont plus Gaussiens, puis les auteurs prouvent qu'il existe une infinité de tels processus. Les processus Gaussiens auto-similaires et stationnaires, tels que le mouvement Brownien fractionnaire (fBm), ont trouvé de nombreuses applications en finance en raison de leur capacité à modéliser des comportements complexes observés sur les marchés financiers, comme le démontre (Burnecki and Weron, 2004). Une application importante est la modélisation et la prévision de la volatilité, qui est cruciale pour la gestion des risques, la tarification des dérivés et l'optimisation de portefeuille. Les travaux de (Fernández-Martínez et al., 2013) montrent empiriquement que les algorithmes, basés sur une approche géométrique (algorithmes GM), sont plus pré-

cis que les algorithmes classiques, notamment avec des séries chronologiques de courte durée. Les auteurs ont vérifié que les algorithmes GM sont efficaces lorsqu'ils travaillent avec des mouvements Browniens (fractionnaires). En particulier, ils prouvent théoriquement que les algorithmes GM sont également valides pour explorer la mémoire longue dans les mouvements (fractionnaires) Lévy stables. Dans (Lavancier et al., 2009), les processus auto-similaires opératoires (os-s) sont étudiés. Dans cet article, un processus stochastique p-varié $X = \{X(t) = (X_1(t), \ldots, X_p(t)), t \in \mathbb{R}\}$ est dit auto-similaire opératoire (os-s) s'il existe une matrice $p \times p$ H (appelée l'exposant de X) telle que pour tout $\lambda > 0$, l'égalité suivante des distributions de dimension finie (fdd) soit vérifiée :

$$X(\lambda t) \overset{fdd}{=} \lambda^H X(t), \tag{4.250}$$

et la matrice $p \times p$ λ^H est définie par la série de puissances $\lambda^H = e^{H \log(\lambda)} = \sum_{k=0}^{\infty} \frac{H^k (\log(\lambda))^k}{k!}$. De plus, un processus aléatoire $X = \{X(t), t \in \mathbb{R}\}$ a des incréments stationnaires (si) si :

$$\{X(t+T) - X(T), t \in \mathbb{R}\} = \{X(t) - X(0), t \in \mathbb{R}\}, \text{ pour tout } T \in \mathbb{R}. \tag{4.251}$$

Les auteurs étudient un processus os-s gaussien avec incréments stationnaires appelé mouvement brownien fractionnaire opératoire (ofBm). Ils expliquent que pour $p = 1$, la classe de ofBm coïncide avec la classe fondamentale des mouvements browniens fractionnaires (fBm) comme prouvé dans (Samorodnitsky and Taqqu, 1994). Ils rappellent qu'un fBm avec un exposant $H \in (0, 1)$ peut être défini alternativement comme un processus gaussien stochastiquement continu $X = \{X(t), t \in \mathbb{R}\}$ à espérance nulle et covariance :

$$\mathbb{E}\left[X(s)X(t)\right] = \frac{\sigma^2}{2}\left(|s|^{2H} + |t|^{2H} - |t-s|^{2H}\right), \quad t, s \in \mathbb{R}, \tag{4.252}$$

où $\sigma^2 = \mathbb{E}\left[X^2(1)\right]$. Selon leurs travaux, la forme de la covariance d'un ofBm général semble inconnue et peut être assez compliquée. La structure du ofBm et les représentations intégrales stochastiques sont étudiées dans (Didier and Pipiras, 2008), (Didier and Pipiras, 2012), et (Didier and Pipiras, 2011). Un cas particulier de processus os-s correspond à la matrice diagonale $H = diag(H_1, \ldots, H_p)$. Dans ce cas, l'Équation (**??**) devient :

$$\left(X_1(\lambda t), \ldots, X_p(\lambda t)\right) \overset{fdd}{=} \left(\lambda^{H_1} X_1(t), \ldots, \lambda^{H_p} X_p(t)\right). \tag{4.253}$$

Un processus p-varié X satisfaisant l'Équation (**??**) pour tout $\lambda > 0$ est appelé auto-similaire vectoriel (vs-s) et un processus gaussien vs-s stochastiquement continu avec incréments stationnaires (si) est appelé mouvement brownien fractionnaire vectoriel (vfBm). À partir de l'Équation (**??**), chaque composant $X_i = \{X_i(t), t \in \mathbb{R}\}$, $i = 1, \ldots, p$ d'un processus vs-s est un processus (scalaire) auto-similaire, ce qui n'est pas vrai pour les processus os-s généraux. Dans (Lavancier et al., 2009), une forme générale de la fonction de (croisement-)covariance d'un processus vs-s si X avec variance finie et exposant $H = diag(H_1, \ldots, H_p)$, $0 < H_i < 1$, est obtenue. En d'autres termes, l'article obtient la forme générale de la fonction de croisement-covariance du mouvement

brownien fractionnaire vectoriel avec des composants corrélés ayant des auto-similarités différentes, appelée fonction de croisement-covariance.

Soit $X = \{X(t), t \in \mathbb{R}\}$ un processus du 2nd ordre avec des valeurs dans $\mathbb{R}^p$. Supposons que X a des incréments stationnaires, une espérance nulle, $X(0) = 0$, et que X est auto-similaire vectoriel avec un exposant $H = diag(H_1, \ldots, H_p)$, $0 < H_i < 1$ $(i = 1, \ldots, p)$. De plus, supposons aussi que pour tout $i, j = 1, \ldots, p$, la fonction $t \mapsto \mathbb{E}\left[X_i(t)X_j(1)\right]$ est continûment différentiable sur $(0, 1) \cup (1, \infty)$. Soit $\sigma_i^2 > 0$ la variance de $X_i(1), i = 1, \ldots, p$.

(i) Si $i = j$, alors pour tout $(s, t) \in \mathbb{R}^2$, nous avons :

$$\mathbb{E}\left[X_i(s)X_i(t)\right] = \frac{\sigma_i^2}{2}\{|s|^{2H_i} + \left|t\right|^{2H_i} - \left|t-s\right|^{2H_i}\}. \tag{4.254}$$

(ii) Si $i \neq j$ et $H_i + H_j \neq 1$, alors sous certaines conditions de régularité, il existe $c_{ij}, c_{ji} \in \mathbb{R}$ tels que pour tout $(s, t) \in \mathbb{R}^2$:

$$\text{Cov}\left(X_i(s), X_j(t)\right) = \frac{\sigma_i\sigma_j}{2}\{c_{ij}(s)\,|s|^{H_i+H_j} + c_{ji}(t)\,\left|t\right|^{H_i+H_j} - c_{ij}(t-s)\,\left|t-s\right|^{H_i+H_j}\} \tag{4.255}$$

où $\sigma_i^2 := \mathbb{V}[X_i(1)]$ et :

$$c_{ij}(t) = \begin{cases} c_{ij} & \text{si } t > 0 \\ c_{ji} & \text{si } t \leq 0. \end{cases} \tag{4.256}$$

(iii) Si $i \neq j$ et $H_i + H_j = 1$, alors il existe $d_{ij}, f_{ij} \in \mathbb{R}$ tels que pour tout $(s, t) \in \mathbb{R}^2$, nous avons :

$$\mathbb{E}\left[X_i(s)X_j(t)\right] = \frac{\sigma_i\sigma_j}{2}\{d_{ij}(|s| + \left|t\right| - \left|s-t\right|) + f_{ij}(t\log\left|t\right| - s\log|s|) - (t-s)\log\left|t-s\right|\}. \tag{4.257}$$

(iv) La matrice $R = (R_{ij})_{i,j=1,\ldots,p}$ est définie positive, où :

$$R_{ij} := \begin{cases} 1 & \text{si } i = j, \\ c_{ij} + c_{ji} & \text{si } i \neq j, H_i + H_j \neq 1, \\ d_{ij} & \text{si } i \neq j, H_i + H_j = 1. \end{cases} \tag{4.258}$$

Les auteurs fournissent également une représentation stochastique du vfBm et la fonction de covariance du vfBm. La représentation stochastique du vfBm est basée sur (Didier and Pipiras, 2008) :

$$X(t) = \int_{\mathbb{R}} \left\{ \left((t-x)_+^{H-\frac{1}{2}} - (-x)_+^{H-\frac{1}{2}} \right) A_+ + \left((t-x)_-^{H-\frac{1}{2}} - (-x)_-^{H-\frac{1}{2}} \right) A_- \right\} W(\mathrm{d}x), \tag{4.259}$$

où $H - \frac{1}{2} := diag\left(H_1 - \frac{1}{2}, \ldots, H_p - \frac{1}{2}\right)$, $x_+ := \max(x, 0)$, $x_- := \max(-x, 0)$, A_+, A_- sont des matrices $p \times p$ réelles et $W(\mathrm{d}x) = \left(W_1(\mathrm{d}x), \ldots, W_p(\mathrm{d}x)\right)$ est un bruit blanc gaussien à espérance nulle, composants indépendants et covariance $\mathbb{E}\left[W_i(\mathrm{d}x)W_j(\mathrm{d}x)\right] =$

$\delta_{ij}\mathrm{d}x$. La fonction de covariance du vfBm $X = \{X(t), t \in \mathbb{R}\}$ donnée par la représentation intégrale stochastique bilatérale est fournie. Soit $a_{ij}^{++} := \sum_{k=1}^{p} a_{ik}^{+} a_{jk}^{+}$, $a_{ij}^{--} := \sum_{k=1}^{p} a_{ik}^{-} a_{jk}^{-}$, $a_{ij}^{+-} := \sum_{k=1}^{p} a_{ik}^{+} a_{jk}^{-}$, $a_{ij}^{-+} := \sum_{k=1}^{p} a_{ik}^{-} a_{jk}^{+}$, où $A_+ = (a_{ij}^{+})$, $A_- = (a_{ij}^{-})$ sont les matrices $p \times p$. Clairement, $A_+ A_+^{\star} = (a_{ij}^{++})$, $A_- A_-^{\star} = (a_{ij}^{--})$, $A_+ A_-^{\star} = (a_{ij}^{+-})$, $A_- A_+^{\star} = (a_{ij}^{-+})$.

Les auteurs montrent que la covariance du processus défini par l'intégrale stochastique bilatérale satisfait les propriétés suivantes :

(i) Pour tout $i = 1, \ldots, p$, la variance de $X_i(1)$ est :

$$\sigma_i^2 = \frac{B\left(H_i + \frac{1}{2}\right), H_i + \frac{1}{2}}{sin(H_i\pi)} \{a_{ii}^{++} + a_{ii}^{--} - 2sin(H_i\pi)a_{ii}^{+-}\}. \tag{4.260}$$

(ii) Si $H_i + H_j \neq 1$, alors pour tout $s, t \in \mathbb{R}$, la croisement-covariance $\mathbb{E}\left[X_i(s)X_j(t)\right]$ du processus défini par l'Équation (**??**) est donnée par l'Équation (**??**) avec :

$$\frac{\sigma_i\sigma_j}{2} c_{ij} := \frac{B\left(H_i + \frac{1}{2}, H_j + \frac{1}{2}\right)}{sin(\left(H_i + H_j\right)\pi)} \left\{ a_{ij}^{++} cos(H_i\pi) + a_{ij}^{--} cos(H_j\pi) - a_{ij}^{+-} sin(\left(H_i + H_j\right)\pi) \right\}. \tag{4.261}$$

(iii) Si $H_i + H_j = 1$, alors pour tout $s, t \in \mathbb{R}$, la croisement-covariance $\mathbb{E}\left[X_i(s)X_j(t)\right]$ du processus défini par l'Équation (**??**) est donnée par l'Équation (**??**) avec :

$$\sigma_i\sigma_j d_{ij} := B\left(H_i + \frac{1}{2}, H_j + \frac{1}{2}\right) \times \left\{ \frac{sin(H_i\pi) + sin(H_j\pi)}{2} \left(a_{ij}^{++} + a_{ij}^{--}\right) - a_{ij}^{+-} - a_{ij}^{-+} \right\} \tag{4.262}$$

$$\sigma_i\sigma_j f_{ij} := (H_j - H_i)(a_{ij}^{++} - a_{ij}^{--}). \tag{4.263}$$

Dans (Coeurjolly, 2000b) et (Amblard and Coeurjolly, 2011b), les auteurs proposent une approche pour l'identification du mouvement brownien fractionnaire multivarié. Ils utilisent une technique de filtrage similaire aux ondelettes. Les paramètres de Hurst, les variances, les coefficients de corrélation et d'asymétrie sont estimés par régression sur les coefficients empiriques de log-variances et de log-correlations. Ils montrent que l'estimateur converge presque sûrement et satisfait un théorème central de la limite. La convergence est illustrée par des simulations, et ils appliquent la procédure d'estimation à des données financières. L'analyse de l'irrégularité des données modélisées par un fBm, l'étude de leur comportement spectral, ainsi que tout problème de prévision basé sur le fBm, impliquent la nécessité d'estimer le paramètre de Hurst. Dans (Coeurjolly, 2000b), les auteurs décrivent les principales méthodes paramétriques pour estimer le paramètre d'auto-similarité H. Ils distinguent quatre approches : les méthodes spectrales (log-périodogramme, une variante de la méthode de Lobato et Robinson), le maximum de vraisemblance (estimateur de Whittle), les méthodes à l'échelle de temps (décomposition en ondelettes du fBm), les méthodes temporelles (nombre de croisements de niveaux, variations discrètes). L'article (Coeurjolly, 2001) développe une classe

d'estimateurs cohérents des paramètres d'un mouvement brownien fractionnaire basés sur le comportement asymptotique du k-ième moment absolu des variations discrètes de ses trajectoires échantillonnées sur une grille discrète de l'intervalle [0, 1]. Les auteurs établissent des taux de convergence explicites pour ces types d'estimateurs, valables sur l'ensemble de l'intervalle $0 < H < 1$ du paramètre d'auto-similarité. Ils établissent également la normalité asymptotique de leurs estimateurs. L'efficacité de leur procédure est étudiée dans une étude de simulation. Dans (Coeurjolly et al., 2013), les auteurs étudient le mouvement brownien fractionnaire multivarié (mfBm) à travers la lentille de la transformée en ondelettes. Ils calculent la structure de corrélation de la transformée en ondelettes du mfBm. Ils étudient le comportement asymptotique de la corrélation, montrant que si l'ondelette d'analyse possède un nombre suffisant de moments d'ordre un nuls, la décomposition élimine toute dépendance (inter)-longue portée possible. La densité croisée spectrale est également considérée. Son existence est prouvée et son évaluation est réalisée à l'aide d'une représentation de type von Bahr-Essen de la fonction $sign(t)\left|t\right|^{\alpha}$. Le comportement de la densité croisée spectrale du champ d'ondelettes à la fréquence nulle est également développé et confirme les résultats fournis par l'analyse asymptotique de la corrélation. En (Jean-franÇois Coeurjolly and Vidakovic, 2014), les auteurs discutent de l'estimation d'un paramètre d'échelle σ^2 lorsque l'exposant de Hurst est connu. Pour estimer σ^2, ils proposent trois approches basées respectivement sur l'estimation du maximum de vraisemblance, l'ajustement des moments et les inégalités de concentration, tout en discutant des caractéristiques théoriques des estimateurs et des lignes directrices de filtrage optimal. Ils justifient l'amélioration de l'estimation de σ^2 lorsque le paramètre de Hurst est connu. À l'aide des trois approches et d'une méthodologie de bootstrap paramétrique dans une étude de simulation, ils comparent les intervalles de confiance de σ^2 en termes de longueurs, de taux de couverture, de complexité computationnelle et discutent des attributs empiriques des approches testées. Ils concluent que l'approche basée sur l'estimation du maximum de vraisemblance est optimale en termes d'efficacité et de précision, mais qu'elle est également coûteuse en termes de calcul. L'approche d'ajustement des moments s'est révélée non seulement efficace et précise de manière comparable, mais aussi rapide sur le plan computationnel et robuste aux écarts par rapport au modèle du mouvement brownien fractionnaire. Dans (Coeurjolly, 2005), les auteurs introduisent une nouvelle classe d'estimateurs consistants de la dimension fractale des processus gaussiens localement auto-similaires. Ces estimateurs sont basés sur des combinaisons linéaires de quantiles empiriques (statistiques L) des variations discrètes d'une trajectoire échantillonnée sur une grille discrète de l'intervalle [0, 1]. Ils démontrent la convergence presque sûre de ces estimateurs et prouvent leur normalité asymptotique. L'ingrédient clé est une représentation de Bahadur pour les quantiles empiriques de fonctions non linéaires de séquences gaussiennes avec une fonction de corrélation décroissant hyperboliquement. Dans (Coeurjolly et al., 2010a), les auteurs étudient la structure de covariance du bruit gaussien fractionnaire multivarié. Ils évaluent plusieurs paramètres du modèle permettant de contrôler la structure de corrélation à un décalage nul entre toutes les composantes du processus multivarié. Ensuite, ils spécifient un algorithme permettant la simulation exacte du bruit gaussien fractionnaire multivarié et donc des mouvements browniens fractionnaires. Les illustrations impliquent l'estimation des exposants de Hurst de chacune des composantes.

Les travaux de (Coeurjolly et al., 2010b) sont consacrés à l'étude de certaines propriétés d'une extension du célèbre mouvement brownien fractionnaire au cas multivarié. Ils étudient la structure de covariance du bruit gaussien fractionnaire multivarié et évaluent plusieurs paramètres du modèle permettant de contrôler la structure de corrélation à un décalage nul entre toutes les composantes du processus multivarié. Ils se concentrent particulièrement sur deux cas pour lesquels ils peuvent relier les paramètres caractéristiques de la fonction de covariance aux paramètres de la représentation stochastique des processus. Ces cas comprennent le cas causal, une généralisation multivariée directe de la représentation de (Mandelbrot and Van Ness, 1968b), et le cas bien équilibré qui ajoute au précédent cas le filtrage anti-causal d'un mouvement brownien. La caractérisation de la fonction de covariance est ensuite utilisée pour étudier le bruit gaussien fractionnaire multivarié, défini comme le processus d'incrément du mouvement brownien fractionnaire multivarié. Les auteurs étudient la structure de covariance ainsi que la structure spectrale de ce processus stationnaire multivarié. Ils exposent le fait intrigant que deux bruits gaussiens fractionnaires peuvent être interdépendants à longue portée alors que seulement l'un d'entre eux est dépendant à longue portée. Ensuite, ils effectuent une analyse en ondelettes du mouvement brownien fractionnaire multivarié et montrent que cette analyse en ondelettes peut détruire l'interdépendance à longue portée si l'ondelette est choisie de manière appropriée. Dans (Garcin, 2019), les auteurs étudient la transformée inverse de Lamperti d'un mouvement brownien fractionnaire et ses propriétés. La transformée inverse de Lamperti d'un mbf est un processus stationnaire. Ils déterminent l'exposant de Hurst empirique d'un tel processus composite à l'aide d'une régression des moments absolus logarithmiques de ses incréments, à diverses échelles, sur les échelles correspondantes. Cet exposant de Hurst perçu sous-estime l'exposant de Hurst du mbf sous-jacent. Ils rencontrent des séries chronologiques ayant un exposant de Hurst perçu inférieur à $\frac{1}{2}$, mais un exposant de Hurst sous-jacent supérieur à $\frac{1}{2}$. Cela ouvre la voie à la prévision à court et moyen terme. En effet, dans de telles séries, la réversion à la moyenne prédomine à grande échelle, tandis que la persistance est prépondérante à des échelles plus petites. Ils proposent une manière de caractériser l'horizon de Hurst, à savoir une échelle limite entre ces comportements opposés. Ils montrent que le mbf délamperisé, qui mélange persistance et réversion à la moyenne, est pertinent pour les séries chronologiques financières, en particulier pour les taux de change à haute fréquence. Dans leur échantillon, l'horizon de Hurst empirique est toujours supérieur à 1 heure et 23 minutes. Les travaux de (Laha and Rohatgi, 1981) introduisent les processus stochastiques d'opérateurs autosimilaires prenant des valeurs dans un espace euclidien de dimension finie et étudient certaines de leurs propriétés. Dans (Hudson and Mason, 1982), une représentation générale pour un processus d'opérateur autosimilaire est obtenue et sa classe d'exposants est caractérisée. Il est montré qu'un tel processus est la limite, dans un certain sens, d'un processus normé par l'opérateur et que toute limite d'un processus normé par l'opérateur est d'opérateur autosimilaire. Les travaux de (Sato, 1991) proposent une étude des processus autosimilaires et d'opérateurs autosimilaires avec des incréments indépendants. Ils démontrent que sous certaines conditions supplémentaires de continuité sur le processus d'opérateur autosimilaire, l'opérateur peut être choisi comme une fonction de puissance avec un exposant présentant certaines propriétés spectrales spéciales. Dans (Maejima

and Mason, 1994), les processus d'opérateurs autosimilaires sont étudiés et plusieurs exemples de processus d'opérateurs autosimilaires et stables (au sens ordinaire ou au sens de l'opérateur) sont construits. Des théorèmes limites pour de tels processus sont également démontrés. Dans (Marinucci and Robinson, 2000), une convergence faible vers une forme de mouvement brownien fractionnaire est établie pour une large classe de processus fractionnellement intégrés non stationnaires. Une extension du mouvement brownien fractionnaire classique, appelée mouvement brownien multifractionnaire (mBm), dont l'exposant de Hurst dépend du temps, est étudiée dans (Stoev and Taqqu, 2006). Les processus de mouvement brownien multifractionnaire (mBm) sont des processus gaussiens localement autosimilaires. Dans la littérature, deux types de processus mBm ont été introduits en utilisant respectivement les représentations intégrales de domaine temporel et de domaine fréquentiel du mbf. Dans cet article, les auteurs montrent que ces deux types de processus ont des structures de corrélation différentes lorsque la fonction $H(t)$ n'est pas constante. Ils se concentrent sur une classe de processus mBm paramétrée par $(a^+, a^-) \in \mathbb{R}^2$, qui contient les deux types de processus précédemment introduits comme cas spéciaux. Ils établissent la connexion entre leurs représentations intégrales de domaine temporel et de domaine fréquentiel et obtiennent des expressions explicites pour leurs covariances. Ils démontrent qu'il existe des fonctions non constantes $H(t)$ pour lesquelles la structure de corrélation des processus mBm dépend non trivialement de la valeur de (a^+, a^-) et donc, même pour une fonction $H(t)$ donnée, il existe un nombre infini de processus mBm avec des distributions essentiellement différentes.

Modélisation de la volatilité : La volatilité (l'écart-type des rendements d'actifs), n'est pas constante mais présente un regroupement (cluster) et une persistance dans le temps, connus sous le nom de regroupement de volatilité. Les processus Gaussiens auto-similaires comme le fBm peuvent capturer efficacement cette caractéristique. En modélisant la volatilité sous la forme d'un mouvement Brownien fractionnaire, les analystes peuvent incorporer une dépendance à long terme et une mémoire dans la dynamique de la volatilité, ce qui est observé dans les données réelles de séries chronologiques financières. Les grands changements de prix ont tendance à se regrouper, ce qui entraîne une persistance des amplitudes des variations de prix. Plusieurs travaux ont été publiés sur ce sujet, comme les travaux de (Cheong, 2010) ou (Cont, 2007) qui expliquent l'origine de ce regroupement de volatilité, en proposant des modèles basés sur des agents qui permettent de passer de régimes d'activité faible à des régimes d'activité élevée avec des durées de régime à queue lourde. Les travaux de (Gatheral et al., 2014) portent sur l'estimation de la volatilité à partir de données récentes à haute fréquence en revisitant la question de la régularité du processus de volatilité. Dans ces travaux, l'hypothèse est la suivante, la log-volatilité se comporte essentiellement comme un mouvement Brownien fractionnaire avec un exposant de Hurst H d'ordre $0,1$, à n'importe quelle échelle de temps raisonnable. Une telle hypothèse permet d'obtenir des prévisions améliorées de la volatilité réalisée. Et un modèle de microstructure de marché quantitatif reliant la rugosité de la volatilité au trading à haute fréquence et à la division des ordres est fourni. Dans (Rostek, 2012), la dynamique du titre sous-jacent est pilotée par un processus de diffusion à sauts où la partie diffusion est un mouvement Brownien fractionnaire tandis que les sauts présentent une distribution double exponen-

tielle. Dans (Chong et al., 2022a), (Chong et al., 2022b) et (Szymanski and Takabatake, 2023), des méthodes d'estimation de l'exposant de Hurst ainsi que des taux de convergence sont fournis. La propriété de stationnarité qui garantit l'invariance par translation du processus permet d'obtenir une stabilité des estimations basées sur les incréments. De plus, la propriété d'auto-similarité qui établit l'invariance par changement d'échelle de temps du processus assure que les estimations sont les mêmes quelle que soit l'échelle de temps choisie.

Tarification des options : Les modèles de tarification des options supposent souvent un certain processus stochastique pour la dynamique des prix des actifs sous-jacents. Comme la volatilité influence significativement les prix des options, la modélisation précise de la volatilité est cruciale. Les processus Gaussiens auto-similaires peuvent être utilisés pour modéliser les processus de volatilité, qui sont à leur tour utilisés dans des modèles de tarification des options tels que le modèle d'Heston ou le modèle de volatilité stochastique. Les travaux de (Rostek, 2009), (Rostek and Rostek, 2009) utilisent des mouvements Browniens fractionnaires qui sont des processus Gaussiens auto-similaires et stationnaires dans le but de tarifer des options. Dans ce livre, les questions suivantes sont abordées : Dans quelle mesure peut-on établir des parallèles entre le mouvement Brownien fractionnaire et le mouvement Brownien classique ? Plus précisément, comme le mouvement Brownien fractionnaire est une extension du mouvement Brownien, est-il possible d'étendre la théorie de la tarification des options ? Les techniques bien développées du calcul stochastique sont-elles transférables au mouvement Brownien fractionnaire ? Serons-nous confrontés à des problèmes conceptuels ? Pouvons-nous obtenir des solutions analytiques ? Dans (Rostek and Schöbel, 2006), sous l'hypothèse que le marché est piloté par un mouvement Brownien fractionnaire, des formules pour les options européennes fractionnaires sont dérivées en utilisant l'idée traditionnelle de l'espérance conditionnelle. Dans les travaux de (Rostek and Schoebel, 2010), les prix des options européennes sont dérivés lorsque la dynamique du titre sous-jacent est pilotée par un mouvement Brownien fractionnaire géométrique. Grâce à la propriété d'auto-similarité du mouvement Brownien fractionnaire, et grâce à la corrélation entre les incréments permise par un exposant de Hurst différent de 0.5, ce dernier est une manière parcimonieuse de capturer la corrélation au sein des séries chronologiques financières. Ils discutent d'un modèle où les participants au marché ont une aversion au risque relativement constante et échangent selon un processus de temps discret. La richesse de l'investisseur et l'action sous-jacente sont supposés être de type fBm et suivent une distribution log-normale bivariée. Ils introduisent une condition d'équilibre et fournissent des solutions analytiques pour les options européennes. Les résultats présentés sont des extensions des formules de tarification de Black-Scholes et contiennent ces dernières comme cas spécifiques.

Optimisation de portefeuille : La théorie moderne de portefeuille repose sur une estimation précise des rendements et des volatilités des actifs. Les processus Gaussiens auto-similaires peuvent améliorer l'estimation de ces paramètres en capturant leurs caractéristiques intrinsèques, telles que la mémoire longue et l'auto-similarité. Cela

conduit à des stratégies d'optimisation de portefeuille plus robustes qui tiennent compte de la nature non linéaire et non Gaussienne des marchés financiers. Voir par exemple (Papenbrock, 2011), (Zlatniczki and Telcs, 2024), (Galloway and Nolder, 2008), (Lunga, 2006).

Les travaux de (Czichowsky et al., 2018), et (Czichowsky and Schachermayer, 2017) proposent une conciliation de deux concepts conflictuels en finance : d'une part, la notion d'absence d'arbitrage, et d'autre part, la prise en compte de processus de prix non semi-martingales, comme le mouvement Brownien fractionnaire. En imposant des coûts de transaction (proportionnels) arbitrairement petits et en considérant des optimiseurs d'utilité logarithmique, l'existence d'un processus de prix fictif semi-martingale et sans friction pour un marché financier exponentiel avec un mouvement Brownien fractionnaire est prouvée. Dans (Jumarie, 2005), le modèle de portefeuille optimal proposé initialement par Merton est considéré avec l'hypothèse supplémentaire que les bruits impliqués dans la dynamique de la richesse sont des mouvements Browniens fractionnaires (au sens de la dérivée fractionnaire des bruits blancs Gaussiens) avec une dépendance à court-terme, c'est-à-dire avec un paramètre de Hurst inférieur à $1/2$. Les travaux de (Sarol et al., 2007) considèrent le problème classique de Merton consistant à trouver le taux de consommation optimal et le portefeuille optimal dans un marché piloté par un mouvement Brownien fractionnaire avec un paramètre de Hurst $H > 1/2$. Les intégrales par rapport au fBm sont dans le sens de Skorohod, et non au sens trajectoriel, ce qui est connu pour conduire à l'arbitrage. Une forme explicite est dérivée pour le taux de consommation optimal et le portefeuille optimal dans un tel marché pour un agent avec des fonctions d'utilité logarithmiques. Un portefeuille autofinancé réel s'avère conduire à un terme de consommation toujours favorable à l'investisseur. Dans (HU et al., 2003), les auteurs présentent un modèle mathématique pour un marché piloté par un mouvement Brownien fractionnaire avec un paramètre de Hurst. L'interprétation des intégrales par rapport au fBm est dans le sens d'Itô (Skorohod-Wick), et non au sens trajectoriel (ce qui est connu pour conduire à l'arbitrage). Ils trouvent explicitement le taux de consommation optimal et le portefeuille optimal dans un tel marché pour un agent avec des fonctions d'utilité de type puissance. Dans (Garcin, 2022), les log-prix suivent un fBm, la nature non Markovienne du fBm est utilisée pour prévoir les états futurs du processus et réaliser des arbitrages statistiques. Certaines questions sur l'optimisation des stratégies de trading dans le cadre du fBm sont abordées. Quels sont les incréments décalés du fBm, observés à temps discret, à prendre en compte ? Si l'incrément prédit est proche de zéro, jusqu'à quel seuil est-il plus rentable de ne pas investir ? Dans (Bauerle and Desmettre, 2020), une version fractionnaire du modèle de volatilité de Heston est considérée. Dans ce modèle, des problèmes d'optimisation de portefeuille pour des fonctions d'utilité de puissance sont traités.

Trading haute fréquence : Dans le Trading Haute Fréquence (HFT), où les décisions sont prises en quelques millisecondes, comprendre et prédire la dynamique du marché est crucial. Les processus Gaussiens auto-similaires peuvent fournir des perpectives sur les comportements à court terme du marché en capturant les motifs complexes et les corrélations présentes dans les données haute fréquence. Voir par exemple (Evertsz,

1995), (Arroum, 2007), (Smith, 2010).

Dans la limite haute fréquence, l'espérance conditionnelle des accroissements du mouvement Brownien fractionnaire converge vers un bruit blanc, perdant ainsi sa dépendance vis-à-vis de l'historique de la trajectoire et de l'horizon de prévision, rendant les problèmes d'optimisation dynamique traitables. Les travaux de (Guasoni et al., 2021) proposent une formule explicite pour les stratégies optimales de moyenne-variance locale et leur performance pour un prix d'actif qui suit un mouvement Brownien fractionnaire. Les travaux de (Guasoni et al., 2019) considèrent un marché avec un prix d'actif décrit par un mouvement Brownien fractionnaire, qui peut être échangé avec un impact temporaire non linéaire sur les prix. Ils trouvent des stratégies asymptotiquement optimales pour la maximisation de la richesse finale espérée. En exploitant l'autocorrélation des accroissements tout en limitant les coûts de transaction, ces stratégies génèrent une richesse finale moyenne qui croît selon une fonction puissance de l'horizon temporel, l'exposant dépendant à la fois des paramètres de Hurst et de l'impact sur les prix. Les ratios de Sharpe résultants sont bornés, insensibles à l'horizon de temps et asymétriques par rapport à l'exposant de Hurst. Ces résultats s'étendent aux processus Gaussiens à mémoire longue et à une classe de processus auto-similaires. Les travaux de (Lim and Muniandy, 2002) étudient certains modèles Gaussiens pour la diffusion anormale, qui comprennent le mouvement Brownien temporellement rescalé, deux types de mouvement Brownien fractionnaires, et des modèles associés au mouvement Brownien fractionnaire basés sur l'équation de Langevin généralisée. Les processus Gaussiens associés à ces modèles satisfont à la relation de diffusion anormale qui exige que le déplacement quadratique moyen varie selon la fonction puissance $t^{\alpha}, ; 0 < \alpha < 2$. Cependant, ces processus ont des propriétés différentes, ce qui indique que la relation de diffusion anormale avec un seul paramètre est insuffisante pour caractériser le mécanisme sous-jacent. Bien que les deux versions du mouvement Brownien fractionnaire et le mouvement Brownien temporellement rescalé aient tous la même fonction de densité de probabilité, le théorème de Slepian peut être utilisé pour comparer leurs distributions de temps de premier passage, qui sont différentes. Enfin, pour modéliser la diffusion anormale avec un exposant variable $\alpha(t)$, il est nécessaire de considérer les extensions multifractionnaires de ces processus Gaussiens.

Gestion des risques : Comprendre et gérer les risques est fondamental en finance. Les processus auto-similaires Gaussiens permettent une évaluation des risques plus précise en capturant la nature complexe et corrélée des séries chronologiques financières. En incorporant la dépendance à long terme et l'auto-similarité dans les modèles de risque, les institutions financières peuvent mieux estimer la Valeur-à-Risque (**VaR**) et la Valeur-à-Risque Conditionnelle (CVaR), qui sont des mesures des pertes potentielles dans des conditions de marché défavorables. C'est le sujet de (Michna et al., 1998). En effet, dans cet article, les auteurs s'intéressent à la théorie du risque collectif. Ils observent que la théorie du risque collectif porte sur les fluctuations aléatoires des actifs totaux et de la réserve de risque d'une compagnie d'assurance. Ils considèrent des processus auto-similaires continus avec des accroissements stationnaires pour le modèle de renouvellement en théorie du risque. Ils construisent un modèle de risque qui montre

un mécanisme de dépendance à long terme des sinistres. Une approximation du processus de risque par un processus auto-similaire avec dérive est fournie. La probabilité de ruine dans un temps fini est estimée pour le mouvement Brownien fractionnaire avec dérive. Un modèle similaire est applicable dans les systèmes de file d'attente, décrivant une dépendance à long terme dans les processus de marche/arrêt et les modèles de fluide associés. Les travaux de (Wesselhöfft, 2021) sont basés sur le fait qu'un processus auto-similaire, capable de rendre compte du comportement à mémoire longue, est le mouvement Brownien fractionnaire, qui a une limite possible non Gaussienne par convolution des accroissements. Les accroissements du mouvement Brownien fractionnaire peuvent présenter une mémoire longue à travers un paramètre H, l'exposant de Hurst. Pour le mouvement Brownien fractionnaire, cet exposant de mise à l'échelle (Hurst) serait constant sur différents ordres de moments, étant unifractal. Mais empiriquement, nous observons des exposants de Hölder variables, le continuum des exposants de Hurst, ce qui implique un comportement multifractal. Les auteurs expliquent le comportement multifractal à travers les indices alpha-stables changeants des distributions alpha-stables sur les fréquences d'échantillonnage en appliquant des filtres pour la saisonnalité et la dépendance temporelle (mémoire longue) sur différentes fréquences d'échantillonnage, à partir de hautes fréquences jusqu'à une minute. En utilisant un filtre pour la mémoire longue, ils montrent que le processus à basse fréquence d'échantillonnage, ne contenant pas la composante de dépendance temporelle, peut être régi par le mouvement alpha-stable. Sous le mouvement alpha-stable, ils proposent une méthode semi-paramétrique appelée Méthodologie de Rééchantillonnage Fréquentiel (FRM), qui permet de rééchantillonner l'ensemble de données à haute fréquence filtré à la fréquence d'échantillonnage plus basse. Les ensembles de données, par exemple les données hebdomadaires obtenues en rééchantillonnant les données à haute fréquence avec le FRM, sont davantage à queue lourde que celles observées empiriquement. Les auteurs montrent qu'en utilisant un sous-ensemble de l'ensemble complet de données, le FRM suffit pour obtenir une meilleure prévision en termes de risque pour l'ensemble complet de données. En particulier, le FRM aurait pu rendre compte des événements extrêmes de la crise financière de 2008.

Différents types de processus auto-similaires : Le mouvement Brownien fractionnaire est largement utilisé dans la modélisation de phénomènes avec une densité spectrale de type loi de puissance. Cependant, le FBM a ses limites car il ne peut décrire que des phénomènes avec une structure monofractale ou un degré d'irrégularité uniforme caractérisé par l'exposant de Hölder constant. Pour une modélisation plus réaliste, il est nécessaire de prendre en compte la variation locale de l'irrégularité, l'exposant de Hölder étant autorisé à varier dans le temps ou l'espace. Dans (Muniandy and Lim, 2001), une extension du fBm standard au mouvement Brownien multifractionnaire (mBm) indexé par un exposant de Hölder qui est une fonction du temps est proposée. Cet article propose une généralisation alternative au mBm basée sur le fBm défini par le type d'intégrale fractionnaire de Riemann-Liouville. Les propriétés locales du RLMBM, Mouvement Brownien Multifractionnaire de Riemann-Liouville, sont étudiées et il est constaté qu'elles sont similaires à celles du MBM standard. Un schéma numérique pour simuler les trajectoires d'échantillon localement auto-similaires du RLMBM pour

différents types d'exposants de Hölder variant dans le temps est donné. Les exposants de mise à l'échelle locaux sont estimés en fonction de la croissance locale de la variance et des méthodes de scalogramme d'ondelettes. Enfin, un exemple des applications possibles du RLMBM dans la modélisation de séries temporelles multifractales est illustré.

Dans (Tudor, 2013), plusieurs processus auto-similaires sont étudiés : mouvements Browniens fractionnaires, bi-fractionnaires et sous-fractionnaires. Les travaux de (Pagnini et al., 2012) étudient l'approche de l'équation maîtresse pour modéliser la diffusion anormale. La diffusion anormale dans des milieux complexes peut être décrite comme le résultat d'un mécanisme de superposition reflétant les propriétés d'inhomogénéité et de non-stationnarité du milieu. Par exemple, lorsque cette superposition est appliquée au processus de diffusion temporelle fractionnaire, l'équation maîtresse résultante s'avère être l'équation gouvernante de la diffusion fractionnaire d'Erdélyi-Kober, qui décrit l'évolution de la distribution marginale du prétendu mouvement Brownien gris généralisé. Ce mouvement est une classe paramétrique de processus stochastiques qui fournit des modèles pour une diffusion anormale rapide et lente : il est composé de processus auto-similaires avec des accroissements stationnaires et dépend de deux paramètres réels. La classe comprend le mouvement Brownien fractionnaire, les processus stochastiques de diffusion temporelle fractionnaire et le mouvement Brownien standard. Dans ce cadre, la fonction M-Wright, également connue sous le nom de fonction de Mainardi, émerge comme une généralisation naturelle de la distribution Gaussienne, récupérant le même rôle clé que la densité Gaussienne pour les mouvements Browniens standards et fractionnaires.

Dans (Pang and Taqqu, 2019), l'attention est portée sur les processus de bruit de tir avec des arrivées de Poisson et des bruits non stationnaires. Les bruits sont conditionnellement indépendants étant donné les temps d'arrivée, mais la distribution de chaque bruit dépend de son temps d'arrivée. Les auteurs établissent des limites d'échelle pour de tels processus de bruit de tir dans deux situations : (a) les fonctions de variance conditionnelles des bruits ont une loi de puissance et (b) les distributions conditionnelles de bruit sont constantes par morceaux. Dans les deux cas, les processus limites sont Gaussiens auto-similaires avec des accroissements non stationnaires. Motivés par ces processus, ils introduisent de nouvelles classes de processus Gaussiens auto-similaires avec des accroissements non stationnaires, via la représentation intégrale en domaine temporel, qui sont des généralisations naturelles des mouvements Browniens fractionnaires.

Mouvement Brownien fractionnaire et gestion de l'arbitrage : Utiliser un mouvement Brownien géométrique fractionnaire pour décrire la dynamique des prix permet de gérer la dépendance à long terme (resp. à court terme), cependant, l'hypothèse fondamentale du marché d'absence d'opportunité d'arbitrage n'est plus satisfaite. C'est l'objet des travaux de (Cheridito, 2003). Les auteurs construisent des stratégies d'arbitrage pour un marché financier composé d'un compte de marché monétaire et d'une action dont le prix actualisé suit un mouvement Brownien fractionnaire avec dérive ou un mouvement Brownien fractionnaire exponentiel avec dérive. Ensuite, ils montrent comment l'arbitrage peut être exclu de ces modèles en restreignant la classe des stratégies de trading. Dans (Cheridito, 2001), les auteurs montrent que la somme d'un

mouvement Brownien et d'un multiple non trivial d'un mouvement Brownien fractionnaire indépendant avec un exposant de Hurst $H \in (0, 1]$ n'est pas une semi-martingale si $H \in (0, 1/2) \cup (1/2, 3/4]$, qu'elle est équivalente à un multiple d'un mouvement Brownien si $H = 1/2$ et qu'elle est équivalente au mouvement Brownien si $H \in (3/4, 1]$. Ils discutent du prix d'une option d'achat européenne sur un actif piloté par une combinaison linéaire d'un mouvement Brownien et d'un mouvement Brownien fractionnaire indépendant.

Intégrale stochastique par rapport au mouvement Brownien fractionnaire : Jusqu'à présent, l'intégrale stochastique était définie par rapport au mouvement Brownien standard. Certains auteurs ont proposé une extension de la notion d'intégrale stochastique par rapport au mouvement Brownien fractionnaire pour différentes plages de valeurs prises par l'exposant de Hurst. Dans (Cheridito and Nualart, 2005), l'intégrale stochastique par rapport au mouvement Brownien fractionnaire avec un exposant de Hurst $H \in (0, 1/2)$ est définie. Cela étend l'intégrale de divergence du calcul de Malliavin. Pour cette intégrale de divergence étendue, un théorème de Fubini est prouvé et des versions des formules d'Itô et de Tanaka sont établies pour tous les $H \in (0, 1/2)$. Les travaux de (Carmona et al., 2003) définissent une intégrale stochastique par rapport au mouvement Brownien fractionnaire pour chaque valeur de l'exposant de Hurst $H \in (0, 1)$. Cela est fait en approximant le mouvement Brownien fractionnaire par des semi-martingales. Ensuite, pour $H > 1/6$, ils établissent une formule d'Itô de changement de variables, qui est plus précise que la formule d'Itô de Privault (1998). Le calcul stochastique pour le mouvement Brownien fractionnaire est également traité dans (Biagini et al., 2008a). Les travaux de (Duncan et al., 2000) donnent un calcul stochastique pour les mouvements Browniens fractionnaires ayant l'exposant de Hurst dans $(1/2, 1)$. Une intégrale stochastique de type Itô est définie pour une famille d'intégrandes de sorte que l'intégrale ait une moyenne nulle et une expression explicite pour le deuxième moment. Cette intégrale utilise le produit de Wick et une dérivée dans l'espace des trajectoires. Quelques formules d'Itô (ou formules de changement de variables) sont données pour les fonctions lisses d'un mouvement Brownien fractionnaire ou certains processus liés à un mouvement Brownien fractionnaire. Une intégrale stochastique de type Stratonovich est définie et les deux types d'intégrales stochastiques sont explicitement liés. Une fonction de carré intégrable d'un mouvement Brownien fractionnaire est exprimée comme une série infinie d'intégrales multiples orthogonales. Le livre (Coutin, 2007) offre une introduction au calcul stochastique par rapport au mouvement Brownien fractionnaire. Différentes approches ont été introduites pour construire des intégrales stochastiques par rapport au fBm : techniques trajectorielles, calcul de Malliavin, approximation par des sommes de Riemann. Dans (Nualart, 2006), les auteurs décrivent ces méthodes et présentent les formules de changement de variable correspondantes. Les travaux de (Decreusefond, 2003), (Decreusefond and Üstünel, 1998), présentent un nouveau résultat théorique sur le mouvement Brownien fractionnaire, comprenant différentes définitions (et leurs relations) de l'intégrale stochastique par rapport à ce processus, le théorème de Girsanov, la formule de représentation de Clark, la formule d'Itô, etc. Une analyse stochastique du mouvement Brownien fractionnaire est également fournie dans (Lin, 1995). Les questions d'intégration liées au

mouvement Brownien fractionnaire sont abordées dans (Pipiras and Taqqu, 2000).

Dans l'ensemble, les processus auto-similaires Gaussiens avec des accroissements stationnaires offrent un cadre puissant pour modéliser divers aspects des marchés financiers, allant de la dynamique de la volatilité à la gestion des risques et aux stratégies de trading. L'invariance en temps et en espace par translation et par changement d'échelle de temps garantie par les propriétés de stationnarité et d'auto-similarité ainsi que la capacité des processus auto-similaires à capturer la dépendance à long terme ou à court terme en font des outils indispensables pour les analystes financiers et les chercheurs.

Dans nos travaux, nous nous intéressons à la formulation minimale des propriétés de stationnarité et d'autosimilarité du carré de la norme des fonctions aléatoires de l'espace $\mathbb{L}^2$. Nous démontrons que la combinaison de la formulation minimale de la stationnarité et de l'autosimilarité dans les espaces $\mathbb{L}^2$ suffit pour caractériser pleinement le noyau interne dépendant uniquement des fonctions puissances d'exposant $\gamma \in (0,1)$, et pour prouver la stationnarité et l'autosimilarité du noyau interne dans $\mathbb{L}^2$. Ensuite, nous introduisons l'hypothèse Gaussienne et obtenons les propriétés de stationnarité et d'autosimilarité non seulement en distribution mais aussi en trajectoire. Nous présentons quelques exemples de processus Gaussiens stationnaires et autosimilaires, appelés mouvement Brownien et mouvement Brownien fractionnaire. Dans cette partie, nous retrouvons les résultats présentés dans (Taqqu, 1994), (Mandelbrot and Van Ness, 1968a), (Das and Pan, 2011, Chap. 3), (Embrechts and Maejima, 2000), (Embrechts, 2009), (Samorodnitsky, 2006), (Chaumont, 2006), (Lamperti, 1962). Enfin, nous proposons une extension au cas du mouvement Brownien fractionnaire multivarié (mfBm) et nous retrouvons les résultats énoncés dans (**?**).

Contributions

Chapter 2 se concentre sur les effets qu'ont les transformations temporelles sur les familles de variables aléatoires et les processus aléatoires dans les espaces $\mathbb{L}^2$. L'objectif de ce chapitre est d'étudier les propriétés caractéristiques liées à ces transformations temporelles et leurs conséquences sur les familles de variables aléatoires et les processus aléatoires dans les espaces $\mathbb{L}^2$. Les deux transformations temporelles d'intérêt sont le changement d'origine temporelle ou la translation temporelle, et le changement d'échelle temporelle. Les propriétés caractéristiques associées sont respectivement la stationnarité et l'auto-similarité.

Dans Chapter 2, nous considérons les sous-espaces de $\mathbb{L}^2$ engendrés par des familles de variables aléatoires (v.a.) $\{X(\theta), \theta \in J\}$, notés $\mathcal{H}_X$. $\mathcal{H}_X$ est le sous-espace de Hilbert contenant toutes les combinaisons linéaires finies $\alpha \cdot X(\theta) = \sum_1^n \alpha_i X(\theta_i)$ et leurs limites dans $\mathbb{L}^2$. Il est caractérisé par les caractéristiques $\mathbb{L}^2$ de $\{X(\theta)\}$, données par la *norme quadratique* des composantes (c'est-à-dire le carré de la norme) $Q_X(\theta) := \mathbf{Q}(X(\theta)) = \|X(\theta)\|_2^2 = \mathbb{E}(|X(\theta)|^2)$ et les produits scalaires $K_X(\theta, \theta') = \langle\langle X(\theta), X(\theta') \rangle\rangle := \mathbb{E}(X(\theta)\, X(\theta'))$, indicateurs de dépendance.

Les familles aléatoires indexées par le temps réel $\{X(t), t \in \mathbb{R}\}$ (ou processus aléatoires) suscitent un intérêt particulier. Leurs caractéristiques $\mathbb{L}^2$, $Q_X(t)$ et $K_X(t,s)$, peuvent être complexes à calculer. Par conséquent, des propriétés supplémentaires sont introduites pour réduire leur complexité. Si les propriétés ne sont vraies que "composante par composante" (t par t), seule la famille des normes quadratiques $\{Q_X(t)\}$ est concernée, et nous parlons de la propriété Q_X.

Q_X et $\mathbb{L}^2$-Stationnarité et Auto-similarité des processus aléatoires La propriété de *stationnarité* est liée aux effets causés par le changement d'origine temporelle sur le processus, $\{X_h(t) = X(t+h) - X(h), (h \in \mathbb{R})\}$. La *stationnarité* dans $\mathbb{L}^2$ exprime le fait que les processus $\{X_h(t)\}$ et $\{X(t)\}$ ont les mêmes caractéristiques $\mathbb{L}^2$, c'est-à-dire la même norme quadratique $\mathbf{Q}(X(t+h) - X(h)) = Q_X(t)$ et le même produit scalaire $\langle\langle X(t+h) - X(h), X(s+h) - X(h)\rangle\rangle = K_X(t,s)$, tandis que la Q_X-*stationnarité* exprime seulement le fait que les variables aléatoires $X(t+h) - X(h)$ et $X(t)$ ont la même norme quadratique Q_X, c'est-à-dire $\mathbf{Q}(X(t+h) - X(h)) = Q_X(t)$. Heureusement, cette hypothèse plus faible est suffisante pour calculer le produit scalaire $K_X(s,t)$, et déduire la stationnarité dans $\mathbb{L}^2$.

Plus précisément, la stationnarité dans $\mathbb{L}^2$ exprime l'invariance par translation temporelle ou changement d'origine temporelle des caractéristiques $\mathbb{L}^2$ du processus.

Definition 4.36. (i) *Un processus $\{X(t)\}$ est dit $\mathbb{L}^2$-stationnaire si, pour tout $h \in \mathbb{R}$, les processus $\{X_h(t) = X(t+h) - X(h)\}$ et $\{X(t)\}$ ont les mêmes caractéristiques $\mathbb{L}^2$, c'est-à-dire même norme quadratique et même produit scalaire :*

$$\mathbf{Q}[X(t+h) - X(h)] = \mathbf{Q}[X(t)] = Q_X(t), \text{ et } \langle\langle X_h(t), X_h(s)\rangle\rangle = K_X(t,s). \quad (4.264)$$

(ii) *Le processus $\{X(t)\}$ est dit Q_X-stationnaire si seule la première condition est satisfaite :*

$$\mathbf{Q}(X_h(t)) = \mathbf{Q}(X(t)) = Q_X(t) \quad .$$

La propriété d'*auto-similarité* est liée à un changement d'échelle temporelle dans le processus tel que $\{X^\lambda(t) = \lambda^{-\frac{1}{2}} X(\lambda t), (\lambda > 0)\}$. La $\mathbb{L}^2$-*auto-similarité* suppose l'existence d'une fonction strictement positive $(\Theta(\lambda), \lambda > 0)$ telle que les processus $\{X(\lambda t)\}$ et $\{\Theta^{\frac{1}{2}}(\lambda) X(t)\}$ aient les mêmes caractéristiques $\mathbb{L}^2$, c'est-à-dire la même norme quadratique $Q_X(\lambda t) = \Theta(\lambda) Q_X(t)$ et le même produit scalaire $\langle\langle X(\lambda t), X(\lambda s)\rangle\rangle = \Theta(\lambda) K_X(t,s)$. La Q_X-auto-similarité suppose seulement que, pour tout t et $\lambda > 0$, $Q_X(\lambda t) = \Theta(\lambda) Q_X(t)$. Ainsi, ces deux concepts ne sont pas équivalents.

Definition 4.37. *Soit $\{\Theta(\lambda), \lambda > 0\}$ une fonction strictement positive et non-constante.* (i) *Un processus $\{X(t)\}$ est dit $\mathbb{L}^2$-auto-similaire de fonction d'auto-similarité Θ si, pour tout $\lambda > 0$, les processus $\{X(\lambda t)\}$ et $\{\Theta^{\frac{1}{2}}(\lambda) X(t)\}$ ont les mêmes caractéristiques $\mathbb{L}^2$, c'est-à-dire,*

$$Q_X(\lambda t) = \Theta(\lambda) Q_X(t) \quad et \quad \langle\langle X(\lambda t), X(\lambda s)\rangle\rangle = \Theta(\lambda) \langle\langle X(t), X(s)\rangle\rangle. \quad (4.265)$$

(ii) *$\{X(t)\}$ est dit Q_X-(Θ)-auto-similaire si seule la condition sur la norme quadratique $Q_X(\lambda t) = \Theta(\lambda) Q_X(t), \forall t$, est vraie.*

Plus précisément, la Q_X-auto-similarité établit une relation de proportionnalité spatiale entre les normes quadratiques évaluées à des instants temporels proportionnels $\{Q_X(\lambda t)\}$ et $\{Q_X(t)\}$, dont le facteur de proportionnalité spatiale est une fonction du facteur de proportionnalité temporelle λ, tel que $\Theta(\lambda) > 0$. Autrement dit, la $\mathbb{L}^2$-auto-similarité établit une relation de proportionnalité spatiale, à la fois entre les normes quadratiques $\{Q_X(\lambda t)\}$ et $\{Q_X(t)\}$, et entre les produits scalaires $\langle\langle X(\lambda t), X(\lambda s)\rangle\rangle$ et $\langle\langle X(t), X(s)\rangle\rangle$, évalués à des instants temporels proportionnels, dont le facteur de proportionnalité spatiale est une fonction du facteur de proportionnalité temporelle λ, tel que $\Theta(\lambda) > 0$.

Avec la seule hypothèse de Q_X-auto-similarité, la forme de la fonction Θ peut être déterminée, et la norme quadratique est caractérisée.

Proposition 4.38. *Une famille $\{X(t)\}$, avec une norme quadratique (rc) continue à droite $\{Q_X(t)\}$ est auto-similaire en norme quadratique* ***uniquement*** *si la fonction Q_X est une fonction puissance, avec un exposant positif γ, $Q_X(t) = |t|^{2\gamma} Q_X(1)$ si $t > 0$, $Q_X(t) = \big|t\big|^{2\gamma} Q_X(-1)$ si $t < 0$, et $Q_X(0) = 0$. En d'autres termes, pour tout $t \in \mathbb{R}$, $Q_X(t) = \big|t\big|^{2\gamma} Q_X(sgn(t))$ où $sgn(t) = 1$ si $t > 0$, $sgn(t) = -1$ si $t < 0$, et $sgn(t) = 0$ si $t = 0$.*

Il est observé que la propriété de Q_X-auto-similarité ne peut être étendue au produit scalaire sans une hypothèse supplémentaire.

Dans Chapter 2, nous nous plaçons à l'échelle des espaces de Hilbert et prouvons que, sans aucune hypothèse de distribution, les seules hypothèses de Q_X-stationnarité et de Q_X-auto-similarité sont suffisantes pour caractériser entièrement le noyau de covariance et montrer que ce noyau de covariance est lui-même stationnaire et auto-similaire, c'est-à-dire obtenir la $\mathbb{L}^2$-stationnarité et $\mathbb{L}^2$-auto-similarité.

La combinaison de la Q_X-auto-similarité et de la Q_X-stationnarité implique de restreindre l'exposant γ de la fonction puissance $Q_X(t)$ à l'intervalle $(0, 1)$, en raison de la sous-linéarité de la norme quadratique Q_X induite par la stationnarité.

Theorem 4.39. (i) *Une condition nécessaire et suffisante pour qu'un processus $\mathbb{L}^2$-rc $\{X(t)\}$ soit $\mathbb{L}^2$-auto-similaire et $\mathbb{L}^2$-stationnaire est l'existence d'une fonction puissance, avec un exposant $0 < \gamma < 1$ telle que les caractéristiques $\mathbb{L}^2$ soient :*

$$Q_X(t) = Q_X(-t) = Q_X(1)|t|^{2\gamma} \quad et \quad K_X(t,s) = \frac{Q_X(1)}{2}(|t|^{2\gamma} + |s|^{2\gamma} - |t-s|^{2\gamma}). \tag{4.266}$$

(ii) *La condition $\gamma = 1/2$ est équivalente à l'orthogonalité des incréments, définie sur les intervalles disjoints. Dans ce cas, $K_X(t,s) = Q_X(1)(|t| \wedge |s|)$.*

> Remarque : − Le produit scalaire est à la fois stationnaire d'après la Proposition **??**, et γ-auto-similaire, ce qui est évident dans l'Équation (4.266).

Stationnarité et auto-similarité dans l'espace de Hilbert Gaussien L'ajout de l'hypothèse Gaussienne aux propriétés de stationnarité et d'auto-similarité conduit à exprimer ces propriétés d'abord en distribution, puis en trajectoire.

Definition 4.40. *Le processus Gaussien* $\{X(t)\}$ *est* γ*-auto-similaire et stationnaire en distribution si et seulement si les deux processus Gaussiens* $\{X(t)\}$ *et* $\{X_h(t) = X(t+h) - X(t)\}, h \in \mathbb{R}$*, respectivement* $\{X(\lambda t)\}$ *et* $\{\lambda^\gamma X(t)\}, \lambda > 0$*, ont la même distribution Gaussienne, c'est-à-dire, la même moyenne et le même produit scalaire.*

Dans ce cadre, nous prouvons que les seules hypothèses de Q_X-stationnarité et de Q_X-auto-similarité permettent d'obtenir la stationnarité et l'auto-similarité en distribution.

Un processus Gaussien stationnaire et auto-similaire en distribution est évidemment un processus $\mathbb{L}^2$-γ-auto-similaire et stationnaire. Ainsi, tous les résultats précédents restent valables. En particulier, sa norme quadratique est la fonction puissance $Q_X(1)|t|^{2\gamma}$ avec $0 < \gamma < 1$, et le produit scalaire est entièrement spécifié. La seule propriété inconnue concerne la moyenne, qui, par la propriété de stationnarité, vérifie $m_X(t+h) = m_X(t) + m_X(h)$. Ainsi, la moyenne est une fonction linéaire $m_X(t) = m_X(1)t$. Cependant, d'après la propriété d'auto-similarité, $m_X(\lambda) = \lambda^\gamma m_X(1)$. Ces deux équations sont contradictoires si $\gamma \neq 1$. Or $\gamma \in (0,1)$, par conséquent, $m_X(t) = 0$.

Theorem 4.41. *Un processus Gaussien* $\{X(t)\}$ *est* γ*-auto-similaire et stationnaire en* **distribution** *si et seulement si* $\{X(t)\}$ *est un processus centré, et est* γ*-auto-similaire et stationnaire au sens de* $\mathbb{L}^2$*. La norme quadratique* $Q_X(t)$ *est la variance* $V_X(t) = V_X(1)|t|^{2\gamma}$*, avec* $\gamma \in (0,1)$*.*

(ii) *Si* $\gamma = \frac{1}{2}$*, le processus Gaussien auto-similaire et stationnaire* $\{X(t)\}$ *a des incréments indépendants et une variance linéaire* $V_B(t) = V_B(1)|t|$*. Ce processus est appelé mouvement Brownien ou processus de Wiener. Par la suite, il sera noté* $\{B(t)\}$*.*

(iii) *Pour tout* $\lambda > 0$ *et* $h \in \mathbb{R}$*, les processus* $\{B_h(t) = B(t+h) - B(h)\}$*,* $\{B^\lambda(t) = \lambda^{-1/2}B(\lambda t)\}$ *et* $\{B_h^\lambda(t) = \lambda^{-1/2}(B(\lambda(t+h)) - B(\lambda h))\}$ *sont également des mouvements Browniens.*

Quelques exemples bien connus Nous présentons quelques exemples bien connus de processus Gaussiens auto-similaires avec des accroissements stationnaires : le mouvement Brownien, dont les accroissements sont indépendants, et le mouvement Brownien dépendant du chemin (PDBM), qui est généré par une intégrale stochastique d'un noyau bivarié déterministe contre un mouvement Brownien, et dont les accroissements sont corrélés.

Mouvement Brownien Le processus Gaussien de variance $|t|$ est connu sous le nom de mouvement Brownien (indexé par $\mathbb{R}$ et non par $\mathbb{R}^+$ comme habituellement) et est noté $\{B(t)\}$. Il s'agit du processus le plus connu de la famille $\mathcal{G}_{ss}$ des processus Gaussiens auto-similaires et stationnaires.

Theorem 4.42. *Soit* $\{B(t)\}$ *un mouvement Brownien standard.*

(i) *L'espace* $\mathbb{L}^2$ *engendré par le mouvement Brownien* $\{B(t)\}$ *est la famille des intégrales stochastiques Gaussiennes* $B(\phi) := \int \phi(u)\mathrm{d}B(u)$ *de fonctions déterministes* ϕ *dans* $\mathbb{L}^2(Leb)$*, avec une variance* $\mathbb{V}[B(\phi)] = \int |\phi|^2 du$*.*

(ii) *La covariance de deux intégrales stochastiques est* $\mathrm{Cov}\left(B(\phi), B(\psi)\right) = \int_{\mathbb{R}} \phi(u)\psi(u)\mathrm{d}u$.

Par conséquent, il existe une isométrie entre $\mathcal{H}_B$ et $\mathbb{L}^2(Leb)$.

Nous prouvons la stationnarité et l'auto-similarité du mouvement Brownien au sens trajectoriel. Ce résultat est bien plus fort que le précédent, car nous obtenons une égalité de processus au lieu d'un résultat en distribution.

Comme ils possèdent les mêmes normes quadratiques Q_X, les processus $\{B_h(t) = B(t+h) - B(t),\ h \in \mathbb{R}\}$ et $\{B^\lambda(t) = \frac{1}{\sqrt{\lambda}} B(\lambda t), \lambda > 0\}$, qui appartiennent également à $\mathcal{G}_{ss}$, sont des mouvements Browniens. L'outil des intégrales stochastiques de fonctions déterministes $\phi \in \mathbb{L}^2(\text{Leb})$, $B(\phi) = \int \phi(u) dB(u)$ permet de décrire l'espace Gaussien $\mathcal{H}_B$, mais fournit également une formule de changement de variable pour la comparaison des trajectoires entre les variables aléatoires $B(\phi), B_h(\phi), B^\lambda(\phi)$.

Theorem 4.43. *Rappelons que pour tout $\lambda > 0$, $h \in \mathbb{R}$ et $t \in \mathbb{R}$, $\{B_t\}$, $\{B^\lambda(t) = \lambda^{-\frac{1}{2}} B(\lambda t)\}$, et $\{B_h(t) = B(t+h) - B(h)\}$ sont des mouvements Browniens.*
Pour tout $\phi \in \mathbb{L}^2(Leb)$, nous avons les représentations trajectorielles suivantes :

$$\int \lambda^{-\frac{1}{2}} \phi\left(\frac{u}{\lambda}\right) \mathrm{d}B(u) = \int \phi(u) \mathrm{d}B^\lambda(u) \text{ et } \int \phi\left(u - h\right) \mathrm{d}B(u) = \int \phi(u) \mathrm{d}B_h(u).$$

Toutes ces variables sont des variables Gaussiennes centrées ayant la même variance $\int |\phi(u)|^2 du$.

Mouvement Brownien fractionnaire (fBm)

Dans une deuxième partie, nous nous intéressons à la construction des processus $\mathcal{G}_{ss}$ sous forme d'intégrales stochastiques de fonctions déterministes $\phi(u)$, contre le mouvement Brownien, où les fonctions déterministes $\phi(u)$ sont remplacées par des noyaux bivariés $\kappa(t,u)$ tels que $\int |\kappa(t,u)|^2 du < \infty$. Le noyau de covariance est défini par $K_{X^\kappa}(t,s) = \int_{\mathbb{R}} \kappa(t,u)\kappa(s,u) \mathrm{d}u$ et $\kappa(0,u) = 0$. En accord avec (Mandelbrot and Van Ness, 1968b), un mouvement Brownien dépendant du chemin (PDBM), est défini comme un processus Gaussien, tel que $X^\kappa(t) = \int \kappa(t,u) dB(u)$.

Nous nous intéressons aux PDBM translatés dans le temps et changés d'échelle temporelle, défini pour tout $\lambda > 0$, $h \in \mathbb{R}$ et presque tous $(t,u) \in \mathbb{R}^2$ par :

$$\begin{cases} X^\kappa(t+h, B) - X^\kappa(h, B) = \int (\kappa(t+h,u) - \kappa(h,u)) \mathrm{d}B(u) \\ X^\kappa(\lambda t, B) = \int \kappa(\lambda t, u) \mathrm{d}B(u). \end{cases} \tag{4.267}$$

La formule de changement de variable suggère de transporter le changement de temps initialement supporté par la variable temporelle t sur la variable d'intégration u. Ensuite, nous prouvons des identités trajectorielles sur les PDBMs changés de temps.

Pour tout $\lambda > 0$, $h \in \mathbb{R}$ et $t \in \mathbb{R}$, $\{B_t\}$, $\{B_h(t) = B(t+h) - B(h)\}$ et $\{B^\lambda(t) = \lambda^{-\frac{1}{2}} B(\lambda t)\}$ sont des mouvements Browniens. Nous définissons les noyaux de carré intégrable pour presque tous $(t,u) \in \mathbb{R}^2$ et pour tout $h \in \mathbb{R}$ et $\lambda > 0$ par :

$$\kappa(t,u), \quad \kappa_h(t,u) = \kappa(t, u-h), \quad \kappa^\lambda(t,u) = \lambda^{-1/2} \kappa(t, u/\lambda). \tag{4.268}$$

Theorem 4.44. *Définissons les PDBMs* $\{Y(t) = X^{\kappa}(t, B)\}, \{Y^{\lambda}(t) = X^{\kappa}(t, B^{\lambda})\}, \{Y_h(t) = X^{\kappa}(t, B_h)\}$.

(i) *Les processus* $\{Y^{\lambda}(t)\}$ *et* $\{Y_h(t)\}$ *ont la même distribution que* $\{Y(t)\}$.

(ii) $\{Y^{\lambda}(t)\}$ *et* $\{Y_h(t)\}$ *satisfont les identités trajectorielles suivantes :*

$$\{Y^{\lambda}(t) = X^{\kappa^{\lambda}}(t, B)\} \quad et \quad \{Y_h(t) = X^{\kappa_h}(t, B)\}. \tag{4.269}$$

Enfin, nous prouvons la stationnarité et l' auto-similarité des PDBMs en termes d'égalité des processus, c'est-à-dire, au sens trajectoriel.

Theorem 4.45. *Soient* $\{B(t)\}$, $\{B^{\lambda}(t)\}$, $\{B_h(t)\}$, *les mouvements Browniens transformés.*

(i) *Si la condition de stationnarité temporelle du noyau bivarié* $\kappa(t, u)$ *est satisfaite,*

$$\kappa(t + h, u) - \kappa(h, u) = \kappa_h(t, u) = \kappa(t, u - h) \tag{4.270}$$

alors l'identité trajectorielle suivante est vérifiée :

$$\{X^{\kappa}(t + h, B) - X^{\kappa}(h, B) = X^{\kappa}(t, B_h)\} \tag{4.271}$$

et le processus $\{X^{\kappa}(t, B)\}$ *est stationnaire.*

(ii) *Si la condition d' auto-similarité temporelle du noyau bivarié* $\kappa(t, u)$ *est satisfaite,*

$$\kappa(\lambda t, u) = \eta(\lambda)\kappa\left(t, \frac{u}{\lambda}\right) = (\sqrt{\lambda}\eta(\lambda))\kappa^{\lambda}(t, u) \text{ avec } \eta(\lambda) = \lambda^{\nu}\mathbb{1}_{\{\lambda>0\}},\ \nu \in \left(-\frac{1}{2}, \frac{1}{2}\right) \tag{4.272}$$

alors l'identité trajectorielle suivante est vérifiée :

$$\{X^{\kappa}(\lambda t, B) = \eta(\lambda)\sqrt{\lambda}X^{\kappa}(t, B^{\lambda})\}. \tag{4.273}$$

et le processus $\{X^{\kappa}(t, B)\}$ *est auto-similaire.*

(iii) *Si les conditions de stationnarité temporelle (Équation (4.270)) et d'auto-similarité temporelle (Équation (4.272))) sont simultanément satisfaites, alors l'identité trajectorielle suivante est vérifiée :*

$$\{X^{\kappa}(\lambda(t + h), B) - X^{\kappa}(\lambda h, B) = \eta(\lambda)\sqrt{\lambda}X^{\kappa}(t, B_h^{\lambda})\} \tag{4.274}$$

et le processus $\{X^{\kappa}(t, B)\}$ *est à la fois stationnaire et self-similaire avec un noyau de variance-covariance défini comme suit :*

$$V_X(t) = V_X(1)\left|t\right|^{2\gamma} \quad et \quad K_X(t, s) = \frac{V_X(1)}{2}(|t|^{2\gamma} + |s|^{2\gamma} - |t - s|^{2\gamma}) \tag{4.275}$$

où $\gamma \in (0, 1)$, $\eta(\lambda) = \lambda^{\gamma - \frac{1}{2}}\mathbb{1}_{\{\lambda>0\}}$.

Pour que le processus PDBM soit auto-similaire et stationnaire, il suffit que le noyau κ satisfasse le système suivant :

$$\begin{cases} \kappa(t+h,u) - \kappa(h,u) = \kappa(t,u-h) \\ \kappa(\lambda t, \lambda u) = \eta(\lambda)\kappa\left(t,u\right) \quad \text{avec} \quad \eta(\lambda) = \lambda^\nu \mathbb{1}_{\{\lambda>0\}},\ \nu \in \left(-\frac{1}{2},\frac{1}{2}\right). \end{cases} \tag{4.276}$$

La dernière étape consiste à trouver un noyau bivarié $\kappa(t,u)$, solution du système ci-dessus, et de carré intégrable en u. La solution proposée par Mandelbrot et Van Ness est le noyau fractionnaire κ^ν, qui est clairement une solution du système précédent.

Proposition 4.46. *Un noyau bivarié proposé par Mandelbrot et Van Ness est donné par :*

$$\kappa^\nu(t,u) = (t-u)^\nu \mathbb{1}_{\{u<t\}} - (-u)^\nu \mathbb{1}_{\{u<0\}}, \quad \textit{avec} \quad \nu = \gamma - \frac{1}{2} \in \left(-\frac{1}{2},\frac{1}{2}\right). \tag{4.277}$$

(i) *Ce noyau présente des problèmes de définition lorsque $u \to 0$, $u \to t$.*

(ii) *Ce noyau satisfait le Système (4.276).*

(iii) *Pour tout $\nu \in \left(-\frac{1}{2},\frac{1}{2}\right)$, $\kappa(t,u) \sim \nu t(-u)^{(\nu-1)}$ est intégrable au carré par rapport à la mesure de Lebesgue lorsque $u \to -\infty$ car $2(\nu-1) < -1$.*

Pour $\nu < 0$, lorsque $u \to 0$ (resp. $u \to t$), $\kappa(t,u) \sim -(-u)^\nu$ (resp. $\kappa(t,u) \sim (t-u)^\nu$), qui est de carré intégrable par rapport à la mesure de Lebesgue car $2\nu + 1 > 0$.

Ainsi, $\kappa(t,u)$ est de carré intégrable par rapport à u.

La carré intégrabilité vérifie que :

– pour toute valeur de ν, lorsque $u \to -\infty$, $\int \left|\kappa(t,u)\right|^2 \mathrm{d}u < \infty$,

– pour $\nu < 0$, pour u au voisinage de 0 et pour u au voisinage de t, $\int \left|\kappa(t,u)\right|^2 \mathrm{d}u < \infty$.

Quelques remarques guident l'intuition pour obtenir une telle solution.

Extension au cadre multidimensionnel Enfin, une extension au cadre multidimensionnel est proposée, d'abord pour le mouvement Brownien de dimension d dont les composantes sont des mouvements Browniens corrélés de matrice de covariance $(\rho_{i,j}(t \wedge s))_{i,j\in[\![1,d]\!]^2}$, puis pour le mouvement Brownien fractionnaire multivarié (mfBm).

Definition 4.47. *$\{\boldsymbol{X}^{\boldsymbol{\kappa}}(t,B) = (X^{\kappa^1}(t,B^1), X^{\kappa^2}(t,B^2), \ldots, X^{\kappa^d}(t,B^d))\}$, appelé mouvement Brownien fractionnaire de dimension d, est un processus Gaussien multidimensionnel défini comme l'intégrale stochastique d'un vecteur de noyaux bivariés déterministes par rapport à un vecteur de transformations linéaires du processus de Wiener, tel que pour presque tous $(t,u) \in \mathbb{R}^2$:*

$$\boldsymbol{X}^{\boldsymbol{\kappa}}(t,B) = \int \boldsymbol{\kappa}(t,u) \odot \mathrm{d}\boldsymbol{B}(u) := \left(\int \kappa^i(t,u)\mathrm{d}B^i(u)\right)_{i\in[\![1,d]\!]} \tag{4.278}$$

où $\odot$ est le produit élément par élément, $\boldsymbol{\kappa}(t,u) = (\kappa^i(t,u))_{i\in[\![1,d]\!]}$ est le vecteur de noyaux proposé par Mandelbrot et Van Ness dans l'Équation (4.277), $\boldsymbol{B}(t) = (B^i(t))_{i\in[\![1,d]\!]}$, et son noyau de covariance est donné par :

$$K_{X^\kappa}((i,t),(j,s)) = \mathrm{Cov}\left(X^{\kappa^i}(t,B^i), X^{\kappa^j}(s,B^j)\right) = \int \kappa^i(t,u)\kappa^j(s,u)\rho_{i,j}\mathrm{d}u. \quad (4.279)$$

Chaque coordonnée du vecteur $\boldsymbol{\kappa}(t,u)$ satisfait les conditions de stationnarité et d'auto-similarité dans le cadre univarié données dans l'Équation (4.276). Ces propriétés sont ensuite obtenues dans le sens vectoriel, c'est-à-dire pour tout $\lambda > 0$, $h \in \mathbb{R}$ et pour presque tous $(t,u) \in \mathbb{R}^2$:

$$\begin{cases} \boldsymbol{\kappa}(\lambda t, \lambda u) = \boldsymbol{\eta}(\lambda)\boldsymbol{\kappa}\left(t,u\right) \quad \text{avec} \quad \boldsymbol{\eta}(\lambda) = (\lambda^{\nu_i}\mathbb{1}_{\{\lambda>0\}})_{i\in[\![1,d]\!]},\ \nu_i \in \left(-\frac{1}{2},\frac{1}{2}\right) \\ \boldsymbol{\kappa}(t+h,u) - \boldsymbol{\kappa}(h,u) = \boldsymbol{\kappa}(t,u-h). \end{cases} \quad (4.280)$$

Nous prouvons que le mfBm est un processus Gaussien multidimensionnel auto-similaire avec des accroissements stationnaires et corrélés.

Theorem 4.48. *Soit $\{\boldsymbol{X}^{\boldsymbol{\kappa}}(t,B)\}$ un mouvement Brownien fractionnaire de dimension d. Pour tout $\lambda > 0$, $h \in \mathbb{R}$ et pour presque tous $(t,u) \in \mathbb{R}^2$:*

$$\begin{cases} \boldsymbol{X}^{\boldsymbol{\kappa}}(t+h) - \boldsymbol{X}^{\boldsymbol{\kappa}}(h) = \int \boldsymbol{\kappa}(t,u) \odot \mathrm{d}\boldsymbol{B}_{\boldsymbol{h}}(u) = \boldsymbol{X}^{\boldsymbol{\kappa}}(t,B_h) \\ \boldsymbol{X}^{\boldsymbol{\kappa}}(\lambda t) = \sqrt{\lambda}\boldsymbol{\eta}(\lambda) \odot \int \boldsymbol{\kappa}(t,u) \odot \mathrm{d}\boldsymbol{B}^{\boldsymbol{\lambda}}(u) = \sqrt{\lambda}\boldsymbol{\eta}(\lambda) \odot \boldsymbol{X}^{\boldsymbol{\kappa}}(t,B^\lambda). \end{cases} \quad (4.281)$$

Le processus vectoriel $\{\boldsymbol{X}^{\boldsymbol{\kappa}}(t,B)\}$ est stationnaire et auto-similaire au sens trajectoriel.

De plus, son noyau de covariance est entièrement caractérisé par des fonctions puissances dont les exposants dépendent des composantes spatiales. Voici la traduction en français du théorème :

Theorem 4.49. *Soit $\{\boldsymbol{X}^{\boldsymbol{\kappa}}(t,B)\}$ un mouvement Brownien fractionnaire d-dimensionnel.*

(i) *Le noyau de covariance entre deux coordonnées distinctes prises au même instant $t \in \mathbb{R}\backslash\{0\}$ est entièrement caractérisé par une fonction puissance d'exposant $\gamma_i + \gamma_j$ où $\gamma_i, \gamma_j \in (0,1)$:*

$$K_{X^\kappa}((i,t),(j,t)) = K_{X^\kappa}((i,1),(j,1))\left|t\right|^{\gamma^i+\gamma^j}. \quad (4.282)$$

(ii) *Le noyau de covariance symétrisé entre deux coordonnées distinctes prises à deux instants distincts $X^{\kappa_i}(t,B^i)$ et $X^{\kappa_j}(s,B^j)$ avec $t,s \in \mathbb{R}\backslash\{0\}$ est entièrement caractérisé par combinaison linéaire de fonctions puissance d'exposant $\gamma_i + \gamma_j$ où $\gamma_i, \gamma_j \in (0,1)$:*

$$K_{X^\kappa}((i,t),(j,s)) + K_{X^\kappa}((i,s),(j,t)) = K_{X^\kappa}((i,1),(j,1))\left(\left|t\right|^{\gamma^i+\gamma^j} + \left|s\right|^{\gamma^i+\gamma^j} - \left|t-s\right|^{\gamma^i+\gamma^j}\right). \quad (4.283)$$

Changement de notations : Dans Chapter 3, le mouvement Brownien fractionnaire $\{X^{\kappa^i}(t,B^i)\}$ est noté $\{B^{H_i}(t)\}$, l'exposant γ_i de la fonction puissance correspond à l'exposant de Hurst H_i, et le noyau bivarié $\kappa^i(t,u)$ est remplacé par $\psi^{H_i}_{0,t}(u)$.

4.8.2 Chapter 3 - Prédiction de la VaR pour un portefeuille d'actifs sous des dynamiques fractionnaires

Etat de l'art

La Valeur à Risque (VaR) est une statistique qui quantifie l'ampleur des pertes financières possibles au sein d'une entreprise, d'un portefeuille ou d'une position sur une période spécifique. Cette mesure est principalement utilisée par les banques d'investissement et commerciales pour déterminer l'ampleur et les probabilités des pertes potentielles dans leurs portefeuilles institutionnels. Les gestionnaires de risques utilisent la VaR pour mesurer et contrôler le niveau d'exposition au risque. On peut appliquer les calculs de la VaR à des positions spécifiques ou à des portefeuilles entiers, ou les utiliser pour mesurer l'exposition au risque de l'entreprise dans son ensemble. La Valeur à Risque (VaR) est la mesure du risque la plus couramment utilisée par les régulateurs. La VaR au niveau de risque $\alpha \in (0,1)$ correspond au quantile d'ordre α de la distribution des profits et pertes ($P\&L$). La littérature sur la VaR est vaste, avec plus de 2700 articles référencés sur Google Scholar. La VaR a été étudiée en détail dans les ouvrages bien connus de (Wipplinger, 2007, Partie II. p105) et (McNeil et al., 2015, Chap.2, p37). Calculer la VaR est un problème complexe car cela nécessite de connaître la distribution du portefeuille P&L, qui est généralement inconnue. Un large éventail d'articles se sont intéressés à la prédiction de la VaR. Pour n'en citer que quelques-uns, voir par exemple (Cheung and Powell, 2012) qui présente une étude pédagogique utilisant le calcul paramétrique et la simulation Monte-Carlo pour calculer la VaR, ou (Feuerverger and Wong, 2000) qui explique le calcul de la VaR pour les portefeuilles non linéaires. Bien que la VaR ne soit pas une mesure de risque parfaite (voir (Delbaen et al., 1998)), elle est toujours couramment utilisée. Il est fondamental d'avoir une estimation précise de la VaR : en cas de sous-estimation, l'institution financière prendra trop de risques, sans être préparée à cela ; en cas de surestimation, le montant d'argent à mettre de côté serait trop élevé, ce qui empêcherait certaines activités bancaires. De plus, son calcul doit être efficace, car son évaluation est fréquente et doit être réalisée selon de nombreux portefeuilles. Dans (Gaivoronski and Pflug, 2005), une méthode de calcul du portefeuille qui donne la plus petite VaR parmi ceux qui donnent au moins un rendement attendu spécifié est présentée. En utilisant cette approche, la frontière efficace complète en moyenne-VaR peut être calculée. La méthode est basée sur l'approximation de la VaR historique par une VaR lissée (SVaR) qui filtre les irrégularités locales. La Valeur à Risque (VaR) et la Valeur à Risque Conditionnelle (CVaR) sont deux mesures largement utilisées en gestion des risques. Les travaux de (Bardou et al., 2009) traitent du problème de l'estimation à la fois de la VaR et de la CVaR en utilisant une approximation stochastique (avec des pas décroissants) : une première procédure Robbins-Monro (RM) basée sur l'identité de Rockafellar-Uryasev est proposée pour la CVaR. L'estimateur fourni par l'algorithme satisfait un théorème central limite Gaussien. En deuxième étape, afin d'accélérer la procédure initiale, une procédure d'échantillonnage d'importance récursive et adaptative (IS) qui induit une réduction significative de la variance des procédures de VaR et de CVaR est présentée. Dans (Manganelli and Engle, 2001), une enquête et une évaluation des performances des

méthodologies de VaR univariées les plus populaires sont effectuées en accordant une attention particulière à leurs hypothèses sous-jacentes et à leurs lacunes logiques. Les travaux de (Jorion, 1996) exposent la méthodologie statistique pour analyser l'erreur d'estimation dans la VaR et montrent comment améliorer la précision des estimations de VaR. Les régulateurs exigent que les banques utilisent la Valeur à Risque (VaR) pour estimer l'exposition de leurs portefeuilles de trading au risque de marché, afin d'établir des exigences en capital. Cependant, l'analyse de la VaR au niveau du portefeuille est un problème de grande dimension et donc intensif en termes de calcul. Les travaux de (Albanese et al., 2002) présentent deux nouvelles approches basées sur le portefeuille pour réduire la dimensionnalité de l'analyse de la VaR. Une approche classique pour prédire la VaR est le modèle de moyenne mobile pondérée, comme présenté dans (Hendricks, 1996). Les travaux de (Lucas and Zhang, 2016) présentent une méthodologie simple pour modéliser la variation temporelle des volatilités et d'autres moments d'ordre supérieur en utilisant un schéma de mise à jour récursive similaire à l'approche RiskMetrics familière. Dans (Gabrielsen et al., 2015), l'auteur propose un aperçu des dynamiques variables dans le temps de la forme de la distribution des séries de rendement financier en proposant un modèle de moyenne mobile pondérée exponentielle qui estime conjointement la volatilité, l'asymétrie et l'aplatissement au fil du temps en utilisant une forme modifiée de la densité de Gram-Charlier dans laquelle l'asymétrie et l'aplatissement apparaissent directement dans la forme fonctionnelle de cette densité. Dans ce cadre, la VaR peut être décrite comme une fonction des moments d'ordre supérieur variables dans le temps en appliquant la série d'expansion de Cornish-Fisher des quatre premiers moments. L'article (Alexander and Dakos, 2023) discute de la contribution à la Valeur à Risque dans le cadre du modèle actif-passif en utilisant l'approche EWMA. On suppose que les rendements des actifs et des passifs sont des séries chronologiques suivant le modèle de moyenne mobile pondérée exponentielle (EWMA). Le rendement excédentaire, qui est la différence entre le rendement des actifs et des passifs, est analysé en utilisant le modèle actif-passif. Dans ce cas, le risque de rendement excédentaire est mesuré à l'aide du modèle de la Valeur à Risque. Lorsque des investissements sont réalisés sur plusieurs actifs, chaque actif contribuera à l'établissement de la Valeur à Risque du portefeuille d'investissement, qui peut être mesurée à l'aide du modèle de contribution à la Valeur à Risque. En utilisant la contribution à la Valeur à Risque, il est possible de voir dans quelle mesure la Valeur à Risque excédentaire du portefeuille d'investissement est influencée, et quelle est la proportion de la contribution à la Valeur à Risque de chaque excédent de l'actif d'investissement. Sur la base du calcul de la contribution à la Valeur à Risque, les investisseurs peuvent envisager d'investir dans certains actifs analysés. Les travaux de (Sukono et al., 2018) dérivent un modèle de couverture combiné et dynamique - moyenne mobile pondérée exponentielle-autorégressif conditionnel généralisé hétéroscédasticité (GARCH)(1,1)-M applicable aux marchés financiers réels sur la base d'études précédentes. Les résultats de cet article montrent que le modèle construit est non seulement excellent pour rechercher la VaR minimale mais aussi pratique pour la situation réelle où les variances des données de prix financiers sont variables dans le temps. L'article (Anantafortuna and Anggono, 2019) vise à utiliser la Valeur à Risque comme outil d'évaluation des risques pour les activités de trading basé sur la moyenne mobile exponentielle et la stratégie de ligne

de rétrogradation. Cet article vise également à informer sur le niveau de risque que prennent les investisseurs avec un certain investissement en calculant la perte potentielle maximale qui se produit chaque jour et à évaluer si les activités de trading ont été effectuées dans ou au-delà de la Valeur à Risque calculée. Des études empiriques ont montré qu'un grand nombre de rendements d'actifs financiers présentent une dépendance à long terme. De plus, une certaine mémoire à long terme ou dépendance à longue portée peut également être observée comme un fait stylisé dans la volatilité du marché, avec un impact significatif sur la tarification et la prévision de la volatilité du marché. Étant donné que les mesures de risque sont basées sur la dynamique des rendements des actifs sous-jacents, alors dans de tels cas, la méthodologie classique utilisée pour prédire les mesures de risque présente de mauvaises performances. Pour remédier à ce problème, des modèles autorégressifs tels que ARCH ou GARCH ont été utilisés pour prédire la VaR. Dans (Anantafortuna and Anggono, 2019), les auteurs montrent que, d'une part, les modèles qui prennent en compte la mémoire à long terme offrent la promesse d'une meilleure prévision de la volatilité à long terme ainsi qu'une tarification précise des contrats à long terme. D'autre part, l'attention récente se porte sur la question de savoir si la mémoire à long terme peut affecter la mesure du risque de marché dans le contexte de la valeur en risque (VaR). Dans cet article, les auteurs évaluent la valeur en risque (VaR) et la perte attendue (ESF) sur les marchés financiers dans de telles conditions. La méthodologie d'estimation classique de la VaR telle que la moyenne mobile exponentielle (EMA) ainsi que l'extension aux cas où la mémoire à long terme est une caractéristique inhérente du système sont investiguées. En particulier, deux modèles de mémoire à long terme sont estimés, le Fractional Integrated Asymmetric Power-ARCH et le Hyperbolic-GARCH avec différentes hypothèses de distribution des erreurs. Ces modèles présentent de meilleures performances dans la prédiction de la VaR que les approches classiques qui ne tiennent pas compte de la dépendance à long terme. Dans (Aloui, 2008), la mémoire à long terme est explorée sur la volatilité des marchés de l'énergie et la valeur en risque (VaR). La question principale est la suivante : pouvons-nous estimer la VaR de manière plus précise si une mémoire à long terme existe ? Pour enquêter sur cette question, plusieurs processus de type GARCH, y compris le processus FIGARCH, ont été mis en œuvre pour les prix quotidiens de certains produits énergétiques principaux (janvier 1986 à juillet 2007). La valeur en risque a été estimée pour les positions de trading à court et à long terme et à divers niveaux de confiance. Les distributions normales, de Student et de Student asymétriques sont suggérées pour divers processus GARCH. La performance de la VaR de type GARCH est évaluée en estimant le taux d'échec de la statistique de test de Kupiec. Conformément aux études précédentes, nos résultats montrent que la volatilité des prix de l'énergie présente une mémoire à long terme. La VaR calculée à l'aide d'un processus FIGARCH de Student-t asymétrique offre les meilleures performances tant pour les positions de trading à court terme que pour les positions de trading à long terme. Les travaux de (Youssef et al., 2015) présentent trois modèles de mémoire à long terme, y compris FIGARCH, HYGARCH et FIAPARCH, pour prévoir la volatilité des matières premières énergétiques en capturant certains faits stylisés sur la volatilité tels que la mémoire à long terme, l'hétéroscédasticité, l'asymétrie et les queues épaisses. L'article (Tang and Shieh, 2006) examine les propriétés de mémoire à long terme pour les prix de clôture de trois marchés

à terme sur indices boursiers. Les modèles FIGARCH (1, d, 1) et HYGARCH (1, d, 1) avec des distributions normales, de Student-t et de Student-t asymétriques pour les prix quotidiens du S&P500, du Nasdaq100 et du Dow Jones sont d'abord estimés. Ensuite, les valeurs en risque sont calculées à l'aide des modèles estimés. Les résultats empiriques montrent que, pour les trois marchés à terme sur indices boursiers, les modèles HYGARCH (1, d, 1) avec distribution de Student-t asymétrique sont plus performants selon les tests de LR de Kupiec. En particulier, pour les prix des contrats à terme sur le S&P500 et le Nasdaq100.

4.8.3 Contributions

Chapter 3 se concentre sur la construction d'un modèle prédictif précis pour la Value-at-Risk (VaR), capable de prendre en compte la dépendance à long-terme (respectivement à court-terme). Des modèles autorégressifs ont déjà été développés pour accomplir cette tâche ; cependant, ces modèles dépendent d'un grand nombre de paramètres à estimer. Nos travaux proposent une autre approche capable de capturer la dépendance à long-terme (respectivement à court-terme) grâce à des processus Gaussiens autosimilaires avec des incréments stationnaires, appelés mouvements Browniens fractionnaires (fBm). Nous considérons un cadre dans lequel la dynamique des prix est décrite par des modèles de Black-Scholes fractionnaires, dépendant de mouvements Browniens fractionnaires paramétrés par leur exposant de Hurst. Ces modèles présentent l'avantage de capturer les corrélations à la fois dans le temps et entre les actifs. Comparé au modèle classique de Black-Scholes, la flexibilité dans le choix de l'exposant de Hurst permet une meilleure description des trajectoires des prix afin de coller à la réalité. Ainsi, le modèle prédictif pour la VaR conditionnelle à l'horizon temporel h (c'est-à-dire, la VaR des profits et pertes sur la prochaine période de longueur h, conditionnellement aux observations disponibles au moment du calcul) sera en mesure de prédire de manière plus précise le montant d'argent à mettre de côté, évitant ainsi une surcharge ou une sous-estimation pour le régulateur. La précision de nos prédictions de VaR est évaluée à l'aide d'une procédure de backtesting basée sur les travaux de (Christoffersen, 1998). L'utilisation des modèles fractionnaires en finance n'est pas nouvelle et a pris de l'importance ces dernières années. Les modèles fractionnaires ont été largement utilisés à diverses fins ; cependant, dans les études mentionnées ci-dessus, la prédiction de la VaR conditionnelle n'a pas été abordée.

En résumé, nous apportons les contributions suivantes. Nous concevons un cadre théorique pour les modèles fractionnaires dans l'espace de Hilbert Gaussien. Nous proposons une formule fermée pour l'approximation Gaussienne de la $\mathbf{VaR}_\alpha$ conditionnelle, à l'horizon temporel h, de la variation future du portefeuille sous des dynamiques fractionnaires. Nous fournissons ensuite une quantification de l'erreur d'approximation Gaussienne. Nous utilisons une méthodologie robuste pour estimer les paramètres du modèle fractionnaire. Nous réalisons une procédure de backtesting pour évaluer la précision de notre approximation. Nous appuyons notre analyse par diverses expériences qui illustrent le comportement de notre nouveau modèle.

Tout d'abord, nous établissons un cadre théorique Gaussien sans aucune application dans le domaine financier. Dans cette partie, nous prouvons la forme spécifique de l'espérance et de la variance conditionnelles d'une combinaison linéaire d'incréments futurs de plusieurs fBms corrélés, étant donné les incréments passés.

Cadre théorique Il n'existe pas de méthode unique pour modéliser le mouvements Brownien fractionnaire multivarié : la façon la plus directe est de donner leur fonction de covariance comme un processus Gaussien multivarié ; cependant, cette méthode est assez peu pratique lorsqu'il s'agit de calculs conditionnels dans le temps, avec diverses filtrations à prendre en compte en fonction des observations disponibles. Par conséquent, nous préférons travailler directement au niveau de l'espace de Hilbert Gaussien indexé par des fonctions dans

$$\mathcal{L}_d^2 := L_2(\mathbb{R}^d, \mathrm{d}u) = \left\{ f : \mathbb{R} \mapsto \mathbb{R}^d \ s.t. \ \left|f\right|_{\mathcal{L}_d^2}^2 := \int_{\mathbb{R}} |f(u)|^2 \mathrm{d}u < +\infty \right\} \tag{4.284}$$

où les quantités de base sont définies par des intégrales de Wiener, et les projections et espérances conditionnelles sont réalisées par le biais de projections de fonctions dans $\mathcal{L}_d^2$.

Nous adoptons le cadre du processus Gaussien isonormal associé à l'espace de Hilbert $\mathcal{L}_d^2$, avec le produit scalaire

$$\langle f, g \rangle_{\mathcal{L}_d^2} = \int_{\mathbb{R}} f(u) \cdot g(u) \mathrm{d}u, \tag{4.285}$$

également appelé espace de Hilbert Gaussien ; ici $f(u) \cdot g(u)$ est simplement le produit scalaire dans $\mathbb{R}^d$ entre les fonctions vectorielles f et g au point u. Voir (Janson, 1997) pour un large aperçu sur l'espace de Hilbert Gaussien. En d'autres termes, nous considérons un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ supportant une famille Gaussienne centrée de variables aléatoires scalaires

$$\mathcal{H} = \{I(f) : f \in \mathcal{L}_d^2\} \tag{4.286}$$

telle que la propriété d'isométrie est en vigueur

$$\mathbb{E}\left[I(f)I(g)\right] = \langle f, g \rangle_{\mathcal{L}_d^2}. \tag{4.287}$$

Une façon de construire cet espace de Hilbert Gaussien indexé par des fonctions de carré intégrable $f \in \mathcal{L}_d^2$ est de supposer que l'espace de probabilité supporte un mouvement Brownien de dimension d indexé par la droite réelle, $(W_t = (W_t^1, \cdots, W_t^d)^\top : t \in \mathbb{R})$, et de définir $I(f)$ par l'intégrale stochastique

$$I(f) = \int_{\mathbb{R}} f(u) \cdot \mathrm{d}W_u = \sum_{i=1}^{d} \int_{\mathbb{R}} f^i(u) \mathrm{d}W_u^i \tag{4.288}$$

(a.k.a. intégrale de Wiener). Voir (Janson, 1997, Chapter 7) où les intégrales sont restreintes à $\mathbb{R}^+$, l'extension à $\mathbb{R}$ est immédiate. Comme montré dans Chapter 2, le mouvement Brownien fractionnaire (fBm) d'exposant de Hurst H correspond à un choix spécifique de fonction f, qui est la représentation dépendant temporellement de la trajectoire proposée par Mandelbrot et Van Ness (Mandelbrot and Van Ness, 1968b).

Definition-Proposition 5. *Soit $H \in (0,1)$ et définissons, pour tout $t \geq 0$,*

$$B_t^H = I((\psi_{0,t}^H(.), 0, \cdots, 0)^\top) = \int_{\mathbb{R}} \psi_{0,t}^H(u) \mathrm{d}W_u^1 \quad avec \quad \psi_{s,t}^H(u) = \frac{1}{c_H} \left((t-u)_+^{H-\frac{1}{2}} - (s-u)_+^{H-\frac{1}{2}} \right), \tag{4.289}$$

où

$$c_H := \sqrt{\frac{1}{2H} \left(\frac{3}{2} - H \right) \mathfrak{B} \left(2-2H, H+\frac{1}{2} \right)} \tag{4.290}$$

et $\mathfrak{B}(x,y) = \int_0^1 u^{x-1}(1-u)^{y-1}\mathrm{d}u$ est la fonction Beta.

Cela définit un fBm scalaire B^H, c'est-à-dire un processus Gaussien centré de covariance

$$\mathrm{Cov}\left(B_t^H, B_s^H \right) = \frac{1}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H}), \quad H \in (0,1). \tag{4.291}$$

Le fBm possède une propriété d'auto-similarité. La propriété d'auto-similarité stipule que, pour tout $\lambda > 0$ fixé, le processus $\{\lambda^{-H} B_{\lambda t}^H\}_{t\in\mathbb{R}}$ est également un fBm d'exposant de Hurst $H \in (0,1)$. Cela est équivalent à dire que pour tout $\lambda > 0$, les processus $\{\lambda^{-H} B_{\lambda t}^H\}_{t\in\mathbb{R}}$ et $B_t^H\}_{t\in\mathbb{R}}$ ont la même distribution. En d'autres termes, la propriété d'auto-similarité établit une relation de proportionnalité spatiale de facteur λ^H, entre les distributions de $\{B_{\lambda t}^H\}$ et $\{B_t^H\}$. Cela conduit à obtenir des relations de proportionnalité spatiale de facteur λ^{2H} entre les variances, respectivement entre les noyaux de covariance, des processus $\{B_{\lambda t}^H\}$ et $\{B_t^H\}$. Ce type de propriété d'échelle jouera un rôle important dans la simplification de certaines formules ultérieures, car elle permet des changements d'échelles temporelles et spatiales dans les processus. La propriété d'auto-similarité est, par exemple, utile pour calculer la fonction de covariance du fBm, car elle permet d'exprimer cette covariance uniquement à l'aide de fonctions puissance d'exposant $H \in (0,1)$ (fonctions auto-similaires), et de la covariance au temps 1.

On définit le mouvement Brownien fractionnaire multivarié (mfBm) dont les composantes sont corrélées. Pour introduire des corrélations entre les mouvements Brownien fractionnaires (fBms), nous ajoutons des corrélations entre les composantes du mouvement Brownien sous-jacent à travers une transformation linéaire du processus de Wiener W. Cela fonctionne comme suit : soit $C = (\rho_{i,j})_{1\leq i,j\leq d}$ une matrice de corrélation, et soit R la racine carrée symétrique de C (qui existe car C est une matrice symétrique semi-définie positive), de sorte que

$$\rho_{i,j} = R_{:,i} \cdot R_{:,j} \tag{4.292}$$

où $R_{:,i}$ est la i-ème colonne de R. Nous définissons alors un mouvement Brownien fractionnaire multivarié, avec des composantes corrélées, chacune ayant son propre paramètre de Hurst.

Definition-Proposition 6. *Soit $H_1, \cdots, H_d$ une séquence d'exposants de Hurst dans $(0,1)$ et définissons, pour tout $t \geq 0$ et tout $i \in \{1, \cdots, d\}$,*

$$B_t^{i,H_i} := I(\psi_{0,t}^{H_i} R_{:,i}) = \int_{\mathbb{R}} \psi_{0,t}^{H_i}(u) R_{:,i} \cdot \mathrm{d}W_u, \tag{4.293}$$

où $R_{:,i}$ est la i-ème colonne de la matrice R. Cela définit un mouvement Brownien fractionnaire multivarié (mfBm) de paramètres $H = (H_i)_i \in (0,1)^d$ dont les coordonnées (B_t^{i,H_i}) sont des fBms d'exposant de Hurst H_i, et sont corrélées de manière à ce que :

$$\forall s,t \in \mathbb{R}, \quad \mathrm{Cov}(B_t^{i,H_i}, B_s^{j,H_j}) + \mathrm{Cov}(B_t^{j,H_j}, B_s^{i,H_i}) \tag{4.294}$$

$$= \mathrm{Cov}(B_1^{i,H_i}, B_1^{j,H_j})\Big(|t|^{H_i+H_j} + |s|^{H_i+H_j} - |t-s|^{H_i+H_j}\Big), \tag{4.295}$$

où $\mathrm{Cov}(B_1^{i,H_i}, B_1^{j,H_j})$ est donnée explicitement dans le Lemme 3.1. On remarque que

$$B_t^i := R_{:,i} \cdot W_t, \quad i \in \{1, \cdots, d\} \tag{4.296}$$

définit un mouvement Brownien de dimension d de corrélation C.

Dans ce qui suit, nous devons généralement calculer l'espérance et la variance conditionnelles d'une certaine intégrale stochastique $I(f)$ conditionnellement aux incréments du fBm sur une certaine grille temporelle. Un résultat général sur la façon de calculer $\mathbb{E}\Big[I(f) \mid \{I(f_l)\}_{(l\in\mathcal{I}}\Big]$ et $\mathbb{V}\Big[I(f) \mid \{I(f_l)\}_{(l\in\mathcal{I}}\Big]$, où $(f_l)_{l\in\mathcal{I}}$ est une famille arbitraire de fonctions dans $\mathcal{L}_d^2$, $\mathcal{L}'$ est le sous-espace fermé de $\mathcal{L}_d^2$ engendré par $(f_l)_{l\in\mathcal{I}}$, et $f^\star \in \mathcal{L}'$ est la projection orthogonale (unique) de f sur $\mathcal{L}'$, est fourni par :

$$\mathbb{E}\Big[I(f) \mid \{I(f_l)\}_{(l\in\mathcal{I}}\Big] = I(f^\star) \quad \text{et} \quad \mathbb{V}\Big[I(f) \mid \{I(f_l)\}_{l\in\mathcal{I}}\Big] = \Big|f - f^\star\Big|^2_{\mathcal{L}_d^2}. \tag{4.297}$$

En appliquant les formules ci-dessus au cadre spécifique du mouvement Brownien fractionnaire, nous obtenons le résultat suivant.

Theorem 4.50. *Considérons la famille d'indices $\mathcal{I} = \{(i,l) \mid i \in [\![1,d]\!],\ l \in [\![1,N]\!]\}$ et l'ensemble de fonctions à valeurs $\mathcal{L}_d^2$*

$$f_l^i(\cdot) := \sigma_i R_{:,i} \psi_{t_l,t_{l+1}}^{H_i}(\cdot). \tag{4.298}$$

Soit

$$f(\cdot) = \sum_{i=1}^d \omega_{t_N}^i \psi_{t_N,t_{N+1}}^{H_i}(\cdot) \tag{4.299}$$

une fonction de $\mathcal{L}_d^2$, où chaque coefficient d-dimensionnel $\omega_{t_N}^i = \omega_i \sigma_i S_{t_N}^i R_{:,i}$ est mesurable par rapport à $\{I(f_l^i)\}_{(i,l)\in\mathcal{I}}$.

Alors, l'espérance de $I(f)$ conditionnellement à $\{I(f_l^i)\}_{(i,l)\in\mathcal{I}}$ peut être écrite comme

$$\mathbb{E}\Big[I(f)|\{I(f_l^i)\}_{(i,l)\in\mathcal{I}}\Big] = \sum_{(i,l)\in\mathcal{I}} a_l^i I(f_l^i) = \sum_{i=1}^d \sum_{l=1}^N a_l^i (B_{t_{l+1}}^{H_i} - B_{t_l}^{H_i}), \tag{4.300}$$

et la variance conditionnelle est donnée par :

$$\mathbb{V}\Big[I(f) \mid \{I(f_l^i)\}_{(i,l)\in\mathcal{I}}\Big] = \sum_{i,j=1}^d (\omega_{t_N}^i \cdot \omega_{t_N}^j) \int_{\mathbb{R}} \psi_{t_N,t_{N+1}}^{H_i}(u) \psi_{t_N,t_{N+1}}^{H_j}(u) \mathrm{d}u \tag{4.301}$$

$$- \sum_{i,j=1}^N \sum_{k,l=1}^N a_l^i a_k^j \rho_{ij} \int_{\mathbb{R}} \psi_{t_l,t_{l+1}}^{H_i}(u) \psi_{t_k,t_{k+1}}^{H_j}(u) \mathrm{d}u, \tag{4.302}$$

où les coefficients de projection $\alpha^\star = (a_1^1, \cdots, a_N^1, a_1^2, \cdots, a_N^2, \cdots, a_1^d, \cdots, a_N^d)^\top$ *sont donnés par :*

$$\alpha^\star = \mathfrak{M}^+ y, \tag{4.303}$$

où $\mathfrak{M}^+$ *est le pseudo-inverse de Moore-Penrose (voir (MacAusland, 2014, Section 5.5.4)) de la matrice*

$$\mathfrak{M} = \left(\langle f_{k_1}^{i_1}, f_{k_2}^{i_2} \rangle_{\mathcal{L}_d^2} \right)_{(i_1-1)\times N+k_1, (i_2-1)\times N+k_2}$$

et où

$$y = \left(\langle f, f_{k_2}^{i_2} \rangle_{\mathcal{L}_d^2} \right)_{(i_2-1)\times N+k_2}.$$

Parmi les coefficients de projection intervenant dans l'espérance conditionnelle, le coefficient $\alpha^\star$ *dans l'Équation (4.303) est celui avec la norme minimale.*

2) Supposons que la discrétisation temporelle soit uniforme, c'est-à-dire, $t_l = lh$ *pour* $l = 1, \ldots, N$*. La propriété d'auto-similarité permet de factoriser le noyau de covariance par une puissance du pas de temps* h*, et la quantité restante dans la factorisation représente le noyau de covariance au temps* 1 *(indépendant de* h*) entre les deux fBms donnés :*

$$\mathfrak{M}_{(i_1-1)*N+k_1,(i_2-1)*N+k_2} = \rho_{i_1,i_2} h^{H_{i_1}+H_{i_2}} \int_{\mathbb{R}} \psi_{0,1}^{H_{i_1}} \left(v - k_1\right) \psi_{0,1}^{H_{i_2}} \left(v - k_2\right) \mathrm{d}v, \tag{4.304}$$

$$y_{(i_2-1)N+k_2} = \sum_{i=1}^{d} (\omega_{t_N}^i \cdot R_{:,i_2}) h^{H_i+H_{i_2}} \int_{\mathbb{R}} \psi_{0,1}^{H_i} \left(v - N\right) \psi_{0,1}^{H_{i_2}} \left(v - k_2\right) \mathrm{d}v. \tag{4.305}$$

Les résultats théoriques sont ensuite appliqués au cadre de la gestion des risques en finance.

Cadre financier Nous présentons le modèle utilisé pour décrire la dynamique des prix. Habituellement, celle-ci est décrite à l'aide d'un modèle standard de Black-Scholes basé sur un mouvement Brownien classique. L'approche présentée dans ce travail est assez différente, car la dynamique des prix est décrite par un modèle de Black-Scholes fractionnaire. Ce choix de modèle est justifié par le fait que les incréments des log-prix restent Gaussiens, auto-similaires et stationnaires. Ainsi, des propriétés intéressantes peuvent être exploitées, d'une part, pour modéliser la trajectoire des incréments des log-prix avec précision et réalisme, en introduisant des corrélations entre eux qui expriment une dépendance à long-terme (respectivement à court-terme) ; et d'autre part, pour assurer la stabilité et la cohérence des estimateurs, ainsi que pour réaliser les calculs théoriques de variances-covariances.

De plus, les actifs que nous décrivons, tels que les taux de change, présentent des corrélations significatives entre les incréments des log-prix, qui peuvent être prises en compte par les propriétés fractales du mouvement Brownien fractionnaire (fBm). Une telle approche a déjà été adoptée dans les travaux de (Garcin, 2020) (en dimension $d = 1$), où le fBm est utilisé pour prendre en compte les corrélations positives entre les

rendements. Ici, nous proposons une extension multivariée – le modèle dit de Black-Scholes fractionnaire multivarié – pour modéliser la dynamique des prix dans un univers composé de plusieurs actifs corrélés à la fois en temps et en espace.

Supposons que nous ayons un univers composé de d actifs dont les prix de marché à l'instant t sont notés $\left(S_t^1, \ldots, S_t^d\right)$ et dont les logarithmes des prix de marché à l'instant t sont notés $\left(X_t^1, \ldots, X_t^d\right)$.

La dynamique des prix est décrite par un mouvement Brownien fractionnaire géométrique :

$$\forall i \in [\![1, d]\!],\ S_t^i = S_0^i e^{c_t^i + \sigma_i B_t^{i,H_i}} \text{ avec } c_t^i = \log\left(\frac{\mathbb{E}[S_t^i]}{S_0^i}\right) - \frac{\sigma_i^2}{2} t^{2H_i}, \tag{4.306}$$

où S_0 est fixé et connu, et c_t^i est le paramètre de centrage dans le modèle. Nous supposons que tous les paramètres sont connus grâce à leurs estimations à partir des données de marché.

Concentrons-nous sur les propriétés du processus des logarithmes des prix :

$$X_t^i = X_0^i + c_t^i + \sigma_i B_t^{i,H_i}. \tag{4.307}$$

Le processus des log-prix est une fonction affine par rapport au mouvement Brownien fractionnaire et, en tant que tel, il bénéficie des propriétés liées à l'auto-similarité et à la stationnarité des incréments, tout en relaxant la propriété d'indépendance des incréments. Pour tout $h > 0$, les incréments des log-prix de longueur h sont définis par (voir Équation (4.293)) :

$$\delta_h X_t^i := X_{t+h}^i - X_t^i = \delta_h c_t^i + \sigma_i I(\psi_{t,t+h}^{H_i} R_{:,i}). \tag{4.308}$$

Dans ce cadre, nous nous intéressons à la détermination de mesures de risque de l'incrément futur du prix d'un portefeuille à l'horizon temporel h. Le portefeuille et ses incréments de taille h sont définis par :

$$P_t = \sum_{i=1}^{d} \omega^i S_t^i \qquad \text{et} \qquad \delta_h P_t := P_{t+h} - P_t. \tag{4.309}$$

L'objectif est de calculer la $\mathbf{VaR}_\alpha$ conditionnelle de l'incrément futur du portefeuille $\delta_h P_{t_N} = P_{t_{N+1}} - P_{t_N}$, étant donné les observations passées $\{S_{t_0}^1, S_{t_1}^1, \ldots, S_{t_N}^1, \ldots, S_{t_0}^d, S_{t_1}^d, \ldots, S_{t_N}^d\}$ prises sur une partition uniforme de l'intervalle de temps $[0, t]$, de sorte que le pas de temps soit constant et égal à h, $t_0 = 0$, $t_N = t$ et pour tout $l \in [\![0, N]\!]$, $t_l = lh$. Comme tous les paramètres sont connus, observer les prix, les log-prix ou les incréments des log-prix fournit la même information, modélisée par la tribu

$$\mathcal{G} := \sigma(\delta_h X_{t_l}^i : (i,l) \in \mathcal{I}) = \sigma(I(f_l^i) : (i,l) \in \mathcal{I}) \qquad \text{où} \qquad f_l^i(\cdot) := \sigma_i R_{:,i} \psi_{t_l, t_{l+1}}^{H_i}(\cdot) \tag{4.310}$$

et $\mathcal{I} = \{(i,l) \mid i \in [\![1, d]\!],\ l \in [\![1, N]\!]\}$ est une famille d'indices permettant à l'indexation à la fois des actifs et des incréments temporels.

Conditionnellement à $\mathcal{G}$, $\delta_h \hat{P}_{t_N} = I(f)$, où la forme de la fonction f est définie par l'Équation (4.299) du Théorème 4.50 avec la pondération spécifique $\omega^i_{t_N} = \omega_i \sigma_i S^i_{t_N} R_{:,i}$:

$$f(\cdot) = \sum_{i=1}^{d} \omega_i \sigma_i S^i_{t_N} R_{:,i} \psi^{H_i}_{t_N, t_{N+1}}(\cdot), \tag{4.311}$$

suit une distribution Gaussienne. Plus précisément, conditionnellement à $\mathcal{G}$, $\delta_h \hat{P}_{t_N}$ suit une distribution Gaussienne caractérisée par son espérance conditionnelle et sa variance conditionnelle.

Modèle de VaR Pour calculer la $\mathbf{VaR}_\alpha$ conditionnelle de $\delta_h P_{t_N}$, nous devons connaître la distribution conditionnelle de l'incrément futur du portefeuille, compte tenu des observations passées. Cependant, $\delta_h P_{t_N}$ est une combinaison linéaire de variables aléatoires log-normales, dont les distributions marginales et conditionnelles sont inconnues. Nous proposons donc une approximation Gaussienne de la $\mathbf{VaR}_\alpha$ conditionnelle de $\delta_h P_{t_N}$, à l'aide de la $\mathbf{VaR}_\alpha$ conditionnelle de $\delta_h \hat{P}_t$, dont la distribution conditionnelle est Gaussienne. La $\mathbf{VaR}_\alpha$ Gaussienne est facilement déterminée, car elle repose seulement sur l'espérance et l'écart-type de la distribution Gaussienne.

$$\delta_h P_{t_N} = P_{t_N+h} - P_{t_N} = \sum_{i=1}^{d} \omega^i \delta_h S^i_{t_N} \simeq \sum_{i=1}^{d} \omega^i S^i_{t_N} \delta_h X^i_{t_N} =: \delta_h \hat{P}_{t_N}. \tag{4.312}$$

Theorem 4.51 (Value-at-Risk conditionnelle Gaussienne)**.** *Si $\delta_h \hat{P}_{t_N} = I(f)$, avec f donnée par l'Équation (??), représente l'incrément futur du portefeuille des log-prix, et si $(\delta_h X^i_{t_l} = I(f^i_l))_{(i,l)\in\mathcal{I}}$, avec f^i_l données par l'Équation (4.298) du Théorème 4.50, sont les incréments passés des log-prix des actifs qui composent l'univers, alors la $\mathbf{VaR}_\alpha$ conditionnelle Gaussienne est donnée par la formule suivante :*

$$\mathbf{VaR}_\alpha(\delta_h \hat{P}_{t_N} | \mathcal{G}) = \sqrt{\mathbb{V}[\delta_h \hat{P}_{t_N} | \mathcal{G}]} \mathcal{N}^{-1}(\alpha) + \mathbb{E}[\delta_h \hat{P}_{t_N} | \mathcal{G}] \tag{4.313}$$

où $\mathbb{E}[\delta_h \hat{P}_{t_N} | \mathcal{G}]$ et $\mathbb{V}[\delta_h \hat{P}_{t_N} - \mathbb{E}[\delta_h \hat{P}_{t_N} | \mathcal{G}]]$ sont données par le Théorème 4.50.

Modèle d'ES L'Expected-Shortfall conditionnel de l'approximation Gaussienne de l'incrément futur du portefeuille est donnée dans le théorème suivant.

Theorem 4.52 (Expected-Shortfall conditionnel Gaussien)**.** *Si $\delta_h \hat{P}_{t_N} = I(f)$, avec f donnée par l'Équation (??), représente l'incrément futur du portefeuille des log-prix, et si $(\delta_h X^i_{t_l} = I(f^i_l))_{(i,l)\in\mathcal{I}}$, avec f^i_l données par l'Équation (4.298) du Théorème 4.50, sont les incréments passés des log-prix des actifs qui forment l'univers, alors l'Expected-Shortfall conditionnel Gaussien est donné par la formule suivante :*

$$\mathbf{ES}_\alpha(\delta_h \hat{P}_{t_N} \mid \mathcal{G}) = \frac{\sqrt{\mathbb{V}[\delta_h \hat{P}_{t_N} | \mathcal{G}]}}{1-\alpha} n(\mathcal{N}^{-1}(\alpha)) + \mathbb{E}[\delta_h \hat{P}_{t_N} | \mathcal{G}] \tag{4.314}$$

où $n(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ et $N(x) = \int_{-\infty}^{x} n(y)\mathrm{d}y$ sont respectivement la fonction de densité Gaussienne centrée réduite et la fonction de répartition.

Quantification de l'erreur Nous fournissons une quantification de l'erreur de notre approximation Gaussienne. Plus précisément, l'objectif est de quantifier la précision de l'approximation de $\mathbf{VaR}_\alpha(\delta_h P_{t_N} \mid \mathcal{G})$ avec $\mathbf{VaR}_\alpha(\delta_h \hat{P}_{t_N} \mid \mathcal{G})$. En d'autres termes, nous cherchons à déterminer la borne supérieure de l'expression suivante :

$$\left|\mathbf{VaR}_\alpha(\delta_h P_{t_N} \mid \mathcal{G}) - \mathbf{VaR}_\alpha(\delta_h \hat{P}_{t_N} \mid \mathcal{G})\right|.$$

Theorem 4.53. *Considérons l'incrément futur du portefeuille $\delta_h P_{t_N}$ et son approximation Gaussienne conditionnelle à $\mathcal{G}$, donnée par le rendement logarithmique futur du portefeuille $\delta_h \hat{P}_{t_N}$. Ainsi, nous avons la borne supérieure suivante :*

$$\Delta = \left\|\delta_h P_{t_N} - \delta_h \hat{P}_{t_N}\right\|_{p,\mathcal{G}} \leq 2 \sum_{i=1}^{d} \left|\omega^i\right| S^i_{t_N} \left(\sigma_i^2 h^{2H_i} C_{2pq}^{1/pq} + M_i^2\right) e^{(M_i)_+ + \frac{pr}{2}\sigma_i^2 h^{2H_i}} \tag{4.315}$$

avec $p > 1$ et $\frac{1}{q} + \frac{1}{r} = 1$.

Si nous supposons que Δ est suffisamment petit, c'est-à-dire $|\Delta| < 1 - \alpha$ p.s., alors :

(i) *La quantification de l'approximation de la $\mathbf{VaR}_\alpha$ conditionnelle de l'incrément futur du prix du portefeuille $\delta_h P_{t_N}$ par la $\mathbf{VaR}_\alpha$ Gaussienne conditionnelle de l'incrément futur du portefeuille des log-prix $\delta_h \hat{P}_{t_N}$ est donnée comme suit :*

$$\left|\mathbf{VaR}_\alpha(\delta_h P_{t_N} \mid \mathcal{G}) - \mathbf{VaR}_\alpha(\delta_h \hat{P}_{t_N} \mid \mathcal{G})\right| \leq \left(\frac{2\sigma(\mathcal{G})}{1-\alpha-\Delta} \sum_{i=1}^{d} \left|\omega^i\right| S^i_{t_N} \right. \tag{4.316}$$

$$\left. \times \left(\sigma_i^2 h^{2H_i} C_{2pq}^{1/pq} + M_i^2\right) e^{(M_i)_+ + \frac{pr}{2}\sigma_i^2 h^{2H_i}}\right), p.s. \tag{4.317}$$

avec $M_i := \mathbb{E}\left[\delta_h X^i_{t_N} \mid \mathcal{G}\right]$, $\sigma(\mathcal{G}) = \sqrt{\mathbb{V}[\delta_h \hat{P}_{t_N} | \mathcal{G}]}$, $G \sim \mathcal{N}(0,1)$, et $C_{2pq} := \mathbb{E}\left[G^{2pq}\right]$.

(ii) *La quantification de l'approximation de l'$\mathbf{ES}_\alpha$ conditionnel de l'incrément futur du prix du portefeuille $\delta_h P_{t_N}$ par l'$\mathbf{ES}_\alpha$ Gaussien conditionnel de l'incrément futur du portefeuille des log-prix $\delta_h \hat{P}_{t_N}$ est donnée comme suit :*

$$\left|\mathbf{ES}_\alpha(\delta_h P_{t_N} \mid \mathcal{G}) - \mathbf{ES}_\alpha(\delta_h \hat{P}_{t_N} \mid \mathcal{G})\right| \tag{4.318}$$

$$\leq \left[\left(1 + \frac{1}{1-\alpha}\right)\frac{2\sigma(\mathcal{G})}{1-\alpha-\Delta} + \frac{2}{1-\alpha}\right] \sum_{i=1}^{d} \left|\omega^i\right| S^i_{t_N} \left(\sigma_i^2 h^{2H_i} C_{2pq}^{\frac{1}{pq}} + M_i^2\right) e^{(M_i)_+ + \frac{pr}{2}\sigma_i^2 h^{2H_i}}. \tag{4.319}$$

Estimation des paramètres Notre modèle prédictif comprend les paramètres suivants : $\sigma = (\sigma_1, \ldots, \sigma_d)$ représentant le vecteur de volatilités de marché associées à chaque actif, $H = (H_1, \ldots, H_d)$ étant le vecteur des exposants de Hurst qui contrôlent la régularité des trajectoires de prix de chaque actif, et $R = \{\rho_{ij}, i, j = 1, \ldots d\}$ désignant la matrice des corrélations entre les actifs i et j. Nous considérons la suite d'observations $(X^1_{t_0}, \ldots, X^1_{t_N}, \ldots, X^d_{t_0}, \ldots, X^d_{t_N})$ représentant les trajectoires des log-prix des d actifs.

Pour faciliter l'estimation, nous supposons que les observations sont uniformément espacées dans le temps, de sorte que $t_j = j\tau$, où τ est le pas de temps minimal. Le modèle de VaR est testé à posteriori, d'abord sur des données synthétiques, puis sur des données réelles. Par conséquent, nous estimons les paramètres du modèle à partir des données synthétiques dans un premier temps, puis à partir des données réelles. La méthodologie d'estimation repose sur trois hypothèses clés : *Gaussianité*, *stationnarité* et *autosimilarité*. Ces hypothèses sont nécessaires pour estimer les paramètres.

L'*hypothèse de Gaussianité* stipule que pour chaque actif, les log-prix sont assimilés à un fBm. Ainsi, les incréments des log-prix présentent le même comportement que les incréments du fBm. Ainsi, les incréments des log-prix forment un vecteur Gaussien et toute combinaison linéaire des incréments est une variable Gaussienne. L'approximation de la variation future du portefeuille des prix des actifs $\delta_h \hat{P}_{t_N}$ par la variation future du portefeuille des log-prix avec une pondération bien choisie $\delta \hat{P}_{t_N}$, conditionnellement aux incréments passés des log-prix, est une variable Gaussienne. Par conséquent, la $\mathbf{VaR}_\alpha$ conditionnelle de $\delta_h \hat{P}_{t_N}$, étant donné les rendements passés des actifs, est le quantile d'une distribution Gaussienne. Plus précisément, $\mathbf{VaR}_\alpha(\delta_h \hat{P}_N \mid \mathcal{G})$ est une fonction affine du quantile α de la distribution Gaussienne centrée réduite, dont l'ordonnée à l'origine est donnée par l'espérance conditionnelle $\mu(\mathcal{G}) = \mathbb{E}\left[\delta_h \hat{P}_N \mid \mathcal{G}\right]$ et dont la pente est la racine carrée de la variance conditionnelle $\sigma(\mathcal{G}) = \sqrt{\mathbb{V}[\delta_h \hat{P}_N \mid \mathcal{G}]}$.

L'*hypothèse de stationnarité* établit l'invariance par changement d'origine temporelle, ou translation, des propriétés et des caractéristiques des familles de variables aléatoires de $\mathbb{L}^2$. En particulier, la propriété de stationnarité établit l'invariance par changement d'origine temporelle des espérances, variances et covariances, du processus des incréments du fBm et donc des log-prix. La propriété de stationnarité implique que les estimateurs des espérances, variances et covariances, appliqués au processus des incréments restent les mêmes lorsque l'origine temporelle du processus des incréments change ; ainsi, les estimations des paramètres effectuées sur le processus des incréments sont stables et fiables.

L'*hypothèse d'autosimilarité* établit une relation de proportionnalité spatiale entre les caractéristiques, en particulier les espérances au carré, variances et covariances, du processus lui-même et entre les caractéristiques du processus des incréments, pris à deux échelles temporelles proportionnelles λt et t, avec comme facteur de proportionnalité spatiale une fonction puissance de λ : λ^{2H_i}, $H_i \in (0, 1)$. Cela implique que les estimateurs des espérances au carré, variances et covariances appliqués au processus des incréments pris à deux échelles temporelles proportionnelles λt et t, sont liés par une relation de proportionnalité spatiale de facteur λ^{2H_i}, ce qui permet une adaptation facile des estimateurs à différentes échelles temporelles. La propriété d'autosimilarité permet d'obtenir des estimations fiables, cohérentes avec les différentes échelles temporelles.

Sous ces hypothèses, la méthode d'estimation des paramètres du mfBm consiste à calculer, pour chaque composante du mfBm, la variance empirique des incréments des log-prix pour différentes tailles d'incréments, en suivant l'estimateur basé sur les moments du fBm introduit par (Istas and Lang, 1994; Kent and Wood, 1997).

La variation quadratique des incréments de taille $m\tau$ avec $m \in \mathbb{N}^\star$ est respectivement donnée pour les incréments disjoints et pour les incréments chevauchants, comme suit :

$$V_m^i = \frac{1}{\left\lfloor N/m \right\rfloor} \sum_{j=0}^{\left\lfloor \frac{N}{m} \right\rfloor - 1} (X^i_{(j+1)m\tau} - X^i_{jm\tau})^2 \quad \text{et} \quad V_m^i = \frac{1}{N-m+1} \sum_{j=0}^{N-m} (X^i_{(j+m)\tau} - X^i_{j\tau})^2. \tag{4.320}$$

Deux méthodes alternatives d'estimation, quasiment équivalentes, ont été étudiées. La première méthode, basée sur les travaux de (Amblard and Coeurjolly, 2011a), consiste en la régression linéaire des log-variances par rapport aux log-échelles temporelles. L'estimation de l'exposant de Hurst est donnée par la moitié de la pente de la régression linéaire, et celle de la volatilité par l'exponentielle de la moitié de l'ordonnée à l'origine. Une autre méthode d'estimation basée sur les niveaux d'énergie, étudiée en détails dans (Chong et al., 2022a, p.12), (Chong et al., 2022b), et (Szymanski and Takabatake, 2023), est également exploitée.

Rétrotest (Backtesting) Nous proposons une procédure de rétrotest pour évaluer la performance de notre modèle à travers des expériences numériques.

Nous considérons un ensemble d'informations $\mathfrak{I}_{t-1} = \{\mathcal{I}_{t-1}, \mathcal{I}_{t-2}, \mathcal{I}_{t-3}, \ldots, \mathcal{I}_1\}$ qui consiste en une séquence de fonctions indicatrices, où la variable de dépassement $\mathcal{I}_t$ à l'instant t est une variable de Bernoulli égale à 1 si le rendement entre les instants t et $t+h$ dépasse la VaR prédite à l'instant t, et à 0 sinon. Plus précisément, nous obtenons les variables de dépassement en utilisant une approche par fenêtres glissantes.

Pour chaque fenêtre de taille w, nous prédisons à l'instant t la VaR conditionnelle à l'horizon h, étant donné l'historique de la fenêtre $[t-w, t]$, et nous la comparons au rendement logarithmique réalisé entre les instants t et $t+h$, générant ainsi une variable de dépassement égale à 1 ou 0.

Selon la littérature traditionnelle sur les rétrotests de la VaR (Christoffersen, 1998; Davis, 2016), une $\mathbf{VaR}_\alpha$ satisfaisante doit être telle que $\mathbb{E}[\mathcal{I}_t|\mathfrak{I}_{t-1}] = \alpha$. En d'autres termes, nous évaluons deux critères sur les VaR générées : la couverture conditionnelle et l'indépendance des variables de dépassement ; il doit être prouvé que $\{\mathcal{I}_t\} \overset{iid}{\sim} \mathcal{B}ern(\alpha)$.

En pratique, le test de couverture consiste à estimer l'espérance conditionnelle de la séquence des variables de dépassement, qui doit correspondre au niveau de risque souhaité α. Cette estimation est réalisée en calculant la moyenne empirique des variables de dépassement. Plus cette dernière est proche du niveau de risque souhaité α, plus le test de couverture conditionnelle est satisfaisant.

Concernant l'indépendance, nous proposons une méthode inspirée de la théorie des valeurs extrêmes, initialement proposée par (Bücher et al., 2020). En effet, deux événements extrêmes deviennent approximativement indépendants s'ils sont séparés par un

laps de temps suffisant. Pour mesurer le degré de dépendance dans une série temporelle, nous utilisons l'indice extrémal.

L'indice extrémal θ est un indicateur capable de quantifier le degré de dépendance dans une série temporelle de variables aléatoires, en comptant le nombre de regroupements de variables au-dessus d'un seuil prédéterminé, à savoir les valeurs extrêmes. Il est égal à un si la séquence est indépendante, ce qui signifie qu'il n'y a pas de regroupement de valeurs extrêmes. Plus l'indice extrémal se rapproche de zéro, plus les regroupements de valeurs extrêmes sont nombreux et plus la dépendance dans la série est grande.

Dans le cadre du rétrotest de la VaR, nous appliquons cette procédure à la séquence des variables de dépassement, en considérant qu'une valeur extrême correspond à un dépassement égal à un. Ainsi, l'indice extrémal quantifie la présence de regroupements de violations de la VaR.

Un estimateur de l'indice extrémal peut être fourni en utilisant soit un algorithme de dégroupement par blocs (block declustering), soit un algorithme de dégroupement par fenêtres glissantes (run declustering). Plus précisément, étant donné une taille de bloc arbitraire b, nous divisons la séquence, composée de n variables de dépassement successives, en $k = \left\lfloor \frac{n}{b} \right\rfloor$ blocs disjoints (dégroupement par blocs) ou en $n - b$ fenêtres glissantes (dégroupement par fenêtres glissantes). Dans l'approche par dégroupement par blocs, chaque bloc contenant des variables de dépassement non-nulles représente un regroupement. Dans l'approche par dégroupement par fenêtres glissantes, nous comptons le nombre de fenêtres sans variable de dépassement non-nulle à partir de la fin d'un regroupement. Avec ces deux méthodes, l'estimateur de l'indice extrémal est le quotient entre le nombre estimé de regroupements et le nombre de variables de dépassement non nulles. Formellement, si nous notons $M_{i,j} = \max\{\mathcal{I}_{i+1}, \ldots, \mathcal{I}_j\}$, les deux estimateurs sont définis comme suit :

$$\hat{\theta}_n^B(b) = \frac{\sum_{i=1}^{k} \mathbb{1}_{\{M_{(i-1)b,ib}=1\}}}{\sum_{i=1}^{kb} \mathbb{1}_{\{\mathcal{I}_i=1\}}} \quad \text{et} \quad \hat{\theta}_n^R(b) = \frac{\sum_{i=1}^{n-b} \mathbb{1}_{\{\mathcal{I}_i=1, M_{i,i+b}=0\}}}{\sum_{i=1}^{n-b} \mathbb{1}_{\{\mathcal{I}_i=1\}}}. \tag{4.321}$$

Le processus de rétrotest permet de comparer trois modèles de VaR : la VaR conditionnelle (VaR modèle), la VaR Gaussienne (nous forçons H à être égal à $0,5$), et la VaR empirique. Dans la plupart des cas, la VaR conditionnelle se distingue en termes de taux de couverture et d'indice extrémal, suivi par la VaR Gaussienne puis la VaR empirique. La VaR la plus basse est presque toujours la VaR empirique, mais son taux de couverture et son indice extrémal sont très éloignés des valeurs souhaitées.

4.8.4 Chapter 4 - Estimation de l'Expected-Shortfall pour des distributions à queues lourdes

Etat de l'art

Le cadre Gaussien est le plus couramment utilisé en finance en raison de ses propriétés pratiques. Cependant, la distribution gaussienne a une queue fine, ce qui suppose que les événements extrêmes sont rares, et tend à sous-estimer la probabilité de tels événements.

En revanche, les distributions à queue lourde comme Pareto attribuent des probabilités plus élevées aux événements extrêmes, ce qui est souvent plus réaliste en finance. L'un des inconvénients de la VaR est que la VaR n'est pas sensible au risque de queue, ce qui conduit souvent à une sous-estimation du risque en ne capturant pas les événements extrêmes. Lorsque la distribution des pertes a une queue fine, comme la distribution gaussienne, les événements extrêmes se produisent avec une probabilité très faible, alors la VaR reste efficace malgré ses lacunes. Cependant, lorsque les pertes sont décrites grâce à une distribution à queue lourde, les événements extrêmes se produisent plus fréquemment, et la VaR n'est plus pertinente. Contrairement à la VaR, le Expected-Shortfall fournit une mesure de l'ampleur moyenne des pertes au-delà du seuil de VaR, en se concentrant sur le risque de queue. Cela est crucial pour les institutions afin de comprendre l'étendue potentielle des pertes lors d'événements extrêmes, que la VaR ne capture pas pleinement. Par conséquent, lorsque les pertes sont décrites par des distributions à queue lourde, comme Pareto, ES est plus approprié que VaR pour fournir une évaluation précise du risque.

Littérature sur l'Expected-Shortfall La littérature sur l'Expected-Shortfall comme solution pour remédier aux lacunes de la Value-at-Risk est vaste. Dans un contexte de renforcement de la réglementation bancaire, de nombreux auteurs se sont intéressés à comparer VaR et ES afin de comprendre leurs forces et faiblesses (Kellner and Rösch, 2016). En particulier, les travaux de (Acerbi et al., 2001), (Acerbi and Tasche, 2002b), (Artzner et al., 1999) sont des références intéressantes.

Certains auteurs se sont concentrés sur le célèbre débat entre VaR et ES. Ce débat découle du fait que la VaR n'est pas une mesure de risque cohérente en raison de son manque de sous-additivité, mais la VaR est élicitable alors que ES est une mesure de risque cohérente mais n'est pas élicitable. D'une part, le manque de sous-additivité dans la VaR peut avoir des conséquences significatives, principalement dans la gestion des risques et les processus de prise de décision au sein des institutions financières ou de toute organisation exposée aux risques de marché. La sous-additivité fait référence à la propriété selon laquelle le risque total d'un portefeuille est inférieur ou égal à la somme des risques de ses composants individuels. Cependant, si la VaR manque de sous-additivité, cela signifie que le risque combiné d'un portefeuille pourrait être supérieur à la somme des risques de ses parties individuelles, ce qui peut conduire à une sous-estimation du risque, des stratégies de couverture inexactes, des réserves de capitaux inadéquates, une allocation d'actifs mal orientée, voire une amplification du risque systémique. C'est le sujet des articles (Embrechts, 2000), (Tibiletti, 2008).

Dans (Garcia et al., 2007), des méthodes sont proposées pour résoudre le manque de sous-additivité de la VaR. Les travaux de (Daníelsson et al., 2013) montrent que la VaR peut être sous-additive dans certaines régions des distributions à queue lourde. Pour résoudre les problèmes conceptuels causés par la VaR, (Artzner et al., 1999), (Tasche, 2002a) présentent une mesure de risque alternative à la VaR, appelée ES, et mettent en évidence les avantages de ES qui permettent de pallier les lacunes de la VaR. L'article (Inui and Kijima, 2005) montre que toute mesure de risque cohérente est donnée par une combinaison convexe de expected shortfalls, et un expected shortfall (ES) est op-

timal en ce sens qu'il donne la valeur minimale parmi la classe de mesures de risque cohérentes plausibles. Une méthode d'extrapolation de l'ES est fournie. Les auteurs discutent des propriétés de l'ES ainsi que de sa généralisation à une classe de mesures de risque cohérentes qui peuvent incorporer des effets de moments plus élevés. Les travaux de (Yamai and Yoshiba, 2005) illustrent comment le risque de queue de la VaR peut poser de sérieux problèmes dans certains cas, cas dans lesquels l'Expected-Shortfall peut servir plus adéquatement à sa place. Ils discutent spécifiquement de deux cas : portefeuille de crédits concentré et taux de change étrangers sous stress de marché. Ils montrent que l'Expected Shortfall nécessite une plus grande taille d'échantillon que la VaR pour fournir le même niveau de précision. Les travaux de (Acerbi and Tasche, 2002a) mettent en évidence le fait que la plupart des définitions de ES conduisent aux mêmes résultats lorsqu'elles sont appliquées à des distributions de pertes continues, mais des différences peuvent apparaître lorsque les distributions de pertes sous-jacentes présentent des discontinuités. Dans ce cas, même la propriété de cohérence de ES peut être perdue à moins de prendre soin des détails dans sa définition. Les auteurs comparent certaines des définitions de ES, soulignant qu'il en existe une qui est robuste en ce sens qu'elle produit une mesure de risque cohérente quelles que soient les distributions sous-jacentes. De plus, cet ES peut être estimé efficacement même dans les cas où les estimateurs habituels de la VaR échouent. Basé sur l'observation que la mesure de performance ajustée au risque du portefeuille implique le calcul de la contribution au risque pour chaque actif qu'il contient. L'article (Fan et al., 2012) utilise des fonctions de Copule multivariées pour modéliser la structure de dépendance entre les actifs d'un portefeuille, puis, sur la base d'une simulation, décompose la VaR du portefeuille et l'Expected Shortfall. De plus, avec cette approche, la contribution au risque calculée à l'aide de l'Expected Shortfall est plus robuste, et son erreur d'estimation peut être réduite en augmentant la taille de l'échantillon de simulation. Une application de l'Expected-Shortfall au risque de crédit est fournie dans (Fan et al., 2010).

Cependant, l'Expected-Shortfall ne satisfait pas la propriété d'élicitabilité. L'élicitabilité est un concept dans le domaine de la mesure du risque et de la théorie de la décision, particulièrement dans le contexte des règles de notation et de l'évaluation des risques. Une règle de notation est une fonction utilisée pour évaluer l'exactitude d'une prévision probabiliste ou d'une estimation. L'élicitabilité se réfère à la propriété d'une règle de notation qui dicte si elle peut évaluer avec précision la qualité des prévisions probabilistes ou des estimations sans fournir d'incitations aux prévisionnistes pour déformer leurs croyances. De plus, l'élicitabilité permet de comparer différents modèles pour classer leur performance grâce à la règle de notation qui est pratique pour tester en arrière les modèles. Le manque d'élicitabilité peut entraîner plusieurs conséquences parmi lesquelles on peut trouver les incitations à la déformation, la difficulté d'évaluation, le potentiel de risque systémique et la perte de confiance. L'opinion était partagée sur le besoin d'élicitabilité dans les tests de rétroaction. Selon certains auteurs, l'élicitabilité n'est utile que mais pas nécessaire pour tester en arrière les modèles tandis que selon d'autres, l'élicitabilité est nécessaire et dans ce cas il est préférable d'utiliser VaR au lieu de ES. Ce débat est par exemple présenté dans (Acerbi and Szekely, 2014). Dans (Fissler et al., 2015), les auteurs montrent que ES est élicitable conjointement avec VaR.

Un enjeu important réside dans le choix de la méthode d'estimation de l'Expected-Shortfall. De nombreux auteurs se sont intéressés à ce sujet. Dans (Brazauskas et al., 2008), les auteurs développent des outils d'inférence statistique pour estimer et comparer les fonctions de conditionnement de l'espérance de queue (CTE), qui présentent un intérêt considérable en science actuarielle. Dans (Chen, 2008), deux estimateurs non paramétriques de l'Expected-Shortfall pour des pertes financières dépendantes sont présentés. L'un est une moyenne échantillonnale des pertes excessives supérieures à une VaR. L'autre est une version lissée par noyau du premier estimateur, espérant que l'estimation soit plus précise par le lissage. Dans (Taylor, 2008), une méthode d'estimation de VaR et ES grâce aux expectiles est fournie. Une estimation non paramétrique de ES est proposée dans (Scaillet, 2004). Une méthode d'estimation robuste de ES est développée dans (Jadhav et al., 2009) et (Pan et al., 2019). Les auteurs développent une approximation normale basée sur la queue avec des formules explicites dérivées en faisant correspondre un quantile spécifique et le carré de l'excès moyen des observations de l'échantillon. Pour améliorer la précision de l'estimation, ils proposent une approximation normale basée sur la queue ajustée en fonction du poids de la queue de l'échantillon. L'estimateur ajusté de l'Expected Shortfall est robuste et efficace en ce sens qu'il peut être appliqué à diverses distributions à queue lourde, telles que Student, lognormale, Gamma et Weibull, et les erreurs sont toutes faibles.

Littérature sur les distributions à queues lourdes Le cadre gaussien est le plus couramment utilisé en finance en raison de ses propriétés pratiques. Cependant, la distribution gaussienne a une queue fine, ce qui suppose que les événements extrêmes sont rares, et tend à sous-estimer la probabilité de tels événements. En revanche, les distributions à queue lourde comme Pareto attribuent des probabilités plus élevées aux événements extrêmes, ce qui est souvent plus réaliste en finance. Les distributions à queue lourde ne sont pas seulement utilisées en économie et en finance mais dans de nombreux autres domaines tels que la physique, la biologie, les sciences de la Terre et planétaires, l'informatique, la démographie et les sciences sociales. De nombreux auteurs se sont intéressés à l'étude des distributions à queue lourde. Par exemple, les distributions des tailles des villes, des tremblements de terre, des incendies de forêt, des éruptions solaires, des cratères lunaires et des fortunes personnelles des individus semblent toutes suivre des distributions à queue lourde. En finance, les distributions à queue lourde sont dédiées à modéliser les distributions de pertes ou de rendements.

Dans (ZINCHENKO, 2001), les auteurs fournissent une synthèse des principales tendances dans les investigations théoriques et les applications pratiques des modèles à queue lourde, en mettant l'accent sur les distributions subexponentielles, de type Pareto et stables. Certains problèmes liés aux théorèmes limites, à l'approximation, à l'estimation, à la simulation numérique pour les queues lourdes sont également traités ainsi que la connexion avec la théorie du risque. Les distributions à queue grasse en économie et en finance sont étudiées en détail dans (Haas and Pigorsch, 2009), (Broda and Paolella, 2011). Basé sur l'observation qu'il est de grande importance pour ceux en charge de la gestion des risques de comprendre comment les rendements des actifs financiers sont distribués. Les praticiens supposent souvent par commodité que la distribution est

normale. Depuis les années 1960, cependant, des preuves empiriques ont conduit beaucoup à rejeter cette hypothèse en faveur de diverses alternatives à queue lourde. Dans une distribution à queue lourde, la probabilité de rencontrer des écarts significatifs par rapport à la moyenne est beaucoup plus grande que dans le cas de la distribution normale. Il est désormais communément admis que les rendements des actifs financiers sont, en fait, à queue lourde. Dans (Bradley and Taqqu, 2003), une enquête est menée pour examiner comment ces queues lourdes affectent plusieurs aspects de la théorie des portefeuilles financiers et de la gestion des risques. Certaines des méthodes que l'on peut utiliser pour traiter les queues lourdes sont présentées et illustrées à l'aide de l'indice composite NASDAQ. De la même manière, l'article de (Guo, 2017) montre que la distribution à queue lourde, qui estime avec précision le risque de queue, améliorerait considérablement la pratique de la gestion quantitative des risques. Dans (Nolan, 2014), une introduction accessible aux distributions stables pour la modélisation financière est fournie. Il est vraiment nécessaire d'utiliser de meilleurs modèles pour les rendements financiers car le modèle normal (ou courbe en cloche / gaussien) ne capture pas les grandes fluctuations observées dans les actifs réels. Les lois stables sont une classe de distributions de probabilité à queue lourde qui peuvent modéliser de grandes fluctuations et permettre des structures de dépendance plus générales. Les travaux de (Peng and Qi, 2017) présentent des méthodes d'inférence pour les données à queue lourde et les appliquent dans le contexte de l'assurance et de la finance. De nombreuses données de sinistres d'assurance sont connues pour être à queue lourde. Dans l'article (Ahn et al., 2012), la classe des distributions Log-phase-type (LogPH) est étudiée comme une alternative paramétrique pour ajuster les données à queue lourde. Transformée de la classe populaire de distribution de type phase, la LogPH introduite par Ramaswami présente plusieurs avantages par rapport à d'autres alternatives paramétriques.

Littérature sur les distributions de Pareto La distribution de Pareto est une distribution bien connue qui appartient à la classe des distributions à queue lourde. La distribution de Pareto est souvent utilisée dans le contexte économique et financier pour modéliser la distribution des pertes ou des rendements lorsque des événements extrêmes se produisent plus fréquemment que dans le cadre gaussien. Une telle distribution présente des propriétés très pratiques telles que la stabilité par conditionnement et les propriétés d'échelle qui seront détaillées dans la suite. Plus spécifiquement, la distribution de Pareto est une distribution de loi de puissance. Lorsque la probabilité de mesurer une valeur particulière d'une quantité varie de manière inverse à une puissance de cette valeur, on dit que la quantité suit une loi de puissance, également connue sous le nom de loi de Zipf ou de distribution de Pareto. Dans (Newman, 2005), les auteurs passent en revue certaines des preuves empiriques de l'existence de formes de lois de puissance et des théories proposées pour les expliquer. Une référence intéressante sur la distribution de Pareto est (Chattamvelli and Shanmugam, 2021). Les distributions de Pareto et les généralisations associées ont historiquement été considérées comme adaptées pour modéliser les distributions de revenus et de richesse. Dans ce contexte, dans (Arnold, 2014), une brève revue de l'histoire de ces modèles, des propriétés de distribution et des procédures d'inférence est effectuée. Diverses distributions connexes, y compris des variantes multivariées, y sont décrites. Les travaux de (Arnold, 2008)

fournissent une synthèse des résultats liés à ces modèles de type Pareto, y compris la discussion de questions de distribution et d'inférence connexes. Les sujets abordés incluent les modèles de Pareto classiques et leurs généralisations, les modèles de revenus stochastiques conduisant à des distributions de revenus de type Pareto, les propriétés de distribution des distributions de Pareto généralisées, les distributions discrètes connexes, les mesures d'inégalité pour les modèles de Pareto, les problèmes d'inférence et les extensions multivariées. La plupart du temps, cette parétianité est inférée de l'observation de certains graphiques, tels que le graphique de Zipf et le graphique du surplus moyen. Si le graphique de Zipf semble presque linéaire, alors tout va bien et les paramètres de la distribution de Pareto sont estimés. Souvent avec les MCO. Malheureusement, ces outils graphiques heuristiques ne sont pas fiables. C'est ce que montrent les auteurs dans (Cirillo, 2013). En effet, ils montrent que seule une combinaison de graphiques peut donner un certain degré de confiance quant à la réelle présence de la parétianité dans les données. Dans (Crovelli and Barton, 1995), les fractales et la distribution de Pareto sont appliquées aux distributions de taille des accumulations pétrolières. Dans (Su and Furman, 2017), une nouvelle distribution multivariée possédant des marges univariées de Pareto paramétrées de manière arbitraire et positivement dépendantes est introduite.

Littérature sur l'estimation de l'indice de Pareto L'indice de Pareto gère le caractère lourd ou léger de la queue de la distribution. Un défi est de l'estimer avec précision.

En statistique des valeurs extrêmes, l'indice de valeurs extrêmes est un paramètre bien connu pour mesurer la lourdeur de la queue d'une distribution. Les distributions de type Pareto, avec un indice de valeurs extrêmes (ou indice de queue) strictement positif, sont considérées. Dans (Rytgaard, 1990), différents estimateurs du paramètre de Pareto sont proposés et comparés les uns aux autres. Tout d'abord, des estimateurs traditionnels tels que l'estimateur du maximum de vraisemblance et l'estimateur par moments sont déduits et leurs propriétés statistiques sont analysées. Il est montré que l'estimateur du maximum de vraisemblance est biaisé mais peut facilement être modifié en un estimateur non biaisé de variance minimale. Mais le coefficient de variation de cet estimateur reste très élevé. Pour des portefeuilles similaires contenant les mêmes types de risques, les valeurs estimées sont censées être au même niveau. Par conséquent, la théorie de la crédibilité est utilisée pour obtenir un estimateur alternatif plus stable et moins sensible aux fluctuations aléatoires des pertes observées. Enfin, un estimateur de la prime de risque pour une couverture de pertes illimitée est proposé. Il est démontré que cet estimateur est un estimateur de variance minimale non biaisé de la prime de risque. Cet estimateur de la prime de risque est comparé aux méthodes plus traditionnelles de calcul de la prime de risque. Dans (Crovella and Taqqu, 1999), les auteurs proposent une méthode (appelée "estimateur de mise à l'échelle") basée sur les propriétés de mise à l'échelle des sommes de variables aléatoires à queue lourde. Plus précisément, pour toute variable aléatoire X, Σ_n est définie comme une variable aléatoire qui est la somme de n variables aléatoires indépendantes ayant chacune la même distribution que X. Pour les distributions à queue lourde avec un indice de queue α, des théorèmes

limites similaires aux théorèmes limites centraux usuels peuvent être formulés montrant que les sommes de telles variables convergent vers des distributions stables avec le même α. Une distribution est stable au sens strict si pour chaque n il existe des constantes $c_n > 0$ telles que :

$$\Sigma_n \stackrel{d}{=} c_n X. \tag{4.322}$$

Dans le cas spécifique de la somme de variables stables indépendantes, $c_n = n^{\frac{1}{\alpha}}$. Cette propriété est la propriété de mise à l'échelle sur laquelle la méthode présentée dans cet article est basée pour obtenir l'estimateur de l'indice à queue lourde α. Il présente l'avantage d'être non paramétrique, facile à appliquer, de fournir une seule valeur et d'être relativement précis sur des ensembles de données synthétiques. Comme la méthode repose sur la mise à l'échelle des sommes, elle mesure une propriété qui est souvent l'un des effets les plus importants du comportement à queue lourde. Plus important encore, ils présentent des preuves selon lesquelles l'estimateur de mise à l'échelle semble augmenter en précision à mesure que la taille de l'ensemble de données augmente. Il est donc particulièrement adapté aux grands ensembles de données. L'un des problèmes les plus importants dans l'estimation des indices de Pareto est la réduction du biais en cas de disparition de la partie à variation lente du modèle de type Pareto à un rythme très lent. Dans d'autres cas, lorsque le problème de biais n'est pas aussi grave, l'application d'estimateurs bien connus tels que l'estimateur de Hill (1975) et l'estimateur par moments (Dekkers et al. (1989)) demande toujours une sélection adaptative de la fraction d'échantillon à utiliser dans de telles procédures d'estimation. Dans (Beirlant et al., 1996), (Beirlant et al., 1999), (Ocran et al., 2022), des estimateurs à biais réduit pour l'estimation de l'indice de queue d'une distribution de type Pareto sont proposés. Cela est réalisé grâce à l'utilisation d'une méthode des moindres carrés pondérés régularisée avec un modèle de régression exponentielle pour les espacements logarithmiques des principales statistiques d'ordre. Les propriétés asymptotiques des estimateurs proposés sont investiguées analytiquement et trouvées être asymptotiquement non biaisées, cohérentes et normalement distribuées. Les travaux de (Finkelstein et al., 2006) proposent un estimateur de l'indice de queue d'une distribution de Pareto basé sur l'utilisation de la transformation intégrale de probabilité. Cet nouvel estimateur offre des performances comparables aux meilleurs estimateurs robustes, tout en conservant une simplicité conceptuelle et computationnelle. Un paramètre de réglage dans le nouvel estimateur peut être ajusté pour contrôler le compromis entre robustesse et efficacité. Un nouvel estimateur de type médiane généralisé est introduit dans (Brazauskas and Serfling, 2000) et comparé au maximum de vraisemblance et à plusieurs estimateurs bien établis associés aux méthodes des moments, de la suppression, des moindres carrés, des quantiles et de l'ajustement de percentile. Un estimateur robuste de l'indice de queue est proposé dans (Vandewalle et al., 2007), en combinant un affinement de l'approximation de Pareto pour la distribution conditionnelle des excès relatifs au-dessus d'un seuil élevé avec une approche d'erreur quadratique intégrée sur l'estimation des composantes de densité partielle. Le document (Brzezinski, 2016) examine les propriétés des petits échantillons des estimateurs robustes les plus populaires pour l'indice de queue de Pareto, y compris l'estimateur B-robuste optimal, l'estimateur du maximum de vraisemblance pondéré, l'estimateur médian généralisé, l'estimateur des composantes de densité partielle et l'estimateur statistique de transformation intégrale

de probabilité (PITSE). Des simulations Monte Carlo montrent que le PITSE offre le compromis souhaité entre facilité d'utilisation et capacité à protéger contre les valeurs aberrantes dans le cadre des petits échantillons. Dans (Beirlant and Goegebeur, 2004), une approche motivée par le fait que dans certaines applications le seuil devrait être autorisé à changer avec les covariables en raison d'effets significatifs sur l'échelle et la localisation des distributions conditionnelles, est proposée. L'approche suivie est basée sur la technique de l'estimation du maximum de vraisemblance polynomiale local. En utilisant les résultats asymptotiques, ils sont en mesure de dériver une expression pour l'erreur quadratique moyenne asymptotique, qui peut être utilisée pour guider la sélection de la largeur de bande et du seuil. Dans (Goegebeur et al., 2008), la relation entre le test d'adéquation des ajustements et la sélection optimale de la fraction d'échantillon pour l'estimation de queue, par exemple en utilisant l'estimateur de Hill, est examinée. Les auteurs considèrent ce problème sous un test d'adéquation au bon d'ajustement pour évaluer si un échantillon est cohérent avec le modèle de type Pareto. Deux cas spéciaux importants de la statistique de test au noyau, la statistique de Jackson et la statistique de Lewis, sont discutés plus en profondeur.

Contributions

Chapter 4 vise à explorer des méthodes robustes pour calculer les espérances dans des distributions à queues lourdes, comme alternative à la simple moyenne. Nous considérons le cas jouet de la distribution de Pareto, car elle est souvent utilisée pour modéliser les pertes en finance et elle présente des propriétés intéressantes. Nous nous intéressons particulièrement à l'estimation de l'Expected-Shortfall (ES) dans la distribution de Pareto.

Expected-Shortfall dans la distribution de Pareto Tout d'abord, nous rappelons la théorie sur l'Expected-Shortfall et sur la distribution de Pareto, puis nous présentons les propriétés caractéristiques de cette dernière. Nous démontrons les propriétés utiles de mise à l'échelle et de stabilité par conditionnement de la distribution de Pareto, et nous prouvons la forme de l'ES au seuil de risque α ($\mathbf{ES}_\alpha$) en nous appuyant sur ces propriétés.

Il existe plusieurs définitions de l'ES, toutes équivalentes. La définition retenue dans ce chapitre est celle selon laquelle $\mathbf{ES}_\alpha$ est l'espérance conditionnelle des pertes excédant la $\mathbf{VaR}_\alpha$. En effet, d'après (Tasche, 2002b, Prop 3.4, Eq 3.3),(Sarykalin et al., 2008, Def.2, p.273), pour une distribution continue, l'Expected-Shortfall au seuil de risque α est défini comme suit :

$$\mathbf{ES}_\alpha(X) = \mathbb{E}[X|X \geq \mathbf{VaR}_\alpha(X)] \tag{4.323}$$

où $\alpha \in (0,1)$ est le niveau de risque, $X \in \mathcal{L}^p(\mathcal{F})$ est une variable aléatoire représentant la perte (comme une quantité positive, c'est-à-dire que nous prenons la convention que de grandes pertes correspondent à des nombres positifs élevés), d'un portefeuille à un instant futur, avec comme fonction de répartition $F_X(x) = \mathbb{P}(X \leq x); \forall x \in \mathbb{R}$.

Pour des raisons de précision et de réalisme dans la modélisation, nous nous intéressons à l'estimation de l'ES dans des distributions à queues lourdes, particulièrement dans la distribution de Pareto. Le choix de la distribution de Pareto est motivé par les propriétés intéressantes de cette distribution, comme l'explique (Arnold, 2014). Plusieurs raisons justifient l'intérêt pour cette distribution.

La distribution de Pareto joue un rôle crucial dans la théorie des valeurs extrêmes (EVT), une branche des statistiques qui s'intéresse au comportement statistique des événements extrêmes. L'EVT est particulièrement pertinente en finance, où les événements extrêmes, tels que les krachs boursiers, ont des implications majeures pour les investisseurs et les institutions financières. En finance, les données présentent souvent un comportement à queue lourde, ce qui signifie que les événements extrêmes, tels que les mouvements de prix importants ou les crises financières, se produisent de manière plus sévère que ce que prédit une distribution normale. Le risque de queue fait référence au risque d'occurrence d'événements extrêmes dans les queues d'une distribution. Bien que rares, ces événements peuvent avoir un impact significatif sur les marchés financiers et les portefeuilles. La distribution de Pareto, avec sa capacité à modéliser des queues lourdes, offre une meilleure adéquation à ces données que les distributions traditionnelles, comme la distribution Gaussienne. En effet, le modèle Gaussien attribue de faibles poids aux queues de distribution, négligeant ainsi les événements extrêmes, ce qui peut conduire à une mauvaise prévision de la VaR. En revanche, la distribution de Pareto attribue plus de poids aux queues de distribution, prenant en compte les événements extrêmes, ce qui permet de meilleures prédictions de la VaR, notamment lorsqu'il s'agit d'actifs ou de portefeuilles présentant un comportement à queue lourde.

Mathématiquement, la distribution de Pareto présente des propriétés intéressantes. Premièrement, cette distribution est relativement simple à comprendre et à manipuler. Elle ne comporte que deux paramètres (les paramètres d'échelle x_m et de forme γ ; ce dernier est aussi appelé index de Pareto), ce qui la rend plus facile à estimer et à interpréter par rapport à d'autres distributions à queues lourdes plus complexes, comme la distribution stable. La fonction de densité de probabilité et la fonction de répartition de la distribution de Pareto ont des formes analytiques relativement simples, ce qui facilite les calculs mathématiques et statistiques par rapport à certaines autres distributions à queues lourdes.

Definition 4.54 (Distribution de Pareto $\mathcal{P}(x_m, \gamma)$)**.** *Si X est une variable aléatoire suivant une distribution de Pareto $\mathcal{P}(x_m, \gamma)$, avec $(x_m > 0, \gamma > 0)$, alors la probabilité que X soit supérieure à un certain nombre x, c'est-à-dire la fonction de survie, également appelée fonction de queue, est donnée par :*

$$\bar{F}_X(x) = \mathbb{P}(X > x) = \mathbb{1}_{\{x<x_m\}} + \left(\frac{x_m}{x}\right)^{\gamma} \mathbb{1}_{\{x\geq x_m\}} \tag{4.324}$$

où x_m est la valeur minimale (nécessairement positive) que X peut prendre, et γ est un paramètre positif. La distribution de Pareto est caractérisée par un paramètre d'échelle x_m et un paramètre de forme γ, également appelé indice de queue.

La fonction de répartition (c.d.f.) d'une variable aléatoire de Pareto de paramètres x_m et γ est :

$$F_X(x) = \left(1 - \left(\frac{x_m}{x}\right)^{\gamma}\right) \mathbb{1}_{\{x \geq x_m\}}. \tag{4.325}$$

Et la fonction de densité de probabilité (p.d.f.) est donnée par :

$$f_X(x) = \frac{\gamma x_m^{\gamma}}{x^{\gamma+1}} \mathbb{1}_{\{x \geq x_m\}}. \tag{4.326}$$

La distribution de Pareto est une loi de puissance et est bien adaptée pour modéliser des données présentant un comportement en loi de puissance. En effet, la queue de la distribution de Pareto suit une fonction linéaire sur un graphique log-log. Cela constitue une preuve supplémentaire que les événements extrêmes surviennent avec une intensité plus sévère que celle prédite par d'autres distributions à queue lourde. Cela est particulièrement utile pour modéliser des phénomènes financiers qui suivent une distribution en loi de puissance, comme la répartition des revenus ou les mouvements de prix extrêmes.

L'une des propriétés les plus intéressantes de la distribution de Pareto est sa **propriété d'échelle**. En effet, elle permet d'établir des liens entre deux distributions de Pareto ayant le même paramètre de forme γ, mais des paramètres d'échelle x_m différents. Par exemple, la propriété d'échelle permet de passer de la distribution de Pareto standardisée $\mathcal{P}(1, \gamma)$ à n'importe quelle distribution de Pareto non-standardisée $\mathcal{P}(x_m, \gamma)$ par une simple multiplication de la distribution de Pareto standardisée par le paramètre d'échelle x_m de la distribution de Pareto non-standardisée. Inversement, cette propriété permet de passer de n'importe quelle distribution de Pareto non-standardisée $\mathcal{P}(x_m, \gamma)$ à la distribution de Pareto standardisée $\mathcal{P}(1, \gamma)$ par une simple division de la distribution non-standardisée par son paramètre d'échelle x_m.

Plus précisément, la propriété d'échelle établit une relation de proportionnalité entre les distributions de Pareto standardisée et non-standardisée, avec un facteur de proportionnalité égal au paramètre d'échelle x_m de la distribution de Pareto non-standardisée. Inversement, cette propriété permet de passer de la distribution de Pareto non-standardisée $\mathcal{P}(x_m, \gamma)$ à la distribution de Pareto standardisée $\mathcal{P}(1, \gamma)$ en divisant la distribution non standardisée par son paramètre d'échelle x_m. Par conséquent, il existe une relation de proportionnalité entre la distribution de Pareto standardisée $\mathcal{P}(1, \gamma)$ et la distribution de Pareto non-standardisée $\mathcal{P}(x_m, \gamma)$, avec un facteur de proportionnalité égal à $\frac{1}{x_m}$.

De manière plus générale, la propriété d'échelle établit un lien entre deux distributions de Pareto non-standardisées $\mathcal{P}(x_m^{(1)}, \gamma)$ et $\mathcal{P}(x_m^{(2)}, \gamma)$ avec $x_m^{(1)} > 0$ et $x_m^{(2)} > 0$. En effet, la distribution de Pareto $\mathcal{P}(x_m^{(1)}, \gamma)$ est proportionnelle à la distribution de Pareto $\mathcal{P}(x_m^{(2)}, \gamma)$, avec un facteur de proportionnalité égal au rapport entre les deux paramètres d'échelle $\frac{x_m^{(1)}}{x_m^{(2)}}$, et vice versa. Le facteur de proportionnalité est un ratio dont le numérateur correspond au paramètre d'échelle de la distribution de Pareto non-standardisée que l'on souhaite atteindre $\mathcal{P}(x_m^{(1)}, \gamma)$ (la distribution cible), et le dénominateur correspond au paramètre d'échelle de la distribution de Pareto non-standardisée initiale $\mathcal{P}(x_m^{(2)}, \gamma)$.

La division de la distribution de Pareto non-standardisée $\mathcal{P}(x_m^{(2)}, \gamma)$ par son propre paramètre d'échelle $x_m^{(2)}$ permet de la standardiser, atteignant ainsi la distribution de Pareto standardisée $\mathcal{P}(1, \gamma)$. Ensuite, la multiplication de la distribution standardisée par $x_m^{(1)}$ permet d'atteindre la distribution de Pareto désirée $\mathcal{P}(x_m^{(1)}, \gamma)$. Inversement, la distribution de Pareto $\mathcal{P}(x_m^{(2)}, \gamma)$ est proportionnelle à la distribution $\mathcal{P}(x_m^{(1)}, \gamma)$ avec un facteur de proportionnalité égal au rapport des deux paramètres d'échelle $\frac{x_m^{(2)}}{x_m^{(1)}}$. Le facteur de proportionnalité est un ratio dont le numérateur correspond au paramètre d'échelle de la distribution de Pareto non-standardisée que l'on souhaite atteindre $\mathcal{P}(x_m^{(2)}, \gamma)$, et le dénominateur correspond au paramètre d'échelle de la distribution de Pareto non standardisée initiale $\mathcal{P}(x_m^{(1)}, \gamma)$.

La division de la distribution de Pareto $\mathcal{P}(x_m^{(1)}, \gamma)$ par son propre paramètre d'échelle $x_m^{(1)}$ permet de la standardiser, atteignant ainsi la distribution de Pareto standardisée $\mathcal{P}(1, \gamma)$. Ensuite, la multiplication de la distribution standardisée par $x_m^{(2)}$ permet d'atteindre la distribution de Pareto désirée $\mathcal{P}(x_m^{(2)}, \gamma)$.

En conséquence, la distribution de Pareto est invariante par changement d'échelle, ce qui signifie que multiplier ou diviser une distribution de Pareto par un paramètre constant ne modifie pas la forme de la distribution (le paramètre γ reste inchangé). La nouvelle distribution mise à l'échelle est toujours une distribution de Pareto avec le même paramètre de forme γ, mais avec un nouveau paramètre d'échelle. La nouvelle distribution de Pareto est obtenue en multipliant la distribution initiale par le rapport entre le paramètre d'échelle de la distribution cible et celui de la distribution initiale.

En plus de la propriété d'échelle, la distribution de Pareto satisfait la *propriété de stabilité par conditionnement*. Cela signifie que la queue de toute distribution de Pareto au-delà d'un certain seuil positif reste une distribution de Pareto avec le même paramètre de forme, mais avec un nouveau paramètre d'échelle.

Par exemple, la queue de toute distribution de Pareto standardisée $\mathcal{P}(1, \gamma)$ au-delà d'un seuil positif donné $s_m^{(1)} > 0$ reste une distribution de Pareto avec le même paramètre de forme γ mais avec un nouveau paramètre d'échelle égal au paramètre de conditionnement $s_m^{(1)}$: $\mathcal{P}(s_m^{(1)}, \gamma)$. En raison de la propriété d'échelle, il existe une relation de proportionnalité liant la distribution de Pareto non-standardisée $\mathcal{P}(s_m^{(1)}, \gamma)$ et la distribution de Pareto standardisée $\mathcal{P}(1, \gamma)$ avec un facteur de proportionnalité égal à $s_m^{(1)}$. Par conséquent, en combinant à la fois les propriétés de stabilité par conditionnement et d'échelle, une relation de proportionnalité est établie entre la distribution de Pareto standardisée conditionnée à ses valeurs supérieures au seuil $s_m^{(1)}$, et la distribution de Pareto marginale $\mathcal{P}(1, \gamma)$, avec un facteur de proportionnalité égal au paramètre de conditionnement $s_m^{(1)}$.

Ainsi, toute distribution de Pareto standardisée $\mathcal{P}(1, \gamma)$ conditionnée à ses valeurs au-delà d'un seuil donné $s_m^{(1)}$ reste une distribution de Pareto avec le même paramètre de forme γ mais avec un nouveau paramètre d'échelle égal au paramètre de conditionnement $s_m^{(1)}$.

Plus généralement, la queue de toute distribution de Pareto non-standardisée $\mathcal{P}(x_m^{(1)}, \gamma)$ au-delà d'un seuil positif donné $s_m^{(1)} > x_m^{(1)}$ reste une distribution de Pareto avec le même paramètre de forme γ, mais avec un nouveau paramètre d'échelle égal au paramètre de conditionnement $s_m^{(1)}$: $\mathcal{P}(s_m^{(1)}, \gamma)$. En raison de la propriété d'échelle, il existe une relation de proportionnalité entre la distribution de Pareto non-standardisée $\mathcal{P}(s_m^{(1)}, \gamma)$ et la distribution de Pareto standardisée $\mathcal{P}(1, \gamma)$, avec un facteur de proportionnalité égal à $s_m^{(1)}$. De plus, la propriété d'échelle établit également une relation de proportionnalité entre les deux distributions de Pareto non standardisées $\mathcal{P}(s_m^{(1)}, \gamma)$ et $\mathcal{P}(x_m^{(1)}, \gamma)$, avec un facteur de proportionnalité égal au rapport entre les deux paramètres d'échelle $\frac{s_m^{(1)}}{x_m^{(1)}}$. Par conséquent, en combinant les propriétés de stabilité par conditionnement et d'échelle, une relation de proportionnalité est établie entre la distribution de Pareto non-standardisée $\mathcal{P}(x_m^{(1)}, \gamma)$ conditionnée à ses valeurs supérieures au seuil $s_m^{(1)}$ et la distribution de Pareto standardisée $\mathcal{P}(1, \gamma)$, avec un facteur de proportionnalité égal au paramètre de conditionnement $s_m^{(1)}$. De plus, une relation de proportionnalité est également établie entre la distribution de Pareto non-standardisée $\mathcal{P}(x_m^{(1)}, \gamma)$ conditionnée à ses valeurs supérieures au seuil $s_m^{(1)}$ et la distribution marginale $\mathcal{P}(x_m^{(1)}, \gamma)$, avec un facteur de proportionnalité égal au rapport $\frac{s_m^{(1)}}{x_m^{(1)}}$, où le numérateur est le paramètre de conditionnement $s_m^{(1)}$ et le dénominateur est le paramètre d'échelle $x_m^{(1)}$ de la distribution marginale. Par conséquent, toute distribution de Pareto non-standardisée $\mathcal{P}(x_m^{(1)}, \gamma)$ conditionnée à ses valeurs supérieures au seuil $s_m^{(1)}$ reste une distribution de Pareto avec le même paramètre de forme γ, mais avec un nouveau paramètre d'échelle égal au paramètre de conditionnement $s_m^{(1)}$.

Ces propriétés sont très pratiques. De plus, elles sont spécifiques à la distribution de Pareto, c'est-à-dire qu'elles ne sont pas partagées par toutes les distributions à queues lourdes. Elles sont formulées mathématiquement dans le théorème suivant.

Theorem 4.55 (Stabilité par conditionnement et changement d'échelle de la distribution de Pareto)**.** (i) *Soit X une variable aléatoire de Pareto standardisée $X \sim \mathcal{P}(1, \gamma)$, $\gamma > 0$. Soit $x_m > 0$ un nouveau paramètre d'échelle. Soit Y une variable aléatoire de Pareto non-standardisée telle que $Y \sim \mathcal{P}(x_m, \gamma)$. Alors, nous avons :*

$$\mathbb{P}(X \leq x \mid X \geq x_m) = \mathbb{P}(x_m X \leq x) = \mathbb{P}(Y \leq x). \tag{4.327}$$

De manière équivalente,

$$X \mid X \geq x_m \overset{d}{=} x_m X \overset{d}{=} Y \quad \text{ou} \quad \frac{Y}{x_m} \overset{d}{=} \frac{X \mid X \geq x_m}{x_m} \overset{d}{=} X. \tag{4.328}$$

En d'autres termes, conditionner une variable aléatoire de Pareto standardisée X à partir d'un seuil donné x_m revient à multiplier la distribution de Pareto standardisée, c'est-à-dire la distribution marginale de X, par le paramètre de conditionnement x_m. Ainsi, la distribution de Pareto standardisée conditionnée à ses valeurs au-delà d'un certain seuil reste une distribution de Pareto avec le même paramètre de forme γ mais avec un nouveau paramètre d'échelle égal au paramètre de conditionnement x_m.

(ii) *Soient Z_1 et Z_2 deux variables aléatoires de Pareto non-standardisées telles que $Z_1 \sim \mathcal{P}(x_m^{(1)}, \gamma)$ et $Z_2 \sim \mathcal{P}(x_m^{(2)}, \gamma)$ avec $x_m^{(1)} > 0, x_m^{(2)} > 0, \gamma > 0$. Alors, nous avons :*

$$Z_1 \overset{d}{=} \frac{x_m^{(1)}}{x_m^{(2)}} Z_2 \overset{d}{=} x_m^{(1)} X \quad \text{ou} \quad Z_2 \overset{d}{=} \frac{x_m^{(2)}}{x_m^{(1)}} Z_1 \overset{d}{=} x_m^{(2)} X. \tag{4.329}$$

En d'autres termes, toute distribution de Pareto non-standardisée peut être exprimée à partir de toute autre distribution de Pareto non-standardisée en divisant la distribution de Pareto initiale non-standardisée par son propre paramètre d'échelle pour atteindre la distribution de Pareto standardisée $\mathcal{P}(1, \gamma)$, puis en la multipliant par le paramètre d'échelle de la distribution cible.

(iii) *Soit $s_m^{(1)}$ un paramètre de conditionnement tel que $s_m^{(1)} > x_m^{(1)} > 0$. Alors, les équations (4.328) et (4.329) conduisent à :*

$$Z_1 \mid Z_1 \geq s_m^{(1)} \overset{d}{=} \frac{s_m^{(1)}}{x_m^{(1)}} Z_1 \overset{d}{=} s_m^{(1)} X. \tag{4.330}$$

En d'autres termes, conditionner toute distribution de Pareto non-standardisée $\mathcal{P}(x_m^{(1)}, \gamma)$ à partir d'un certain seuil $s_m^{(1)}$ revient à multiplier la distribution de Pareto standardisée $\mathcal{P}(1, \gamma)$ par le paramètre de conditionnement $s_m^{(1)}$, ou de manière équivalente à multiplier la distribution marginale $\mathcal{P}(x_m^{(1)}, \gamma)$ par le rapport entre le paramètre de conditionnement et le paramètre d'échelle de la distribution marginale $\frac{s_m^{(1)}}{x_m^{(1)}}$.

De plus, les moments de la distribution de Pareto n'existent que pour certaines plages de valeurs du paramètre de forme γ. En particulier, le k-ième moment existe si et seulement si $\gamma > k$. Cela signifie que pour de petites valeurs de γ, les moments peuvent ne pas exister, ce qui est lié à la nature des distributions à queues lourdes. À mesure que le paramètre de forme γ diminue, la queue de la distribution devient de plus en plus lourde, accordant de plus en plus de poids aux événements extrêmes. Cela conduit à des moments finis uniquement pour des ordres très faibles. C'est une caractéristique des distributions à queues lourdes, où les observations extrêmes ont un impact significatif sur les moments supérieurs, ce qui est moins marqué sur les distributions à queues plus légères.

Par exemple, la distribution de Pareto a une moyenne infinie pour un paramètre de forme $\gamma \leq 1$. Pour $\gamma \leq 2$, la distribution de Pareto a une moyenne finie mais une variance infinie. Pour $\gamma > 2$, la variance de la distribution de Pareto existe et est finie. Cela signifie que si le paramètre de forme est supérieur à 2, la distribution est caractérisée par à la fois une moyenne finie et une variance finie, ce qui la rend plus gérable dans certaines analyses statistiques.

La propriété d'échelle de la distribution de Pareto stipule que toute distribution de Pareto non-standardisée $\mathcal{P}(x_m, \gamma)$ est proportionnelle à la distribution de Pareto standardisée $\mathcal{P}(1, \gamma)$ avec un facteur de proportionnalité égal à x_m. Cela implique que toutes les quantités (statistiques) calculées sur la distribution de Pareto non-standardisée

$\mathcal{P}(x_m, \gamma)$ sont proportionnelles à celles calculées sur la distribution de Pareto standardisée $\mathcal{P}(1, \gamma)$, avec un facteur de proportionnalité égal à x_m.

Par conséquent, les calculs peuvent être effectués sur la distribution de Pareto standardisée, et les quantités équivalentes sur toute distribution de Pareto non standardisée $\mathcal{P}(x_m, \gamma)$ peuvent être récupérées en multipliant les quantités standardisées par le paramètre d'échelle adéquat x_m.

Par exemple, la propriété d'échelle de la distribution de Pareto permet de définir le k-ième moment de toute distribution de Pareto non-standardisée $\mathcal{P}(x_m, \gamma)$ grâce à une simple multiplication du k-ième moment de la distribution de Pareto standardisée $\mathcal{P}(1, \gamma)$ par le paramètre d'échelle x_m élevé à la puissance k.

Ces propriétés sont formulées mathématiquement dans le lemme suivant.

Lemma 4.56 (Moments de la distribution de Pareto)**.** *Soient $X \sim \mathcal{P}(1, \gamma)$ et $Y \sim \mathcal{P}(x_m, \gamma)$ deux variables aléatoires de Pareto, avec $\gamma > k$, $k \in \mathbb{N}^\star$. Par conséquent, les variables aléatoires de Pareto X et Y admettent des moments finis jusqu'à l'ordre k.*

(i) *Moments d'ordre k ($\gamma > k$):*

$$\mathbb{E}\left[X^k\right] = \frac{\gamma}{\gamma - k} \quad et \quad \mathbb{E}\left[Y^k\right] = x_m^k \mathbb{E}\left[X^k\right]. \tag{4.331}$$

(ii) *Moments centraux d'ordre k ($\gamma > k$):*

$$\mathbb{E}\left[(X - \mathbb{E}[X])^k\right] = \sum_{j=0}^{k} \binom{k}{j} \frac{\gamma^{j+1}}{(1-\gamma)^j(\gamma + j - k)} \quad et \quad \mathbb{E}\left[(Y - \mathbb{E}[Y])^k\right] = x_m^k \mathbb{E}\left[(X - \mathbb{E}[X])^k\right]. \tag{4.332}$$

Ces propriétés s'appliquent à toutes les statistiques calculées sur la distribution de Pareto, en particulier à la $\mathbf{VaR}_\alpha$ et l'$\mathbf{ES}_\alpha$. Soit $X \sim \mathcal{P}(1, \gamma)$ une variable aléatoire de Pareto standardisée et $Y \sim \mathcal{P}(x_m, \gamma)$, $Z_1 \sim \mathcal{P}(x_m^{(1)}, \gamma)$ et $Z_2 \sim \mathcal{P}(x_m^{(2)}, \gamma)$ des variables aléatoires de Pareto non-standardisées.

La propriété d'échelle stipule que la distribution de Pareto non-standardisée $\mathcal{P}(x_m, \gamma)$ est proportionnelle à la distribution de Pareto standardisée $\mathcal{P}(1, \gamma)$, avec un facteur de proportionnalité égal à x_m. Cela implique que la $\mathbf{VaR}_\alpha$ de la distribution de Pareto non standardisée $\mathcal{P}(x_m, \gamma)$ est proportionnelle à la $\mathbf{VaR}_\alpha$ de la distribution de Pareto standardisée, avec un coefficient de proportionnalité égal au paramètre d'échelle x_m.

De même, la propriété d'échelle de la distribution de Pareto énonce que la distribution de Pareto non-standardisée $\mathcal{P}(x_m^{(1)}, \gamma)$ est proportionnelle à la distribution de Pareto non-standardisée $\mathcal{P}(x_m^{(2)}, \gamma)$, avec un facteur de proportionnalité égal au rapport entre les deux paramètres d'échelle $\frac{x_m^{(1)}}{x_m^{(2)}}$. Cela implique que la $\mathbf{VaR}_\alpha$ de la distribution de Pareto non-standardisée $\mathcal{P}(x_m^{(1)}, \gamma)$ est proportionnelle à la $\mathbf{VaR}_\alpha$ de la distribution de Pareto non-standardisée $\mathcal{P}(x_m^{(2)}, \gamma)$, avec un coefficient de proportionnalité égal à $\frac{x_m^{(1)}}{x_m^{(2)}}$.

D'autre part, la propriété de stabilité par conditionnement stipule que la distribution de Pareto standardisée $\mathcal{P}(1, \gamma)$, conditionnée à ses valeurs supérieures à un certain seuil, est toujours une distribution de Pareto avec le même paramètre de forme γ mais un nou-

veau paramètre d'échelle égal au paramètre de conditionnement. De plus, la propriété d'échelle implique que la distribution de Pareto standardisée $\mathcal{P}(1,\gamma)$, conditionnée à ses valeurs supérieures à un certain seuil, est proportionnelle à la distribution de Pareto standardisée avec un facteur de proportionnalité égal au seuil de conditionnement.

L'Expected-Shortfall de la distribution de Pareto standardisée $\mathcal{P}(1,\gamma)$ est l'espérance conditionnelle étant donné que la variable aléatoire de Pareto standardisée X est supérieure à $\mathbf{VaR}_\alpha(X)$: $\mathbf{ES}_\alpha(X) = \mathbb{E}\left[X \mid X \geq \mathbf{VaR}_\alpha(X)\right]$ (définie par (4.3)).

La combinaison des propriétés de stabilité par conditionnement et d'échelle conduit à une relation de proportionnalité entre la distribution de Pareto standardisée $\mathcal{P}(1,\gamma)$, conditionnée à ses valeurs supérieures à $\mathbf{VaR}_\alpha(X)$, et la distribution marginale qui est la distribution de Pareto standardisée $\mathcal{P}(1,\gamma)$, avec un facteur de proportionnalité égal au seuil de conditionnement $\mathbf{VaR}_\alpha(X)$.

En conséquence, l'Expected-Shortfall de la distribution de Pareto standardisée $\mathbf{ES}_\alpha(X)$ est proportionnel à son espérance $\mathbb{E}\left[X\right]$, avec un facteur de proportionnalité égal à $\mathbf{VaR}_\alpha(X)$, qui est supposé être connu. Par conséquent, $\mathbf{ES}_\alpha(X)$ peut être facilement récupéré grâce à la simple multiplication par $\mathbf{VaR}_\alpha$ de l'espérance de la distribution de Pareto standardisée.

De même, la propriété de stabilité par conditionnement stipule que la distribution de Pareto non-standardisée $Z_1 \sim \mathcal{P}(x_m^{(1)},\gamma)$, conditionnée à ses valeurs supérieures à un certain seuil, est toujours une distribution de Pareto avec le même paramètre de forme γ mais un nouveau paramètre d'échelle égal au paramètre de conditionnement. De plus, la propriété d'échelle implique que la distribution de Pareto non-standardisée $\mathcal{P}(x_m^{(1)},\gamma)$, conditionnée à ses valeurs supérieures à un certain seuil, est proportionnelle à la distribution de Pareto standardisée $\mathcal{P}(1,\gamma)$ avec un facteur de proportionnalité égal au paramètre de conditionnement ; et est également proportionnelle à la distribution marginale de Pareto $\mathcal{P}(x_m^{(1)},\gamma)$ avec un facteur de proportionnalité égal au rapport entre le paramètre de conditionnement et le paramètre d'échelle $x_m^{(1)}$.

La combinaison des propriétés de stabilité par conditionnement et d'échelle implique que l'Expected-Shortfall de la distribution de Pareto non-standardisée $\mathcal{P}(x_m^{(1)},\gamma)$, qui est l'espérance de la distribution $\mathcal{P}(x_m^{(1)},\gamma)$ conditionnée à ses valeurs supérieures à $\mathbf{VaR}_\alpha(Z_1)$, est proportionnel à l'espérance de la distribution de Pareto standardisée $\mathcal{P}(1,\gamma)$ avec un facteur de proportionnalité égal au paramètre de conditionnement $\mathbf{VaR}_\alpha(Z_1)$, et est proportionnel à l'espérance de la distribution marginale de Pareto $\mathcal{P}(x_m^{(1)},\gamma)$ avec un facteur de proportionnalité égal au rapport $\frac{\mathbf{VaR}_\alpha(Z_1)}{x_m^{(1)}}$.

Par conséquent, l'Expected-Shortfall de toute distribution de Pareto non-standardisée $\mathcal{P}(x_m^{(1)},\gamma)$ peut être récupéré grâce à une simple multiplication de l'espérance de la distribution de Pareto standardisée $\mathcal{P}(1,\gamma)$ par le seuil de conditionnement $\mathbf{VaR}_\alpha(Z_1)$, et grâce à une simple multiplication de l'espérance de la distribution marginale de Pareto $\mathcal{P}(x_m^{(1)},\gamma)$ par le rapport $\frac{\mathbf{VaR}_\alpha(Z_1)}{x_m^{(1)}}$.

Toutes ces propriétés sont formulées mathématiquement dans la proposition suivante.

Proposition 4.57 (**VaR**$_\alpha$ et **ES**$_\alpha$ dans une distribution de Pareto)**.** *Si $X \sim \mathcal{P}(1,\gamma)$ et $Y \sim \mathcal{P}(x_m,\gamma)$, avec $x_m > 0$ et $\gamma > 0$, alors leurs* **VaR**$_\alpha$ *respectives sont définies comme suit :*

$$\begin{cases} \mathbf{VaR}_\alpha(X) &= F_X^{-1}(\alpha) = (1-\alpha)^{-\frac{1}{\gamma}}, \\ \mathbf{VaR}_\alpha(Y) &= F_Y^{-1}(\alpha) = x_m \mathbf{VaR}_\alpha(X), \end{cases} \quad \alpha \in (0,1), \tag{4.333}$$

et les **ES**$_\alpha$ *respectifs sont donnés par :*

$$\begin{cases} \mathbf{ES}_\alpha(X) = \mathbf{VaR}_\alpha(X)\mathbb{E}\left[X\right] = \frac{\gamma}{\gamma-1}(1-\alpha)^{-\frac{1}{\gamma}}, \\ \mathbf{ES}_\alpha(Y) = x_m \mathbf{ES}_\alpha(X), \end{cases} \quad \alpha \in (0,1). \tag{4.334}$$

Comme mentionné précédemment, en raison de la propriété d'échelle de la distribution de Pareto, les statistiques dans toute distribution de Pareto non-standardisée sont proportionnelles aux statistiques correspondantes dans la distribution de Pareto standardisée, avec un facteur de proportionnalité égal au paramètre d'échelle de la distribution de Pareto cible. Par conséquent, par la suite, toutes les statistiques seront calculées dans la distribution de Pareto standardisée. Une simple multiplication de ces statistiques standardisées par le paramètre d'échelle approprié de la distribution de Pareto cible permettra de récupérer les statistiques correspondantes dans la distribution de Pareto non-standardisée souhaitée.

Comme expliqué précédemment, en raison de la propriété de stabilité par conditionnement, l'Expected-Shortfall au seuil de risque α de la distribution de Pareto standardisée est proportionnel à son espérance, avec un facteur de proportionnalité égal à la Value-at-Risk au seuil de risque α de la distribution de Pareto standardisée, qui est supposée être connu. Par conséquent, l'étape principale dans le calcul de l'Expected-Shortfall de la distribution de Pareto standardisée est de calculer son espérance. L'Expected-Shortfall de la distribution de Pareto standardisée peut être récupéré grâce à une simple multiplication de son espérance par **VaR**$_\alpha$, qui est supposée être connue.

Problématique Nous nous interrogeons sur la manière d'estimer avec précision l'Expected-Shortfall. La présence de valeurs extrêmes avec une faible probabilité dans les distributions à queues lourdes peut influencer l'estimateur (moyenne empirique) de l'espérance et entraîner de mauvaises performances. Dans ce contexte, l'objectif est de trouver un estimateur de l'ES qui soit plus robuste à la présence de valeurs extrêmes. Dans ce chapitre, nous supposons que **VaR**$_\alpha$ est connue et nous nous concentrons sur les méthodes d'estimation de l'**ES**$_\alpha$. Différents estimateurs de moyenne sont étudiés en fonction de leurs propriétés de concentration et de fluctuation.

En raison de la propriété d'échelle de la distribution de Pareto qui établit une relation de proportionnalité entre la distribution de Pareto non-standardisée $\mathcal{P}(x_m,\gamma)$ et la distribution de Pareto standardisée $\mathcal{P}(1,\gamma)$ avec un facteur de proportionnalité égal à x_m, il suffit d'estimer l'ES sur la distribution de Pareto standardisée $\mathcal{P}(1,\gamma)$; nous pouvons

ensuite récupérer l'estimation de l'ES sur la distribution de Pareto non-standardisée $\mathcal{P}(x_m, \gamma)$ grâce à une simple multiplication de l'ES standardisé par x_m.

De plus, en raison des propriétés de stabilité par conditionnement et d'échelle de la distribution de Pareto qui établissent une relation de proportionnalité entre la distribution de Pareto standardisée $\mathcal{P}(1, \gamma)$ conditionnée à ses valeurs supérieures à un certain seuil et la distribution de Pareto standardisée, avec un facteur de proportionnalité égal au seuil de conditionnement, estimer l'espérance de la distribution de Pareto standardisée est suffisant pour obtenir l'estimation de l'$\mathbf{ES}_\alpha$ standardisé, car nous pouvons récupérer l'estimation de l'$\mathbf{ES}_\alpha$ standardisé grâce à une simple multiplication de l'estimation de l'espérance standardisée par $\mathbf{VaR}_\alpha$.

Idéalement, un estimateur est censé satisfaire les contraintes suivantes : être robuste aux distributions à queues lourdes, c'est-à-dire à la présence de valeurs extrêmes, et atteindre un haut niveau de précision avec un haut niveau de confiance. Il est à noter que ce critère est non-asymptotique.

Ces contraintes sont satisfaites par des estimateurs sous-Gaussiens. Malgré les avantages des estimateurs sous-Gaussiens, ils présentent également certaines limitations. En effet, prouver qu'un estimateur est sous-Gaussien n'est pas si évident, car cela repose sur des hypothèses restrictives sur les modèles, ou nécessite des astuces pour développer un nouvel estimateur tel que la Médiane-des-Moyenne (voir plus loin dans ce chapitre). De plus, certains estimateurs sous-Gaussiens peuvent avoir une complexité computationnelle plus élevée que des estimateurs plus simples. Cela peut être un inconvénient lors de la gestion de grands ensembles de données ou dans des applications en temps réel.

Pour ces raisons, nos travaux proposent plusieurs estimateurs de moyenne robustes qui ne sont pas sous-Gaussiens, pour lesquels une étude de leur biais avec le taux de convergence est réalisée et des inégalités de concentration sont développées.

Estimateurs de moyenne L'attention est principalement portée sur trois estimateurs : l'estimateur des Médianes-des-Moyennes (MoM), l'estimateur des Moyennes Tronquées (TM) et l'estimateur de Lee-Valiant (LV). Les performances de ces estimateurs sont ensuite comparées entre elles et avec l'estimateur de référence, à savoir la moyenne empirique, qui atteint des performances sous-Gaussiennes dans le cadre asymptotique.

Rappelons que le vrai $\mathbf{ES}_\alpha$ est défini comme la perte excédentaire moyenne au-delà de la $\mathbf{VaR}_\alpha$ théorique :

$$\mathbf{ES}_\alpha(X) = \mathbb{E}[X | X \geq \mathbf{VaR}_\alpha] = \frac{\mathbb{E}[X \mathbb{1}_{\{X \geq \mathbf{VaR}_\alpha\}}]}{\mathbb{P}(X \geq \mathbf{VaR}_\alpha)}. \tag{4.335}$$

Un estimateur de l'$\mathbf{ES}_\alpha$, appelé l'$\mathbf{ES}_\alpha$ empirique, est la perte excédentaire moyenne au-delà de l'estimateur de $\mathbf{VaR}_\alpha$, qui est le quantile empirique α-quantile ($\mathbf{q}_{n,\alpha}$), comme indiqué dans la définition suivante.

Definition 4.58 (Expected-Shortfall empirique)**.** *Soit* $\mathbf{X} = (X_1, \dots, X_n)$ *une séquence de variables aléatoires i.i.d. de Pareto telles que* $\forall i \in [\![1, n]\!], X_i \sim \mathcal{P}(1, \gamma)$*, avec* $\gamma > 2$*. Notons* $\boldsymbol{X}^\star = (X_1^\star, \dots, X_n^\star)$ *la séquence de statistiques d'ordre associée à* $\mathbf{X}$*. Le quantile empirique d'ordre* α *est donné par* $q_\alpha^n = X^\star_{\lceil n\alpha \rceil}$*, et l'Expected-Shortfall empirique au seuil de risque* α *est défini comme suit :*

$$\mathbf{ES}_\alpha^n = \frac{\frac{1}{n}\sum_{i=1}^n X_i \mathbb{1}_{\{X_i \geq q_\alpha^n\}}}{\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{X_i \geq q_\alpha^n\}}} = \frac{1}{n - \lceil n\alpha \rceil} \sum_{i=1}^n X_i \mathbb{1}_{\{X_i \geq q_\alpha^n\}}. \tag{4.336}$$

Dans notre étude, nous distinguons deux cadres pour l'estimation de l'$\mathbf{ES}_\alpha$ dans la distribution de Pareto standardisée $\mathcal{P}(1, \gamma)$: le premier est appelé *cas idéalisé* et l'autre est appelé *cas réaliste*.

Cas idéalisé : $q_\alpha^n = \mathbf{VaR}_\alpha$ Nous considérons n échantillons indépendants et identiquement distribués (i.i.d.) $X_1, \dots, X_n$ suivant une distribution de Pareto standardisée $\mathcal{P}(1, \gamma)$. Le *cas idéalisé* suppose que le quantile empirique d'ordre α noté q_α^n correspond à la vraie $\mathbf{VaR}_\alpha$: bien que cela ne soit vrai que dans le cadre asymptotique et non dans la pratique pour des tailles d'échantillon réduites, nous utiliserons cette hypothèse dans certaines parties de notre étude pour simplifier l'analyse. Dans ce cas, l'estimation de l'$\mathbf{ES}_\alpha$ dans la distribution de Pareto standardisée $\mathcal{P}(1, \gamma)$ correspond à la moyenne empirique de la perte excédentaire au-delà de la vraie $\mathbf{VaR}_\alpha$. En d'autres termes, l'estimation de l'$\mathbf{ES}_\alpha$ dans la distribution de Pareto standardisée $\mathcal{P}(1, \gamma)$ correspond à la moyenne empirique de la distribution de Pareto standardisée $\mathcal{P}(1, \gamma)$ conditionnée à ses valeurs au-delà de la vraie $\mathbf{VaR}_\alpha$. Dans ce cas, le seuil de conditionnement ne dépend pas de l'échantillon sous-jacent. Cela implique que les échantillons supérieurs à la vraie $\mathbf{VaR}_\alpha$ sont toujours indépendants et identiquement distribués (i.i.d.), et que les propriétés de stabilité par conditionnement et de mise à l'échelle sont valides. Grâce à la propriété de stabilité par conditionnement, la queue de la distribution de Pareto standardisée au-delà de la vraie $\mathbf{VaR}_\alpha$ reste une distribution de Pareto avec le même paramètre de forme γ mais un nouveau paramètre d'échelle égal au seuil de conditionnement $\mathbf{VaR}_\alpha$: $\mathcal{P}(\mathbf{VaR}_\alpha, \gamma)$. De plus, la propriété de mise à l'échelle établit une relation de proportionnalité entre la distribution de Pareto non-standardisée $\mathcal{P}(\mathbf{VaR}_\alpha, \gamma)$ et la distribution de Pareto standardisée $\mathcal{P}(1, \gamma)$, avec un facteur de proportionnalité égal à $\mathbf{VaR}_\alpha$. Par conséquent, il existe une relation de proportionnalité entre la distribution de Pareto standardisée conditionnée à ses valeurs étant au-delà de la vraie $\mathbf{VaR}_\alpha$ et la distribution marginale $\mathcal{P}(1, \gamma)$, avec un facteur de proportionnalité égal au seuil de conditionnement $\mathbf{VaR}_\alpha$. Cela implique que l'espérance de la distribution de Pareto standardisée conditionnée à ses valeurs au-delà de la vraie $\mathbf{VaR}_\alpha$ (c'est-à-dire, l'$\mathbf{ES}_\alpha$) est proportionnelle à l'espérance de la distribution de Pareto standardisée $\mathcal{P}(1, \gamma)$, avec un facteur de proportionnalité égal à $\mathbf{VaR}_\alpha$. De même, la moyenne empirique de la distribution de Pareto standardisée conditionnée à ses valeurs au-delà de la vraie $\mathbf{VaR}_\alpha$ (c'est-à-dire l'$\mathbf{ES}_\alpha$ empirique) est proportionnelle à la moyenne empirique de la distribution de Pareto standardisée $\mathcal{P}(1, \gamma)$, avec un facteur de proportionnalité égal à $\mathbf{VaR}_\alpha$. Par conséquent, le biais entre l'$\mathbf{ES}_\alpha$ empirique et le vrai $\mathbf{ES}_\alpha$ de la distribution de Pareto standardisée est proportionnel au biais entre l'espérance et la moyenne empirique de la distribution de Pareto standardisée, avec un facteur de proportionnalité

égal à $\mathbf{VaR}_\alpha$. Ainsi, dans le cas idéalisé, il suffit de déterminer une formule-fermée pour le biais entre la moyenne empirique et l'espérance de la distribution de Pareto standardisée $\mathcal{P}(1,\gamma)$. Le biais entre l'$\mathbf{ES}_\alpha$ empirique et le vrai $\mathbf{ES}_\alpha$ de la distribution de Pareto standardisée peut être récupéré grâce à une simple multiplication du biais entre la moyenne empirique et l'espérance de la distribution de Pareto standardisée par le seuil de conditionnement $\mathbf{VaR}_\alpha$. Or, le quantile empirique d'ordre α ne correspond à la vraie $\mathbf{VaR}_\alpha$ que dans le cas asymptotique (c'est-à-dire lorsque la taille de l'échantillon est suffisamment grande), mais dans la plupart des cas, lorsque la taille de l'échantillon s'éloigne du cas asymptotique, le quantile empirique d'ordre α ne correspond pas à la vraie $\mathbf{VaR}_\alpha$.

Cas réaliste : $q_\alpha^n \neq \mathbf{VaR}_\alpha$ Dans le *cas réaliste*, le quantile empirique d'ordre α ne correspond pas à la $\mathbf{VaR}_\alpha$ théorique. Dans ce cas, l'estimation de l'$\mathbf{ES}_\alpha$ dans la distribution de Pareto standardisée $\mathcal{P}(1,\gamma)$ correspond à la moyenne empirique de la perte excédentaire au-delà du quantile empirique d'ordre α. En d'autres termes, l'estimation de l'$\mathbf{ES}_\alpha$ dans la distribution de Pareto standardisée $\mathcal{P}(1,\gamma)$ correspond à la moyenne empirique de la distribution de Pareto standardisée conditionnée à ses valeurs supérieures au quantile empirique d'ordre α. Dans ce cas, le seuil de conditionnement dépend de l'échantillon sous-jacent. Cela implique que les échantillons supérieurs au quantile empirique d'ordre α ne sont plus indépendants et identiquement distribués (i.i.d.), et les propriétés de stabilité par conditionnement et de mise à l'échelle de la distribution de Pareto ne sont plus valides. Par conséquent, la queue de la distribution de Pareto standardisée au-delà du quantile empirique d'ordre α n'est pas nécessairement une distribution de Pareto. Ainsi, la distribution des échantillons supérieurs au quantile empirique d'ordre α est inconnue, et il est difficile de dériver une formule analytique pour le biais entre l'$\mathbf{ES}_\alpha$ empirique et l'$\mathbf{ES}_\alpha$ théorique. Pour cette raison, dans le cas réaliste, nous proposons une étude empirique du biais entre l'$\mathbf{ES}_\alpha$ empirique et l'$\mathbf{ES}_\alpha$ théorique, ainsi que du taux de convergence. Dans le cas réaliste, le biais entre l'$\mathbf{ES}_\alpha$ empirique et l'$\mathbf{ES}_\alpha$ théorique est plus important que dans le cas idéalisé, car il existe un terme d'erreur supplémentaire lié à l'estimation du biais entre la vraie $\mathbf{VaR}_\alpha$ et le quantile empirique d'ordre α.

Dans tout le chapitre, les estimateurs de l'$\mathbf{ES}_\alpha$ sont étudiés à la fois dans le *cas idéalisé*, où le quantile empirique d'ordre α correspond à la vraie $\mathbf{VaR}_\alpha$, et dans le *cas réaliste* où les deux quantités diffèrent.

Étant donné que le biais d'un estimateur est défini par la différence entre la valeur cible et l'espérance de l'estimateur, le défi dans le calcul du biais réside dans le calcul de l'espérance de l'estimateur.

Médiane des Moyennes (MoM)

Nous avons tenté de dériver une formule explicite pour le biais de l'estimateur des MoM. Ce problème est complexe et reste une question ouverte. Dans ce contexte, nous développons une méthodologie pour dériver une formule explicite pour le biais de

l'estimateur des MoM, et nous prouvons les résultats intermédiaires dans cette procédure, tels que la dérivation d'une formule explicite pour le biais de la médiane empirique.

L'estimateur des MoM combine les estimateurs de la moyenne et de la médiane empiriques. L'estimateur de la Médiane des Moyennes nécessite de partitionner les données en k blocs de taille à peu près égale, de calculer la moyenne empirique dans chaque bloc, et de prendre la médiane de la séquence composée par les k moyennes empiriques.

Definition 4.59 (Médiane des Moyennes (**MoM**))**.** *Soit $X_1, \ldots, X_n$ une séquence de n tirages aléatoires indépendants et identiquement distribués (i.i.d.) provenant de la distribution de X. Soit m et k deux entiers positifs. Nous supposons que n est un multiple de k tel que $n = mk$. La moyenne empirique de chaque bloc est définie comme suit :*

$$\forall j \in [\![1, k]\!], \quad \bar{\mu}_{B_j} = \frac{1}{\left|B_j\right|} \sum_{i \in B_j} X_i. \tag{4.337}$$

L'estimateur des **MoM** *est alors défini par $\widehat{MoM}_n = M(\bar{\mu}_1, \ldots, \bar{\mu}_k)$.*

En raison de la manière dont l'estimateur est construit, l'estimateur des MoM peut être considéré comme un bon estimateur de la moyenne.

Première étape dans la construction de l'estimateur des MoM : Il s'agit de partitionner l'ensemble composé de n tirages aléatoires i.i.d. en k blocs disjoints et de calculer la moyenne empirique $\bar{\mu}_j$ de chacun des k blocs. Pour des raisons de simplicité, supposons que le nombre de tirages aléatoires n est un multiple du nombre de blocs k, de sorte que $n = mk$, avec m, k, n étant des entiers positifs. Pour chacun des k blocs, la moyenne empirique est un estimateur sans biais de l'espérance de la distribution. En effet, comme les tirages aléatoires X_i sont i.i.d. de moyenne $\mu = \mathbb{E}[X]$, alors $\mathbb{E}[\bar{\mu}_j] = \mu$. De plus, puisque les tirages aléatoires sont i.i.d. de variance $\mathbb{V}[X_i] = \sigma^2$, l'écart type des moyennes empiriques $\bar{\mu}_j$ est donné par $\frac{\sigma}{\sqrt{m}}$. Cela signifie que, dans chacun des k blocs, la moyenne empirique ne s'écarte pas de l'espérance de la distribution de plus d'une quantité égale à $\frac{\sigma}{\sqrt{m}}$.

Deuxième étape : Il s'agit de prendre la médiane de la séquence composée des k moyennes empiriques par bloc.

Pourquoi pouvons-nous dire que la médiane des moyennes par bloc est un bon estimateur de l'espérance de la distribution ?

En réalité, la médiane empirique ne s'écarte pas de la moyenne empirique de plus d'une quantité égale à l'écart type empirique, c'est ce qu'énonce le lemme suivant.

Lemma 4.60. *Soit X une variable aléatoire scalaire, avec un écart type σ et une médiane M. Alors, l'inégalité suivante est vérifiée :*

$$\left|\mathbb{E}\left[X\right] - M\right| \leq \sigma. \tag{4.338}$$

En supposant que σ soit suffisamment petit, nous pouvons dire que la médiane est suffisamment proche de l'espérance de la distribution ce qui en fait un bon estimateur.

Dans notre cas, si $\mathbf{X} = (X_1, \ldots, X_n)$ est un échantillon composé de n variables aléatoires indépendantes et identiquement distribuées (i.i.d.), alors $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ est la moyenne empirique associée. Notons par B_j, pour tout $j \in [\![1, k]\!]$, les k blocs composés de m variables aléatoires. La moyenne empirique par bloc $\bar{\mu}_{B_j} = \frac{1}{|B_j|}\sum_{i \in B_j} X_i$ est un estimateur sans biais de l'espérance μ de la distribution. De plus, puisque les tirages aléatoires initiaux sont i.i.d. et que les blocs sont disjoints, les moyennes par bloc sont des variables aléatoires i.i.d.. La moyenne empirique des moyennes par bloc est égale à celle de l'échantillon initial $\hat{\mu}_n$. La variance empirique est notée $\hat{\sigma}_n^2 = \frac{1}{k}\sum_{j=1}^{k}(\bar{\mu}_{B_j} - \hat{\mu}_n)^2$. Ensuite, à partir de l'Équation (4.136), l'inégalité de concentration suivante est vérifiée :

$$\left|\hat{\mu}_n - \widehat{\mathbf{MoM}}_n\right| \leq \hat{\sigma}_n. \tag{4.339}$$

Si l'écart type empirique de l'échantillon est petit, nous pouvons conclure que l'estimateur des MoM et l'estimateur de la moyenne empirique sont proches l'un de l'autre, et donc, ils devraient tous deux estimer l'espérance théorique. Néanmoins, en général, les fluctuations de l'estimateur des MoM sont plus petites que celles de l'estimateur de la moyenne empirique.

Avantages de l'estimateur des MoM : Cet estimateur permet de combiner deux estimateurs de centralité : la moyenne empirique et la médiane empirique. La moyenne empirique n'est pas robuste aux valeurs extrêmes et est significativement influencée par leur présence. Par conséquent, les k moyennes empiriques par bloc tiennent compte des valeurs extrêmes ; elles sont sans biais par rapport à l'espérance de la séquence, mais elles ont un écart type élevé. La médiane empirique est robuste aux valeurs extrêmes car elle dépend uniquement des valeurs centrales de la séquence, et l'écart type de la médiane empirique est censé être beaucoup plus faible que celui de la moyenne empirique ; cependant, la médiane peut présenter un biais par rapport à l'espérance de la séquence. Cela implique que l'estimateur final combinant moyenne et médiane empiriques peut être sensible aux données tout en n'étant pas trop influencé par les valeurs extrêmes de la distribution. De plus, comme la deuxième couche de construction de l'estimateur consiste à prendre la médiane empirique de la séquence des moyennes empiriques par bloc, un biais est introduit par rapport à l'espérance de la séquence.

Des inégalités de concentration sont fournies par les travaux de (Lugosi and Mendelson, 2019, Théorème 2 p.7, Théorème 4 p.9), et prouvent que, dans des conditions qui ne sont pas si restrictives, l'estimateur des MoM est un bon estimateur de la moyenne.

Après avoir choisi la bonne paramétrisation de l'estimateur, nous quantifions le biais de l'estimateur MoM. Étant donné que l'estimateur des MoM résulte de la combinaison des estimateurs de la moyenne empirique et de la médiane empirique, nous démontrons dans un premier temps une formule fermée pour le biais de la médiane empirique.

Theorem 4.61 (Espérance de la médiane empirique lorsque n est impair)**.** (i) *Soit $X_1, \ldots, X_{2k+1}$ une séquence composée de $2k+1$ variables aléatoires i.i.d. suivant la même distribution que $X \sim \mathcal{P}(1, \gamma)$. Soit $X_1^\star, \ldots, X_{2k+1}^\star$ la séquence associée des statistiques d'ordre. La densité de la médiane empirique $X_{k+1}^\star$ est donnée par :*

$$f_{\hat{M}_{2k+1}^{(1)}}(x) = \frac{(2k+1)!}{k!k!}\left(1 - \frac{1}{x^\gamma}\right)^k \frac{1}{x^{k\gamma}}\left(\frac{\gamma}{x^{\gamma+1}}\right)\mathbb{1}_{\{x \geq 1\}}. \tag{4.340}$$

L'espérance de la médiane empirique est définie comme suit :

$$\mathbb{E}[\hat{M}_{2k+1}^{(1)}] = \frac{(2k+1)!}{k!k!}\mathfrak{B}\left(k+1, k+1-\frac{1}{\gamma}\right) \tag{4.341}$$

où $\mathfrak{B}(x,y) = \int_0^1 (1-t)^{x-1}t^{y-1}\mathrm{d}t = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$ est la fonction Beta, et $\Gamma(z) = \int_0^{+\infty} t^{z-1}e^{-t}\mathrm{d}t$ est la fonction Gamma.

(ii) *Soit $Y_1, \ldots, Y_{2k+1}$ une séquence composée de $2k+1$ variables aléatoires i.i.d. suivant la même distribution que $Y \sim \mathcal{P}(x_m, \gamma)$. Alors, d'après la propriété d'échelle de la distribution de Pareto (Théorème 4.55), nous obtenons la relation de proportionnalité suivante entre les distributions non-standardisée $\mathcal{P}(x_m, \gamma)$ et standardisée $\mathcal{P}(1, \gamma)$ de Pareto :*

$$(Y_1, \ldots, Y_{2k+1}) \overset{d}{=} (x_m X_1, \ldots, x_m X_{2k+1}). \tag{4.342}$$

En particulier, une relation de proportionnalité de facteur x_m relie la médiane empirique de la distribution de Pareto non-standardisée $\mathcal{P}(x_m, \gamma)$ à celle de la distribution de Pareto standardisée $\mathcal{P}(1, \gamma)$ telle que $\hat{M}_{2k+1}^{(x_m)} \overset{d}{=} x_m \hat{M}_{2k+1}^{(1)}$, ainsi que leurs espérances :

$$\mathbb{E}[\hat{M}_{2k+1}^{(x_m)}] = x_m \mathbb{E}[\hat{M}_{2k+1}^{(1)}]. \tag{4.343}$$

Theorem 4.62 (Espérance de la médiane empirique lorsque n est pair)**.** (i) *Soit $X_1, \ldots, X_{2k}$ $2k$ variables aléatoires indépendantes et identiquement distribuées (i.i.d.) suivant la même distribution que $X \sim \mathcal{P}(1, \gamma)$. Soit $X_1^\star, \ldots, X_{2k}^\star$ la séquence associée de statistiques ordonnées. Nous définissons*

$$\hat{M}_{2k} := \frac{1}{2}(X_k^\star + X_{k+1}^\star). \tag{4.344}$$

L'espérance de la médiane empirique est donnée par :

$$\mathbb{E}\left[\hat{M}_{2k}\right] = \frac{1}{2}\left[\frac{(2k)!}{(k-1)!k!}\mathfrak{B}\left(k, k+1-\frac{1}{\gamma}\right) + \frac{(2k)!}{k!(k-1)!}\mathfrak{B}\left(k+1, k-\frac{1}{\gamma}\right)\right] \tag{4.345}$$

(ii) *Soit $Y_1, \ldots, Y_{2k}$ $2k$ variables aléatoires i.i.d. suivant la même distribution que $Y \sim \mathcal{P}(x_m, \gamma)$. Alors, d'après la propriété d'échelle de la distribution de Pareto (Théorème 4.55), nous obtenons la relation de proportionnalité suivante entre les distributions $\mathcal{P}(x_m, \gamma)$ et $\mathcal{P}(1, \gamma)$:*

$$(Y_1, \ldots, Y_{2k}) \overset{d}{=} (x_m X_1, \ldots, x_m X_{2k}). \tag{4.346}$$

En particulier, une relation de proportionnalité de facteur x_m relie la médiane empirique de la distribution de Pareto non standardisée $\mathcal{P}(x_m, \gamma)$ à celle de la distribution de Pareto standardisée $\mathcal{P}(1, \gamma)$ telle que $\hat{M}_{2k}^{(x_m)} \overset{d}{=} x_m \hat{M}_{2k}^{(1)}$, ainsi que leurs espérances :

$$\mathbb{E}[\hat{M}_{2k}^{(x_m)}] = x_m \mathbb{E}[\hat{M}_{2k}^{(1)}]. \tag{4.347}$$

Theorem 4.63. *Lorsque $n \to +\infty$,*

$$\mathbb{E}\left[\hat{M}_n\right] = \sqrt[\gamma]{2}\left(1 + \frac{1}{2\gamma n} + \frac{1}{2\gamma^2 n} + o\left(\frac{1}{n}\right)\right). \tag{4.348}$$

Le défi dans le calcul du biais de l'estimateur des MoM réside dans le fait que la densité de l'estimateur est nécessaire pour calculer l'espérance des MoM. Cependant, cet estimateur est défini comme la médiane empirique des moyennes par blocs. La distribution d'une statistique d'ordre est basée sur la distribution de l'échantillon sous-jacent, c'est-à-dire sur l'échantillon des moyennes par blocs dont la distribution est inconnue. Pour aborder ce problème, nous proposons une formule analytique pour la fonction caractéristique des moyennes par blocs.

Theorem 4.64 (Fonction caractéristique de la loi de Pareto standardisée)**.** *Soit $X \sim \mathcal{P}(1, \gamma)$ avec $\gamma > 1$. La fonction caractéristique de X, définie par $\phi_X(t) = \mathbb{E}\left[e^{itX}\right] = \int_1^{+\infty} \frac{\gamma}{x^{\gamma+1}} \mathrm{d}x$, est une solution de l'Équation Différentielle Ordinaire (EDO) :*

$$\frac{\partial}{\partial t} J_\gamma(t) - \frac{\gamma}{t} J_\gamma(t) = -\frac{\gamma}{t} e^{it}, \tag{4.349}$$

et prend la forme suivante :

$$\forall t \in \mathbb{R}, \phi_X(t) = \gamma(-it)^\gamma \Gamma(-\gamma, -it). \tag{4.350}$$

où $\Gamma(z) = \int_0^{+\infty} u^{z-1} e^{-u} \mathrm{d}u$, et satisfait la relation suivante $\Gamma(z+1) = z\Gamma(z)$.

À partir de l'Équation (**??**), la fonction caractéristique des moyennes par bloc $\bar{\mu}_j$ peut être définie comme suit :

$$\phi_{\bar{\mu}_j}(t) = \gamma^m \left(-i\frac{t}{m}\right)^{m\gamma} \left(\Gamma\left(-\gamma, -i\frac{t}{m}\right)\right)^m. \tag{4.351}$$

Ensuite, la fonction de densité des moyennes par bloc peut être récupérée à l'aide de la transformée de Fourier inverse, de la manière suivante :

$$f_{\bar{\mu}_j}(x) = \int_{\mathbb{R}} \frac{1}{2\pi} e^{-itx} \phi_{\bar{\mu}_j}(t) \mathrm{d}t. \tag{4.352}$$

Cependant, comme la fonction caractéristique des moyennes par bloc $\bar{\mu}_j$ dépend de la fonction Gamma, qui est déjà une intégrale, il est très compliqué de déterminer une forme analytique explicite pour la densité des moyennes par bloc. Dans ce cadre, il est difficile d'obtenir une forme analytique du biais entre l'estimateur des **MoM** appliqué

à la distribution de Pareto standardisée et son espérance. C'est encore un travail en cours.

Une étude numérique du biais de l'estimateur des $\widehat{\mathbf{MoM}}_n$ est réalisée.

Le taux de convergence du biais entre l'estimateur des **MoM** appliqué à la distribution de Pareto standardisée et son espérance semble dépendre de l'indice de Pareto γ. En effet, le taux de convergence varie entre $\frac{1}{n^{0.7}}$ et $\frac{1}{n}$ et se rapproche de $\frac{1}{n}$ lorsque la queue de la distribution devient de plus en plus fine, c'est-à-dire lorsque γ augmente.

Rappelons que l'objectif est d'étudier le biais entre l'estimateur des MoM appliqué à la queue de la distribution de Pareto standardisée et l'$\mathbf{ES}_\alpha$. Dans le cas idéalisé, le quantile empirique d'ordre α correspond à la vraie $\mathbf{VaR}_\alpha$, et l'$\mathbf{ES}_\alpha$ empirique correspond à la moyenne empirique de la distribution de Pareto standardisée conditionnée à ses valeurs supérieures à la vraie $\mathbf{VaR}_\alpha$. Le seuil de conditionnement est indépendant de l'échantillon sous-jacent. Cela implique que les échantillons supérieurs à la $\mathbf{VaR}_\alpha$ théorique sont indépendants et identiquement distribués (i.i.d.) et que les propriétés de stabilité par conditionnement et de mise à l'échelle sont valides. Par conséquent, la distribution de Pareto standardisée conditionnée à ses valeurs supérieures à la vraie $\mathbf{VaR}_\alpha$ est toujours une distribution de Pareto avec le même paramètre de forme γ, mais un nouveau paramètre d'échelle égal au seuil de conditionnement $\mathbf{VaR}_\alpha$: $\mathcal{P}(\mathbf{VaR}_\alpha, \gamma)$. De plus, la propriété de mise à l'échelle établit une relation de proportionnalité entre la distribution de Pareto non-standardisée $\mathcal{P}(\mathbf{VaR}_\alpha, \gamma)$ et la distribution de Pareto standardisée $\mathcal{P}(1, \gamma)$, avec un facteur de proportionnalité égal au paramètre de conditionnement $\mathbf{VaR}_\alpha$. Par conséquent, en combinant les propriétés de stabilité par conditionnement et de mise à l'échelle, la distribution de Pareto standardisée conditionnée à ses valeurs supérieures à la vraie $\mathbf{VaR}_\alpha$ est proportionnelle à la distribution marginale $\mathcal{P}(1, \gamma)$, avec un facteur de proportionnalité égal à $\mathbf{VaR}_\alpha$.

Cela implique que l'estimateur des MoM appliqué à la distribution de Pareto standardisée conditionnée à ses valeurs supérieures à $\mathbf{VaR}_\alpha$ est proportionnel à l'estimateur des MoM appliqué à l'ensemble de la distribution de Pareto standardisée, avec un facteur de proportionnalité égal à $\mathbf{VaR}_\alpha$. De même, l'$\mathbf{ES}_\alpha$ est proportionnel à l'espérance de la distribution de Pareto standardisée, avec un facteur de proportionnalité égal à $\mathbf{VaR}_\alpha$. Ainsi, le biais entre l'estimateur $\widehat{\mathbf{LV}}_n$ appliqué à la distribution de Pareto standardisée conditionnée à ses valeurs supérieures à $\mathbf{VaR}_\alpha$ et l'$\mathbf{ES}_\alpha$ est proportionnel au biais entre l'estimateur des MoM appliqué à la distribution de Pareto standardisée complète et l'espérance, avec un facteur de proportionnalité égal à $\mathbf{VaR}_\alpha$. De plus, parce que le seuil de conditionnement est indépendant de l'échantillon sous-jacent, la vitesse de convergence du biais entre l'estimateur des MoM appliqué à la distribution de Pareto standardisée conditionnée à ses valeurs supérieures à $\mathbf{VaR}_\alpha$ et le vrai $\mathbf{ES}_\alpha$ est la même que celle du biais entre l'estimateur des MoM appliqué à l'ensemble de la distribution de Pareto standardisée et son espérance.

Dans le cas réaliste, la quantile empirique d'ordre α ne correspond pas à la vraie $\mathbf{VaR}_\alpha$, et l'$\mathbf{ES}_\alpha$ empirique correspond à la moyenne empirique de la distribution de Pareto standardisée conditionnée à ses valeurs supérieures au quantile empirique d'ordre α. Le seuil de conditionnement est une statistique d'ordre et dépend de l'échantillon sous-

jacent. Cela implique que les échantillons supérieurs au quantile empirique d'ordre α ne sont plus indépendants et identiquement distribués (i.i.d.), et que les propriétés de stabilité par conditionnement et de mise à l'échelle ne sont plus valides. Par conséquent, la distribution de Pareto standardisée conditionnée à ses valeurs supérieures au quantile empirique d'ordre α n'est pas nécessairement une distribution de Pareto. Ainsi, la distribution des échantillons supérieurs au quantile empirique d'odre α est inconnue, et il est difficile d'établir une formule fermée pour le biais entre le vrai $\mathbf{ES}_\alpha$ et l'$\mathbf{ES}_\alpha$ empirique. Pour cette raison, les expériences numériques donnent une idée de la vitesse de convergence du biais et montrent si la vitesse de convergence du biais dans le cas réaliste est éloignée ou proche de la vitesse de convergence du biais dans le cas idéal. L'étude numérique montre que le taux de convergence est très proche de celui du biais dans le cas idéalisé.

Moyennes Tronquées (TM)

Nous fournissons une formule explicite pour le biais de l'estimateur des Moyennes Tronquées tant dans le cas non-asymptotique que dans le cas asymptotique.

L'estimateur des Moyennes Tronquées est basé sur l'idée suivante. L'idée la plus naturelle pour améliorer la performance de l'estimateur de la moyenne empirique est de supprimer d'éventuels points aberrants en tronquant l'échantillon. L'estimateur des Moyennes Tronquées est défini en éliminant une fraction de l'échantillon ϵn des points les plus grands et les plus petits pour un certain paramètre $\epsilon \in (0, 1)$, puis en moyennant le reste.

L'estimateur des Moyennes Tronquées est construit comme suit. Nous divisons les données en deux parties égales. Une moitié est utilisée pour déterminer le seuil de troncature approprié. Les points de l'autre moitié sont moyennés, à l'exception des points de données qui tombent en dehors de la région de troncature, qui sont seuillés.

Pour des raisons de commodité, nous supposons que les données sont composées de $2n$ copies indépendantes de la variable aléatoire $X \sim \mathcal{P}(1, \gamma)$, notées $X_1, \ldots, X_n, Y_1, \ldots, Y_n$. Nous notons respectivement par $X_1^\star, \ldots, X_n^\star$ et $Y_1^\star, \ldots, Y_n^\star$ les séquences de statistiques d'ordre liées aux échantillons $X_1, \ldots, X_n =: \mathbf{X}_n$ et $Y_1, \ldots, Y_n =: \mathbf{Y}_n$. L'espérance de la distribution est notée $\mu = \mathbb{E}[X] = \frac{\gamma}{\gamma-1}$. La troncature des deux côtés de la distribution est pertinente lorsque les deux queues de la distribution sont non-bornées et contiennent des valeurs extrêmes. Cependant, la distribution de Pareto présente la particularité que sa queue gauche est bornée, tandis que sa queue droite ne l'est pas. Cela implique que, pour éliminer les valeurs aberrantes, il n'est pas nécessaire de seuiller la queue gauche de la distribution, mais seulement la queue droite.

Dans ce cas spécifique, la fonction de troncature est donnée par :

$$\phi_{\mathbf{Y}_n}(x) = \begin{cases} Y^\star_{\lceil (1-\epsilon)n \rceil} & \text{si } x > Y^\star_{\lceil (1-\epsilon)n \rceil}, \\ x & \text{si } x \in \left[1, Y^\star_{\lceil (1-\epsilon)n \rceil}\right]. \end{cases} \tag{4.353}$$

De plus, il est important de noter que les seuils de troncature sont fixés sur un échantillon qui est indépendant de l'échantillon sur lequel l'estimateur est construit. Avec cette notation en place, l'estimateur est défini comme suit :

(i) Étant donné le niveau de confiance $\delta \geq 8e^{-3n/16}$, on définit le seuil de troncature de la manière suivante :

$$\epsilon_n = \frac{16\log(8/\delta)}{3n}. \tag{4.354}$$

(ii) Puis, l'estimateur des Moyennes Tronquées est donné par :

$$\widehat{TM}_{2n} = \frac{1}{n}\sum_{i=1}^{n} \phi_{\mathbf{Y}_n}(X_i). \tag{4.355}$$

L'inconvénient de cet estimateur est que, pour un niveau de confiance élevé, la quantité de données requise pour obtenir un seuil de troncature raisonnable est très importante.

Dans (Lugosi and Mendelson, 2019, p.14), une inégalité de concentration de l'estimateur des Moyennes Tronquées est donnée pour quantifier la performance de l'estimateur.

Dans ce contexte, nous fournissons une formule fermée pour le biais de l'estimateur des Moyennes Tronquées dans le cas non-asymptotique.

Nous étudions l'évolution du biais entre l'estimateur des Moyennes Tronquées appliqué à la queue de la distribution de Pareto standardisée et l'$\mathbf{ES}_\alpha$ dans deux cadres différents. Le premier cadre fait référence au cas idéalisé où le quantile empirique d'ordre α correspond à la $\mathbf{VaR}_\alpha$ théorique. Le second cadre correspond au cas réaliste où le quantile empirique d'ordre α ne correspond pas à la $\mathbf{VaR}_\alpha$ théorique. Pour ce faire, nous étudions d'abord le biais entre l'estimateur des Moyennes Tronquées appliqué à la distribution de Pareto standardisée $\mathcal{P}(1,\gamma)$ complète et l'espérance.

(i) Le biais entre l'estimateur des Moyennes Tronquées et l'espérance de la distribution de Pareto est défini par :

$$B_\mu(\widehat{TM}_{2n}) = \mathbb{E}\left[\widehat{TM}_{2n}\right] - \mu. \tag{4.356}$$

Proposition 4.65 (Biais de l'estimateur des Moyennes Tronquées dans la distribution de Pareto standardisée)**.** *Soient $X_1,\ldots,X_n,Y_1,\ldots,Y_n$ et $X_1^\star,\ldots,X_n^\star,Y_1^\star,\ldots,Y_n^\star$, respectivement $2n$ copies indépendantes de la variable aléatoire X qui suit une distribution de Pareto standardisée $\mathcal{P}(1,\gamma)$, et les statistiques d'ordre associées. Soit $\epsilon_n \in (0,1)$ le seuil de troncature qui satisfait l'Équation (4.175). Le biais entre l'estimateur des Moyennes Tronquées appliqué à la distribution de Pareto standardisée $\mathcal{P}(1,\gamma)$ complète et l'espérance est alors donné comme suit :*

$$B_\mu(\widehat{TM}_{2n}) = -\frac{1}{\gamma-1}\mathbb{E}\left[\left(Y^\star_{\lceil(1-\epsilon)n\rceil}\right)^{1-\gamma}\right] = -\frac{1}{\gamma-1}\kappa_{\epsilon,n}\mathfrak{B}\left(n-\left\lceil(1-\epsilon)n\right\rceil+2-\frac{1}{\gamma},\left\lceil(1-\epsilon)n\right\rceil\right). \tag{4.357}$$

où $\kappa_{\epsilon,n} = \frac{n!}{\left(\left\lceil(1-\epsilon)n\right\rceil-1\right)!\left(n-\left\lceil(1-\epsilon)n\right\rceil\right)!}$ *et* $\mathfrak{B}(x,y) = \int_0^1 u^{x-1}(1-u)^{y-1}\mathrm{d}u = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$.

Theorem 4.66 (Biais asymptotique des Moyennes Tronquées dans la distribution de Pareto standardisée)**.** *Lorsque* $n \to +\infty$*,*

$$B_\mu(\widehat{TM}_{2n}) = -\frac{1}{\gamma - 1} \frac{\Gamma\left(\lfloor C_\delta \rfloor + 2 - \frac{1}{\gamma}\right)}{\Gamma(\lfloor C_\delta \rfloor + 1)} n^{\frac{1}{\gamma}-1} + o\left(n^{\frac{1}{\gamma}-1}\right) \tag{4.358}$$

où pour tout $x \in \mathbb{R}$*,* $\Gamma(x) = \int_0^{+\infty} u^{x-1} e^{-u} \mathrm{d}u$ *est la fonction Gamma.*

Une formule fermée pour le biais entre l'estimateur des Moyennes Tronquées à droite appliqué à la distribution de Pareto standardisée $\mathcal{P}(1, \gamma)$ complète et l'espérance est fournie. Rappelons que l'objectif est d'étudier le biais entre l'estimateur des Moyennes Tronquées à droite appliqué à la queue de la distribution de Pareto standardisée et l'$\mathbf{ES}_\alpha$. Dans le cas idéalisé, le quantile empirique d'ordre α correspond à la vraie $\mathbf{VaR}_\alpha$, et l'$\mathbf{ES}_\alpha$ empirique correspond à la moyenne empirique de la distribution de Pareto standardisée conditionnée à ses valeurs supérieures à la vraie $\mathbf{VaR}_\alpha$. Le seuil de conditionnement est indépendant de l'échantillon sous-jacent. Cela implique que les échantillons supérieurs à la $\mathbf{VaR}_\alpha$ théorique sont indépendants et identiquement distribués (i.i.d.) et que les propriétés de stabilité par conditionnement et de mise à l'échelle sont valides. Par conséquent, la distribution de Pareto standardisée conditionnée à ses valeurs supérieures à la vraie $\mathbf{VaR}_\alpha$ est toujours une distribution de Pareto avec le même paramètre de forme γ, mais un nouveau paramètre d'échelle égal au seuil de conditionnement $\mathbf{VaR}_\alpha$: $\mathcal{P}(\mathbf{VaR}_\alpha, \gamma)$. De plus, la propriété de mise à l'échelle établit une relation de proportionnalité entre la distribution de Pareto non-standardisée $\mathcal{P}(\mathbf{VaR}_\alpha, \gamma)$ et la distribution de Pareto standardisée $\mathcal{P}(1, \gamma)$, avec un facteur de proportionnalité égal au paramètre de conditionnement $\mathbf{VaR}_\alpha$. Par conséquent, en combinant les propriétés de stabilité par conditionnement et de mise à l'échelle, la distribution de Pareto standardisée conditionnée à ses valeurs supérieures à la vraie $\mathbf{VaR}_\alpha$ est proportionnelle à la distribution marginale $\mathcal{P}(1, \gamma)$, avec un facteur de proportionnalité égal à $\mathbf{VaR}_\alpha$.

Cela implique que l'estimateur des Moyennes Tronquées appliqué à la distribution de Pareto standardisée conditionnée à ses valeurs supérieures à $\mathbf{VaR}_\alpha$ est proportionnel à l'estimateur des Moyennes Tronquées appliqué à l'ensemble de la distribution de Pareto standardisée, avec un facteur de proportionnalité égal à $\mathbf{VaR}_\alpha$. De même, l'$\mathbf{ES}_\alpha$ est proportionnel à l'espérance de la distribution de Pareto standardisée, avec un facteur de proportionnalité égal à $\mathbf{VaR}_\alpha$. Ainsi, le biais entre l'estimateur des Moyennes Tronquées appliqué à la distribution de Pareto standardisée conditionnée à ses valeurs supérieures à $\mathbf{VaR}_\alpha$ et l'$\mathbf{ES}_\alpha$ est proportionnel au biais entre l'estimateur des Moyennes Tronquées appliqué à la distribution de Pareto standardisée complète et l'espérance, avec un facteur de proportionnalité égal à $\mathbf{VaR}_\alpha$. De plus, parce que le seuil de conditionnement est indépendant de l'échantillon sous-jacent, la vitesse de convergence du biais entre l'estimateur des Moyennes Tronquées appliqué à la distribution de Pareto standardisée conditionnée à ses valeurs supérieures à $\mathbf{VaR}_\alpha$ et le vrai $\mathbf{ES}_\alpha$ est la même que celle du biais entre l'estimateur des Moyennes Tronquées appliqué à l'ensemble de la distribution de Pareto standardisée et son espérance. La formule fermée du biais fournit la vitesse de convergence comme indiqué dans le Théorème **??**. La vitesse de convergence varie

en fonction de l'indice de Pareto γ et satisfait la règle $n^{\frac{1}{\gamma}-1}$, multipliée par la constante appropriée.

Dans le cas réaliste, la quantile empirique α ne correspond pas à la vraie $\mathbf{VaR}_\alpha$, et l'$\mathbf{ES}_\alpha$ empirique correspond à la moyenne empirique de la distribution de Pareto standardisée conditionnée à ses valeurs supérieures au quantile empirique d'ordre α. Le seuil de conditionnement est une statistique d'ordre et dépend de l'échantillon sous-jacent. Cela implique que les échantillons supérieurs au quantile empirique d'ordre α ne sont plus indépendants et identiquement distribués (i.i.d.), et que les propriétés de stabilité par conditionnement et de mise à l'échelle ne sont plus valides. Par conséquent, la distribution de Pareto standardisée conditionnée à ses valeurs supérieures au quantile empirique d'ordre α n'est pas nécessairement une distribution de Pareto. Ainsi, la distribution des échantillons supérieurs au quantile empirique d'odre α est inconnue, et il est difficile d'établir une formule fermée pour le biais entre le vrai $\mathbf{ES}_\alpha$ et l'$\mathbf{ES}_\alpha$ empirique. Pour cette raison, les expériences numériques donnent une idée de la vitesse de convergence du biais et montrent si la vitesse de convergence du biais dans le cas réaliste est éloignée ou proche de la vitesse de convergence du biais dans le cas idéal.

L'étude numérique montre que le taux de convergence du biais entre l'$\mathbf{ES}_\alpha$ empirique et le vrai $\mathbf{ES}_\alpha$ dans le cas réaliste semble être très proche de celui du biais entre l'$\mathbf{ES}_\alpha$ empirique et le vrai $\mathbf{ES}_\alpha$ dans le cas idéalisé.

Lee-Valiant (LV)

L'estimateur de Lee-Valiant est une version améliorée de l'estimateur des MoM, qui inclut un terme de correction. Etant donné qu'une formule fermée pour le biais de l'estimateur des MoM n'est pas disponible, il est difficile d'en déduire une formule fermée pour le biais de l'estimateur LV. Pour cette raison, nous proposons une étude empirique qui nous permet de comprendre le comportement du biais entre l'estimateur LV appliqué à la queue de la distribution de Pareto standardisée et le vrai $\mathbf{ES}_\alpha$, et de déterminer si le taux de convergence du biais dépend de l'indice de Pareto.

Definition 4.67 (Estimateur de Lee-Valiant)**.** *Pour un niveau de confiance δ donné, on définit l'estimateur des MoM $\widehat{\mathbf{MoM}}_n = \mathbf{MoM}(X_1, \ldots, X_n)$, calculé sur $k = \log\left(\frac{1}{\delta}\right) \leq n$ blocs avec $\delta \geq e^{-n}$, et k un entier positif. L'estimateur de Lee-Valiant est alors défini de la manière suivante :*

$$\widehat{LV}_n = \widehat{\mathbf{MoM}}_n + \frac{1}{n}\sum_{i=1}^{n}(X_i - \widehat{\mathbf{MoM}}_n)(1 - \min(\alpha(X_i - \widehat{\mathbf{MoM}}_n)^2, 1)) \tag{4.359}$$

où le paramètre α est la solution de l'équation monotone, linéaire par morceaux :

$$\sum_{i=1}^{n}\min(\alpha(X_i - \widehat{\mathbf{MoM}}_n)^2, 1) = \frac{1}{3}\log\left(\frac{1}{\delta}\right). \tag{4.360}$$

Le zéro de l'équation :

$$\sum_{i=1}^{n}\min(\alpha(X_i - \widehat{\mathbf{MoM}}_n)^2, 1) - \frac{1}{3}\log\left(\frac{1}{\delta}\right) = 0 \tag{4.361}$$

peut être déterminé en utilisant un algorithme de dichotomie.

Dans (Gobet et al., 2022, p.14, Thm. 2.5), un résultat de déviation est fourni. Voir (Lee and Valiant, 2022) pour plus de détails sur l'estimateur de Lee-Valiant.

La présente étude vise à analyser l'évolution du biais entre l'estimateur de Lee-Valiant appliqué à la queue de la distribution de Pareto standardisée et l'$\mathbf{ES}_\alpha$ dans deux cadres différents. Le premier cadre fait référence au cas idéalisé où la quantile empirique α correspond à la $\mathbf{VaR}_\alpha$ théorique. Le second cadre correspond au cas réaliste où la quantile empirique α ne correspond pas à la $\mathbf{VaR}_\alpha$ théorique.

Pour ce faire, nous étudions d'abord le biais entre l'estimateur de Lee-Valiant appliqué à la distribution de Pareto standardisée $\mathcal{P}(1,\gamma)$ complète et l'espérance. L'étude numérique montre que la vitesse de convergence du biais entre l'estimateur LV appliqué à la distribution de Pareto standardisée $\mathcal{P}(1,\gamma)$ complète et son espérance est supérieure à $\frac{1}{n}$ et inférieure à $\frac{1}{\sqrt{n}}$.

Une fois que l'évolution du biais entre l'estimateur $\widehat{\mathbf{LV}}_n$ et l'espérance de la distribution de Pareto standardisée $\mathcal{P}(1,\gamma)$ complète, ainsi que sa vitesse de convergence sont étudiées, nous nous intéressons à l'évolution du biais entre l'estimateur $\widehat{\mathbf{LV}}_n$ appliqué à la queue de la distribution de Pareto standardisée et l'$\mathbf{ES}_\alpha$ dans deux cas : un *cas idéalisé* dans lequel le seuil de conditionnement correspond à la $\mathbf{VaR}_\alpha$ théorique, et un *cas réaliste* dans lequel le seuil de conditionnement correspond au quantile empirique d'ordre α.

Cas idéalisé : Dans le cas idéalisé, le quantile empirique d'ordre α correspond à la vraie $\mathbf{VaR}_\alpha$, qui est supposée connue. Dans ce cas, l'estimateur de l'$\mathbf{ES}_\alpha$ dans la distribution de Pareto standardisée $\mathcal{P}(1,\gamma)$ est la moyenne empirique de la distribution de Pareto standardisée conditionnée à ses valeurs supérieures à la vraie $\mathbf{VaR}_\alpha$. Le seuil de conditionnement est indépendant de l'échantillon sous-jacent. Cela implique que les échantillons supérieurs à la vraie $\mathbf{VaR}_\alpha$ sont indépendants et identiquement distribués (i.i.d.) et que les propriétés de stabilité par conditionnement et de mise à l'échelle, comme indiqué dans le Théorème 4.55, restent valides. La propriété de stabilité par conditionnement implique que la distribution de Pareto standardisée $\mathcal{P}(1,\gamma)$ conditionnée à ses valeurs supérieures à $\mathbf{VaR}_\alpha$ est toujours une distribution de Pareto, avec le même paramètre de forme γ, mais un nouveau paramètre d'échelle égal au paramètre de conditionnement $\mathbf{VaR}_\alpha$: $\mathcal{P}(\mathbf{VaR}_\alpha,\gamma)$. La propriété d'échelle stipule que cette distribution de Pareto non-standardisée $\mathcal{P}(\mathbf{VaR}_\alpha,\gamma)$ est proportionnelle à la distribution marginale de Pareto $\mathcal{P}(1,\gamma)$, avec un facteur de proportionnalité égal au paramètre de conditionnement $\mathbf{VaR}_\alpha$. Par conséquent, la distribution de Pareto standardisée conditionnée à ses valeurs supérieures à $\mathbf{VaR}_\alpha$ est proportionnelle à la distribution de Pareto standardisée, avec un facteur de proportionnalité égal à $\mathbf{VaR}_\alpha$.

Cela implique que l'estimateur $\widehat{\mathbf{LV}}_n$ appliqué à la distribution de Pareto standardisée conditionnée à ses valeurs étant supérieures à $\mathbf{VaR}_\alpha$ est proportionnel à l'estimateur $\widehat{\mathbf{LV}}_n$ appliqué à l'ensemble de la distribution de Pareto standardisée, avec un facteur de proportionnalité égal à $\mathbf{VaR}_\alpha$. De même, l'$\mathbf{ES}_\alpha$ est proportionnel à l'espérance de la

distribution de Pareto standardisée, avec un facteur de proportionnalité égal à $\mathbf{VaR}_\alpha$. Ainsi, le biais entre l'estimateur $\widehat{\mathbf{LV}}_n$ appliqué à la distribution de Pareto standardisée conditionnée à ses valeurs supérieures à $\mathbf{VaR}_\alpha$ et l'$\mathbf{ES}_\alpha$ est proportionnel au biais entre l'estimateur $\widehat{\mathbf{LV}}_n$ appliqué à la distribution de Pareto standardisée complète et l'espérance, avec un facteur de proportionnalité égal à $\mathbf{VaR}_\alpha$. De plus, parce que le seuil de conditionnement ne dépend pas de l'échantillon sous-jacent, la vitesse de convergence du biais entre l'estimateur LV appliqué à la distribution de Pareto standardisée conditionnée à ses valeurs supérieures à $\mathbf{VaR}_\alpha$ et le vrai $\mathbf{ES}_\alpha$ est la même que celle du biais entre l'estimateur LV appliqué à l'ensemble de la distribution de Pareto standardisée et son espérance.

Cas réaliste : Dans le cas réaliste, le quantile empirique d'ordre α ne correspond pas à la vraie $\mathbf{VaR}_\alpha$. Dans ce cas, l'estimateur de l'$\mathbf{ES}_\alpha$ dans la distribution de Pareto standardisée $\mathcal{P}(1,\gamma)$ est la moyenne empirique de la distribution de Pareto standardisée conditionnée à ses valeurs supérieures au quantile empirique d'ordre α. Le seuil de conditionnement est une statistique d'ordre et dépend de l'échantillon sous-jacent. Cela implique que les échantillons supérieurs au quantile empirique d'ordre α ne sont plus indépendants et identiquement distribués (i.i.d.). Par conséquent, la distribution des échantillons supérieurs au quantile empirique d'ordre α n'est pas nécessairement une distribution de Pareto. Les propriétés de stabilité par conditionnement et de mise à l'échelle de la distribution de Pareto, énoncées dans le Théorème 4.55, ne sont plus valides. Ainsi, la distribution des échantillons supérieurs au quantile empirique α est inconnue et il est difficile d'établir une formule analytique pour le biais entre l'estimateur LV appliqué à la queue de la distribution de Pareto au-delà de ce quantile empirique d'orddre α et le vrai $\mathbf{ES}_\alpha$.

Pour cette raison, nous fournissons une étude expérimentale pour donner un aperçu de la vitesse de convergence du biais entre l'estimateur LV appliqué à la queue de la distribution de Pareto standardisée au-delà du quantile empirique d'ordre α et le vrai $\mathbf{ES}_\alpha$. La vitesse de convergence du biais entre l'estimateur LV appliqué à la distribution de Pareto standardisée conditionnée à ses valeurs supérieures au quantile empirique d'ordre α et le vrai $\mathbf{ES}_\alpha$ semble similaire à celle du biais dans le cas idéalisé, c'est-à-dire entre $\frac{1}{n}$ et $\frac{1}{\sqrt{n}}$. Plus précisément, des tests ont été effectués pour trois distributions de Pareto : $\mathcal{P}(1,2.5)$, $\mathcal{P}(1,3.5)$ et $\mathcal{P}(1,5)$, et la vitesse de convergence se situe entre $\frac{1}{n^{0.9}}$ et $\frac{1}{n^{0.7}}$.

Étude comparative Enfin, nous soutenons l'analyse théorique par des expériences. Nous comparons également les performances des différents estimateurs.

Biais entre l'$\mathbf{ES}_\alpha$ empirique et le vrai $\mathbf{ES}_\alpha$ dans le cas réaliste (travail en cours) Comme mentionné précédemment, dans le cas réaliste, le quantile empirique d'ordre α ne correspond pas à la $\mathbf{VaR}_\alpha$ théorique. Dans ce cas, l'estimateur de l'$\mathbf{ES}_\alpha$ dans la distribution de Pareto standardisée $\mathcal{P}(1,\gamma)$ est la moyenne empirique de la distribution de Pareto standardisée conditionnée à ses valeurs supérieures au quantile

empirique d'ordre α. Le seuil de conditionnement est une statistique d'ordre et dépend de l'échantillon sous-jacent. Cela implique que les échantillons supérieurs au quantile empirique d'ordre α ne sont plus indépendants et identiquement distribués (i.i.d.). Par conséquent, la distribution des échantillons supérieurs au quantile empirique d'ordre α n'est pas nécessairement une distribution de Pareto. Les propriétés de stabilité par conditionnement et de mise à l'échelle de la distribution de Pareto, énoncées dans le Théorème 4.55, ne sont plus valides. Ainsi, la distribution des échantillons supérieurs au quantile empirique d'ordre α est inconnue et il est plus difficile d'établir une formule analytique pour le biais entre l'$\mathbf{ES}_\alpha$ empirique et le vrai $\mathbf{ES}_\alpha$ dans la distribution de Pareto standardisée. Nous présentons un travail en cours concernant l'expression de ce biais.

La formule de l'$\mathbf{ES}_\alpha$ empirique est donnée comme suit :

$$\mathbf{ES}_\alpha^n = \frac{1}{n - \lceil \alpha n \rceil} \sum_{i=1}^{n} X_i \mathbb{1}_{\{X_i \geq q_\alpha^n\}}. \tag{4.362}$$

Pour simplifier les notations, notons $r = \lceil \alpha n \rceil$ le rang du quantile empirique. Alors, l'**ES** empirique correspond à la moyenne empirique des échantillons supérieurs à la statistique d'ordre de rang r et peut être réécrit de la manière suivante:

$$\mathbf{ES}_r^n = \frac{1}{n - r} \sum_{i=1}^{n} X_i \mathbb{1}_{\{X_i \geq X_r^\star\}}.$$

L'objectif est de calculer le biais entre l'**ES** empirique et l'**ES** théorique dans la distribution de Pareto standardisée $\mathcal{P}(1, \gamma)$.

$$B_{\mathbf{ES}_r}[\mathbf{ES}_r^n] = \mathbb{E}\left[\mathbf{ES}_r^n\right] - \mathbf{ES}_r \quad \text{où} \quad \mathbb{E}\left[\mathbf{ES}_r^n\right] = \frac{1}{n-r} \sum_{i=1}^{n} \mathbb{E}\left[X_i \mathbb{1}_{\{X_i \geq X_r^\star\}}\right] = \frac{n}{n-r} \mathbb{E}\left[X_1 \mathbb{1}_{\{X_1 \geq X_r^\star\}}\right]$$

puisque l'échantillon est i.i.d..

Sous la distribution uniforme, nous prouvons la forme de la distribution jointe :

$$\mathbb{P}(X_r^\star \leq u, X_i \leq v) = v \sum_{k=r}^{n-1} \binom{n-1}{k} u^k (1-u)^{n-1-k} + \min(u, v) \binom{n-1}{r-1} u^{r-1} (1-u)^{n-r}.$$

En procédant à la transformation du quantile suivante $\mathbf{VaR}_\alpha(X_i) = Y_i$ dans l'Équation (4.246), nous adaptons la formule ci-dessus pour la distribution de Pareto standardisée :

$$\mathbb{P}(X_r^\star \leq u, X_i \leq v) = \mathbb{P}(Y_r^\star \leq \mathbf{VaR}_\alpha(u), Y_i \leq \mathbf{VaR}_\alpha(v)).$$

"Mathematics is the art of giving the same name to different things."

(**Henri Poincaré**, Science and Method, 1908.)

Bibliography

C. Acerbi and B. Szekely. Back-testing expected shortfall. *Risk*, 27(11):76–81, 2014. pages 116, 370

C. Acerbi and D. Tasche. On the coherence of expected shortfall. *Journal of Banking & Finance*, 26(7):1487–1503, 2002a. ISSN 0378-4266. doi: https://doi.org/10.1016/S0378-4266(02)00283-2. URL https://www.sciencedirect.com/science/article/pii/S0378426602002832. pages 115, 370

C. Acerbi and D. Tasche. On the coherence of expected shortfall. *Journal of banking & finance*, 26(7):1487–1503, 2002b. pages 115, 369

C. Acerbi, C. Nordio, and C. Sirtori. Expected shortfall as a tool for financial risk management. *arXiv preprint cond-mat/0102304*, 2001. pages 74, 115, 369

J. A. Adamu. Jameel's criterion and jameel's advanced stressed models: an ideas that lead to non-normal stocks brownian motion models. *Noble International Journal of Business and Management Research*, 1(10):136–154, 2017. page 180

S. Ahn, J. H. Kim, and V. Ramaswami. A new class of models for heavy tailed distributions in finance and insurance risk. *Insurance: Mathematics and Economics*, 51 (1):43–52, 2012. pages 117, 372

C. Albanese, K. Jackson, and P. Wiberg. Dimension reduction in the computation of value-at-risk. *The Journal of Risk Finance*, 3(4):41–53, 2002. pages 102, 356

C. Alexander. *Market risk analysis, value at risk models*, volume 4. John Wiley & Sons, 2009. page 67

C. Alexander and M. Dakos. Assessing the accuracy of exponentially weighted moving average models for value-at-risk and expected shortfall of crypto portfolios. *Quantitative Finance*, 23(3):393–427, 2023. pages 103, 356

M. Allouche, S. Girard, and E. Gobet. A generative model for fbm with deep relu neural networks. *Journal of Complexity*, page 101667, 2022. page 204

C. Aloui. Value-at-risk analysis for energy commodities: long-range dependencies and fat-tails in return innovations. *Journal of Energy Markets*, 1(1):31–63, 2008. pages 104, 357

P.-O. Amblard and J.-F. Coeurjolly. Identification of the multivariate fractional brownian motion. *IEEE Transactions on Signal Processing*, 59(11):5152–5168, 2011a. pages 113, 211, 212, 367

P.-O. Amblard and J.-F. Coeurjolly. Identification of the Multivariate Fractional Brownian Motion. *IEEE Transactions on Signal Processing*, 59(11):5152–5168, Nov. 2011b. ISSN 1053-587X, 1941-0476. doi: 10.1109/TSP.2011.2162835. URL http://ieeexplore.ieee.org/document/5960799/. pages 86, 337

P.-O. Amblard, J.-F. Coeurjolly, F. Lavancier, and A. Philippe. Basic properties of the multivariate fractional brownian motion. *arXiv preprint arXiv:1007.0828*, 2010. page 204

A. A. Anantafortuna and A. H. Anggono. Risk measurement for trading activities based on exponential moving average and count back line strategy. *Review of Integrative Business and Economics Research*, 8:198–205, 2019. pages 103, 356, 357

B. C. Arnold. Pareto and generalized pareto distributions. In *Modeling income distributions and Lorenz curves*, pages 119–145. Springer, 2008. pages 118, 372

B. C. Arnold. Pareto distribution. *Wiley StatsRef: Statistics Reference Online*, pages 1–10, 2014. pages 118, 120, 246, 372, 376

W. Arroum. *Time-changed self-similar processes: an application to high frequency financial data*. PhD thesis, University of Southampton, 2007. pages 90, 343

P. Artzner. Thinking coherently. *Risk*, 10:68–71, 1997. page 67

P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical finance*, 9(3):203–228, 1999. doi: doi.org/10.1111/1467-9965.00068. pages 67, 75, 115, 369

V. Bally. Introduction to malliavin calculus. *Lecture Notes, available at http://perso-math. univ-mlv. fr/users/bally. vlad*, 2007. page 143

V. Bally and C. Rey. Approximation of Markov semigroups in total variation distance. *Electronic Journal of Probability*, 21(none):1 – 44, 2016. doi: 10.1214/16-EJP4079. URL https://doi.org/10.1214/16-EJP4079. page 143

V. Bally and D. Talay. The law of the euler scheme for stochastic differential equations: I. convergence rate of the distribution function. *Probability theory and related fields*, 104:43–60, 1996. page 143

J. C. A. Barata and M. S. Hussein. The moore–penrose pseudoinverse: A tutorial review of the theory. *Brazilian Journal of Physics*, 42:146–165, 2012. page 191

J.-M. Bardet, G. Lang, G. Oppenheim, A. Philippe, and M. S. Taqqu. Generators of long-range dependent processes: a survey. *Theory and applications of long-range dependence*, 1:579–623, 2003. page 204

O. Bardou, N. Frikha, and G. Pages. Recursive computation of value-at-risk and conditional value-at-risk using mc and qmc. In *Monte Carlo and Quasi-Monte Carlo Methods 2008*, pages 193–208. Springer, 2009. pages 102, 355

O. Bardou, N. Frikha, and G. Pagès. Cvar hedging using quantization-based stochastic approximation algorithm. *Mathematical Finance*, 26(1):184–229, 2016. page 196

O. E. Barndorff-Nielsen and V. Pérez-Abreu. Stationary and self-similar processes driven by lévy processes. *Stochastic processes and their applications*, 84(2):357–369, 1999. pages 83, 334

N. Bauerle and S. Desmettre. Portfolio optimization in fractional and rough heston models. *SIAM Journal on Financial Mathematics*, 11(1):240–273, 2020. pages 90, 342

J. Beirlant and Y. Goegebeur. Local polynomial maximum likelihood estimation for pareto-type distributions. *Journal of Multivariate Analysis*, 89(1):97–118, 2004. pages 119, 375

J. Beirlant, P. Vynckier, and J. L. Teugels. Tail index estimation, pareto quantile plots regression diagnostics. *Journal of the American statistical Association*, 91(436): 1659–1667, 1996. pages 119, 374

J. Beirlant, G. Dierckx, Y. Goegebeur, and G. Matthys. Tail index estimation and an exponential regression model. *Extremes*, 2:177–200, 1999. pages 119, 374

A. Ben-Israel and T. N. Greville. *Generalized inverses: theory and applications*. Springer Science & Business Media, 2006. page 191

J. Bertoin. Self-similar fragmentations. In *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, volume 38, pages 319–340. Elsevier, 2002. pages 83, 334

P. Best. *Implementing value at risk*. John Wiley & Sons, 2000. page 67

F. Biagini, Y. Hu, B. Øksendal, and T. Zhang. *Stochastic calculus for fractional Brownian motion and applications*. Springer Science & Business Media, 2008a. pages 93, 346

F. Biagini, Y. Hu, B. Øksendal, and T. Zhang. *Stochastic Calculus for Fractional Brownian Motion and Applications*. Probability and Its Applications. Springer London, London, 2008b. ISBN 978-1-85233-996-8 978-1-84628-797-8. doi: 10.1007/978-1-84628-797-8. URL http://link.springer.com/10.1007/978-1-84628-797-8. page 180

B. O. Bradley and M. S. Taqqu. Financial risk and heavy tails. In S. T. Rachev, editor, *Handbook of Heavy Tailed Distributions in Finance*, volume 1 of *Handbooks in Finance*, pages 35–103. North-Holland, Amsterdam, 2003. doi: https://doi.org/10.1016/B978-044450896-6.50004-2. URL https://www.sciencedirect.com/science/article/pii/B9780444508966500042. pages 117, 372

V. Brazauskas and R. Serfling. Robust and efficient estimation of the tail index of a single-parameter pareto distribution. *North American Actuarial Journal*, 4(4):12–27, 2000. pages 119, 374

V. Brazauskas, B. L. Jones, M. L. Puri, and R. Zitikis. Estimating conditional tail expectation with actuarial applications in view. *Journal of Statistical Planning and Inference*, 138(11):3590–3604, 2008. ISSN 0378-3758. doi: https://doi.org/10.1016/j.jspi.2005.11.011. URL https://www.sciencedirect.com/science/article/pii/S0378375808001018. Special Issue in Honor of Junjiro Ogawa (1915 - 2000): Design of Experiments, Multivariate Analysis and Statistical Inference. pages 116, 371

J. Brehmer. Elicitability and its application in risk management. *arXiv preprint arXiv:1707.09604*, 2017. page 76

S. A. Broda and M. S. Paolella. Expected shortfall for distributions in finance. In *Statistical tools for finance and insurance*, pages 57–99. Springer, 2011. pages 117, 371

M. Brzezinski. Robust estimation of the pareto tail index: a monte carlo analysis. *Empirical Economics*, 51:1–30, 2016. pages 119, 374

A. Bücher, P. N. Posch, and P. Schmidtke. Using the extremal index for value-at-risk backtesting. *Journal of Financial Econometrics*, 18(3):556–584, 2020. pages 113, 214, 367

K. Burnecki and A. Weron. Levy stable processes. from stationary to self-similar dynamics and back. an application to finance. *Acta Physica Polonica Series B*, 35(4): 1343–1358, 2004. pages 83, 334

P. Carmona, L. Coutin, and G. Montseny. Stochastic integration with respect to fractional brownian motion. In *Annales de l'IHP Probabilités et statistiques*, volume 39, pages 27–68, 2003. pages 93, 346

P. Chabardès and F. Delclaux. *Les produits dérivés*. FeniXX, 1997. page 31

G. Chan and A. Wood. Simulation of stationary gaussian vector fields. *Statistics and computing*, 9(4):265–268, 1999. page 204

R. Chattamvelli and R. Shanmugam. Pareto distribution. In *Continuous Distributions in Engineering and the Applied Sciences–Part II*, pages 179–188. Springer, 2021. pages 118, 372

L. Chaumont. Introduction aux processus auto-similaires. *Cours du DEA de Proabilités et Applications. Université de Paris VI*, 2006. pages 82, 94, 334, 347

S. X. Chen. Nonparametric estimation of expected shortfall. *Journal of financial econometrics*, 6(1):87–107, 2008. pages 116, 371

C. W. Cheong. Self-similarity in financial markets: A fractionally integrated approach. *Mathematical and computer modelling*, 52(3-4):459–471, 2010. pages 88, 340

P. Cheridito. Mixed Fractional Brownian Motion. *Bernoulli*, 7(6):913, Dec. 2001. ISSN 13507265. doi: 10.2307/3318626. URL https://www.jstor.org/stable/3318626?origin=crossref. pages 93, 143, 345

P. Cheridito. Arbitrage in fractional Brownian motion models. *Finance and Stochastics*, 7(4):533–553, Oct. 2003. ISSN 0949-2984, 1432-1122. doi: 10.1007/s007800300101. URL http://link.springer.com/10.1007/s007800300101. pages 93, 143, 345

P. Cheridito and D. Nualart. Stochastic integral of divergence type with respect to fractional brownian motion with hurst parameter $h\backslash$in (0,\frac {1}{2}) . *InAnnales de l'IHP Probabilités et statistiques, volume* 41, *pages*1049 − −1081, 2005. *pages* 93, 346

Y. H. Cheung and R. J. Powell. Anybody can do value at risk: a nonparametric teaching study. *Australasian Accounting, Business and Finance Journal*, 6(1):111–123, 2012. pages 102, 180, 355

C. Chong, M. Hoffmann, Y. Liu, M. Rosenbaum, and G. Szymanski. Statistical inference for rough volatility: Central limit theorems. *arXiv preprint arXiv:2210.01216*, 2022a. pages 89, 113, 144, 213, 234, 341, 367

C. H. Chong, M. Hoffmann, Y. Liu, M. Rosenbaum, and G. Szymanski. Statistical inference for rough volatility: Minimax theory. *Available at SSRN 4236905*, 2022b. pages 89, 113, 144, 213, 234, 341, 367

P. Christoffersen. Evaluating interval forecasts. *International economic review*, pages 841–862, 1998. pages 105, 113, 180, 214, 358, 367

P. Cirillo. Are your data really pareto distributed? *Physica A: Statistical Mechanics and its Applications*, 392(23):5947–5962, 2013. pages 118, 373

J.-F. Coeurjolly. Simulation and identification of the fractional brownian motion: a bibliographical and comparative study. *Journal of statistical software*, 5:1–53, 2000a. pages 204, 211

J.-F. Coeurjolly. Simulation and Identification of the Fractional Brownian Motion: A Bibliographical and Comparative Study. *Journal of Statistical Software*, 5(7), 2000b. ISSN 1548-7660. doi: 10.18637/jss.v005.i07. URL http://www.jstatsoft.org/v05/i07/. pages 86, 337

J.-F. Coeurjolly. Estimating the parameters of a fractional brownian motion by discrete variations of its sample paths. *Statistical Inference for stochastic processes*, 4:199–227, 2001. pages 86, 144, 211, 212, 337

J.-F. Coeurjolly. L-type estimators of the fractal dimension of locally self-similar gaussian processes. *Preprint. Available at hal. ccsd. cnrs. fr/docs/00/03/13/77/PDF/robustHurstHAL. pdf*, 2005. pages 87, 338

J.-F. Coeurjolly and A. Philippe. Basic properties of the multivariate fractional brownian motion. In *Workshop on Random Fields*. Citeseer, 2010. page 144

J.-F. Coeurjolly, P.-O. Amblard, and S. Achard. On multivariate fractional brownian motion and multivariate fractional gaussian noise. In *2010 18th European Signal Processing Conference*, pages 1567–1571. IEEE, 2010a. pages 87, 338

J.-F. Coeurjolly, P.-O. Amblard, and S. Achard. Normalized causal and well-balanced multivariate fractional brownian motion. *arXiv preprint arXiv:1007.2109*, 33, 2010b. pages 87, 339

J.-F. Coeurjolly, P.-O. Amblard, and S. Achard. Wavelet analysis of the multivariate fractional brownian motion. *ESAIM: Probability and Statistics*, 17:592–604, 2013. pages 86, 213, 338

F. Comte and E. Renault. Long memory in continuous-time stochastic volatility models. *Mathematical finance*, 8(4):291–323, 1998. page 213

R. Cont. Long range dependence in financial markets. In *Fractals in engineering: New trends in theory and applications*, pages 159–179. Springer, 2005. page 234

R. Cont. Volatility clustering in financial markets: empirical facts and agent-based models. In *Long memory in economics*, pages 289–309. Springer, 2007. pages 88, 340

L. Coutin. An introduction to (stochastic) calculus with respect to fractional brownian motion. In *Séminaire de Probabilités XL*, pages 3–65. Springer, 2007. pages 94, 346

M. E. Crovella and M. S. Taqqu. Estimating the heavy tail index from scaling properties. *Methodology and computing in applied probability*, 1:55–79, 1999. pages 118, 373

R. A. Crovelli and C. C. Barton. Fractals and the pareto distribution applied to petroleum accumulation-size distributions. In *Fractals in petroleum geology and earth processes*, pages 59–72. Springer, 1995. pages 118, 373

C. Czichowsky and W. Schachermayer. Portfolio optimisation beyond semimartingales: Shadow prices and fractional Brownian motion. *The Annals of Applied Probability*, 27 (3):1414 – 1451, 2017. doi: 10.1214/16-AAP1234. URL https://doi.org/10.1214/16-AAP1234. pages 90, 342

C. Czichowsky, R. Peyre, W. Schachermayer, and J. Yang. Shadow prices, fractional brownian motion, and portfolio optimisation under transaction costs. *Finance and Stochastics*, 22:161–180, 2018. pages 90, 342

J. Daníelsson, B. N. Jorgensen, G. Samorodnitsky, M. Sarma, and C. G. de Vries. Fat tails, var and subadditivity. *Journal of Econometrics*, 172(2):283–291, 2013. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2012.08.011. URL https://www.sciencedirect.com/science/article/pii/S0304407612001959. Latest Developments on Heavy-Tailed Distributions. pages 115, 369

S. Das and I. Pan. *Fractional order signal processing: introductory concepts and applications*. Springer Science & Business Media, 2011. pages 82, 94, 334, 347

M. H. A. Davis. Verification of internal risk measure estimates. *Statistics & Risk Modeling*, 33(3-4):67–93, 2016. doi: doi:10.1515/strm-2015-0007. URL https://doi.org/10.1515/strm-2015-0007. pages 76, 113, 214, 367

L. Decreusefond. Stochastic integration with respect to fractional brownian motion. *Theory and applications of long-range dependence*, pages 203–226, 2003. pages 94, 346

L. Decreusefond and A. S. Üstünel. Fractional brownian motion: Theory and applications. In *ESAIM: proceedings*, volume 5, pages 75–86. Citeseer, 1998. pages 94, 346

F. Delbaen, J. Eber, D. Heath, et al. Coherent measures of risk. *Math Financ*, 4:203–228, 1998. pages 67, 102, 180, 355

G. Didier and V. Pipiras. Gaussian stationary processes: adaptive wavelet decompositions, discrete approximations, and their convergence. *Journal of Fourier Analysis and Applications*, 14:203–234, 2008. pages 84, 85, 335, 336

G. Didier and V. Pipiras. Integral representations and properties of operator fractional Brownian motions. *Bernoulli*, 17(1):1 – 33, 2011. doi: 10.3150/10-BEJ259. URL https://doi.org/10.3150/10-BEJ259. pages 84, 335

G. Didier and V. Pipiras. Exponents, symmetry groups and classification of operator fractional brownian motions. *Journal of Theoretical Probability*, 25(2):353–395, 2012. pages 84, 335

D. E. Dominici*. The Inverse of the Cumulative Standard Normal Probability Function. *Integral Transforms and Special Functions*, 14(4):281–292, Aug. 2003. ISSN 1065-2469, 1476-8291. doi: 10.1080/1065246031000081698. URL http://www.tandfonline.com/doi/abs/10.1080/1065246031000081698. page 201

K. Dowd. *Beyond value at risk: The new science of risk management.* 1998. page 67

D. Duffie and J. Pan. An overview of value at risk. *Journal of derivatives*, 4(3):7–49, 1997. page 67

T. E. Duncan, Y. Hu, and B. Pasik-Duncan. Stochastic calculus for fractional brownian motion i. theory. *SIAM Journal on Control and Optimization*, 38(2):582–612, 2000. doi: 10.1137/S036301299834171X. URL https://doi.org/10.1137/S036301299834171X. pages 93, 346

R. Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019. page 287

P. Embrechts. Extreme value theory: Potential and limitations as an integrated risk management tool. *Derivatives Use, Trading & Regulation*, 6(1):449–456, 2000. pages 115, 369

P. Embrechts. *Selfsimilar processes.* Princeton University Press, 2009. pages 82, 94, 334, 347

P. Embrechts and M. Maejima. An introduction to the theory of self-similar stochastic processes. *International journal of modern physics B*, 14(12n13):1399–1420, 2000. pages 82, 94, 334, 347

P. Embrechts, T. Mao, Q. Wang, and R. Wang. Bayes risk, elicitability, and the expected shortfall. *Mathematical Finance*, 31(4):1190–1217, 2021. doi: https://doi.org/10.1111/mafi.12313. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/mafi.12313. page 76

C. J. Evertsz. Self-similarity of high-frequency usd-dem exchange rates. In *Proceedings of the First International Conference on High Frequency Data in Finance*, volume 3. Citeseer, 1995. pages 90, 342

F. J. Fabozzi, F. Gupta, and H. M. Markowitz. The legacy of modern portfolio theory. *The journal of investing*, 11(3):7–22, 2002. page 61

G. Fan, Y. Zeng, and W. K. Wong. Decomposition of portfolio var and expected shortfall based on multivariate copula simulation. *International Journal of Management Science and Engineering Management*, 7(2):153–160, 2012. pages 115, 370

J. Y. Fan, K. Hamza, and F. Klebaner. Mimicking self-similar processes. 2015. pages 83, 334

Y. Fan, G. Li, and M. Zhang. Expected shortfall and it's application in credit risk measurement. In *2010 Third International Conference on Business Intelligence and Financial Engineering*, pages 359–363, 2010. doi: 10.1109/BIFE.2010.90. pages 116, 370

M. Fernández-Martínez, M. Sánchez-Granero, and J. Trinidad Segovia. Measuring the self-similarity exponent in lévy stable processes of financial time series. *Physica A: Statistical Mechanics and its Applications*, 392(21):5330–5345, 2013. ISSN 0378-4371. doi: https://doi.org/10.1016/j.physa.2013.06.026. URL https://www.sciencedirect.com/science/article/pii/S0378437113005244. pages 83, 334

A. Feuerverger and A. Wong. Computation of value-at-risk for nonlinear portfolios. *The Journal of Risk*, 3(1):37–55, Oct. 2000. ISSN 14651211. doi: 10.21314/JOR.2000.035. URL http://www.risk.net/journal-of-risk/technical-paper/2161085/computation-value-risk-nonlinear-portfolios. pages 102, 180, 355

M. Finkelstein, H. G. Tucker, and J. Alan Veeh. Pareto tail index estimation revisited. *North American actuarial journal*, 10(1):1–10, 2006. pages 119, 374

T. Fissler and J. F. Ziegel. Higher order elicitability and Osband's principle. *The Annals of Statistics*, 44(4):1680 – 1707, 2016. doi: 10.1214/16-AOS1439. URL https://doi.org/10.1214/16-AOS1439. page 76

T. Fissler, J. F. Ziegel, and T. Gneiting. Expected shortfall is jointly elicitable with value at risk - implications for backtesting, 2015. pages 116, 370

T. Fissler, J. Hlavinová, and B. Rudloff. Elicitability and identifiability of set-valued measures of systemic risk. *Finance and Stochastics*, 25:133–165, 2021. page 76

J. C. Francis and D. Kim. *Modern portfolio theory: Foundations, analysis, and new developments*. John Wiley & Sons, 2013. page 61

N. Frikha and L. Huang. A multi-step richardson–romberg extrapolation method for stochastic approximation. *Stochastic Processes and their Applications*, 125(11):4066–4101, 2015. page 143

A. Gabrielsen, A. Kirchner, Z. Liu, and P. Zagaglia. Forecasting value-at-risk with time-varying variance, skewness and kurtosis in an exponential weighted moving average framework. *Annals of Financial Economics*, 10(01):1550005, 2015. pages 103, 356

A. A. Gaivoronski and G. Pflug. Value-at-risk in portfolio optimization: properties and computational approach. *Journal of risk*, 7(2):1–31, 2005. pages 102, 355

M. L. Galloway and C. A. Nolder. Subordination, self-similarity, and option pricing. *Journal of Applied Mathematics & Decision Sciences*, 2008(1), 2008. pages 90, 342

R. Garcia, E. Renault, and G. Tsafack. Proper conditioning for coherent var in portfolio management. *Management Science*, 53(3):483–494, 2007. doi: 10.1287/mnsc.1060.0632. URL https://doi.org/10.1287/mnsc.1060.0632. pages 115, 369

M. Garcin. Estimation of time-dependent hurst exponents with variational smoothing and application to forecasting foreign exchange rates. *Physica A: statistical mechanics and its applications*, 483:462–479, 2017. page 234

M. Garcin. Hurst exponents and delampertized fractional brownian motions. *International journal of theoretical and applied finance*, 22(05):1950024, 2019. pages 87, 339

M. Garcin. Fractal analysis of the multifractality of foreign exchange rates. *Mathematical methods in economics and finance*, 13(1):49–73, 2020. pages 109, 193, 234, 362

M. Garcin. Forecasting with fractional brownian motion: a financial perspective. *Quantitative finance*, 22(8):1495–1512, 2022. pages 90, 342

J. Gatheral, T. Jaisson, and M. Rosenbaum. Volatility is rough. *arXiv:1410.3394 [q-fin]*, Oct. 2014. URL http://arxiv.org/abs/1410.3394. arXiv: 1410.3394. pages 88, 340

J. Gatheral, M. Fukasawa, T. Jaisson, and M. Rosenbaum. Rough volatility: An overview. *Global Derivatives*, 2017. pages 142, 180

J. Gatheral, T. Jaisson, and M. Rosenbaum. Volatility is rough. In *Commodities*, pages 659–690. Chapman and Hall/CRC, 2022. page 213

G. Giambartolomei. *The Karhunen-Loève Theorem*. PhD thesis, Università di Bologna, 2015. URL http://amslaurea.unibo.it/10169/. page 160

E. Gobet, M. Lerasle, and D. Métivier. Mean estimation for randomized quasi monte carlo method. 2022. pages 139, 313, 396

Y. Goegebeur, J. Beirlant, and T. de Wet. Linking pareto-tail kernel goodness-offit statistics with tail index at optimal threshold and second order estimation. *REVSTAT-Statistical Journal*, 6(1):51–69, 2008. pages 120, 375

R. A. Groeneveld and G. Meeden. Measuring skewness and kurtosis. *Journal of the Royal Statistical Society Series D: The Statistician*, 33(4):391–399, 1984. page 61

P. Guasoni, Z. Nika, and M. Rásonyi. Trading fractional brownian motion. *SIAM Journal on Financial Mathematics*, 10(3):769–789, 2019. doi: 10.1137/17M113592X. URL https://doi.org/10.1137/17M113592X. pages 90, 343

P. Guasoni, Y. Mishura, and M. Rásonyi. High-frequency trading with fractional brownian motion. *Finance and stochastics*, 25(2):277–310, 2021. pages 90, 343

Z. Guo. Heavy-tailed distributions and risk management of equity market tail events. *Journal of Risk & Control*, 4(1):31–41, 2017. pages 117, 372

F. Gupta, R. Prajogi, and E. Stubbs. The information ratio and performance. *Journal of Portfolio Management*, 26(1):33, 1999. page 66

L. Haan and A. Ferreira. *Extreme value theory: an introduction*, volume 3. Springer, 2006. page 245

M. Haas and C. Pigorsch. Financial economics, fat-tailed distributions. *Encyclopedia of Complexity and Systems Science*, 4(1):3404–3435, 2009. pages 117, 371

D. Hainaut. Fractional hawkes processes. *Physica A: Statistical Mechanics and its Applications*, 549:124330, 2020. ISSN 0378-4371. doi: https://doi.org/10.1016/j.physa.2020.124330. URL https://www.sciencedirect.com/science/article/pii/S0378437120301096. page 143

X. D. He, S. Kou, and X. Peng. Risk measures: Robustness, elicitability, and backtesting. *Annual Review of Statistics and Its Application*, 9(Volume 9, 2022):141–166, 2022. ISSN 2326-831X. doi: https://doi.org/10.1146/annurev-statistics-030718-105122. URL https://www.annualreviews.org/content/journals/10.1146/annurev-statistics-030718-105122. page 76

D. Hendricks. Evaluation of value-at-risk models using historical data. *Economic policy review*, 2(1), 1996. pages 102, 356

Y. HU, B. ØKSENDAL, and A. SULEM. Optimal consumption and portfolio in a black–scholes market driven by fractional brownian motion. *Infinite Dimensional Analysis, Quantum Probability and Related Topics*, 06(04):519–536, 2003. doi: 10.1142/S0219025703001432. URL https://doi.org/10.1142/S0219025703001432. pages 90, 342

W. Huang, C.-A. Lehalle, and M. Rosenbaum. Simulating and analyzing order book data: The queue-reactive model. *Journal of the American Statistical Association*, 110(509):107–122, 2015. page 143

G. Hübner. The generalized treynor ratio. *Review of Finance*, 9(3):415–435, 2005. page 64

W. N. Hudson and J. D. Mason. Operator-self-similar processes in a finite-dimensional space. *Transactions of the American Mathematical Society*, 273(1):281–297, 1982. pages 88, 339

K. Inui and M. Kijima. On the significance of expected shortfall as a coherent risk measure. *Journal of Banking & Finance*, 29(4):853–864, 2005. ISSN 0378-4266. doi: https://doi.org/10.1016/j.jbankfin.2004.08.005. URL https://www.sciencedirect.com/science/article/pii/S037842660400144X. Risk Measurement. pages 115, 369

J. Istas and G. Lang. Variations quadratiques et estimation de l'exposant de Hölder local d'un processus Gaussien. 1994. pages 112, 211, 367

D. Jadhav, T. V. Ramanathan, and U. V. Naik-Nimbalkar. Modified estimators of the expected shortfall. *Journal of Emerging Market Finance*, 8(2):87–107, 2009. pages 116, 371

S. Janson. *Gaussian Hilbert Spaces*. Number 129 in Cambridge Tracts in Mathematics. Cambridge university press, 1997. ISBN 9780511526169. doi: 10.1017/CBO9780511526169. pages 105, 106, 182, 189, 359

K. L. Jean-franÇois Coeurjolly and B. Vidakovic. Variance estimation for fractional brownian motions with fixed hurst parameters. *Communications in Statistics - Theory and Methods*, 43(8):1845–1858, 2014. doi: 10.1080/03610926.2012.677087. URL https://doi.org/10.1080/03610926.2012.677087. pages 86, 213, 338

P. Jorion. Risk2: Measuring the risk in value at risk. *Financial analysts journal*, 52(6): 47–56, 1996. pages 67, 102, 356

P. Jorion. *Value at risk: the new benchmark for managing financial risk*. McGraw-Hill, 2007. page 67

G. Jumarie. Merton's model of optimal portfolio in a black-scholes market driven by a fractional brownian motion with short-range dependence. *Insurance: Mathematics and Economics*, 37(3):585–598, 2005. ISSN 0167-6687. doi: https://doi.org/10.1016/j.insmatheco.2005.06.003. URL https://www.sciencedirect.com/science/article/pii/S0167668705000703. pages 90, 342

R. Kellner and D. Rösch. Quantifying market risk with value-at-risk or expected shortfall? – consequences for capital requirements and model risk. *Journal of Economic Dynamics and Control*, 68:45–63, 2016. ISSN 0165-1889. doi: https://doi.org/10.1016/j.jedc.2016.05.002. URL https://www.sciencedirect.com/science/article/pii/S0165188916300641. pages 115, 369

J. T. Kent and A. T. Wood. Estimating the fractal dimension of a locally self-similar Gaussian process by using increments. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 679–699, 1997. pages 112, 211, 367

A. Kuritzkes, T. Schuermann, and S. M. Weiner. Risk measurement, risk management, and capital adequacy in financial conglomerates. *Brookings-Wharton Papers on Financial Services*, 2003(1):141–193, 2003. page 40

R. Laha and V. Rohatgi. Operator self similar stochastic processes in rd. *Stochastic Processes and their Applications*, 12(1):73–84, 1981. ISSN 0304-4149. doi: https://doi.org/10.1016/0304-4149(81)90012-0. URL https://www.sciencedirect.com/science/article/pii/0304414981900120. pages 88, 142, 175, 176, 177, 339

J. Lamperti. Semi-stable stochastic processes. *Transactions of the American mathematical Society*, 104(1):62–78, 1962. pages 82, 94, 176, 334, 347

F. Lavancier, A. Philippe, and D. Surgailis. Covariance function of vector self-similar processes. *Statistics & probability letters*, 79(23):2415–2421, 2009. pages 83, 84, 142, 173, 174, 177, 188, 335

J. C. Lee and P. Valiant. Optimal sub-gaussian mean estimation in **R**. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 672–683. IEEE, 2022. pages 139, 313, 396

F. Lillo and J. D. Farmer. The long memory of the efficient market. *Studies in nonlinear dynamics & econometrics*, 8(3), 2004. page 234

S. C. Lim and S. V. Muniandy. Self-similar gaussian processes for modeling anomalous diffusion. *Physical Review E*, 66(2):021114, 2002. pages 91, 343

S. Lin. Stochastic analysis of fractional brownian motions. *Stochastics: An International Journal of Probability and Stochastic Processes*, 55(1-2):121–140, 1995. pages 94, 346

T. J. Linsmeier and N. D. Pearson. Value at risk. *Financial analysts journal*, 56(2):47–67, 2000. page 67

A. Lucas and X. Zhang. Score-driven exponentially weighted moving averages and value-at-risk forecasting. *International Journal of Forecasting*, 32(2):293–302, 2016. pages 102, 356

G. Lugosi and S. Mendelson. Mean Estimation and Regression Under Heavy-Tailed Distributions: A Survey. *Foundations of Computational Mathematics*, 19(5):1145–1190, Oct. 2019. ISSN 1615-3375, 1615-3383. doi: 10.1007/s10208-019-09427-x. URL http://link.springer.com/10.1007/s10208-019-09427-x. pages 132, 137, 264, 271, 284, 300, 388, 393

W. D. Lunga. *Time series analysis using fractal theory and online ensemble classifiers with application to stock portfolio optimization*. PhD thesis, University of the Witwatersrand, 2006. pages 90, 342

R. MacAusland. The moore-penrose inverse and least squares. Math 420: Advanced Topics in Linear Algebra, 2014. URL http://buzzard.ups.edu/courses/2014spring/420projects/math420-UPS-spring-2014-macausland-pseudo-inverse.pdf. pages 108, 190, 191, 362

M. Maejima and J. Mason. Operator-self-similar stable processes. *Stochastic Processes and their Applications*, 54(1):139–163, 1994. ISSN 0304-4149. doi: https://doi.org/10.1016/0304-4149(94)00010-7. URL https://www.sciencedirect.com/science/article/pii/0304414994000107. pages 88, 204, 339

M. Maejima and C. A. Tudor. Selfsimilar processes with stationary increments in the second wiener chaos. *Probab. Math. Statist*, 32(1):167–186, 2012. pages 83, 334

B. B. Mandelbrot and J. W. Van Ness. Fractional brownian motions, fractional noises and applications. *SIAM review*, 10(4):422–437, 1968a. doi: 10.1137/1010093. pages 82, 94, 98, 99, 142, 333, 347

B. B. Mandelbrot and J. W. Van Ness. Fractional Brownian motions, fractional noises and applications. *SIAM review*, 10(4):422–437, 1968b. Publisher: SIAM. pages 87, 106, 149, 160, 182, 339, 351, 359

S. Manganelli and R. F. Engle. Value at risk models in finance. 2001. pages 102, 355

D. Marinucci and P. Robinson. Weak convergence of multivariate fractional processes. *Stochastic Processes and their Applications*, 86(1):103–120, 2000. ISSN 0304-4149. doi: https://doi.org/10.1016/S0304-4149(99)00088-5. URL https://www.sciencedirect.com/science/article/pii/S0304414999000885. pages 88, 340

A. J. McNeil, R. Frey, and P. Embrechts. *Quantitative risk management: concepts, techniques and tools-revised edition*. Princeton university press, 2015. pages 102, 180, 355

Z. Michna et al. Self-similar processes in collective risk theory. *International Journal of Stochastic Analysis*, 11:429–448, 1998. pages 91, 343

S. Muniandy and S. C. Lim. Modeling of locally self-similar processes using multifractional brownian motion of riemann-liouville type. *Physical Review E*, 63(4):046104, 2001. pages 92, 142, 344

M. E. Newman. Power laws, pareto distributions and zipf's law. *Contemporary physics*, 46(5):323–351, 2005. pages 118, 372

J. P. Nolan. Financial modeling with heavy-tailed stable distributions. *Wiley Interdisciplinary Reviews: Computational Statistics*, 6(1):45–55, 2014. pages 117, 372

D. Nualart. Stochastic calculus with respect to fractional brownian motion. In *Annales de la Faculté des sciences de Toulouse: Mathématiques*, volume 15, pages 63–78, 2006. pages 94, 346

E. Ocran, R. Minkah, G. Kallah-Dagadu, K. Doku-Amponsah, et al. Estimation of the tail index of pareto-type distributions using regularisation. *Journal of Mathematics*, 2022, 2022. pages 119, 374

G. Pagnini, A. Mura, and F. Mainardi. Generalized fractional master equation for self-similar stochastic processes modelling anomalous diffusion. *International Journal of Stochastic Analysis*, 2012, 2012. pages 92, 345

Z. Pan, T. Pang, and Y. Zhao. A simple and robust approach for expected shortfall estimation. *Journal of Computational Finance*, 25(1), 2019. pages 116, 371

G. Pang and M. S. Taqqu. Nonstationary self-similar gaussian processes as scaling limits of power-law shot noise processes and generalizations of fractional brownian motion. *High Frequency*, 2(2):95–112, 2019. pages 92, 345

J. Papenbrock. *Asset clusters and asset networks in financial risk management and portfolio optimization*. PhD thesis, Dissertation, Karlsruhe, Karlsruher Institut für Technologie (KIT), 2011, 2011. pages 90, 342

J. Pardo. A brief introduction to self-similar processes. *Department of Mathematical Sciences, University of Bath, United Kingdom*, 9:305–316, 2007. pages 82, 334

L. Peng and Y. Qi. *Inference for heavy-tailed data: applications in insurance and finance*. Academic press, 2017. pages 117, 372

P. Penza and V. K. Bansal. *Measuring market risk with value at risk*, volume 17. John Wiley & Sons, 2001. page 67

J. Picard. Representation Formulae for the Fractional Brownian Motion. In C. Donati-Martin, A. Lejay, and A. Rouault, editors, *Séminaire de Probabilités XLIII*, volume 2006, pages 3–70. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011. ISBN 978-3-642-15216-0 978-3-642-15217-7. doi: 10.1007/978-3-642-15217-7_1. URL http://link.springer.com/10.1007/978-3-642-15217-7_1. Series Title: Lecture Notes in Mathematics. page 184

V. Pipiras and M. S. Taqqu. Integration questions related to fractional brownian motion. *Probability theory and related fields*, 118:251–291, 2000. pages 94, 347

R.-D. Reiss. *Approximate distributions of order statistics: with applications to nonparametric statistics*. Springer science & business media, 2012. pages 271, 287, 329

J. Resin. From classification accuracy to proper scoring rules: Elicitability of probabilistic top list predictions. *Journal of Machine Learning Research*, 24(173):1–21, 2023. page 76

E. Rio. About the constants in the fuk-nagaev inequalities. 2017. page 263

R. T. Rockafellar, S. Uryasev, et al. Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42, 2000. page 196

T. N. Rollinger and S. T. Hoffman. Sortino: a 'sharper'ratio. *Chicago, Illinois: Red Rock Capital*, 2013. page 63

S. Rostek. *Option pricing in fractional Brownian markets*. Number 622 in Lecture notes in economics and mathematical systems. Springer, Berlin Heidelberg, 1. aufl edition, 2009. ISBN 978-3-642-00330-1. pages 89, 341

S. Rostek. Explaining the volatility surface: a closed-form solution to option pricing in a fractional jump-diffusion market. In *Paris December 2012 Finance Meeting EUROFIDAI-AFFI Paper*, 2012. pages 89, 340

S. Rostek and S. Rostek. Risk preference based option pricing in a continuous time fractional brownian market. *Option Pricing in Fractional Brownian Markets*, pages 79–110, 2009. pages 89, 341

S. Rostek and R. Schöbel. Risk preference based option pricing in a fractional brownian market. Technical report, Tübinger Diskussionsbeiträge, 2006. pages 89, 341

S. Rostek and R. Schoebel. Equilibrium pricing of options in a fractional brownian market. *Available at SSRN 1701596*, 2010. pages 89, 341

S. Rostek and R. Schöbel. A note on the use of fractional brownian motion for financial modeling. *Economic Modelling*, 30:30–35, 2013. ISSN 0264-9993. doi: https://doi.org/10.1016/j.econmod.2012.09.003. URL https://www.sciencedirect.com/science/article/pii/S0264999312002787. page 143

M. Rytgaard. Estimation in the pareto distribution. *ASTIN Bulletin: The Journal of the IAA*, 20(2):201–216, 1990. pages 118, 373

G. Samorodnitsky. Long memory and self-similar processes. In *Annales de la Faculté des sciences de Toulouse: Mathématiques*, volume 15, pages 107–123, 2006. pages 82, 94, 334, 347

G. Samorodnitsky and M. S. Taqqu. Levy Measures of Infinitely Divisible Random Vectors and Slepian Inequalities. *The Annals of Probability*, 22(4):1930 – 1956, 1994. doi: 10.1214/aop/1176988490. URL https://doi.org/10.1214/aop/1176988490. pages 84, 335

G. Samorodnitsky, M. S. Taqqu, and R. Linde. Stable non-gaussian random processes: stochastic models with infinite variance. *Bulletin of the London Mathematical Society*, 28(134):554–555, 1996. pages 83, 334

Y. Sarol, F. G. Viens, and T. Zhang. Portfolio optimization with consumption in a fractional black-scholes market. *Communications on Stochastic Analysis*, 1(3):2, 2007. pages 90, 342

S. Sarykalin, G. Serraino, and S. Uryasev. Value-at-risk vs. conditional value-at-risk in risk management and optimization. In *State-of-the-art decision-making tools in the information-intensive age*, pages 270–294. Informs, 2008. pages 73, 120, 243, 375

K.-i. Sato. Self-similar processes with independent increments. *Probability Theory and Related Fields*, 89:285–300, 1991. pages 83, 88, 334, 339

O. Scaillet. Nonparametric estimation and sensitivity analysis of expected shortfall. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 14(1):115–129, 2004. pages 116, 371

W. F. Sharpe. The capital asset pricing model: a "multi-beta" interpretation. In *Financial Dec Making Under Uncertainty*, pages 127–135. Elsevier, 1977. page 61

W. F. Sharpe. The sharpe ratio. *Streetwise–the Best of the Journal of Portfolio Management*, 3:169–185, 1998. page 63

O. Sheluhin, S. Smolskiy, and A. Osin. *Self-similar processes in telecommunications*. John Wiley & Sons, 2007. pages 83, 334

F. Shokrollahi, A. Kılıçman, et al. Pricing currency option in a mixed fractional brownian motion with jumps environment. *Mathematical Problems in Engineering*, 2014, 2014. page 180

K. Sigman. A primer on heavy-tailed distributions. *Queueing systems*, 33(1-3):261, 1999. page 246

R. D. Smith. Is high-frequency trading inducing changes in market microstructure and dynamics? *arXiv preprint arXiv:1006.5490*, 2010. pages 90, 343

S. A. Stoev and M. S. Taqqu. How rich is the class of multifractional brownian motions? *Stochastic Processes and their Applications*, 116(2):200–221, 2006. ISSN 0304-4149. doi: https://doi.org/10.1016/j.spa.2005.09.007. URL https://www.sciencedirect.com/science/article/pii/S0304414905001377. pages 88, 340

J. Su and E. Furman. A form of multivariate pareto distribution with applications to financial risk measurement. *ASTIN Bulletin: The Journal of the IAA*, 47(1):331–357, 2017. pages 118, 373

E. L. Sukono, B. Subartini, Y. Hidayat, and A. T. Bon. Value-at-risk contribution under asset liability models by using expenential weighted moving average approaches. In *International Conference on Industrial Engineering and Operations Management*, pages 2920–2928, 2018. pages 103, 356

D. Surgailis, G. Teyssière, and M. Vaičiulis. The increment ratio statistic. *Journal of multivariate analysis*, 99(3):510–541, 2008. page 234

G. Szymanski and T. Takabatake. Asymptotic efficiency for fractional Brownian motion with general noise. *arXiv preprint arXiv:2311.18669*, 2023. pages 89, 113, 144, 213, 341, 367

D. Talay and L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic analysis and applications*, 8(4):483–509, 1990a. page 143

D. Talay and L. Tubaro. Romberg extrapolations for numerical schemes solving stochastic differential equations. *Structural Safety*, 8(1-4):143–150, 1990b. page 143

T.-L. Tang and S.-J. Shieh. Long memory in stock index futures markets: A value-at-risk approach. *Physica A: Statistical Mechanics and its Applications*, 366:437–448, 2006. pages 104, 357

M. S. Taqqu. A representation for self-similar processes. *Stochastic Processes and their Applications*, 7(1):55–64, 1978. ISSN 0304-4149. doi: https://doi.org/10.1016/0304-4149(78)90037-6. URL https://www.sciencedirect.com/science/article/pii/0304414978900376. pages 82, 333

M. S. Taqqu. *Stable non-Gaussian random processes: stochastic models with infinite variance*. Chapman & Hall, 1994. pages 94, 163, 174, 210, 347

D. Tasche. Expected shortfall and beyond. *Journal of Banking & Finance*, 26(7):1519–1533, 2002a. ISSN 0378-4266. doi: https://doi.org/10.1016/S0378-4266(02)00272-8. URL https://www.sciencedirect.com/science/article/pii/S0378426602002728. pages 115, 369

D. Tasche. Expected shortfall and beyond. *Journal of Banking & Finance*, 26(7):1519–1533, 2002b. pages 72, 73, 120, 243, 375

J. W. Taylor. Estimating value at risk and expected shortfall using expectiles. *Journal of Financial Econometrics*, 6(2):231–252, 2008. pages 116, 371

L. Tibiletti. Value-at-risk: Is lacking in sub-additivity just an annoying technicality? *International Journal of Risk Assessment and Management - Int J Risk Assess Manag*, 9, 07 2008. doi: 10.1504/IJRAM.2008.019312. pages 115, 369

C. Tudor. Analysis of variations for self-similar processes: a stochastic calculus approach. 2013. pages 92, 345

B. Vandewalle, J. Beirlant, A. Christmann, and M. Hubert. A robust estimator for the tail index of pareto-type distributions. *Computational Statistics & Data Analysis*, 51 (12):6252–6268, 2007. pages 119, 374

L. Viitasaari. Representation of stationary and stationary increment processes via langevin equation and self-similar processes. *Statistics & probability letters*, 115:45–53, 2016. pages 82, 333

N. Wesselhöfft. Utilizing self-similar stochastic processes to model rare events in finance. 2021. pages 91, 344

E. Wipplinger. Philippe jorion: value at risk-the new benchmark for managing financial risk. *Financial Markets and Portfolio Management*, 21(3):397, 2007. pages 102, 180, 355

A. Xu, Y. Hu, and P. Tang. Asymptotic expansions for the gamma function. *Journal of Number Theory*, 169:134–143, 2016. page 275

Y. Yamai and T. Yoshiba. Value-at-risk versus expected shortfall: A practical perspective. *Journal of Banking & Finance*, 29(4):997–1015, 2005. ISSN 0378-4266. doi: https://doi.org/10.1016/j.jbankfin.2004.08.010. URL https://www.sciencedirect.com/science/article/pii/S0378426604001499. Risk Measurement. pages 115, 370

M. Youssef, L. Belkacem, and K. Mokni. Value-at-risk estimation of energy commodities: A long-memory garch–evt approach. *Energy Economics*, 51:99–110, 2015. pages 104, 357

J. F. Ziegel. Coherence and elicitability. *Mathematical Finance*, 26(4):901–918, 2016. doi: https://doi.org/10.1111/mafi.12080. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/mafi.12080. page 76

N. M. ZINCHENKO. Heavy-tailed models in finance and insurance: a survey. *International School on Mathematical and Statistical Applications in Economics*, page 346, 2001. pages 117, 371

A. Zlatniczki and A. Telcs. Application of portfolio optimization to achieve persistent time series. *Journal of Optimization Theory and Applications*, pages 1–23, 2024. pages 90, 342

ECOLE
DOCTORALE
DE MATHEMATIQUES
HADAMARD

Titre : Mesures de risque en finance, Rétrotest, Sensibilité et Robustesse

Mots clés : Auto-similarité, Stationnarité, Processus Gaussiens, Mouvement Brownien fractionnaire univarié (fBm) et multivarié (mfBm), Mesures de risque, Distributions à queues lourdes

Résumé : Dans le chapitre 2, nous étudions deux transformations temporelles des processus aléatoires : le changement d'origine temporelle, et le changement d'échelle de temps ainsi que les propriétés associées, la stationnarité et l'auto-similarité. Nous prouvons d'abord les propriétés de stationnarité et d'auto-similarité des processus dans le cadre très général des espaces de Hilbert puis dans le cadre plus spécifique de l'espace de Hilbert Gaussien où les propriétés sont prouvées en distribution (au sens faible) et en un sens trajectoriel (au sens strict). Des exemples de tels processus comme le mouvement Brownien et mouvement Brownien fractionnaire (fBm) sont fournis, dans les cadres univariés et multivariés (mfBm). Dans le Chapitre 3, nous décrivons les trajectoires de prix en utilisant des mouvements Browniens géométriques fractionnaires. Cela permet d'ajouter des corrélations entre les rendements logarithmiques pour exprimer la dépendance à long-terme. Les rendements logarithmiques sont alors décrits par des processus Gaussiens auto-similaires à accroissements stationnaires et corrélés, les fBm et mfBm. Dans ce cadre, les mesures de risque basées sur la distribution des pertes sont alors prédites avec précision en tenant compte de la dépendance à long-terme. Nous considérons la mesure de risque la plus couramment utilisée par les régulateurs, la Valeur-à-Risque (VaR). Nous introduisons un modèle qui fournit une approximation Gaussienne de la VaR conditionnelle pour un portefeuille d'actifs sous dynamiques fractionnaires (mfBm). Nous fournissons une quantification de l'erreur d'approximation et nous effectuons un rétrotest sur des données simulées et de marché. Dans le Chapitre 4, nous proposons de modéliser la distribution des pertes avec une distribution à queue lourde qui prend mieux en compte les événements extrêmes, appelée la distribution de Pareto. La distribution de Pareto présente des propriétés intéressantes de changement d'échelle et de stabilité par conditionnement. Nous remplaçons la VaR par l'Expected-Shortfall (ES), plus sensible au risque de queue. Nous proposons des méthodes robustes non-asymptotiques pour estimer l'ES dans des distributions à queues lourdes telles que la Médiane-des-Moyennes (MoM), les Moyennes-Élaguées (TM), et l'estimateur de Lee-Valiant (LV) que nous comparons à l'estimateur de moyenne empirique (asymptotique). Nous étudions leur biais et leur taux de convergence.

Title : Risk measures in finance, Backtesting, Sensitivity and Robustness

Keywords : Self-similarity, Stationarity, Gaussian processes, univariate and multivariate fractional Brownian motions (fBm and mfBm), risk-measures, heavy-tailed distributions

Abstract : In Chapter 2, we focus on two time transformations of the random processes : the time-origin change and the time-scaling and on the related properties called the stationarity and the self-similarity. We prove the stationarity and self-similarity properties of the processes first in a the very general framework of the Hilbert spaces ; then in a the more specific framework of the the Gaussian Hilbert space where the properties are proved in distribution (weak sense) and in a trajectory sense (strict sense). We also provide examples of such processes called standard Brownian motion and fractional Brownian motion (fBm), in the univariate and multivariate frameworks (mfBm). In Chapter 3, we propose to describe price trajectories using fractional geometric Brownian motions. This allows adding correlations between logarithmic returns to express long-range dependency. Logarithmic returns are then described using self-similar Gaussian processes with stationary and correlated increments, the fBm and mfBm. In this framework, risk measures that are based on the loss distribution, can be accurately predicted taking into account the long-range dependency. We focus on predicting the most commonly used risk measure by regulators, called Value-at-Risk (VaR). We introduce a model that provides a Gaussian approximation of Value-at-Risk (VaR) for the assets portfolio under fractional dynamics (mfBm). We provide a quantification of the error of approximation and we carry out backtesting experiments on simulated and market data. In Chapter 4, we propose to model the loss distribution with a heavy-tailed distribution that better takes into account the extreme events, called the Pareto distribution that presents interesting properties of scaling and stability by conditioning and to replace VaR by Expected-Shortfall (ES) which is more sensitive to the tail risk. The objective is to explore non-asymptotic robust methods for estimating ES in heavy-tailed distributions such that the Median-of-Means (MoM), the Trimmed-Means (TM), and the Lee-Valiant (LV) estimators that we compare to the empirical mean estimator (asymptotic). We study their bias and their convergence rate.

Institut Polytechnique de Paris
91120 Palaiseau, France