



**HAL**  
open science

# Modélisation statistique de l'impact du risque climatique sur la solvabilité des banques

Guillaume Flament

► **To cite this version:**

Guillaume Flament. Modélisation statistique de l'impact du risque climatique sur la solvabilité des banques. Mathématiques générales [math.GM]. Ecole Nationale de la Statistique et de l'Analyse de l'Information [Bruz], 2024. Français. ⟨NNT : 2024NSAIM002⟩. ⟨tel-04963935⟩

**HAL Id: tel-04963935**

**<https://theses.hal.science/tel-04963935v1>**

Submitted on 24 Feb 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization



## Remerciements

En premier lieu il me faut remercier mon directeur de thèse, Valentin Patilea, sans qui rien n'aurait pu aboutir. Ta dévotion, ton engagement et ta rigueur ont été des piliers essentiels tout au long de ce parcours. Je te suis profondément reconnaissant pour tout ce que tu as fait pour moi et pour l'inspiration que tu m'as apportée. Je mesure toute la chance que j'ai eu d'avoir un directeur de thèse comme toi.

Je remercie Emmanuel Flachaire, et Christian Hafner d'avoir accepté d'être les rapporteurs de ma thèse. Je remercie aussi Anne Vanhems et Myriam Vimond qui me font l'honneur de participer à mon jury. Enfin, je remercie Christophe Hurlin et Nicolas Klutchnikov pour avoir fait parti du comité de suivi de ma thèse.

Bien-sûr cette thèse n'aurait pas pu aboutir sans le soutien constant des équipes de Square-Management, en particulier Cyril Regnier et David Alcaud, qui m'ont fait confiance depuis le début et ont su m'encadrer de bout en bout. David d'abord, pour ton encadrement et ta confiance, tu as su construire le SRC. C'est aujourd'hui un environnement particulièrement bienveillant pour les doctorants. La liberté que tu nous as laissée est un atout considérable, merci ! Cyril, pour avoir assuré la pertinence métier de ce que nous développons, ton expertise et tes conseils ont été précieux pour aligner nos travaux avec les besoins réels du marché. Votre soutien et votre engagement ont été déterminants pour la réussite de ce projet. Enfin, je voudrais remercier Marc et Franck qui ont su animer (entre bien d'autres choses) les expertises sur le risque climatique dans le cabinet. Je vous remercie tous pour les échanges passionnés et passionnants que nous avons pu avoir sur le sujet du risque climatique.

Aussi, je remercie les consultants Square avec qui j'ai été amené à travailler. Tout d'abord, Sarah ma référente qui m'a aidé et guidé dans les moments les moins évidents de cette aventure. Margaux pour l'engagement que tu as envers ce cabinet, pour ces fresques, ta connaissance sur l'ensemble des problématiques liées à la transition climatique ainsi que ta diplomatie hors pair, ce fut un plaisir de pouvoir échanger avec toi. Charles pour ces discussions toujours passionnantes sur le risque physique, sur les problématiques métiers, le niveau de tes réalisations est impressionnant. Enfin, celui sans qui cette thèse n'aurait peut-être pas abouti non plus, Quentin, qui a toujours su me remotiver, trouver les mots justes, et avec qui j'ai eu la chance de pouvoir collaborer. Je vais voler les mots de David : Tu es le (et pas un) chercheur modèle.

Aussi, la partie appliquée de ce manuscrit n'aurait pas pu être le même sans les partenaires bancaires avec qui j'ai eu la chance d'intégrer, pour une question d'anonymisation des données utilisées, je ne dévoilerai pas leurs noms, mais ils sauront, j'espère, tous se reconnaître.

Cette thèse n'aurait pas non plus été la même sans les doctorants de l'ENSAI que j'ai pu croiser, déjà Edouard, qui dès ma deuxième année à l'ENSAI m'a accueilli dans le bureau des doctorants pour me présenter ce qu'était la thèse, puis que j'ai recroisé lors

---

de mes premiers pas de jeune doctorant. Daphné, qui a du partager son bureau avec moi (n'oublie pas de passer le bonjour à Günther !), Sunny qui a du subir mon anglais, enfin, merci à Omar et Hassan pour l'aide sur les TD !

Je tiens également à exprimer ma profonde gratitude envers mes parents, qui ont toujours été d'un soutien indéfectible tout au long de mon parcours académique. Leur amour, leur patience et leurs encouragements constants ont été essentiel dans ce long parcours. Leur confiance en mes capacités et leur soutien inconditionnel ont été des sources inépuisables de motivation et de réconfort. Je leur suis infiniment reconnaissant pour tout ce qu'ils ont fait pour moi et pour m'avoir donné la force de poursuivre.

Enfin, je voudrais remercier Camille, j'admire ta force de travail, ta volonté, ton envie de bien faire, tu es absolument admirable et j'aimerais un peu plus te ressembler sur ces points, malgré la note en stat math, tu resteras ma majeure ! Ta patience a été essentielle durant ces années, merci pour ton amour inconditionnel. Sache qu'il est profondément réciproque.



# Contents

Résumé en français	5
English abstract	11
<b>1 General introduction and motivation</b>	<b>17</b>
1.1 Basel I : The first international regulation	17
1.1.1 Some historical context	17
1.1.2 The need for international regulations	18
1.1.3 Examples of modelisation of extreme risks in banking institution	21
1.2 Standardized approach in Basel II	23
1.2.1 Foundation Internal Rating Based Approach (FIRB)	26
1.2.2 Advanced Internal Rating Based Approach (AIRB)	27
1.2.3 The backbone of the IRB methodology : Merton-Vašíček model	28
1.3 Some implications of Basel accord	30
<b>2 A semiparametric location-scale model with application to credit risk</b>	<b>33</b>
2.1 Motivation for conditional Merton-Vašíček models	34
2.1.1 Overview	35
2.2 A general statistical model	37
2.2.1 Location and scale functions	38
2.2.2 Error distribution function estimators	40
2.2.3 Simulation of location-scale models with Gaussian response	42
2.3 Semi-parametric location-scale models	44
2.3.1 Homoscedastic semiparametric single-index regression	45
2.3.2 Nonparametric location-scale regression	53
2.3.3 Location-scale single-index modeling	55
2.3.4 Heteroscedastic partially linear single-index model	60
2.3.5 Semiparametric location-scale models with dependent data	65
2.4 Theoretical grounds	70
2.4.1 Technical lemmas	70
2.4.2 From nonlinear to semiparametric single-index regression	74
2.4.3 Asymptotics for semiparametric single-index regression	76

---

2.4.4	Asymptotics for semiparametric location-scale regression . . . . .	81
2.4.5	Asymptotics for semiparametric autoregression . . . . .	84
2.4.6	Error distribution estimation . . . . .	87
2.4.7	Quantile estimation . . . . .	91
2.5	Conclusion . . . . .	93
<b>3</b>	<b>Density estimation based on quantile regressions applied to credit risk</b>	<b>97</b>
3.1	Introduction . . . . .	98
3.2	Presentation of the methodology . . . . .	99
3.3	aR approach . . . . .	101
3.3.1	A general framework . . . . .	101
3.3.2	Quantile regressions . . . . .	103
3.4	Application on real data . . . . .	107
3.4.1	Application on S&P500 data . . . . .	109
3.4.2	Application on private banking data . . . . .	114
3.5	Conclusion . . . . .	118
<b>4</b>	<b>Macroeconomic scenarios for climate risk assessment</b>	<b>121</b>
4.1	From regulatory stress-testing to climate stress-testing . . . . .	122
4.1.1	Some definitions and regulatory constraints . . . . .	122
4.1.2	Differences between climate and traditional stress-tests . . . . .	124
4.1.3	Current practices for integrating climate change into stress-testing . . . . .	125
4.2	Introduction to climate integrated macroeconomic models . . . . .	127
4.3	Economic model assumptions . . . . .	129
4.3.1	The DICE model . . . . .	129
4.3.2	GDP dynamic . . . . .	131
4.3.3	TFP dynamic . . . . .	132
4.4	Modeling the impact of <i>exergy</i> on the TFP . . . . .	133
4.4.1	Data . . . . .	133
4.4.2	Bayesian auto-regressive model . . . . .	133
4.4.3	Bayesian model with an additional exogenous variable. . . . .	137
4.4.4	Forecast values during the Covid-19 crisis and the recovery . . . . .	138
4.5	Future <i>exergy</i> availability impact on real GDP . . . . .	140
4.5.1	Model for future <i>exergy</i> availability . . . . .	140
4.5.2	Exogenous trajectory . . . . .	141
4.6	The climate module . . . . .	142
4.7	Real GDP forecasts . . . . .	144
4.8	Conclusion . . . . .	146
	<b>General conclusion and perspectives</b>	<b>153</b>

# Résumé en français

Le risque climatique devient un enjeu majeur pour les institutions financières. D'un côté, la société civile réclame de ces institutions une certaine forme d'exemplarité en la matière, un manque d'engagement concret sur ce volet pourrait pousser des clients à quitter certaines banques ou certains assureurs. D'un autre côté, le régulateur estime dorénavant que le risque climatique est devenu un enjeu financier important et que la matérialisation de ce risque peut résulter en des pertes financières. En plus de ces pertes potentielles, les engagements climatiques sont aujourd'hui concrets avec un engagement politique fort qui s'est matérialisé lors des accords de Paris. Il serait dangereux que les institutions financières entravent ces objectifs climatiques. Elles doivent au contraire jouer pleinement leur rôle de financement de l'économie. Ainsi, le régulateur veille à ce que ces institutions anticipent et participent à ces changements économiques et sociaux.

Les banques centrales, et en particulier celle de la zone euro, ont pour mandat de garantir la stabilité monétaire et veiller à la résilience du système financier. Pour ce faire, depuis les années 1980 (Tarullo, 2008) des exercices dits de "stress-tests" sont mis en place afin de vérifier l'adéquation des fonds propres des banques et des assureurs face au risque financier de leurs portefeuilles. En 2015, Mark Carney, alors directeur de la banque centrale d'Angleterre, déclare que le changement climatique représente déjà, un enjeu majeur pour la stabilité financière. Il n'a donc pas fallu attendre longtemps pour que plusieurs banques centrales, dont la banque centrale européenne, lancent un exercice dit de "stress-test climatique", afin de vérifier l'adéquation des fonds propres des institutions européennes face au risque climatique. Ces exercices très récents comparativement aux exercices de stress-tests classiques, ont enclenchés un effort de recherche pour d'un côté, développer des modèles de risque incluant le risque climatique (Battiston & Monasterolo, 2020; Battiston et al., 2017) et d'un autre côté proposer des scénarios macroéconomiques cohérents vis-à-vis de la transition vers une économie bas carbone (Krey et al., 2020; Luderer et al., 2015). Cette thèse s'inscrit dans cet effort de recherche.

Dans le cas d'un exercice de stress-test, les modèles utilisés se basent sur le modèle de Merton-Vašíček. Ce modèle suppose que la probabilité de défaut pour tout emprunteur, d'un portefeuille homogène, est expliquée par un risque idiosyncratique et un risque systémique. Pour les grandes banques le risque idiosyncratique représente un terme d'erreur qu'il est possible de négliger de par la grande diversification de leurs activités.

Ainsi, pour ces acteurs, et dans le cadre du modèle de Merton-Vašíček le risque de crédit au niveau d'un portefeuille (de crédits par exemple) est exprimé uniquement en fonction du risque systémique.

Cependant, dans le cas du modèle de Merton-Vašíček, ce risque systémique est supposé être une variable aléatoire gaussienne. Par conséquent, les pertes extrêmes sont traditionnellement modélisées<sup>1</sup> (N. Tarashev, 2010) en fonction d'un quantile de cette variable aléatoire gaussienne. Il est alors contraignant d'introduire des variables explicatives. Les deux premières contributions de cette thèse, présentent deux méthodes pour étendre ce modèle réglementaire en estimant la distribution du facteur systémique conditionnellement à des variables explicatives. Ensuite, nous proposons de calculer la probabilité de défaut stressée par un quantile de cette distribution conditionnelle et non plus par un quantile de la distribution gaussienne.

Notons par ailleurs que le vecteur de variables explicatives à partir duquel est modélisée la probabilité de défaut, est généralement composé de variables macroéconomiques ou financières : croissance du PIB, taux de chômage, inflation, ... L'intégration du risque climatique à cette étape peut se faire de deux manières. Une première méthode consiste à ajouter dans ce vecteur d'autres variables liées aux risques environnementaux : un prix du carbone, des émissions de gaz à effet de serre, des coûts liés à des catastrophes naturelles etc. Cette approche ne semble cependant pas satisfaisante et peu utilisée en pratique, puisqu'elle revient à négliger l'impact du changement climatique sur les autres variables macroéconomiques, or il semble bien établi par les économistes, comme montré par la célèbre meta-analyse de N. Stern (2007), ou plus récemment les travaux de Pottier (2014), que les contraintes environnementales ont une influence sur ces variables macroéconomiques.

C'est pourquoi les travaux académiques, Albano et al. (2021) ou Cartellier (2022), et les exercices de stress-tests climatique proposés par le régulateur, Alogoskoufis et al. (2021), se sont concentrés sur une deuxième approche qui consiste à supposer que ces variables macroéconomiques sont projetées en prenant directement en compte le risque climatique. Dans ce cadre, la réalisation d'un exercice de stress-test climatique devient un cas particulier d'exercice de stress-test où les variables explicatives intègrent directement l'enjeu climatique. Le déroulement d'un exercice de stress-test mesurant le risque de crédit pourrait alors se résumer ainsi :

- A partir d'un portefeuille de crédits, une banque construit un modèle pour expliquer la probabilité de défaut à partir d'un vecteur de variables macroéconomiques et financières.
- Le régulateur projette ces variables pour différents scénarios (s'il intègre une transition vers une économie bas carbone, nous parlerons alors de stress-test climatique,

---

<sup>1</sup>Voir le paragraphe 272 de l'accord de Bâle II disponible à cette adresse : <https://www.bis.org/publ/bcbs107.pdf>

sinon non).

- La banque projette alors ses probabilités de défaut à partir de ces scénarios afin de déterminer ses besoins en fonds propres.

Cette deuxième approche est bien plus confortable pour les institutions financières puisqu'il suffit alors de considérer des modèles déjà utilisés habituellement, ou une extension comme développé dans ce manuscrit, puis de calculer les probabilités de défaut stressées conditionnellement à des scénarios macroéconomiques futurs intégrant les risques environnementaux.

Le premier apport de cette thèse aborde la première étape de construction des modèles statistiques pour modéliser la probabilité de défaut, sachant les valeurs des variables explicatives. Nous proposons plusieurs modèles de type *location-scale*<sup>2</sup>, à directions révélatrices qui permettent d'estimer les quantiles conditionnels du facteur commun sachant les valeurs des variables explicatives. Un tel type de modèle semi-paramétrique offre plusieurs avantages. Premièrement, il permet de modéliser des interactions potentiellement non linéaires entre les variables explicatives et le facteur commun, car nous proposons d'estimer ces relations de manière flexible en combinant des hypothèses paramétriques (l'existence des directions révélatrices) et non paramétrique (sans hypothèses sur la forme de l'effet des directions révélatrices). Notamment, il présente la possibilité d'estimer l'effet (qu'il soit linéaire ou non) des variables explicatives sur la variance conditionnelle. Cela présente l'avantage de pouvoir interpréter l'effet d'une variable explicative sur la queue de distribution du facteur commun. De plus, les prédicteurs présents dans la moyenne conditionnelle peuvent inclure des retards des valeurs du facteur commun. Grâce au modèle de Merton-Vašíček et à notre modélisation prédictive de l'évolution du facteur commun, il devient alors possible de calculer une probabilité de défaut "stressée" qui évolue en fonction de la valeur prise par ce vecteur de variables explicatives. La flexibilité offerte par ce modèle et les interprétations qu'il permet sont particulièrement intéressantes dans le cadre d'un exercice de stress-test, qu'il soit climatique ou non. Aussi, nous présentons quelques garanties théoriques basées sur des premiers résultats de Akritas and Van Keilegom (2001), Carroll et al. (1997), Delecroix et al. (2006), Koul et al. (2017), and Neumeyer and Van Keilegom (2010). Enfin, nous présentons des études de simulations pour montrer la pertinence de cette approche par rapport au modèle réglementaire et l'appliquons à des données réelles issues d'une banque française.

Dans le Chapitre 3, nous proposons une deuxième méthode pour déterminer les quantiles conditionnelles du facteur commun, basée sur une méthode d'estimation en deux étapes, une première où une régression quantile est effectuée puis une deuxième étape de lissage paramétrique permet d'estimer des distributions conditionnelles. L'objectif de cette deuxième étape est de minimiser l'écart entre les paramètres d'une distribution appartenant à une famille paramétrique flexible et les quantiles estimés par régressions quantiles. Cela permet d'éviter le problème dit de quantile-crossing et cela permet aux

---

<sup>2</sup>*Moyenne-écart type*, en français

praticiens d'interpréter aisément la distribution conditionnelle obtenue. Nous appliquons cette méthode d'estimation sur données réelles et comparons les résultats à ceux obtenus avec le modèle dit location-scale à directions révélatrices.

Cette deuxième approche peut apparaître comme un compromis entre une estimation classique de la probabilité de défaut, et le modèle location-scale proposé. En effet, dans ce cas, la relation entre les quantiles de la distribution et les variables explicatives est supposée linéaire et n'est plus estimée non paramétriquement. Cette contrainte supplémentaire pourrait sembler restrictive, mais elle pourrait en revanche permettre d'obtenir des résultats plus cohérents sur des échantillons de petites tailles mais aussi d'avoir un comportement plus facilement interprétable pour les praticiens.

Cette méthode a été implémentée dans un package publié sur le CRAN où il est aussi proposé d'utiliser ces approches pour déterminer la distribution conditionnelle de variables macroéconomiques utilisées comme prédicteurs dans les modèles de risque de crédit. Cette étape est primordiale dans la réalisation d'un stress-test puisqu'un scénario jugé non cohérent, trop optimiste, voire trop pessimiste devient dès lors inutile. Que ce soit pour le régulateur, qui ne peut juger de l'adéquation des fonds propres des banques face au risque climatique, ou, pour les institutions bancaires pour qui un stress-test est aussi un exercice d'auto-critique sur la pertinence de sa stratégie d'investissement et d'octroi de crédits.

Par exemple, lors de la diffusion des résultats des exercices de stress-test climatique de 2020<sup>3</sup> l'ACPR, a noté que les participants de ce stress-test étaient sceptiques quant à la faible variabilité des scénarios. Les scénarios de transition considérés par le régulateur allaient de l'absence totale de transition, avec une augmentation forte de la fréquence et de l'intensité des catastrophes météorologiques, à une transition douce et rapide vers une économie bas carbone. Or, la projection des différentes variables macroéconomiques (PIB, chômage, inflation, etc.) n'a montré qu'une variation marginale malgré des hypothèses de transition opposées. Puisque le calcul des fonds propres nécessaires a été conduit sur la base de ces scénarios, il a été naturellement constaté que les banques et assurances avaient une exposition minimale au risque climatique, dans le sens où, que la transition ait lieu ou non, les pertes potentielles sont similaires. Cette conclusion pourrait être due à la non exposition de ces institutions aux risques environnementaux, mais elle pourrait aussi résulter de la faible variabilité des scénarios proposés par le régulateur. Une autre critique rapportée par les participants à ce premier exercice est que ces scénarios sont déterministes et ne prennent pas en compte la possible occurrence d'une récession ou d'une crise économique. Ces hypothèses sont hautement improbables, et des politiques climatiques trop, ou pas assez, ambitieuses pourraient conduire à une crise financière.

Nous pensons alors que la génération de scénarios est un exercice particulièrement important dans le cadre de l'évaluation de l'exposition aux risques environnementaux, mais,

---

<sup>3</sup>Rapport disponible à cette adresse : [https://acpr.banque-france.fr/sites/default/files/medias/documents/20210602\\_as\\_exercice\\_pilote\\_english.pdf](https://acpr.banque-france.fr/sites/default/files/medias/documents/20210602_as_exercice_pilote_english.pdf)

cet exercice est aussi particulièrement problématique pour plusieurs raisons. D’abord, les stress-tests climatiques ont un horizon temporel très long, et il est impossible de prédire le moment et la gravité d’une crise financière due à une politique climatique inconnue. Ensuite, les modèles qui incluent les risques environnementaux, par exemple, Krey et al. (2020), Luderer et al. (2015), utilisés par le régulateur, ou W. Nordhaus (2014), supposent souvent que la croissance économique est due à une augmentation exogène de la productivité des facteurs. Dans ce cas, la croissance économique est largement déterminée par cette trajectoire exogène. D’autres modèles proposent une modélisation endogène de cette quantité, mais ils supposent des variables explicatives qui peuvent être difficilement liées au changement climatique. Par exemple, Alestra et al. (2022) supposent des facteurs tels que l’éducation, la stabilité politique, le développement des infrastructures publiques, etc.

Dans le cadre de ces réflexions, le Chapitre 4 se focalise sur le développement d’un modèle permettant de modéliser la variation de la productivité totale des facteurs en fonction de la variation de l’*exergie*, cette quantité physique représente la quantité d’énergie réellement disponible pour le système économique. Cela nous permet de concevoir de nouveaux scénarios économiques intégrant des contraintes liées à la transition vers une économie bas carbone. Ce modèle permet de réconcilier différentes projections. D’une part, si l’*exergie* augmente, nous obtenons des projections similaires à celles des modèles économiques utilisés, par exemple, par les régulateurs. D’autre part, si l’*exergie* diminue, nous obtenons des projections plus proches de celles d’autres modèles d’évaluation intégrée (IAM), où la production économique décline, comme par exemple Capellán-Pérez et al. (2020), ou d’autres modèles, comme Meadows et al. (1972). Cette variabilité est cruciale pour les institutions financières afin de comprendre précisément quel scénario pourrait matérialiser des pertes financières importantes.

Pour résumer, le manuscrit de thèse s’organise de la façon suivante. Le Chapitre 1 présente l’historique des réglementations bâloises, quelques contraintes résultantes de ces réglementations, puis, le modèle du régulateur, ou de Merton-Vašíček, est présenté. Ensuite, le Chapitre 2 présente plusieurs modèles dits location-scale avec quelques résultats théoriques ainsi que des études de simulations pour justifier de son comportement par rapport au modèle réglementaire. Nous présentons dans le Chapitre 3, la méthode at-risk ainsi que des résultats sur données réelles. Finalement, le Chapitre 4, un modèle bayésien hiérarchique modélisant la productivité totale des facteurs en fonction de l’*exergy*. Nous intégrons ensuite cette modélisation dans le modèle DICE afin de générer des scénarios de croissance économique. Le manuscrit est complété par un chapitre sur les perspectives.



# English abstract

This manuscript primarily focuses on incorporating climate-related risks into bank risk management and estimating the resulting extreme financial losses. Typically, existing approaches, such as those reviewed in Acharya et al. (2023) on climate stress-testing, involve two key steps: first, developing a climate-economy model, and second, integrating a financial module to evaluate potential financial losses. Our work contributes to both of these research areas. The initial two chapters concentrate on refining methods for assessing financial losses, while the final chapter discusses the integration of climate-related risks into economic models. By addressing both aspects, we aim to provide a comprehensive approach to understanding and managing climate-related financial risks. The following paragraphs provide additional context of these contributions.

The main contribution of the Chapter 2 concerns credit risk modeling. The regulatory model, the so-called *Merton-Vašíček model*, is commonly used in credit risk management to estimate unexpected losses from a credit portfolio. This static model assumes that the probability of default over a given period for any borrower is explained by an idiosyncratic and a systematic risk. The former is diversified away for very large banks and the credit risk at a portfolio level is explained by the systematic (or market) risk. In the case of the Merton-Vašíček model this systematic risk is assumed to be a Gaussian random variable. Therefore extreme losses are traditionally modeled, see for instance N. Tarashev (2010) or the regulatory model<sup>4</sup>, as a quantile of this Gaussian random variable. Introducing explanatory variables is therefore not allowed, and this is a major drawback, especially when modeling tail risks. Chapter 2 presents a method to extend this regulatory model. We propose several semiparametric location-scale models to estimate the conditional distribution of the common factor given a vector of covariates. In particular, we are able to estimate the conditional quantiles of the common factor. We then propose an adaptation of the regulatory framework to this conditional case. Moreover, the proposed semiparametric models offer a flexible approach, allowing for nonlinear relationships between credit risk and macroeconomic variables. Also, the presented approach allows to model both a mean function and the conditional variance of the error terms. This is particularly important for financial institutions conducting stress-tests, as it enables them to more precisely monitor their risks. Location-scale regression models have been extensively studied in

---

<sup>4</sup>See paragraph 272 of the Basel II accord available : <https://www.bis.org/publ/bcbs107.pdf>

---

the statistical literature, see for example, Fan and Gijbels (1996), Koul et al. (2017), Neumeyer and Van Keilegom (2010). We here propose several extensions of the existing location-scale models, such as considering a single-index assumption in both location and scale functions, and allowing for lags of the response variable (the common factor). We define semiparametric estimators for these extensions and provide some theoretical grounds. Extensive simulation studies are also presented to assess the performance of the proposed estimators compared with the regulatory one. We also propose an application to real data. This chapter was presented in the Journées de Statistique in 2024<sup>5</sup>.

The Chapter 3 also proposes to model the conditional distribution of a random variables. Most econometric modeling forecasts are expressed as a point prediction which represents the average point and is the expectation of the conditional distribution of interest. However, the most complete information is to derive the complete distribution of this variable of interest. This is particularly important when designing stress-test scenarios, as point forecasts that target the most probable outcome may be overly optimistic. Adrian et al. (2019) illustrate this argument with GDP forecasts. They estimated the conditional distribution of the GDP given financial covariates and showed that the evolution of GDP growth depends heavily on the quantile of interest for instance the 90<sup>th</sup> percentile, (representing optimistic economic growth) does not depend on financial condition and is constant at around 7% yearly growth, however the 10<sup>th</sup> percentile (e.g. economic crisis) of the distribution depends heavily on these covariates, varying between -15% to slightly below 0% yearly growth. This highlights the need to consider the entire distribution rather than relying solely on point forecasts to accurately generate economic forecasts. Such methods are currently used by the regulator to generate short term stress-test scenarios, for instance the scenario for IMF member countries are generated as such<sup>6</sup>.

However to the best of our knowledge Adrian et al. (2019) and the following authors proposed economic forecasts. López-Salido and Loria (2020) proposed a report on Inflation using similar estimating strategies, Barbaglia et al. (2023) studied the influence of the sentiment in newspaper to predict several macroeconomic variables. Gelos et al. (2022) studied the impact of several policies on capital flows. We argue that this methodology also apply to the case of credit risk modeling and is especially interesting when modeling for extreme financial losses. Therefore, we propose a second methodology to estimate conditional distribution and apply this methodology in the case of credit risk, in the Merton-Vašíček model previously discussed. This methodology is a two step estimation method : first a quantile regression is fitted then a parametric distribution is fitted accorded to the previously estimated quantiles.

We applied this model on real data from a French bank and developed an R package, atrisk, available on the CRAN where results from Adrian et al. (2019) are reproduced and

---

<sup>5</sup>The presentation abstract is available at this url : <https://jds2024.sciencesconf.org/529551/document>

<sup>6</sup>See Chapter 5 Scenario Design :

<https://www.imf.org/en/Publications/Departmental-Papers-Policy-Papers/Issues/2020/01/31/Stress-Testing-at-the-IMF-48825>

## CONTENTS

---

a probability default modeled is illustration on public data, see Lajaunie et al. (2023)<sup>7</sup>. This chapter is derived from an under review article, Flament et al. (2023), where, the application section has been modified to illustrate our results on private data obtained from a French bank.

Depending on the context any methodology could be preferred to the other. First, the interpretation of single index location-scale model may be complicated not only for the risk management team but also for validation teams and the regulator who supervise the validity of the models used by banks. The complexity arises from several factors. The lack of closed-form expressions for the estimated functions can make it difficult to communicate and understand the model's behavior. Forecasting on extreme or out-of-sample data, can lead to uncertainty and skepticism regarding the model's reliability. Some financial institutions may not have the necessary number of observations to accurately estimate the non-parametric models of Chapter 2. However, in all these cases this methodology may be more suitable, however it has more assumptions. First, it assumes that the relationship between the quantiles of the conditional distribution of interest and the covariates is linear. The conditional variance function is non-parametrically estimated in the location-scale model. Moreover, the second step assumes that the distribution is known up to a vector of parameters. This assumption may be incorrect and could lead to modeling errors. In the case of the location-scale model no such parametric assumption is made.

Finally, we discuss the scenarios used in the climate stress-tests developed by regulators. The main exercises conducted in Europe were combined exercises that assessed both physical and long-term transition risks. These exercises generally consist of a two-step procedure: first, the regulator provides macroeconomic scenarios that assess the impact of the transition towards a green economy on the economy. Then, based on these scenarios, financial institutions must derive the impacts on their portfolios in term of financial losses. A methodological review has been published by Cartellier (2022). However, many critics are concerned about these scenarios. First the ACPR<sup>8</sup> noted participants in the stress-test were skeptical with the low variability among the scenarios. The transition scenarios considered by the regulator ranged from no transition at all and a high impact of meteorological catastrophes to a rapid and abrupt transition with high carbon prices in the short term. However, the differences in economic outcomes remained marginal. As a result, since the stress-test was conducted based on these economic variables, it was found that financial companies had minimal exposure to the transition pathway. Moreover, these scenarios are deterministic and do not account for any recession. These assumptions are highly improbable, and stringent climate policies may lead to a financial crisis. This outcome should be considered, and the probability of such an outcome should be quantified. This aspect should be even more important in the framework of climate stress-testing exercises.

---

<sup>7</sup>The documentation is available at the following URL : <https://cran.r-project.org/web/packages/atRisk/index.html>

<sup>8</sup>See the results of the 2020 climate stress-testing exercises : [https://acpr.banque-france.fr/sites/default/files/medias/documents/20210602\\_as\\_exercice\\_pilote\\_english.pdf](https://acpr.banque-france.fr/sites/default/files/medias/documents/20210602_as_exercice_pilote_english.pdf)

However, modeling such outcomes is ambitious for several reasons. First the climate stress-test have a very long time horizon and predicting the timing and severity of a financial crisis due to unknown climate policy is impossible. Moreover traditional model that includes climate related risks, for instance Integrated Assessment Models (IAM) (see Krey et al. (2020), Luderer et al. (2015), and W. Nordhaus (2014)) that are used by the the regulator, often assumes that economic growth is due to exogenous increase of the factor productivity in this case the economic growth is largely determined by this exogenous path. Other models proposed an endogenous modelisation of the factor productivity. However, they assume covariates that are hardly linked to climate change. Alestra et al. (2022) assumes factor such as education, political stability, public infrastructure development etc.

Therefore in the Chapter 4, we propose a hierarchical Bayesian model that models the variation of the total factor productivity by the variation of exergy. This allows us to design new economic scenarios that respect the Paris Agreement. Moreover, our model is capable to reconcile different projections by the introduction of environmental constraints. On one hand, if exergy increases, we have similar projections with standard economic models used for instance by the regulator. On the other hand, if exergy decreases, we obtain projections closer to other IAM, where economic output declines, see for instance Capellán-Pérez et al. (2020) and Meadows et al. (1972). This variability is crucial for financial institutions to fully understand where their risks may arise. Additionally, Bayesian modeling enables us to calculate the probability of any economic growth scenario given a specific transition pathway. This chapter has been published in Structural Change and Economic Dynamics, for more details see Flament (2023).

## References

- Acharya, V. V., Berner, R., Engle, R., Jung, H., Stroebel, J., Zeng, X., & Zhao, Y. (2023). Climate stress testing. *Annual Review of Financial Economics*, 15(1), 291–326.
- Adrian, T., Boyarchenko, N., & Giannone, D. (2019). Vulnerable growth. *Am. Econ. Rev.*, 109(4), 1263–1289.
- Alestra, C., Cette, G., Chouard, V., & Lecat, R. (2022). Growth impact of climate change and response policies: The advanced climate change long-term (ACCL) model. *J. Policy Model.*, 44(1), 96–112.
- Barbaglia, L., Consoli, S., & Manzan, S. (2023). Forecasting with Economic News. *J. Bus. Econ. Stat.*, 41(3), 708–719.
- Capellán-Pérez, I., de Blas, I., Nieto, J., de Castro, C., Miguel, L. J., Carpintero, Ó., Mediavilla, M., Lobejón, L. F., Ferreras-Alonso, N., Rodrigo, P., et al. (2020). Medeas: A new modeling framework integrating global biophysical and socioeconomic constraints. *Energy Environ. Sci.*, 13(3), 986–1017.
- Cartellier, F. (2022). Climate stress testing, an answer to the challenge of assessing climate-related risks to the financial system? [Available at SSRN: <https://ssrn.com/abstract=4179311> or <http://dx.doi.org/10.2139/ssrn.4179311>].
- Fan, J., & Gijbels, I. (1996). *Local polynomial modelling and its applications* (Vol. 66). Chapman & Hall, London.
- Flament, G. (2023). Impact of the energy transition on long-term factor productivity. *Struct. Change Econ. Dyn.*, 66(100), 393–406.
- Flament, G., Hurlin, C., & Lajaunie, Q. (2023). The at-risk approach: A new tool for stress tests and overlays [Article available at: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4637823](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4637823) or <https://dx.doi.org/10.2139/ssrn.4637823>].
- Gelos, G., Gornicka, L., Koepke, R., Sahay, R., & Sgherri, S. (2022). Capital flows at risk: Taming the ebbs and flows. *J. Int. Econ.*, 134(100).
- Koul, H. L., Müller, U. U., & Schick, A. (2017). Estimating the error distribution in a single-index model. *From Statistics to Mathematical Finance: Festschrift in Honour of Winfried Stute*, 209–233.
- Krey, V., Havlik, P., Kishimoto, P. N., Fricko, O., Zilliacus, J., Gidden, M., Strubegger, M., Kartasasmita, G., Ermolieva, T., Forsell, N., Gusti, M., Johnson, N., Kikstra, J., Kindermann, G., Kolp, P., Lovat, F., McCollum, D. L., Min, J., Pachauri, S., ... Riahi, K. (2020). Messageix-globiom documentation – 2020 release. <https://doi.org/10.22022/iacc/03-2021.17115>
- Lajaunie, Q., Flament, G., Hurlin, C., Kazemi, S., & Lajaunie, M. Q. (2023). Package ‘atrisk’ [Package available on the CRAN at: <https://cran.r-project.org/web/packages/atRisk/index.html>].
- López-Salido, D., & Loria, F. (2020). Inflation at risk [Article available at: <https://www.federalreserve.gov/econres/feds/files/2020013pap.pdf> or <https://doi.org/10.17016/FEDS.2020.013>]. *Finance and Economics Discussion Series*, (2020-013).
- Luderer, G., Leimbach, M., Bauer, N., Kriegler, E., Baumstark, L., Bertram, C., Giannousakis, A., Hilaire, J., Klein, D., Levesque, A., Mouratiadou, I., Pehl, M.,

- Pietzcker, R., Piontek, F., Roming, N., Schultes, A., Schwanitz, V. J., & Strefler, J. (2015, November). Description of the remind model (version 1.6) [Available at SSRN: <https://ssrn.com/abstract=2697070> or <http://dx.doi.org/10.2139/ssrn.2697070>].
- Meadows, D. H., Randers, J., & Meadows, D. L. (1972). The limits to growth (1972) [First edition of the book available at: <https://www.library.dartmouth.edu/digital/digital-collections/limits-growth>].
- Neumeyer, N., & Van Keilegom, I. (2010). Estimating the error distribution in nonparametric multiple regression with applications to model testing. *J. Multivariate Anal.*, *101*(5), 1067–1078.
- Nordhaus, W. (2014). Estimates of the social cost of carbon: Concepts and results from the dice-2013r model and alternative approaches. *J. Assoc. Environ. Resour. Econ.*, *1*(1/2), 273–312.
- Tarashev, N. (2010). Measuring portfolio credit risk correctly: Why parameter uncertainty matters. *J. Bank. Finance*, *34*(9), 2065–2076.

# Chapter 1

## General introduction and motivation

This chapter aims to lay the foundation of the Basel regulation, providing a historical perspective and examining its implications for the financial industry. Although this Chapter is not directly linked to climate risk issues, understanding this context is crucial for grasping the broader scope of risk management and its traditional modeling. This Chapter draws on the history of the Basel regulation. We will begin with an overview of Basel I, discuss its shortcomings, and trace the transition to Basel II and eventually Basel III. Our aim is not to provide an exhaustive account of the regulation, rather, we seek to introduce key concepts that are essential for the remainder of the manuscript. Additionally, we will present the model primarily used by practitioners and regulators.

### 1.1 Basel I : The first international regulation

#### 1.1.1 Some historical context

Firstly, it is important to note that the Basel Accord does not constitute international law. Instead, it serves as a working document commonly used by institutions responsible for financial stability around the world. In Europe, this role is fulfilled by the European Banking Authority (EBA) or the central bank, while in the United States, banking supervision is conducted by the Federal Reserve. Historically, the Basel accord was geared towards achieving two main objectives. First, the goal was to close the gap between all the different international regulations to promote best practices around the world and also to increase the banking stability by increasing capital buffer of all international banks. The core idea is that increasing capital buffers enhances bank resilience, which should stabilize the financial system and prevent banking crises that could severely impact economic activities. For instance, Oulton and Sebastia-Barriel (2017) investigated the effects of the subprime crisis on GDP, employment rates, and productivity levels. They estimated that for each year the crisis persisted, the long-term negative impact on per capita GDP was 1.1%.

The Basel accord initiative was propelled by the collapse of Herstatt Bank in 1975. A few months later, at a G10 summit, the precursor to the Basel Committee was established. The first document released was a set of principles for sound and responsible banking practices. However, the rules for capital standards were published a decade later, in 1988. This publication, known as Basel I, laid the groundwork for all subsequent regulations on banking capital standards, e.g. Basel II, III and its subsequent updates.

Concurrently, the Basel Accord began to be implemented across various jurisdictions, initially in G-10 countries and later globally following the Lyon Summit in 1996. These two events laid the foundation for risk-based regulation and capital adequacy rules. From 1988 onward, most banking institutions were required to monitor and report the riskiness of their portfolios and maintain adequate capital to prevent catastrophic failures. The main idea was to classify all assets held by a bank into risk categories and provide a capital buffer for each asset. This capital is often called the required capital, or minimum capital requirements and they are now part of the first pillar of the Basel accord. The rationale behind this regulation is straightforward: if a bank is sufficiently large and diversified, it is unlikely that all its assets will lose their value simultaneously. Therefore, by applying a capital buffer to each component of the bank's assets, the bank can cover significant decreases in the value of its portfolio, thereby promoting overall banking stability.

Initially, the methodology was very conservative and simple, with broad risk categories and large capital buffer for each categories. However, as regulations evolved and as internal data management and risk modeling framework emerged, the banks were allowed to use more sophisticated models for classifying risk and determining the appropriate level of capital for each asset. In the Basel II accord, the internal rating based approach was introduced. This accord allows very large banks to develop internal risk models and calculate their own capital buffers based on these models. While these models must be reviewed by regulators, the use of internal ratings has led to the creation of risk modeling teams within banking institutions. The role of these teams is to effectively monitor the riskiness of the bank's portfolio and determine the capital requirements by developing internal models. Since regulators must approve the models used by banks, to the best of our knowledge, most banks use the so-called regulatory model, which is based on the Merton-Vašíček model and will be discussed later.

### 1.1.2 The need for international regulations

The Basel I accord on capital requirements was a result of the risk posed by very low capital level (Tarullo (2008)) of international banks on global financial stability. This low capital level might become a risk as it represents the maximum loss a bank could face before going bankrupt. In other words the capital of a bank could be understand as a buffer against financial losses before a banking institution goes bankrupt. The higher the ratio, the less risky a bank is. Historically, capital ratio where as low as 0.95% for Crédit Lyonnais (or 1.28% for BNP) in the 80s which means that if its total assets value drop by more than 0.95% the bank could go bankrupt. Due to the intricate connections

## CHAPTER 1. GENERAL INTRODUCTION AND MOTIVATION

---

among financial institutions in the monetary market, regulators grew concerned about the potential domino effect of defaults triggered by a single bank's failure.

It became all too real during the 2008 financial crisis. Lehman Brothers, in particular, faced significant challenges as its assets were heavily invested in subprime products. The root of the crisis could be traced back to 2000 where there was a chronic underestimation of the credit risk by rating agencies, new financial product designed to spread risk better and a general increase in leverage ratio. The crisis was sparked by the realization that rating agencies had assigned high grades to highly risky assets due to an apparent diversification. This mistake made the valuation of these financial products incorrect. Consequently, the valuation of Lehman Brothers' assets also became extremely challenging. Consequently, Lehman Brothers' creditors became wary of its creditworthiness, leading to a complete freeze in the availability of short-term loans to the bank. Unfortunately, the bank required daily funding of around 5 billion (Wiggins and Metrick (2019)), the inability to secure short-term loans proved to be fatal for the bank, ending in its bankruptcy in September 2008. This bankruptcy, in turn, instigated widespread panic in the monetary market, resulting in government intervention to avoid the potential collapse of multiple banks.

For instance the largest bank in Iceland, Kaupthing, closed its subsidiary in foreign countries due to financial struggle, and on the 9th October of 2008, the domestic operation where seized by the domestic regulator. In France, Crédit agricole raised 5.8 billion in capital in July 2008. Whereas government intervened massively, in the US the Federal reserve launched the TARP (trouble asset relief program) of which almost 250 billions was targeted to stabilize banking institutions. In France the SPPE (société de prise de participation de l'état) landed 20 billions to different banking institution to stabilize the market<sup>1</sup>. Despite the severity of the crisis on the banking system, it underscored the critical importance of promoting banking risk management processes and teams. It provided two significant arguments to promote the regulation. First, it made banks managing directors acutely aware that even large banking institutions could fail and trigger devastating economic consequences. This realization helped them understand that a risk management team was not merely a cost center but a crucial defense against bankruptcy. Secondly, the financial crisis served as a real-world benchmark for risk management teams. For the first time, the losses computed by their models were tested in reality, moving from theoretical pessimistic scenarios to actual outcomes. Additionally, according to internal document from a French bank, these models proved to be somewhat accurate and effectively helping the bank managing their loss and activities during the crisis.

It is important to remember that banks with lower leverage are considered less risky but also less profitable. Consequently, banking institutions were concerned about potential competitive distortions due to varying regulations across countries. Although Basel I primarily aimed to increase capital requirements, it also reduced banks' leverage ratios,

---

<sup>1</sup>€5.1 billion to BNP Paribas, €3.4 billion to Société Générale, €3 billion to Crédit Agricole, €1.2 billion to Crédit Mutuel, and €7.05 billion to Banque Populaire-Caisse d'Épargne (BPCE), see <https://questions.assemblee-nationale.fr/q13/13-41957QE.htm>

thereby impacting profitability. Regulators and governments overseeing banks were worried about the competitive position of their national banks if a consensus was not reached, which would lead to different regulatory approaches in each country. These concerns, however, were largely alleviated by the widespread adoption of the Basel Accord by many, if not all, developed countries where major banks operate.

Hence, Basel I accord was an international accord signed in 1988 on capital requirements for banks. We will largely focus our attention on credit risk and the capital adequacy for debt related products. We will not investigate market risk, liquidity risk or operational risk even though they could be of major importance for a banking institution.

This capital requirements is centered around the so called Cooke ratio which represents the minimum capital a bank must hold with respect to its book risk (also noted RWA), they are defined as follows, Roncalli (2020):

$$\begin{aligned} \text{Tier 1 ratio} &= \frac{C_1}{RWA} \geq 4\%, \\ \text{Tier 2 ratio} &= \frac{C_1 + C_2}{RWA} \geq 8\%. \end{aligned}$$

Where  $C_1$  and  $C_2$  respectively correspond to tier-1 and tier-2 capital that are defined in Table 1.1. These thresholds might appear somewhat arbitrary, as they could have been set to 3% and 6% or 5% and 10%, respectively. Indeed, there is no universally accepted capital adequacy ratio that definitively protects the financial system from the failure of one of its participants. As noted by Goodhart (2005), this is due to the lack of an economic theory to determine such thresholds that would protect the entire financial system. Hence, instead of creating rules to protect the financial system it was decided to implement rules on individual banks capital requirements. Indeed, in this case, models do exist to calculate individual risk (e.g., RWA), and a pragmatic rule has been established to ensure the stability of a single financial institution, with the assumption that safeguarding one entity sufficiently reduces the probability of systemic collapse. Therefore, we will focus on the modelisation of the risk of a single actor.

Risk-weighted assets (RWAs) are calculated by assigning weights to the risks associated with each security held by a bank. According to the Basel I accord, these risk weights are classified, and the specific weights are detailed in Table 1.2.

This classification disregards the creditworthiness of borrowers, assigning all companies the same risk weight of 100%. Such a uniform approach was already recognized as limited when introduced in 1988. For instance, whether the borrower is a well established multinational company or a very young startup, the risk weight is the same. Therefore banks asked for more permissive modelisation of credit worthiness which were allowed under Basel II.

**Example 1.** *Before detailing the refinement of Basel II, let us illustrates the calculation of the RWA in a simple case. First let us assume that a bank has 100m€(read 100 millions)*

Table 1.1: Definitions of Tier 1 and Tier 2 Capital

<b>Tier 1 Capital</b>	CET1 is the highest capital quality, it is the sum of investor investment and retained earnings. For instance, as of 12 december 2022, Société Générale ad a Tier 1 Capital of 66.5bn\$ <sup>2</sup>
<b>Tier 2 Capital</b>	Includes undisclosed funds not appearing on a bank's financial statements, revaluation reserves, hybrid capital instruments, subordinated term debt (also known as junior debt securities), and general loan-loss or uncollected reserves.

*of fully secured by mortgage on residential property, then the risk weights of that loans portfolio is  $100m\text{€} \times 50\% = 50m\text{€}$ . Then, if that bank only has that portfolio in its book, the Cooke ratio indicates that the minimum Tier 1 capital is,  $0.04 \times 50m\text{€} = 2m\text{€}$ .*

### 1.1.3 Examples of modelisation of extreme risks in banking institution

In this section, we aim to identify the key quantities of interest for banking institutions and the variables for which they develop statistical models. The primary objective of these models, as discussed earlier, is to mitigate the risk of bank failure. While it is impossible to eliminate this risk entirely, we seek to quantify the likelihood—with a confidence level of  $\alpha$ —that a bank has sufficient capital to withstand a loss  $L_\alpha$ . Specifically,  $L_\alpha$  is defined as the threshold such that  $\mathbb{P}(L < L_\alpha) = 1 - \alpha$ , where  $L$  represents the yearly losses arising from various risks. Our goal is to model this critical quantity  $L_\alpha$ . For example, market risk, as evidenced by the Kerviel incident, cost Société Générale over 5 billion euros in 2008. Operational risks are another source, such as disruptions caused by the COVID-19 pandemic, which can halt normal operations and incur financial losses. Additionally, credit risk, resulting from the default of numerous borrowers, can also lead to significant losses. The primary focus of this manuscript will be on credit risk, with the aim of accurately modeling extreme losses associated with it. In most applications, these various risks are assumed to be independent and are therefore analyzed separately. Consequently, the total loss is considered to be the sum of the individual losses from each risk category.

To monitor credit risk, we must model three key variables. The first is the EAD, or Exposure At Default. The EAD quantifies the size of the balance sheet, representing the remaining balance of the loan for each borrower if they default instantly. In such a case, the EAD would be a preliminary estimate of the bank's loss. This estimate is only approximate because the bank would then employ a debt collection agency to recover as much as possible by selling the borrower's assets. The proportion of the debt that

<b>Risk Weight</b>	<b>Securities</b>
0%	<p>Cash</p> <p>Claims on central governments and central banks denominated in national currency and funded in that currency</p> <p>Other claims on OECD central governments and central banks</p> <p>Claims collateralized by cash of OECD government securities</p>
20%	<p>Claims collateralized by cash of OECD government securities</p> <p>Claims on multilateral development banks</p> <p>Claims on banks incorporated in the OECD and claims guaranteed by OECD incorporated banks</p> <p>Claims on securities firms incorporated in the OECD subject to comparable supervisory and regulatory arrangements</p> <p>Claims on banks incorporated in countries outside the OECD with a residual maturity of up to one year</p> <p>Claims on non-domestic OECD public-sector entities</p> <p>Cash items in process of collection</p>
50%	<p>Loans fully secured by a mortgage on residential property</p>
100%	<p>Claims on the private sector</p> <p>Claims on banks incorporated outside the OECD with a residual maturity of over one year</p> <p>Claims on central governments outside the OECD and non-denominated in national currency</p> <p>All other assets</p>

Table 1.2: Risk weights for some securities as defined by Roncalli (2020) Chapter 3

remains unrecovered by the agency is called the Loss Given Default (LGD). Therefore, in the event of a default, the loss for a single defaulted borrower is usually modeled as the product of the *LGD* and the *EAD*. This is still an approximation of the actual loss, as the debt collection agency likely incurs a cost, and the funds are not recovered instantly, leading to additional opportunity costs where inflation could also be considered. However, to the best of my knowledge these costs are usually ignored in most modeling strategies and the focus is on modeling *EAD* that is known today but unknown in the future due to possible early repayment or preapprove credit lines that could be drawn at any given time. Note that preapproved credit lines are a component of what is commonly referred to as the Credit Conversion Factor (CCF), which is widely modeled by financial institutions, particularly banks that offer revolving credit to customers.

Hence, let  $L$  be the loss of a bank for some fiscal year, our goal is to find  $L_\alpha$  such that :  $\mathbb{P}(L < L_\alpha) = \alpha$ . Where,

$$L = - \sum_{\text{borrower}} EAD \times LGD \times \mathbb{1}_{\text{borrower defaulted}}.$$

If the distribution of all the random variables were known, this exercise would be straightforward. We could use Monte Carlo simulation techniques or even explicit expressions to accurately determine the distribution of the loss. Unfortunately, the distributions of these variables, especially the default indicator, are unknown. This is of primary concern for both the risk modeling teams within banking institutions and the regulators that both focus their efforts into proposing the best framework to model the probability of default.

Figure 1.1 represents the loss as if the distribution of all credit risk components was known. The resemblance of this illustration to any real data is not important, however this illustration allows us to introduce two fundamental concepts in the banking literature. First, the mean of the loss distribution is called the expected loss, the expected loss should be covered by so called provision. The Basel Committee on Banking Supervision (2004) focuses on the unexpected loss that are on top of the expected losses (see paragraph 212), usually the unexpected losses are the 0.001 quantile of the loss distribution. The rational being that a financial institution should have enough capital to withdraw a loss that should occur only once every milenia.

## 1.2 Standardized approach in Basel II

The Basel II accord is an extension of Basel I. The same capital requirements hold, and the Cooke ratio is still mandatory. However, banks are now allowed to compute their own risk weights using probability of default and loss given defaults models. However, banks that do not desire to develop such model could still use the predefined risk weights. The second approach is named the standardized approach (SA) whereas the first is referred as the Internal Rating Based approach (IRB).

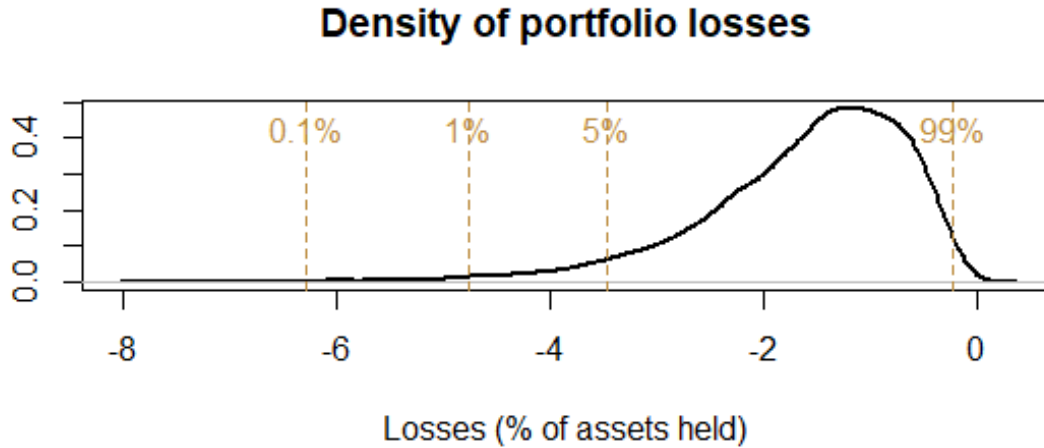


Figure 1.1: Illustration of the distribution of a yearly loss (as a proportion of the size of the portfolio). For arbitrary distributions of the Probability of Default, the Loss Given Default and the Exposure At Default. Note that the loss is here represented as a negative random variable.

Table 1.3: Credit Assessment and Risk Weights

ratings	AAA to	AAA+ to A-	BBB+ to BB-	Below BB	Unrated
<b>Risk Weight</b>	20%	50%	100%	150%	100%

Table 1.4: RWA for rated corporate firms as mentioned in Basel Committee on Banking Supervision (2004), paragraph 67.

As mentioned the risk weights still exist and the ratings are drastically refined. The idea behind this is to classify the assets of the bank depending on the type of claim and then classify each claim based on external credit assessments, for instance, a claim on rated corporate firms is presented in Table 1.4 and a claim on sovereign entity or their central bank (for instance debt of a country) is presented in Table 1.6, see Basel Committee on Banking Supervision (2004) paragraph 66.

Many categories are defined, including international banks, public corporations, sovereign bonds, residential properties, and commercial properties, among others. While we will not delve into all the assigned weights here, it's noteworthy that during this period, the risk-weighted assets (RWA) for residential properties, whether owner-occupied or not, were reduced from 50% to 35%. This reduction in RWA weights, proposed in 2004, likely influenced the increase in debt used for real estate purchases. Interestingly, the Basel Committee observed that, in contrast to claims secured by residential property, claims secured by commercial real estate carried a weight of 100% due to past crises that had

Table 1.5: Credit Assessment and Risk Weights for sovereign entities.

ratings	AAA to AA-	A+ to BBB-	BBB+ to BBB-	BB+ to B-	Below B-	Unrated
<b>Risk Weight</b>	0%	20%	50%	100%	150%	100%

Table 1.6: RWA for sovereign entities as mentioned in Basel Committee on Banking Supervision (2004) paragraph 53.

impacted this sector. Note that the previously presented risks weights are now outdated and have been regularly updated to be both more granular and representative to the true risk. In Europe the last update was published in the CRR3 and CRD6 in march 2024<sup>3</sup>. Below, let us compute the RWA for several portfolio.

**Example 2.** *Let's compute the Risk-Weighted Assets (RWA) for a hypothetical bank with the following assets:*

- *Loans to AAA-rated corporate firms: €100 million*
- *Loans to BBB-rated corporate firms: €150 million*
- *Sovereign bonds rated AA: €100 million*

*According to the Tables 1.4 and 1.6 the respective risk weight for each assets (e.g, RWA) are :*

- *Loans to AAA-rated corporate firms: €50 million*
- *Loans to BBB-rated corporate firms: €150 million*
- *Sovereign bonds rated AA: €0 million*

*Hence, the total RWA is €200millions, therefore, the core tier 1 capital should be at least of €8millions according to the Cooke ratio. This example is an exemple of a very safe assets structure hence the very low capital requirements, however, assume that the first loan is instead to a company rated B and the sovereign bond is to a BB country. In this second case, the core tier 1 capital should be at least of €15 million an increase resulting from the riskier portfolio of the bank.*

---

<sup>3</sup> For a brief overview of this regulation, see : <https://acpr.banque-france.fr/de-bale-iii-au-paquet-bancaire-crr3crd6-un-accord-final-pour-renforcer-la-resilience-du-secteur> .

### 1.2.1 Foundation Internal Rating Based Approach (FIRB)

Recall that we defined Risk-Weighted Assets (RWA) as a measure of the riskiness of a portfolio, specifically a credit portfolio in our case. The internal rating-based approach allows a bank to develop its own measure of risk. Therefore, we must define a way to characterize the RWA using measurable sources of credit risk.

To characterize the Risk-Weighted Assets (RWA) using measurable risk drivers, we consider the factors outlined by the Basel Accord. These include Probability of Default (PD), Loss Given Default (LGD) and Exposure At Default (EAD). The Foundation Internal Rating-Based Approach (FIRB) allows banks to use internal models and data only for the probability of default, while the values for EAD and LGD, are predefined (or given by regulatory formulae). However, implementing this method requires prior approval from the supervisor. In contrast, the Advanced Internal Rating-Based (AIRB) framework in Basel III allows banks to use internal models for all risk components, e.g, PD, LGD, and EAD.

It is interesting to note that starting from 2025, the differences between the regulatory model and the internal rating-based model will be limited. Specifically, in 2025, internal rating-based models may output a required capital of at most 50% of that of the regulatory model, increasing to 75% in 2029. This is known as the input floor rule. Moreover, distinct input floor rules will be introduced for specific risks driver, e.g, PD, LGD, and EAD.

Regardless, the purpose of all these frameworks is to model unexpected losses (UL) that financial institutions may face under financial crisis. Instead of focusing on expected values, these frameworks model the quantile of a distribution, such as the loss. According to Basel Committee on Banking Supervision (2004), the required capital for a borrower under the FIRB framework is as follows:

$$\begin{aligned}
 \xi &= 0.12 \left( \frac{1 - \exp(-50PD)}{1 - \exp(-50)} \right) + 0.24 \left( 1 - \frac{1 - \exp(-50PD)}{1 - \exp(-50)} \right), \\
 b &= (0.11852 - 0.05478 \cdot \ln(PD))^2, \\
 K &= \text{LGD} \left[ \Phi \left( \frac{\Phi^{-1}(PD) + \sqrt{\xi} \Phi^{-1}(1 - \alpha)}{\sqrt{1 - \xi}} \right) - PD \right] \frac{(1 + (M - 2.5)b)}{(1 - 1.5 \times b)}, \\
 \text{RWA} &= K \times 12.5 \times \text{EAD}.
 \end{aligned} \tag{1.1}$$

Where  $\Phi$  is the cumulative distribution function of a standard Gaussian random variable and  $M$  the maturity of the loan for which the RWA is to be calculated. We have deliberately omitted the hat notation, despite each quantity being an estimate, to maintain consistency with the notations used in the Basel II framework Basel Committee on Banking Supervision (2004). A few comments will be made about this framework. First, let us note that the goal is to model the capital requirements due to credit risk. This capital requirement is on top of so-called provisions that could be seen on any publicly traded banks balance sheet. This corresponds to the loss a bank typically encounter in

a normal business environment, in fact provisions should correspond to the expected loss in a typical business year. However the unexpected loss is aimed at modeling extreme events and losses and this capital requirement comes on top of the provisions. Hence the quantity :

$$\Phi \left( \frac{\Phi^{-1}(\text{PD}) + \sqrt{\xi} \Phi^{-1}(1 - \alpha)}{\sqrt{1 - \xi}} \right) - \text{PD},$$

corresponds to the difference between an estimates of the probability of default and a more extreme estimates depending on a value  $\alpha$ . One remarks that if  $\xi = 0$  (which is however not possible under FIRB formula) then, the additional capital charges is null. Also note that all other factors being equal the RWA for a longer maturity will be higher than that for a loan with shorter maturity, indeed the RWA grows linearly with the maturity of the loans at rate :

$$\text{LGD} \left[ \Phi \left( \frac{\Phi^{-1}(\text{PD}) + \sqrt{\xi} \Phi^{-1}(1 - \alpha)}{\sqrt{1 - \xi}} \right) - \text{PD} \right] \frac{b}{1 - 1.5b}.$$

This quantity can be assumed to be positive. The only two exceptions are first when the Loss Given Default (LGD) is negative, which would imply, the not so realistic case of the bank gaining money when a counterparty defaults. Or when one consider  $\alpha$  such that :

$$\alpha > 1 - \Phi \left( \frac{\sqrt{1 - \xi} \Phi^{-1}(\text{PD}) - \Phi^{-1}(\text{PD})}{\sqrt{\xi}} \right). \quad (1.2)$$

If  $\xi$  tends to 0, then then right-hand side of (1.2) tends to 0.5. Consequently, this second case is not of particular interest because it represents an estimate of the probability of default during economic expansion, which is not relevant for stress-testing where  $\alpha = 0.001$  is generally considered. Figure 1.2c displays an illustration of the value of  $K$  with respect to  $\alpha$ .

Assume we have only one loan with an EAD of 1, Figure 1.2 illustrates the value of  $K$ , with several assumptions on the risk drivers (PD, LGD, maturity, ...) and the regulatory formulae (1.1). If we assume that PD = 5%, LGD = 50%, the maturity is five years, and consider  $\alpha = 0.001$ , then  $K = 0.17$  and the risk weighted asset is RWA = 2.11. However, this value is drastically reduced to RWA = 1.34 in the case where  $\alpha = 0.1$ . Finally, if we only consider a one year maturity and still consider  $\alpha = 0.1$ , then the risk weighted assets are again reduced to 0.90.

## 1.2.2 Advanced Internal Rating Based Approach (AIRB)

The FIRB approach only suppose that a probability of default model is developed internally. The EAD, LGD and maturity are supposed to be given. This is not the case in the Advanced Internal Rating Based Approach (AIRB) where all of these risks factors are modeled. The Basel committee promotes the adoption of a FIRB approach by even mid-sized banks as it was intended to be the most widely used method for determining capital

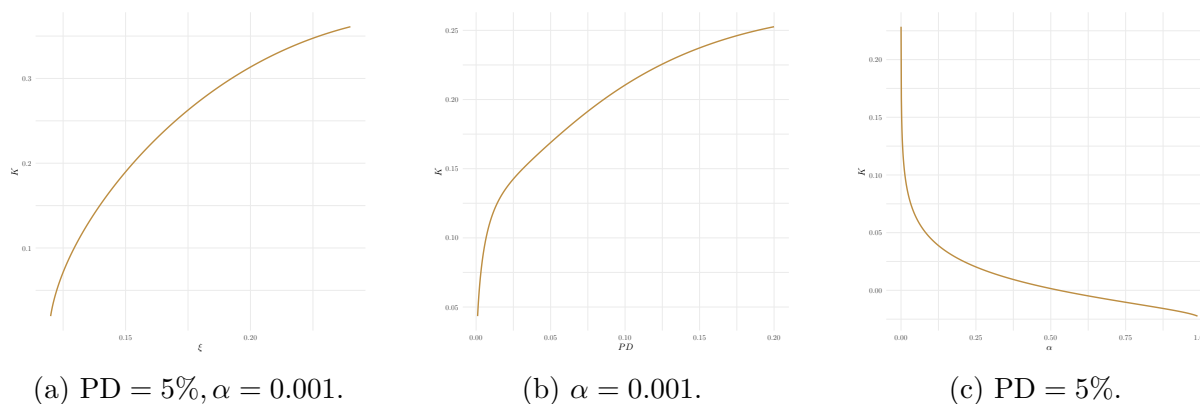


Figure 1.2: Illustrations of  $K$  computed as defined by (1.1), for LGD = 50% and a five year maturity. In Figures 1.2b and 1.2c,  $\xi$  is computed according to (1.1). The regulatory formula is assumed for all displays in the case of  $b$ .

requirements in the banking sector. However, for the largest banks that meet additional criteria, they can choose to adopt the AIRB methodology. The motivation for a bank to do so is evident: the better it models these random variables, the more accurately it reflects the risks associated with its activities, thereby enhancing its profitability.

We briefly presented the general rules for computing capital requirements however, we never presented how such model were obtained. The following section is devoted to presenting this methodology, the so called Merton-Vašíček model.

### 1.2.3 The backbone of the IRB methodology : Merton-Vašíček model

The Merton (1974) model is based on economic reasoning. Let's consider a scenario where you own a company that has a debt with some bank. The debt, with value  $D$ , has a fixed maturity period, represented by  $T$ . Until time  $T$ , no payments are required, but at that point, the entire debt must be repaid. Your company has assets that are currently valued at  $A_0$ , and these assets will generate wealth that can be used to repay the debt at maturity. However, the value generated by your company depends on both your intrinsic competence and the macroeconomic environment. In the case of an economic crisis or if you are not suitable for your role, the value generated will likely be lower. Therefore, the Merton model assumes that the value of the assets fluctuates according to a Wiener process with drift  $\mu$  and variance  $\sigma^2$ . At time  $T$ , the owner of the company can evaluate the value of the assets and has two choices. First, if the assets are valued higher than the debt, the owner can sell these assets to repay the bank and keep the remaining cash. On the other hand, if the value of the assets is lower than the debt, the best choice for the owner is to default and let the bank sell the assets and reimburse as much as it can. Therefore, in this model, default is defined as the event where the value of the assets of the company at time  $T$  is lower than the debt, i.e.,  $A_T < D$ .

This implies some assumptions. First, it is assumed that the firm's capital structure is independent of its value. This simplification aligns with the assumption of a constant debt value used in this section. However, more advanced models, such as those proposed by Longstaff and Schwartz (1995), relax this constraint, allowing the value of debt to vary with the firm's characteristics. Additionally, in the event of a default, the firm is assumed to default on all its financial obligations.

Since we assumed a Wiener process for the company's assets value, the probability of default ( $p$ ) can be explicitly written. Let first denote  $B_t$  a Brownian motion with Gaussian increments, therefore we have :

$$\forall t \in [0, T], \quad dA_t = \mu A_t dt + \sigma A_t dB_t,$$

$$A_T = A_0 \exp \left( \sigma B_T + \left( \mu - \frac{1}{2} \sigma^2 \right) T \right).$$

If the maturity of the loan is one year ( $T = 1$ ), then one can compute the probability of default of the company :

$$p = \Phi \left( \frac{\log(A_0/D) - (\mu - \sigma^2/2)}{\sigma} \right) = \Phi(C_D).$$

Now, let us assume that a bank lend money to  $n$  companies and want to use the Merton (1974) model to estimates their respective probability of default (say  $p_i$  for  $i \in \{1, \dots, n\}$ ). The bank could easily find their respective probability of default by knowing the current values of each company's assets ( $A_{i,0}$ ), its indebtedness ( $D_i$ ) and their respective growth potential and risk ( $\mu_i, \sigma_i$ ). These estimates are far from easily retrieved but could be obtained by a bank assessment during the lending process. However, in the current formulation of the model no direct dependencies between each probability of default is explicit. However, it is known that companies tend to default simultaneously during economic crisis (see for instance Duffie et al. (2007)). Hence, the FIRB framework incorporates the extension proposed by Vašíček (2002) which asserts that the default of each company is conditionally independent of one another, given a common factor. This common factor represents the macroeconomic environment, to which each company is reliant. However, within the macroeconomic environment, each default is independent due to the inherent competence of the owner, staff, business model etc. Formally speaking, let us assume that we want to compute the increase in the value of the assets  $B$  of a company, then the extension of Vašíček (2002) states that :

$$B = \sqrt{\xi} Y + \sqrt{1 - \xi} \epsilon, \quad Y, \epsilon \sim \mathcal{N}(0, 1) \text{ and } \xi \in [0, 1].$$

Note that the time dependencies have been dropped because we are mainly interested in the case of the probability of default at maturity which is supposed to be 1 period of time. In this case, the probability of default for each company conditionally to  $Y = y$  is :

$$\pi(y) = \mathbb{P}(A < D \mid Y = y) = \mathbb{P} \left( \epsilon < \frac{C_D - \sqrt{\xi} y}{\sqrt{1 - \xi}} \mid Y = y \right).$$

Note that  $Y$  is a common factor identical for all companies in the portfolio. Generally banks create so called homogeneous portfolio where each company is similarly affected by this common factor. Therefore, within this homogeneous portfolio, each company is modeled as having the same conditional probability of default depending only on the common factor. To the best of my knowledge these portfolios are generally constructed based on the economic sector of the company. For instance a credit portfolio composed of companies in the construction sector, a second one in the automotive industry, a third one composed of information technology firms etc.

We see that conditionally to  $Y = y$ , all events are independent as  $D$  and  $\xi$  are considered constants. Given that  $p = \Phi(C_D)$ , then we find the said, stressed probability of default under the FIRB framework. This leads to the previously stated capital requirements (1.1) :

$$\pi(\Phi^{-1}(\alpha)) = \Phi \left( \frac{\Phi^{-1}(p) + \sqrt{\xi} \Phi^{-1}(1 - \alpha)}{\sqrt{1 - \xi}} \right).$$

However, several key modeling assumptions remain unexplored. For example, the selection of the correlation function  $\xi$  is not well-defined. Furthermore, N. Tarashev (2010) emphasizes the importance of accurately measuring the variability of this parameter to correctly estimate capital requirements. Failure to account for this uncertainty can lead to an underestimation of unexpected losses.

Moreover, this model significantly simplifies reality. No company has a single debt with a single maturity, and neither companies nor banks can accurately assess the value of all their assets. For instance, valuing a company like Apple would involve estimating the worth of intangible assets such as trademarks, patents, employee quality, and organizational structure. Precise valuation of these assets is extremely challenging. Additionally, estimating the variance of asset values over time is even more difficult due to the lack of frequent observations. Some extensions attempt to address these issues. For example, see N. A. Tarashev (2005) for a brief review of the literature during the discussions of the Basel II accord.

### 1.3 Some implications of Basel accord

Goodhart (2005) and more briefly Roncalli (2020) discussed the implication of the FIRB and AIRB framework on bank capital requirements. A major criticism of these regulations is that they are fundamentally procyclical. During an economic crisis, the probability of default for borrowers is expected to rise, requiring banks, under IRB regulations, to increase their capital to meet the criteria. However, raising capital during a crisis is inherently contradictory to lending money to companies. Consequently, this regulation restricts access to credit during economic downturns, potentially exacerbating the crisis. Although these discussions are beyond the scope of this manuscript, they are extensively debated in the context of the IFRS9 framework. Additionally, note that in the standard

approach, a heavy weight is applied if a counterparty (or borrower) is unrated. Unfortunately, most borrowers in developing countries are unrated, which discourages lending to these companies or governments, despite them potentially posing no greater risk than their rated counterparts in developed countries. This not only restricts credit access for these entities but also poses a challenge for banking institutions that are discouraged from diversifying their portfolios in these countries.

Moreover, the Basel regulation pushed financial institutions, especially banks to internally monitor and model their extreme losses. This regulation drove risk management teams to develop either internal statistical models or specific internal macroeconomic scenarios to monitor the risk of their portfolios. More recently, the regulator advised banks to monitor their climate-related risks (see the CRR3/CRD6 directive<sup>4</sup>) and include environmental risk in their risk assessment programs. These regulatory directives provide the foundation for the work of this thesis.

---

<sup>4</sup>The directive is available : [https://finance.ec.europa.eu/news/latest-updates-banking-package-2023-12-14\\_en](https://finance.ec.europa.eu/news/latest-updates-banking-package-2023-12-14_en)

## References

- Basel Committee on Banking Supervision. (2004, June). *International convergence of capital measurement and capital standards: A revised framework* [Regulation available at: <https://www.bis.org/publ/bcbs107.htm>]. Bank for International Settlements.
- Duffie, D., Saita, L., & Wang, K. (2007). Multi-period corporate default prediction with stochastic covariates. *J. Financ. Econom.*, 83(3), 635–665.
- Goodhart, C. A. (2005). Financial regulation, credit risk and financial stability. *Natl. Inst. Econ. Rev.*, 192, 118–127.
- Longstaff, F. A., & Schwartz, E. S. (1995). A simple approach to valuing risky fixed and floating rate debt. *The Journal of Finance*, 50(3), 789–819.
- Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *J. Finance*, 29(2), 449–470.
- Oulton, N., & Sebastia-Barriel, M. (2017). Effects of financial crises on productivity, capital and employment. *Rev. Income Wealth*, 63, S90–S112.
- Roncalli, T. (2020, April). *Handbook of financial risk management* [Book available at: <https://www.routledge.com/Handbook-of-Financial-Risk-Management/Roncalli/p/book/9781138501874>]. Chapman & Hall/CRC Financial Mathematics Series.
- Tarashev, N. (2010). Measuring portfolio credit risk correctly: Why parameter uncertainty matters. *J. Bank. Finance*, 34(9), 2065–2076.
- Tarashev, N. A. (2005). An empirical evaluation of structural credit risk models [Article available at: <https://www.bis.org/publ/work179.htm>].
- Tarullo, D. K. (2008, September). *Banking on Basel: The Future of International Financial Regulation* [Book available at the Peterson Institute: <https://www.piie.com/bookstore/banking-basel-future-international-financial-regulation>].
- Vašíček, O. (2002). The distribution of loan portfolio value. *Risk*, 15(12), 160–162.
- Wiggins, R. Z., & Metrick, A. (2019). The lehman brothers bankruptcy e: The effects on lehman’s us broker-dealer. *Journal of Financial Crises*, 1(1), 124–137.

## Chapter 2

# A semiparametric location-scale model with application to credit risk

### Abstract :

In this chapter we revisit the regulatory model, the so-called Merton-Vašíček model, that is commonly used in credit risk management to estimate unexpected losses from a credit portfolio. This static model assumes that the probability of default over a given period for any borrower is explained by an idiosyncratic and a systematic risk. The systematic risk is captured by a common (latent) factor, assumed to be a Gaussian random variable. Thus in the regulatory model the explanatory variables, such as macroeconomic variables, are not taken into account. We therefore propose a semiparametric location-scale regression approach allowing to compute the conditional distribution of the common factor given a vector of covariates. We then propose an adaptation of the regulatory framework to this conditional case, based on the conditional quantiles. The proposed semiparametric models offer a flexible approach, allowing for nonlinear relationships between credit risk and macroeconomic variables. We define semiparametric estimators and provide theoretical grounds. Extensive simulation studies are also presented to assess the performance of the proposed estimators. We also propose an application to real data.

## 2.1 Motivation for conditional Merton-Vašiček models

In this chapter our goal is to propose methods for understanding and predicting estimating extreme financial losses due to credit risks. In the case of the credit portfolio of a bank, these losses can arise from three factors. First, it can be due to the outstanding balance on its customers, the higher the riskier is the portfolio. This quantity is called EAD, or ‘Exposure At Default’, and most of the time the EAD is assumed to be known. In fact, it is known up to the unknown number of early repayment.

The second source of risk occurs when a customer defaults on its loan. In this case the bank either reposes the assets of its customer or tries to recover money from insurance if the customer was insured. Most of the time and in both cases the amount of money recovered does not cover the outstanding of the loan the ratio between the amount recover and the loan outstanding is called the recovery rate  $\mathcal{R}$ . The ‘Loss Given Default’, or LGD is then defined as the proportion of the money lost due to some credit defaults. During the 2008 financial crisis, the financial system gave mortgage to many subprime borrowers who defaulted on their loan. By itself the default was not the main problem indeed if the loan is secured by enough capital and the default is only due to a liquidity problem (think of not enough cash on hand) then the banks just has to sell the assets and recoup the outstanding money. However, in the case of the great financial crisis most of the mortgages were backed by real estate which values plunged in some area of the US. [The Case-Shiller US Home Price Index](#) was around 180 during the summer 2006, by the end of 2009 it was around 150 and was as low as 134 in february 2012. It means that all borrowers that bought at the peak and could not meet their financial obligations left their home to the lending bank which then tried to cover their losses by selling the property on the market which was saturated by all the reposed properties driving down the prices and reducing the liquidity. This phenomenon drives up the loss given default and ultimately the banks financial loses. Albeit of great interest, we will consider in this manuscript the EAD and LGD to be either given or at least independent of our quantity of interest. On a side note, remark that the regulation discussed in this manuscript mainly arises on the aftermath of such crisis to prevent any future financial catastrophe and to prevent any mistrust in the financial system.

The third component of credit risks is the probability of default which is the probability that a customer with some outstanding loan defaults to meet its financial obligation during the next period of time, typically a year. Most of the time a customer is said to default if it fails to reimburse the financial institution for 3 months. This credit risk is often correlated with economic difficulties of the clients coming either from idiosyncratic mismanagement such as a failing business model, lack of diversification, incompetency... But some times these defaults occur because of financial crisis that impact the whole lending portfolio of a bank. The idiosyncratic risk is not of great importance for large banking institution whose role is to finance the whole economy they hence have a very well diversified portfolio and these idiosyncratic risks are diversified away. However, the event of a large financial crisis is of high importance for these large banking institutions

and regulators often requires frequent reporting on these risks. One mandatory reporting are called stress-testing exercises. Of which the most common is the supervisory review and evaluation process (SREP). Sometimes stress-testing exercises are thematic which means that the regulator wants to assess the vulnerability of each bank or the banking system to some predefined risks. These exercises could be both quantitative such as the latest *climate related stress-test*<sup>1</sup> from the ECB was in 2022. Or qualitative, such as the *stress-test on cyber attacks*<sup>2</sup> conducted in 2024 to monitor the banks resiliency on a successful cyberattacks. The goal is to test banks resiliency to an interruption to their daily activities and assess their plans to restore normal operations.

Generally speaking these exercises contains two main phases. First the regulator defines some macroeconomic scenario that are both pessimistic but plausible. Then banks assess their credit risk on these scenarios and report their findings to the regulator. Both phases are data intensive and require several modelling techniques.

Moreover, in the broader context of financial risk management, the Merton-Vašíček model represents only one use case for monitoring extreme losses due to credit risk. But these extreme losses, should be dependent on macroeconomic forecasts, or scenarios. These scenarios are often forecasts

### 2.1.1 Overview

Financial institutions are required to manage their unexpected losses, which refer to financial losses that cannot be accurately forecast using conventional risk management models. Unexpected losses can result from a large range of factors, such as incidents on financial markets (e.g., the subprime crisis), natural disasters (e.g., the tidal waves following earthquakes and the Fukushima nuclear power plant failure), economic crises (e.g., the Asian crisis in the 90's), counterparty failures, sudden regulatory changes, etc. These losses are typically challenging to predict due to their complex and often unique nature. To prevent financial institutions defaults as consequence of such unexpected losses, for each borrower<sup>3</sup> all banks are required to have a minimum amount of capital (see also, Roncalli (2020)).

This capital requirement depends on the *EAD* which is *Exposure at Default* and represents the amount of money a lender is exposed to, or stands to lose, if a borrower fails to fulfill their financial obligations. It also depends on the maturity of the loan, the *Losses Given Defaults* (LGD) which is the amount that could not be retrieved by the bank after all assets are liquidated. It finally depends on the probability of default over a given period

---

<sup>1</sup>[https://www.bankingsupervision.europa.eu/ecb/pub/pdf/ssm.climate\\_stress\\_test\\_report.20220708~2e3cc0999f.et.pdf](https://www.bankingsupervision.europa.eu/ecb/pub/pdf/ssm.climate_stress_test_report.20220708~2e3cc0999f.et.pdf)

<sup>2</sup><https://www.bankingsupervision.europa.eu/press/pr/date/2024/html/ssm.pr240103~a26e1930b0.en.html>

<sup>3</sup>The regulation is available here : <https://www.bis.org/publ/bcbs107.pdf>. Formulae for the Internal Rating based, or Foundation Internal Rating Based Ratings (FIRB), method are stated in part 2 section 3, paragraph 272.

(e.g., one year). Let  $p$  be the unconditional probability of default of a lender.

In the Merton (1974) model, a company defaults when the value of its assets  $A$  falls below a given threshold  $D$ , where  $D$  is the debt of the company. Let  $I$  be a default indicator such that :

$$\{I = 1\} \iff \{A \leq D\}. \quad (2.1)$$

Therefore, assuming that  $B$ , the variation of the company's assets valuation, is standard Gaussian random variable, the unconditional probability of default is  $p = \mathbb{P}(I = 1) = \Phi(C_D)$  where  $\Phi(\cdot)$  is a standard Gaussian distribution function and  $C_D$  is derived in Chapter 1. Furthermore one can assume that  $B$  depends on an idiosyncratic factor  $\epsilon$  and a random common factor  $Y$  both supposed to be standard Gaussian variables. The common factor is then interpreted as the state of the global economy, the lower its value the worst the state of the economy, and higher the probability of default. This leads to the so-called Merton-Vašíček model, see Vašíček (2002). In that model, assuming further that the correlation between  $B$  and  $Y$  is equal to  $\sqrt{\xi}$ , the random variable  $B$  is decomposed as

$$B = \sqrt{\xi} Y + \sqrt{1 - \xi} \epsilon. \quad (2.2)$$

Given that  $C_D = \Phi^{-1}(p)$ , the conditional probability of default given  $Y = y$  can be obtained by plugin (2.2) in (2.1) which yields :

$$\pi(y) = \mathbb{P}(\sqrt{\xi} y + \sqrt{1 - \xi} \epsilon \leq \Phi^{-1}(p)) = \Phi\left(\frac{\Phi^{-1}(p) - \sqrt{\xi} y}{\sqrt{1 - \xi}}\right). \quad (2.3)$$

By construction, we have  $p = \mathbb{E}(\pi(Y))$ . Moreover,  $\pi(y)$  is a decreasing function of  $y$ . Using the relationship  $\Phi^{-1}(1 - \alpha) = -\Phi^{-1}(\alpha)$ ,  $\alpha \in (0, 1)$ , and replacing  $y$  by minus the  $(1 - \alpha)$ -th quantile of the standard Gaussian distribution, or, equivalently, the  $\alpha$ -th quantile, we can redefine  $\pi(\cdot)$  in (2.3) as the conditional probability of default given  $Y = \Phi^{-1}(\alpha)$ . With an abuse of notation, we get

$$\pi(\alpha) = \Phi\left(\frac{\Phi^{-1}(p) + \sqrt{\xi} \Phi^{-1}(1 - \alpha)}{\sqrt{1 - \xi}}\right) = \Phi\left(\frac{\Phi^{-1}(p) - \sqrt{\xi} \Phi^{-1}(\alpha)}{\sqrt{1 - \xi}}\right), \quad \alpha \in (0, 1). \quad (2.4)$$

Here, the function  $\pi(\cdot)$  is defined on  $(0, 1)$ , while the function in (2.3) is defined on the whole real line.

To prevent banks' failure due to extreme event, the regulator is interested by the evaluation the conditional probabilities corresponding to extreme left quantiles  $\Phi^{-1}(\alpha)$ . In the case of the Basel II regulation,  $\alpha = 0.001$ . The required capital due to credit risk for a banking institution is then depending on  $\pi(\alpha)$ . The higher the probability of default, the higher the required capital.

The classical Merton-Vašíček model presents several pitfalls. The most important one is that it assumes that the required capital does not directly depends on any covariates. Indeed, the regulatory formula (2.3) simply evaluates the conditional probability of default

given  $Y = y$ , without any reference to observed covariates. In the short term, this is perhaps less problematic for the calculation of the required capital. Indeed, fixing this required capital under the scenario that a financial crisis will occur in the next year seems reasonable, since stress-testing exercises aim at protecting the bank from a catastrophic event. However, in the long run, for example for climate stress-testing exercises, the practitioners would prefer to take into account the dynamics of macroeconomic, financial and other types of predictors.

Therefore we construct a statistical model that links the common factor  $Y$  to a vector of covariates  $W$  (e.g., macroeconomic variables, financial variables). More precisely, our goal will be to estimate the conditional distribution function of  $Y$  given the covariates  $W$ . We will then be able to estimate the conditional quantiles of  $Y$  given  $W$ . In particular, this will allow practitioners to include covariates in the regulatory stress-testing exercise by simply replacing the unconditional (marginal) quantiles  $\Phi^{-1}(\alpha)$  by their conditional versions in (2.4).

## 2.2 A general statistical model

Let  $\tau \in (0, 1)$  and define the conditional quantile function of order  $\tau$  given  $W = w$ , *i.e.*,

$$q_\tau(w) = \inf\{y : \Psi_{Y|W}(y | w) \geq \tau\} \quad \text{where} \quad \Psi_{Y|W}(y | w) = \mathbb{P}(Y \leq y | W = w).$$

Here,  $\Psi_{Y|W}$  is the conditional cumulative distribution function of  $Y$  given the covariate vector value, and we have

$$\Phi_Y(\cdot) = \mathbb{E}[\Psi_{Y|W}(\cdot | W)],$$

where  $\Phi_Y$  is the unconditional (marginal) distribution of  $Y$ . Under the Gaussianity assumptions like in Merton-Vašiček's model,  $\Phi_Y = \Phi$  and the conditional version of (2.3) given  $W = w$  is

$$\pi(\tau | w) = \Phi\left(\frac{\Phi^{-1}(p) - \sqrt{\xi} q_\tau(w)}{\sqrt{1 - \xi}}\right), \quad \tau \in (0, 1). \quad (2.5)$$

In our statistical analysis,  $\tau$  can be any fixed value between 0 and 1. For the bank regulation rules,  $\tau = \alpha = 0.001$  is generally considered.

It is worth noting that there is no reason to expect the conditional distribution of the common  $Y$  given the covariate value to be symmetric. Therefore, hereafter we will consider the setup of (2.5).

We here consider that the value  $p$  is given. In the credit risk analysis, the probability of default  $p$  can be accurately estimated from external data, such as the so-called transition matrices obtained from the reports of the rating agencies.

A flexible model for the conditional distribution of  $Y$  given  $W$ , and thus for the conditional quantile function, can be constructed as a location-scale regression model

$$Y = m(Z) + \sigma(X)\varepsilon, \quad \text{with} \quad \varepsilon \perp \{Z, X\} \quad \text{and} \quad \mathbb{E}(\varepsilon) = 0, \quad (2.6)$$

where  $Z \in \mathbb{R}^{d_Z}$  and  $X \in \mathbb{R}^{d_X}$  are subvectors of  $W \in \mathbb{R}^{d_W}$ , with possibly common components in which case  $d_Z + d_X \geq d_W$ . (Here and in the following, the symbol  $\perp$  is used for stochastic independence.) In general, the functions  $m$  and  $\sigma$  are unknown. The model (2.6) is a general regression model with heteroscedastic errors where the function  $\sigma^2(\cdot)$  is the conditional variance of the error terms. For identification purposes, the variance  $\text{Var}(\varepsilon)$  has to be set to some fixed value.

We will present the models in different forms and under different sets of hypothesis in the following sections. The location-scale regression model was extensively studied in the statistically literature. See, for example, Akritas and Van Keilegom (2001), Neumeier and Van Keilegom (2010), Racine and Van Keilegom (2020), and the textbooks Fan and Gijbels (1996), Ruppert et al. (2003), Härdle et al. (2004). In the location-scale model, the conditional quantile function of  $Y$  given  $W = w$ , for the order  $\tau \in (0, 1)$ , is

$$q_\tau(w) = \inf \left\{ y : F_\varepsilon \left( \frac{y - m(z)}{\sigma(x)} \right) \geq \tau \right\}, \quad (2.7)$$

where  $F_\varepsilon(\cdot)$  denotes the distribution function of the standardized error term  $\varepsilon$ . The vectors  $w$ ,  $z$  and  $x$  belong to the support of  $W$ ,  $Z$  and  $X$ , respectively. It is worth noting that

$$q_\tau(w) = m(z) + \sigma(x)q_\varepsilon(\tau),$$

where  $q_\varepsilon(\tau) = \inf \{t : F_\varepsilon(t) \geq \tau\}$  is the quantile of order  $\tau$  of  $\varepsilon$ .

In the following,  $(Y_i, W_i^\top)^\top = (Y_i, Z_i^\top, X_i^\top)^\top$ ,  $1 \leq i \leq n$  will be a sample of  $(Y, W^\top)^\top$ , where  $W$  is a random vector of covariates (possibly including discrete predictors), and  $Y$  is defined by (2.6). (Herein, for a matrix  $A$ ,  $A^\top$  denotes its transpose. Moreover, the vectors are column matrices.) Although for most parts of our study we consider an independent sample of  $(Y, W^\top)^\top$ , for now we do not exclude the cases of stationary vector-valued time series or  $W$  including lagged values of  $Y$ .

### 2.2.1 Location and scale functions

By construction,

$$\mathbb{E}(Y | W) = \mathbb{E}(Y | Z) = m(Z),$$

and thus the function  $m(\cdot)$  can be estimated by existing regression techniques. Let us also note that

$$\mathbb{E}(Y) = \mathbb{E}[m(Z)].$$

Regarding the scaling function  $\sigma(\cdot)$  in (2.6), two different estimation strategies can be found in the literature. The first one considers the problem sequentially : after estimating  $m(\cdot)$ , use the squared residuals to estimate  $\sigma^2(\cdot)$ . Let us call this the *residual-based approach*. See, e.g., Härdle and Tsybakov (1997) and Fan and Yao (1998). In general, the estimation of the location function  $m(\cdot)$  is expected to deteriorate the attainable rates for estimating the scale function  $\sigma$ . In the case  $X = Z = W \in \mathbb{R}$ , Hall and Carroll

(1989) investigate the issue and notice that for instance in nonparametric regression with two derivatives on  $\sigma(\cdot)$ , the ‘classical’ rates of convergence are possible if and only if the unknown  $m(\cdot)$  satisfies a Hölder-continuity condition of a certain order. They show that optimal rates of convergence for  $\sigma(\cdot)$  are attained by kernel estimators. Wang et al. (2008) reconsidered and corrected the results of Hall and Carroll (1989), namely the order of the Hölder-continuity condition. Wang et al. (2008) point out that, contrary to the common practice, it is often not desirable to build the variance function estimator using the residuals from an optimal nonparametric fit of the mean function. Instead it is more desirable to use estimators of the mean function with minimal bias. They explain that main reason is that the bias and variance of mean function estimator have quite different effects on the estimation of variance function. Indeed, the bias of the nonparametric mean function estimator cannot be removed or even reduced in the second stage smoothing, while the effect of its variance can be mitigated easily. A similar idea will appear below where, for our final estimation purpose, that is the conditional quantiles, we consider undersmoothing for the unknown nonparametric functions (that means, we consider a smaller bandwidth rate than  $n^{-1/5}$ , the one providing the optimal rate of convergence  $n^{-2/5}$  for twice differentiable functions; the reason is to make the squared bias to be negligible compared to the parametric rate).

The study was further extended to more general settings. For instance, Härdle and Tsybakov (1997) and Fan and Yao (1998) propose local polynomial estimators for  $\sigma(\cdot)$  with dependent, stationary data. Using a same bandwidth for the first and second order conditional moments given the lag-1 observation, Härdle and Tsybakov (1997) focus on the case of nonparametric autoregressive models of order 1, particularly well-tailored for studying the volatility in financial time series. They derive the asymptotic normality of their estimator. Finally, under regularity conditions similar to Hall and Carroll (1989) they show asymptotic properties of the estimates of  $m$  and  $\sigma$ .

A second approach for the estimation of the scale function, designed for the case of univariate or constant  $\sigma(\cdot)$  and known as the *difference-based approach* estimator, focuses on directly estimating  $\sigma(\cdot)$  without requiring an estimate for  $m$ . See Muller and Stadtmuller (1987) and Shen et al. (2020) and their references. To briefly recall the idea, let us first consider the case homoscedastic sample. A simple method for estimating the variance  $\sigma^2$  is then obtained by computing the mean of all squared pairwise differences within the sample. However, when dealing with heteroscedasticity, it becomes necessary to locally average these differences around a fixed point, employing for instance kernel smoothing. Muller and Stadtmuller (1987) showed that the resulting estimator is uniformly consistent, asymptotically normal. Recently, Shen et al. (2020) showed that this type of estimator achieves the minimax rates with both random and non-random design. With finite samples, the difference-based approach is expected to achieve smaller bias and mean squared error. However, the residual based approach seems more convenient for extensions to multivariate scale functions. Indeed, the literature on the case where  $\sigma^2(\cdot)$  is a multivariate function is scarcer, and focused on the difference-based approach. See, for example, Cai et al. (2009), Liitiäinen et al. (2010) and the references therein.

Let us now discuss an identification aspect. Usually, in location-scale models, the variance of the variable  $\varepsilon$  in (2.6), the unscaled error term, is set equal to 1. However, for the application we keep in mind to the Merton-Vašíček model, the variance of  $Y$  is 1. Thus, the usual choice for  $\text{Var}(\varepsilon)$  is no longer admissible. Hence, unless stated differently, we consider  $\text{Var}(\varepsilon) = \sigma_Y^2/2 = 1/2$ .

## 2.2.2 Error distribution function estimators

The final estimation step concerns the distribution function (df)  $F_\varepsilon$  of  $\varepsilon$ , which is required in order to build estimates of the conditional quantile function in (2.7). Theoretical results for that distribution functions are somehow more involved. Akritas and Van Keilegom (2001) proved asymptotic properties in the unidimensional case. Neumeyer and Van Keilegom (2010) extended the results to the multidimensional case, considering only the empirical distribution function of the residuals. Racine and Van Keilegom (2020) recently showed that these asymptotic properties extends to the case of kernel distribution function estimator, extending an estimation idea studied by Azzalini (1981). These approaches and few variants are detailed in the following.

Given the data  $(Y_i, W_i)$ ,  $1 \leq i \leq n$ , and the estimators  $\hat{m}$  and  $\hat{\sigma}^2$ , let

$$\hat{\varepsilon}_i = \frac{Y_i - \hat{m}(Z_i)}{\hat{\sigma}(X_i)}, \quad (2.8)$$

be the *standardized residuals* obtained in model (2.6). We here distinguish the standardized residuals from the residuals, which usually refers to the quantities  $Y_i - \hat{m}(Z_i)$  in mean regression models.

Definition (2.8) may be refined depending on the proposed estimator for  $m$  and  $\sigma$ . Based on the standardized residuals  $\hat{\varepsilon}_i$ , we study four methods to estimate  $F_\varepsilon$ . The first method is using the empirical df of the standardized residuals :

$$\hat{F}_{\varepsilon, \text{emp}}(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{\hat{\varepsilon}_i \leq t\}}, \quad t \in \mathbb{R}. \quad (2.9)$$

See Koul et al. (2017) for a similar estimator in the case of a single-index model for the location function, and without scale function. See also Neumeyer and Van Keilegom (2010) for a general setup of location-scale regression.

This empirical distribution may be a simple and quite effective estimator of  $F_\varepsilon(\cdot)$  if the interest lies in quantiles  $q_\tau(\cdot)$  with  $\tau$  away from 0 and 1. It is however expected to behave poorly for extreme quantiles. We therefore also consider smoothed versions of the empirical df. Let  $k(\cdot)$  be a symmetric kernel of order 2 on the real line, that is

$$\forall u \in \mathbb{R}, \quad k(-u) = k(u) \geq 0, \quad \text{and} \quad \int_{\mathbb{R}} k(u) du = 1.$$

We then consider the smooth empirical df :

$$\widehat{F}_{\varepsilon,1}(u) = \frac{1}{n} \sum_{i=1}^n K((\widehat{\varepsilon}_i - u)/h_\varepsilon) \quad \text{with} \quad K(u) = \int_{-\infty}^u k(v)dv, \quad (2.10)$$

and  $h_\varepsilon$  some bandwidth devoted to the estimation of  $F_\varepsilon$ . In the case of the Epanechnikov kernel, we have

$$k(u) = (3/4)(1 - u^2)\mathbb{1}_{\{|u| \leq 1\}} \quad \text{and} \quad K(u) = (1/4)(3u - u^3 + 2)\mathbb{1}_{\{|u| \leq 1\}} + \mathbb{1}_{\{u \geq 1\}}.$$

The smooth empirical distribution is the df obtained by integrating the Parzen-Rosenblatt density estimator. See Azzalini (1981) for the properties of the smooth empirical distribution estimator in the case where the sample of the variable (here  $\varepsilon$ ) is observed. See also Racine and Van Keilegom (2020) for the case where  $\widehat{\varepsilon}_i$  are obtained after fitting a nonparametric location-scale model.

Finally, keeping in mind the application to stress-testing where the interest focuses on the accurate estimation of the left tail of the distribution, we also investigate a version of the smoothed empirical distribution with asymmetric kernel. More precisely, we replace  $k(u)$  by  $k_l(u) = 2k(u)\mathbb{1}_{\{u \leq 0\}}$ , and thus  $K(u)$  by

$$K_l(u) = 2 \int_{-\infty}^{\min(0,u)} k(v)dv,$$

and construct a third estimator under the form

$$\widehat{F}_{\varepsilon,2}(u) = \frac{1}{n} \sum_{i=1}^n K_l((\widehat{\varepsilon}_i - u)/h_\varepsilon) = \frac{1}{n} \sum_{i=1}^n \{2K((\widehat{\varepsilon}_i - u)/h_\varepsilon)\mathbb{1}_{\{\widehat{\varepsilon}_i \leq u\}} + \mathbb{1}_{\{\widehat{\varepsilon}_i > u\}}\}.$$

In our empirical study, we consider this idea with  $k(\cdot)$  the Epanechnikov kernel.

Finally, in order to diminish the bias of the smoothed empirical distribution estimator, we consider it with a higher order kernel. For simplicity, we only investigate the case of the asymmetric kernel. More precisely, we define

$$\widehat{F}_{\varepsilon,3}(u) = \frac{1}{n} \sum_{i=1}^n \{2K_6((\widehat{\varepsilon}_i - u)/h_\varepsilon)\mathbb{1}_{\{\widehat{\varepsilon}_i \leq u\}} + \mathbb{1}_{\{\widehat{\varepsilon}_i > u\}}\},$$

with

$$K_6(u) = C_1 [u - 7u^3/3 + 63u^5/25 - 33u^7/25 - C_2] \mathbb{1}_{\{|u| \leq 1\}} + \mathbb{1}_{\{u > 1\}},$$

where  $C_1 = 525/256$  and  $C_2 = 128/525$ .

Given the estimators (parametric, semiparametric single-index...) of the location and scale functions, as well as an estimator  $\widehat{F}_\varepsilon(\cdot)$  of the distribution function  $F_\varepsilon(\cdot)$ , among

the four types we described above (that are,  $\widehat{F}_{\varepsilon, \text{emp}}$ ,  $\widehat{F}_{\varepsilon, 1}$ ,  $\widehat{F}_{\varepsilon, 2}$ ,  $\widehat{F}_{\varepsilon, 3}$ ), the estimator of the conditional quantile function of order  $\tau \in (0, 1)$  given  $W = w$  defined in (2.7) is

$$\widehat{q}_\tau(w) = \widehat{q}_\tau(z, x) = \inf \left\{ y : \widehat{F}_\varepsilon \left( \frac{y - \widehat{m}(z)}{\widehat{\sigma}(x)} \right) \geq \tau \right\}, \quad (2.11)$$

where  $z$  and  $x$  belong to the support of  $Z$  and  $X$ , respectively.

Let us recall the purpose of our semiparametric location-scale modeling. We want to get estimates for

$$q_\tau(w) = \inf \left\{ y : F_\varepsilon \left( \frac{y - m(z)}{\sigma(x)} \right) \geq \tau \right\},$$

where  $w, z, x$  are values in the supports of  $W, Z, X$ , where  $W$  is a vector of covariates and  $Z, X$  are subvectors of  $W$ . With the notation introduced above, we can write

$$\widehat{q}_\tau(w) = \widehat{m}(z) + \widehat{\sigma}(x)\widehat{q}_\varepsilon(\tau),$$

where  $\widehat{q}_\varepsilon(\tau) = \inf\{t : \widehat{F}_\varepsilon(t) \geq \tau\}$ . By simple algebra, we get

$$\widehat{q}_\tau(w) - q_\tau(w) = \{\widehat{m}(z) - m(z)\} + \{\widehat{\sigma}(x) - \sigma(x)\}q_\varepsilon(\tau) + \sigma(x)\{\widehat{q}_\varepsilon(\tau) - q_\varepsilon(\tau)\}.$$

The last equation reveals that the rate of convergence and the asymptotic distribution of  $\widehat{q}_\tau(w)$  will be determined by the estimator with the slowest rate of convergence among  $\widehat{m}(z)$ ,  $\widehat{\sigma}(x)$  and  $\widehat{q}_\varepsilon(\tau)$ . An important consequence of this is that fully nonparametric models for the location and scale functions would suffer from the curse of dimensionality. Semiparametric models including a dimension reduction assumption will be preferable.

### 2.2.3 Simulation of location-scale models with Gaussian response

In the following, we will investigate several models providing estimates  $\widehat{m}$  and  $\widehat{\sigma}$  to be used in (2.11). Special attention will be given to the finite sample performance which will be investigated by numerical simulations. We therefore need a synthetic data generator. Generally speaking, the simulation in the model (2.6) seems rather simple. One simply has to consider some functions  $m(\cdot)$  and  $\sigma(\cdot)$ , and some distribution function for the standardized error term  $\varepsilon$ .

However, we here study a specific class of problems, that is rather popular in the financial theory. Indeed, a very popular model is that of Merton (1974), originally used for pricing options, but also widely used for predicting the probability of firms defaulting. This second use case has become very popular, as it is the standard regulatory model (see Roncalli (2020)). This means that financial institutions must perform external audits of the regulator in order to use other models. Recall that a firm is said to default if the value of its assets fall below its amount of debt. In the spirit of Merton (1974), the value of the firms' assets are modeled by a Wiener process, and this naturally causes the common factor to be a standard Gaussian random variable. Hence, we are interested in the specific

case of modeling a location-scale model (2.6) where  $Y \sim \mathcal{N}(0, 1)$ . In particular, we want to generate synthetic samples where the marginal distribution of  $Y$  is guaranteed to be standard normal.

The method for generating samples used in our numerical study, based on the so-called Box and Muller (1958) method for generating a pair of independent  $\mathcal{N}(0, 1)$  variables, is described in the following. First, Box and Muller (1958) noticed that, if  $U_1, U_2$  are independent uniform random variables on  $[0, 1]$ , the random variables  $\sqrt{-2\ln(U_1)} \cos(2\pi U_2)$  and  $\sqrt{-2\ln(U_1)} \sin(2\pi U_2)$  are independent standard Gaussians. Therefore, for generating sample according to the model

$$Y = m(Z) + \sigma(X)\varepsilon, \quad Y \sim \mathcal{N}(0, 1),$$

we propose to simulate our data as follows: for some  $\rho \in (0, 1)$ , and  $U_1, U_2 \sim U[0, 1]$  with  $U_1 \perp U_2$ , let

$$Y = \underbrace{\sqrt{\rho}\sqrt{-2\ln(U_1)} \sin(2\pi U_2)}_{m(Z)} + \underbrace{\sqrt{1-\rho}\sqrt{-2\ln(U_1)}}_{\sigma(X)} \underbrace{\cos(2\pi U_2)}_{\varepsilon}. \quad (2.12)$$

The user parameter  $\rho$  controls the signal-to-noise ratio (SNR) in the regression model. By construction,

$$\varepsilon \perp X, Z \quad \text{and} \quad Y \sim \mathcal{N}(0, 1).$$

The next step is to generate  $Z$  and  $X$  after choosing suitable functions  $m(\cdot)$  and  $\sigma(\cdot)$ , such that  $m(Z)$  and  $\sigma(X)$  match the location and scale terms in (2.12), respectively. It is worth noting that the range of  $\sigma(X)$  is the whole positive half-line. Therefore, when  $X \in \mathbb{R}$ , the conditional standard deviation function  $\sigma(\cdot)$ , should be a one-to-one mapping with positive values. In our setups we mainly use the exponential function. The same one-to-one mapping condition will be imposed to  $m(\cdot)$ , which will set to be a polynomial function.

The case of multi-dimensional  $Z \in \mathbb{R}^{d_z}$  and  $X \in \mathbb{R}^{d_x}$  is slightly more involved. In order to propose user friendly and interpretable tools for the practitioners, in all the example we consider in the following the location and scale functions<sup>4</sup> will have the form  $m(Z^\top \gamma)$  and  $\sigma(X^\top \beta)$ , respectively, with  $\gamma \in \mathbb{R}^{d_z}$ ,  $\beta \in \mathbb{R}^{d_x}$  and  $m(\cdot)$ ,  $\sigma(\cdot)$  univariate functions. Depending on the model, the functions  $m(\cdot)$  and  $\sigma(\cdot)$  are given (this is the case of parametric models) or considered unknown for the estimation (as is the case in semiparametric models). However, for generating synthetic samples for simulation studies, the user knows these univariate functions and can proceed as follows. Let

$$X = (X_1, \dots, X_{d_x})^\top = (X_1, \tilde{X}^\top)^\top, \quad \beta = (1, \beta_2, \dots, \beta_{d_x})^\top = (1, \tilde{\beta}^\top)^\top,$$

---

<sup>4</sup>We here use a slight abuse of notation when we replace  $m(Z)$  and  $\sigma(X)$  by  $m(Z^\top \gamma)$  and  $\sigma(X^\top \beta)$ , respectively. The former functions are possibly multivariate while the latter are defined on an interval on the real line.

and

$$Z = (Z_1, \dots, Z_{d_Z})^\top = (Z_1, \tilde{Z}^\top)^\top, \quad \gamma = (1, \gamma_2, \dots, \gamma_{d_Z})^\top = (1, \tilde{\gamma}^\top)^\top.$$

With one-to-one functions  $m(\cdot)$  and  $\sigma(\cdot)$ ,  $\rho \in (0, 1)$ , given the sub-vectors  $\tilde{\beta}$  and  $\tilde{\gamma}$ , given the random draws  $U_1$  and  $U_2$  obtained as described above, and given the draws from the random vectors  $\tilde{X}$  and  $\tilde{Z}$  according to some distribution, the components  $X_1$  and  $Z_1$  are uniquely defined as the solutions of the equations

$$\begin{aligned} \sigma(X_1 + \tilde{X}^\top \tilde{\beta}) &= \sqrt{1 - \rho} \sqrt{-2 \ln(U_1)} \\ m(Z_1 + \tilde{Z}^\top \tilde{\gamma}) &= \sqrt{\rho} \sqrt{-2 \ln(U_1)} \sin(2\pi U_2). \end{aligned} \tag{2.13}$$

For simplicity, the components of  $\tilde{X}$  and  $\tilde{Z}$  are drawn independently. The equations (2.13) allows for vectors  $\tilde{X}$  and  $\tilde{Z}$  with some, all or no common components, while  $X_1$  and  $Z_1$  are necessarily different. In the case of constant variance, in the scale part of the model the simulation scheme (2.12) becomes

$$\sigma = \sqrt{1 - \rho} \quad \text{and} \quad \varepsilon_t = \sqrt{-2 \ln(U_{1,t})} \cos(2\pi U_{2,t}).$$

The simulation scheme proposed above can be used in several semiparametric location-scale models. We will also show that it accommodates to the case where  $\tilde{X}$  and  $\tilde{Z}$  are time dependent or the covariates include lags of  $Y$ .

## 2.3 Semi-parametric location-scale models

For our modeling, we consider a general setup where the functions  $m(\cdot)$  (location) and  $\sigma(\cdot)$  (scale) are unknown and depend on possibly different covariate vectors  $X$  and  $Z$ . In order to avoid poor accuracy induced by nonparametric estimation in several dimensions, we impose some dimension reduction assumptions commonly used in the semiparametric modeling, that are index regression and additive functions. The idea of the index regressions is that the information carried by a covariate vector can be summarized by one (single-index) or a small number (multi-index) of linear combinations. The linear combinations are given by unknown vectors of parameters which have to be estimated. In a semiparametric context, the conditional expectation of the response given the indices (the so-called *link function*) is unknown and has to be estimated nonparametrically. The common parametric regression models (linear model, generalized linear model...) satisfy the assumption of index regression with a given link function. In additive regressions, the nonparametric regression is assumed to have an additive structure, that is it is the sum of functions, each of them depending on a subvector of covariates or linear combinations of the covariates. In this way, the estimation of the regression only requires nonparametric estimation in lower dimensions.

In the following, we combine these type of assumptions resulting in several types of semi-parametric location-scale models. For the nonparametric estimation, we focus on kernel smoothing.

### 2.3.1 Homoscedastic semiparametric single-index regression

Let  $Y \in \mathbb{R}$  and  $Z \in \mathbb{R}^{dz}$ , and assume  $\mathbb{E}(Y^2) < \infty$ . Let us assume the model

$$Y = m(Z^\top \gamma) + \varepsilon, \quad \varepsilon \perp Z \quad \text{and} \quad \mathbb{E}(\varepsilon) = 0, \quad \mathbb{E}(\varepsilon^2) = \sigma^2, \quad (2.14)$$

where  $m(\cdot)$  is some function,  $\gamma \in \Gamma \subset \mathbb{R}^{dz}$  and  $\sigma > 0$ . Let  $\gamma_0$  denote the true value of the parameter. In this section, we discuss the model (2.14), starting from the simplest case of a parametric model where  $m(\cdot)$  is given before considering the semiparametric setup where  $m(\cdot)$  has to be estimated nonparametrically. Let us notice that the case of a nonparametric, univariate regression model corresponds to the case where  $Z \in \mathbb{R}$  and there is no  $\gamma$ . That case will be considered in more detail in Section 2.3.2, where a heteroscedastic error term will be allowed.

In a parametric modeling context where the function  $m(\cdot)$  is given, the parameter vector  $\gamma_0$  can be estimated by nonlinear least squares (NLS), that is by

$$\hat{\gamma} = \arg \min_{\gamma} \sum_{i=1}^n \{Y_i - m(X_i^\top \gamma)\}^2. \quad (2.15)$$

See, for example, Newey and McFadden (1994) for a general presentation of NLS. In particular, the estimator is well defined, consistent and  $\sqrt{n}$ -asymptotically normal under mild conditions on the function  $m(\cdot)$  (satisfied in the common regression models) and data generating processes (DGP) for the  $(\varepsilon_i, W_i)$ ,  $1 \leq i \leq n$  (in particular with i.i.d. data).

Since  $\sigma^2 = \mathbb{E}\{Y - m(Z^\top \gamma)\}^2$  by the definition of the model (2.14), using the estimate  $\hat{\gamma}$  in (2.15), one can compute the variance estimator

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \{Y_i - m(Z^\top \hat{\gamma})\}^2. \quad (2.16)$$

The asymptotic properties of the variance estimator  $\hat{\sigma}^2$  in (2.16) can be derived from standard asymptotic results, as presented for example in Newey and McFadden (1994) and van der Vaart (1998). Under mild assumptions on the regularity of the function  $m(\cdot)$  and the DGP for the  $(\varepsilon_i, W_i)$ 's, the variance estimator is consistent and  $\sqrt{n}$ -asymptotically normal.

Let us now extend the framework and consider that the function  $m(\cdot)$  is no longer given, and has to be estimated *nonparametrically*. This is the framework of the so-called *semi-parametric single-index (mean regression) model*. In such models, it is assumed that, concerning the regression function, the information carried by the vector of covariates, here  $Z$ , is the same as that carried by a linear combination. More formally, there exists a vector  $\gamma_0 \in \mathbb{R}^{dz}$  such that<sup>5</sup>

$$\mathbb{E}(Y \mid Z) = \mathbb{E}(Y \mid Z^\top \gamma_0). \quad (2.17)$$

---

<sup>5</sup>The model defined by (2.17) is more general than the one in (2.14) because (2.17) does not impose the conditional variance of  $Y$  given  $Z$  to be constant.

The random variable  $Z^\top \gamma_0$  is called the *index*. See, for example, Powell et al. (1989), Härdle et al. (1993), Ichimura (1993), Hristache et al. (2001), Delecroix et al. (2006), Horowitz (2009) and the references therein.

The model defined by (2.17) combines the flexibility of a nonparametric regression and a *dimension reduction* assumption which prevent from estimating multivariate functions. Indeed, the form of the conditional expectation  $\mathbb{E}(Y \mid Z^\top \gamma_0)$  is unknown, but this is a univariate function, while  $Z$  is a (multidimensional) vector of covariates. It is worth noting that the components of  $Z$  do not need to be continuous (as it would have been the case in the case of a multivariate nonparametric regression context). In fact, only the indices  $Z^\top \gamma$ , for some set of  $\gamma$ , are required to admit a density. This latter condition can be achieved with only a continuous component of  $Z$ .

In the model defined by (2.17), the vector  $\gamma = (\gamma_1, \dots, \gamma_{d_Z})^\top$  is not identifiable, because for any scalar  $a \neq 0$  and  $b$ ,  $\mathbb{E}(Y \mid Z^\top \gamma) = \mathbb{E}(Y \mid Z^\top (a\gamma) + b)$ . First, this reveals that  $Z$  cannot have a constant component (corresponding to the intercept in parametric regression). Regarding  $\gamma$ , the usual identification constraints are either

$$\|\gamma\| = 1 \quad \text{and} \quad \gamma_1 > 0,$$

or

$$\gamma = (1, \tilde{\gamma}^\top)^\top \quad \text{with} \quad \tilde{\gamma} = (\gamma_2, \dots, \gamma_{d_Z})^\top \in \tilde{\Gamma} \subset \mathbb{R}^{d_Z-1}. \quad (2.18)$$

The true value of the parameter is then  $\gamma_0 = (1, \tilde{\gamma}_0^\top)^\top$ . We use the later identification condition in the following, and we suppose that  $\tilde{\gamma}_0$  is an interior point of the compact set  $\tilde{\Gamma}$ . Let us notice that both types of identification restrictions implicitly impose a sign for the effect of the first component of the covariate vector  $Z$ . This should be taken into account by the practitioner by choosing as the first covariate a component for which the sign of the effect is not in doubt. Finally, it should be noted that in a single index semiparametric model only the relative levels of the coefficients  $\gamma_2, \dots, \gamma_{d_Z}$  can be identified.

Given a value  $\gamma$  satisfying condition (2.18), we consider the regression function

$$m(t; \gamma) = \mathbb{E}[Y \mid Z^\top \gamma = t], \quad t \in \mathbb{R}.$$

The single-index condition (2.17) then implies that there exists  $\gamma_0 \in \{1\} \times \tilde{\Gamma}$  such that

$$\mathbb{E}(Y \mid Z) = m(Z^\top \gamma_0; \gamma_0).$$

Since the conditional expectation  $\mathbb{E}(Y \mid Z)$  is also the closest squared integrable function to  $Y$ , in terms of the mean squared error, it is natural to assume that there exists a unique solution for

$$\min_{\gamma \in \{1\} \times \tilde{\Gamma}} \mathbb{E} [\{Y - m(Z^\top \gamma; \gamma)\}^2 \tau(Z)], \quad (2.19)$$

which is  $\gamma_0$ . Here,  $\tau(\cdot)$  is a weight and/or a *trimming function*. On the one hand, it can help improve efficiency and robustness for the estimators defined by minimizing a sample

counterpart of (2.19). On the other hand, it is a mostly theoretical device allowing to control erratic behavior of the nonparametric estimates of the function  $m(\cdot; \gamma)$  in areas with low index design. See also discussions in the following.

The condition (2.19) reveals a simple estimation strategy, similar to that in (2.15), based on least squares. The function  $m(\cdot; \gamma)$  can be now replaced by a nonparametric regression estimator, such as the kernel estimator, provided the variable  $Z^\top \gamma$  admits a density. Here, we consider local linear smoothing, see Fan and Gijbels (1996). More precisely, for each  $\gamma \in \{1\} \times \tilde{\Gamma}$ , we define

$$\hat{m}_h(t; \gamma) = \arg \min_{a, b \in \mathbb{R}^2} \sum_{k=1}^n \{Y_k - a - b(Z_k^\top \gamma - t)\}^2 k\left(\frac{Z_k^\top \gamma - t}{h}\right), \quad (2.20)$$

where  $k(\cdot)$  is a second order symmetric kernel and  $h = h_n$  is the bandwidth which decreases with  $n$ . The index  $\gamma$  is then estimated by least squares as

$$\hat{\gamma} = \arg \min_{\gamma \in \{1\} \times \tilde{\Gamma}} \sum_{i=1}^n \{Y_i - \hat{m}_h(Z_i^\top \gamma; \gamma)\}^2 \tau_n(Z_i). \quad (2.21)$$

Finally, for  $z$  in the support of  $Z$ , the regression function value  $\mathbb{E}(Y \mid Z = z)$  is estimated by  $\hat{m}(z^\top \hat{\gamma}; \hat{\gamma})$ , obtained as in (2.20) with  $\gamma$  replaced by  $\hat{\gamma}$  and  $t = z^\top \hat{\gamma}$ . See, for example, Ichimura (1993), Härdle et al. (1993), Carroll et al. (1997), Delecroix et al. (2006). The quantity  $\tau_n(Z_i)$  is the finite sample version of the so-called *trimming function*, and has the role to avoid unstable behavior of the kernel estimator  $\hat{m}(\cdot; \gamma)$  in areas of low design, that is where the density of  $Z^\top \gamma$  is close to zero. The role of trimming is mainly theoretical and often ignored in practice. Several ideas of trimming have been proposed in the literature, such for instance and indicator function for the variable  $Z$  over a set where the densities of  $Z^\top \gamma$  are uniformly bounded away from zero for  $\gamma \in \{1\} \times \tilde{\Gamma}$ . See Härdle et al. (1993). The set increase to the whole support of  $Z$  as the sample size increases, but this requires quite lengthy technical arguments. See Ichimura (1993). Another trimming idea could be

$$\tau_n(Z_i) = \mathbb{1}_{\{\hat{f}(Z_i^\top \tilde{\gamma}; \tilde{\gamma}) \geq c\}}, \quad (2.22)$$

for some small constant  $c > 0$ , where  $\tilde{\gamma}$  is a consistent, *pilot estimator* of  $\gamma_0$  (obtained in a preliminary step) and  $\hat{f}(\cdot; \tilde{\gamma})$  is a nonparametric estimator of the density of  $Z_i^\top \tilde{\gamma}$  (such as, for example, the Parzen-Rosenblatt estimator). The trimming in (2.22) is a finite sample version of the trimming

$$\tau(Z) = \mathbb{1}_{\{f(Z^\top \gamma_0; \gamma_0) \geq c\}},$$

which is convenient for deriving the asymptotic results because it is a function of the true index. See Delecroix et al. (2006), see also Section 2.4 below.

The bandwidth choice in (2.21) is sensitive. The nonparametric estimator  $\hat{m}(\cdot; \gamma)$  is a kind of *nuisance* parameter for the scope of the estimation of  $\gamma$ . However, it can significantly affect the result. Härdle et al. (1993), see also Delecroix et al. (2006), propose to select

a bandwidth at the same time as the parameter estimate  $\hat{\gamma}$ . This is particularly easy to implement with a kernel estimator of the regression function, as it amounts to removing the constraint  $\gamma_1 = 1$  in the optimization (2.21). The optimal bandwidth is then the inverse of the absolute value of the first component of the optimum, and this optimal value can be used to rescale the other components of the optimum. The idea of simultaneous optimization with respect to restricted  $\gamma$  and  $h$  (which is thus equivalent to optimization with no constraints) can lead to a degenerate solution ( $h = 0$ ). Therefore, a leave-one-out version of the kernel estimator  $\hat{m}_h(\cdot; \gamma)$  should be used, which is achieved by simply imposing  $k(0) = 0$ .

Theoretical properties of  $\hat{\gamma}$  and  $\hat{m}_h(z^\top \hat{\gamma}; \hat{\gamma})$  are quite well known. Under standard regularity conditions on the regression function and the moments of the variables  $Z$  and  $Y$ , with i.i.d. data, it can be shown that  $\hat{\gamma}$  is consistent in probability and  $\sqrt{n}(\hat{\gamma} - \gamma)$  is asymptotically normal. Moreover, for any  $z$  in the support of  $Z$ , under suitable condition on the regularity of the regression and the rate of the bandwidth,

$$\sqrt{nh} \{ \hat{m}_h(z^\top \hat{\gamma}; \hat{\gamma}) - \mathbb{E}(Y \mid Z = z) \}, \quad (2.23)$$

is asymptotically normal, with the mean determined by the nonparametric bias term. In particular, when the regression  $m$  admits a Lipschitz continuous derivative of order two, the asymptotic normality is achieved with the data-driven choice of the bandwidth described above, which yields a bandwidth with the standard rate  $O_{\mathbb{P}}(n^{-1/5})$ . See Ichimura (1993), Härdle et al. (1993), Delecroix et al. (2006) for formal statements on the theoretical properties. It is worth noting that  $Z$  can have a discrete component, only  $Z^\top \gamma$  is required to have a density for all  $\gamma \in \{1\} \times \tilde{\Gamma}$ , a requirement which can be satisfied if the first component of  $Z$  is a continuous variable.

Note that while  $\hat{\gamma}$  can be obtained with an ‘optimal’ bandwidth rule, as for instance the one proposed by Härdle et al. (1993) and studied by Delecroix et al. (2006) which has the asymptotic rate  $O_P(n^{-1/5})$ , when estimating the location and scale functions with the purpose of estimating  $F_\varepsilon$ , we cannot use the bandwidth rate  $O_P(n^{-1/5})$ . The reason comes from the fact that the estimation of the df  $F_\varepsilon$  of the error term in (2.6) requires a negligible bias for the ‘nuisance’ parameters, that are the functions  $m(\cdot)$  and  $\sigma(\cdot)$ . The negligible bias condition rules out the rates of  $O_{\mathbb{P}}(n^{-1/5})$  for the bandwidth. This explains why, in the following we will impose a faster decrease rate bandwidth for estimating the single-index regression function given the estimate  $\hat{\gamma}$ . As a result, the bias term appearing in the limit distribution in (2.23) is asymptotically negligible.

Let us end this section with a simulation study illustrating the performance of the estimators discussed above in the context of the single-index model with homoscedastic errors. In this simulation experiment we take the mean function to be

$$m(t) = \{(t + 2)/2\}^3.$$

The data has been generated according to the methodology presented in Section 2.2.3, with  $\rho \in \{0.3, 0.5, 0.7\}$ . We also considered  $d_Z = 2$ ,  $\gamma_0 = (\gamma_{01}, \gamma_{02})^\top = (1, -0.4)^\top$  and we

computed performance metrics for the quantile level  $\tau \in \{0.01, 0.05, 0.1, 0.2\}$ , for sample sizes of 50, 150 and 250. In this setup, the only single-index parameter to be estimated is  $\gamma_{02}$ . We replicated this experiment 200 times.

Furthermore, let us note that, given the value  $\gamma$ , the optimization problem (2.20) is a weighted least squares problem with an explicit solution, depending on  $\gamma$ , while (2.21) is a nonlinear optimization problem based on the solution in (2.20). As a result, in practice the estimator of  $\gamma$  can be calculated iterating between (2.20) and (2.21). In our implementation, to iterate between (2.20) and (2.21) we use the L-BFGS-B algorithm available in the `optim` package<sup>6</sup> in R, with the default stopping rule.

Since the estimate of  $m(t; \gamma)$ , with a suitable  $\gamma$  value, is further used to estimate the error distribution function, the bandwidth for smoothing  $m(t; \gamma)$  for that purpose has to be sufficiently small in order to get a negligible bias. Following Neumeyer and Van Keilegom (2010), we then consider a bandwidth with the rate of decrease  $n^{-2/7}$  for smoothing  $m(t; \gamma)$ . This rate is also admissible for guaranteeing the  $\sqrt{n}$ -asymptotic normality for the estimator of  $\gamma$  obtained from the iteration between (2.20) and (2.21). In order to avoid handling several bandwidth rules, at each iteration step of the L-BFGS-B algorithm, we simply take in (2.20) the bandwidth  $h$  of the form

$$h = Cn^{-2/7}, \quad (2.24)$$

where  $C$  is a constant depending on the iteration. More precisely, at each iteration step, the value of  $C$  is set equal to the empirical standard deviation of the  $Z_i^\top \tilde{\gamma}$ ,  $1 \leq i \leq n$ , where  $\tilde{\gamma}$  is the value of the parameter obtained in the previous iteration.

The results for the estimation of  $\gamma_2$  are presented in Figure 2.1 under the form of QQ-plots. One can notice that  $\sqrt{n}(\hat{\gamma}_2 - \gamma_{02})$  gets close to a Gaussian random variable, as the theory indicates. It is worth noting that by construction, the parameter  $\rho$  controls the signal-to-noise ratio: larger values of  $\rho$  yields a larger signal-to-noise ratio; see (2.12). One should therefore expect  $\rho$  to influence the accuracy of the estimator of  $\gamma$ . This is indeed what Figure 2.1 reveals: the slopes of the theoretical lines in the QQ-plots decrease as moving from left (small  $\rho$ ) to the right (large  $\rho$ ).

We next proceed to the estimation of the conditional quantiles of  $Y$  given  $Z = z$ . In this simplified, homoscedastic context, the conditional quantile of order  $\tau \in (0, 1)$  is

$$q_\tau(z) = \inf \left\{ y : F_\varepsilon(y - m(z^\top \gamma)) \geq \tau \right\},$$

where  $F_\varepsilon(\cdot)$  denotes the distribution function of the error term  $\varepsilon$  in (2.14). For this experiment, we consider  $\tau = 0.05$  and  $\tau = 0.1$ . The vector  $z$  belongs to the support of  $Z$ . The conditional quantile function is then estimated by

$$\hat{q}_\tau(z) = \inf \left\{ y : \hat{F}_\varepsilon(y - \hat{m}_h(z^\top \hat{\gamma}; \hat{\gamma})) \geq \tau \right\}, \quad (2.25)$$

---

<sup>6</sup>The documentation of the package used is available on this URL : <https://cran.r-project.org/web/packages/optim.functions/index.html>

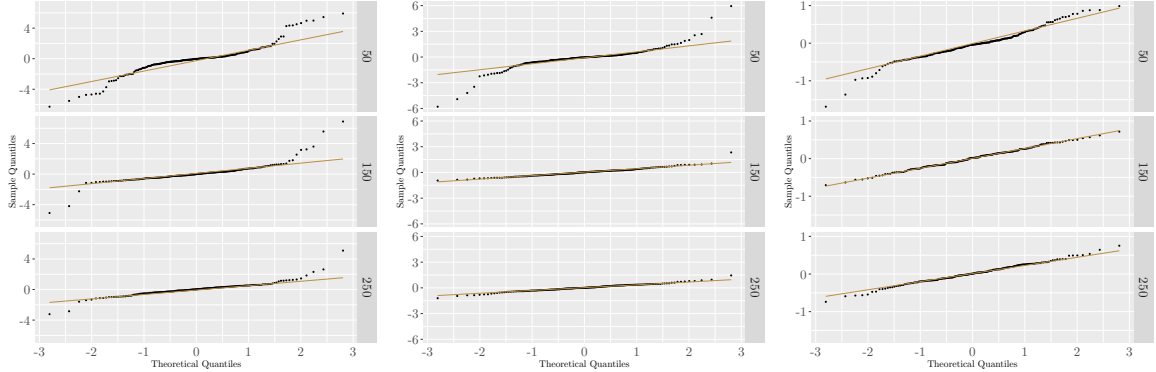


Figure 2.1: QQ-plot of  $\sqrt{n}(\hat{\gamma}_2 - \gamma_{02})$  for different signal-to-noise ratios:  $\rho = 0.3$  (left panels),  $\rho = 0.5$  (middle panels) and  $\rho = 0.7$  (right panels). Results obtained from 200 replications,  $\hat{\gamma}$  is obtained as in (2.21).

where  $\hat{\gamma} = (1, \hat{\gamma}_2)^\top$  is the estimator defined in (2.21), and  $\hat{m}_h$  is the local linear estimator obtained with the bandwidth defined as in (2.24). Here, to compute  $\hat{F}_\varepsilon$  we use the estimators  $\hat{F}_{\varepsilon, \text{emp}}$ ,  $\hat{F}_{\varepsilon, 1}$ ,  $\hat{F}_{\varepsilon, 2}$ ,  $\hat{F}_{\varepsilon, 3}$  defined in Section 2.2.2. For the smoothing based estimates of  $F_\varepsilon$ , we set  $h_\varepsilon = C_\varepsilon n^{-1/3}$ , with the constant  $C_\varepsilon$  equal to the empirical standard deviation of the residuals. This rule follows the optimal rate (in the sense of the mean squared error) derived by Azzalini (1981).

For assessing the behavior of our estimator of the conditional quantile, for  $\tau \in \{0.05, 0.1\}$ , we illustrate the distribution of the squared errors (SEs)  $|\hat{q}_\tau(z) - q_\tau(z)|^2$ , on a logarithmic scale, with three values  $z$ , denoted  $z^{(1)}, z^{(2)}, z^{(3)}$ , using 200 replications of each setup. The values  $z^{(j)} \in \mathbb{R}^2$ ,  $1 \leq j \leq 3$ , are selected such that each component of  $z^{(1)}, z^{(2)}$  and  $z^{(3)}$  is equal to the estimated 0.25, 0.5 and 0.75-th quantile of the component, respectively. That means,  $z^{(1)}$  denotes the 2-dimensional vector with the estimated 0.25-th quantiles of the components of  $Z$ , and so on.

In this simulation study of the homoscedastic single-index model, it is worth noting that the SEs at  $z^{(1)}$  are consistently lower than the SEs at  $z^{(3)}$ , with the components given by the estimated 0.75-th quantiles of the components of  $Z$ . This can be explained by the fact that the mean function  $m$  is increasing and takes negative values in the region corresponding to  $\gamma^\top z^{(1)}$  and positive values in  $\gamma^\top z^{(3)}$ . Moreover, our results in Figures 2.2 and 2.3 shows that the unconditional quantile is a poor approximation of the conditional quantile defined by the single-index location-scale model, regardless of the covariates value considered. The unconditional quantile is the quantity used in the regulatory model. The simulation experiment in this section provides a first evidence of the significant differences between the conditional and the unconditional quantile paradigms.

Let us finally investigate the asymptotic distribution of the estimates  $\hat{q}_\tau(z)$  estimated with (2.25) where the estimator of the distribution function is given by (2.10). Following the lines of the results derived by Neumeyer and Van Keilegom (2010), it can be proved that

## CHAPTER 2. A SEMIPARAMETRIC LOCATION-SCALE MODEL WITH APPLICATION TO CREDIT RISK

---

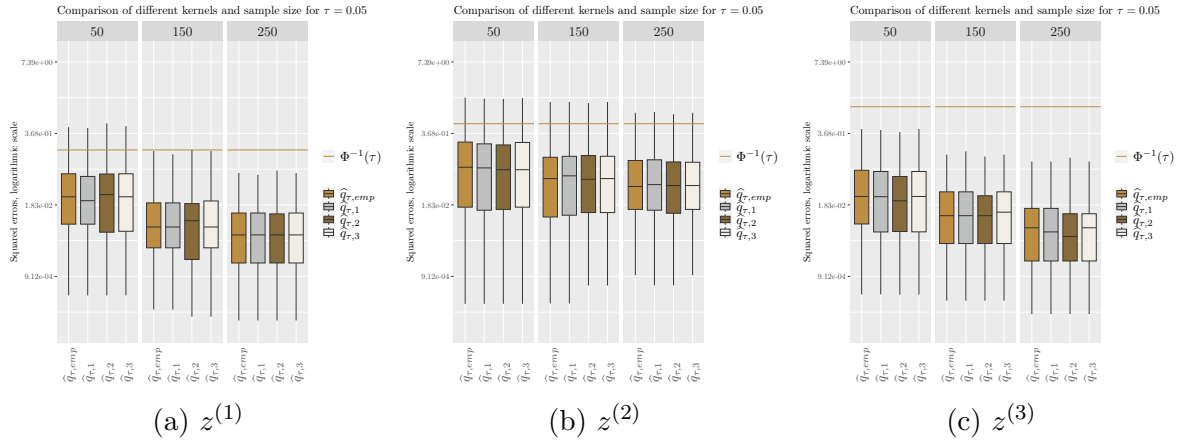


Figure 2.2: Box-plots of the squared errors of the estimated conditional quantile. Here  $\rho = 0.7$ , and 200 replications are performed. From the left panel (a) to the right,  $\ln|\widehat{q}_{0.05}(z^{(j)}) - q_{0.05}(z^{(j)})|^2$ , for  $j \in \{1, 2, 3\}$ , and for each type of estimator of  $F_\varepsilon$ , are presented. The horizontal line on the left side of each panel represents the error of the regulatory model, followed by the box-plots of the squared errors obtained with each of the four conditional quantile estimators corresponding to the four distribution function estimators presented in Section 2.2.2.

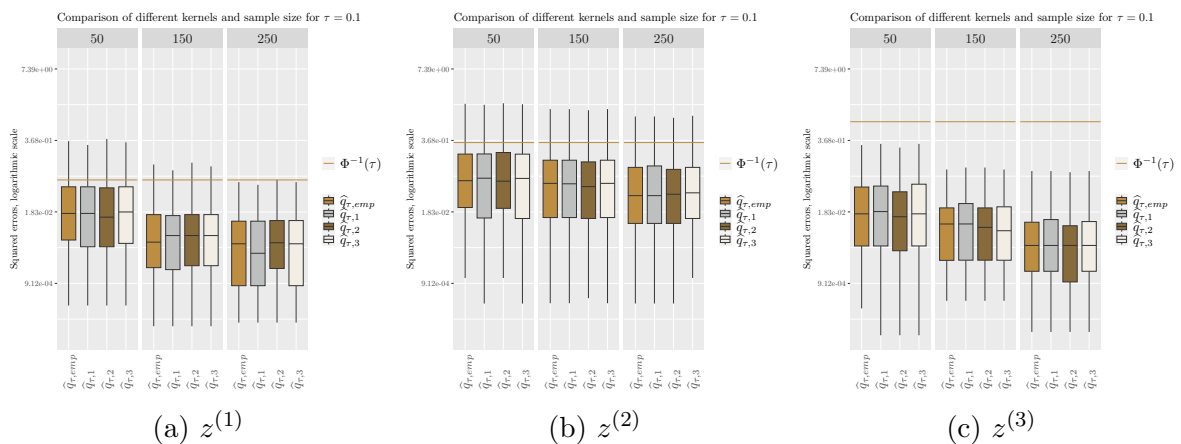


Figure 2.3: Box-plots of the squared errors of the estimated conditional quantile on a logarithmic scale. The same setup as in Figure 2.2, here with  $\tau = 0.1$ .

the process  $\sqrt{n}(\widehat{F}_\varepsilon(\cdot) - F_\varepsilon(\cdot))$  has a Gaussian limit, provided  $\widehat{F}_\varepsilon = \widehat{F}_{\varepsilon, \text{emp}}$  defined in (2.9). See also Section 2.4 below for details. Following the lines of the results proved by Racine and Van Keilegom (2020), the same can be shown for the smoothed estimator  $\widehat{F}_\varepsilon^\times = \widehat{F}_{\varepsilon, 1}$ . By the delta-method, one can get the asymptotic normality for the quantile process

$$\sqrt{n}(\widehat{q}_\varepsilon(\tau) - q_\varepsilon(\tau)), \quad \tau \in (0, 1),$$

with  $\widehat{q}_\varepsilon(\tau) = \inf\{t : \widehat{F}_\varepsilon(t) \geq \tau\}$ , for  $\widehat{F}_\varepsilon = \widehat{F}_{\varepsilon, \text{emp}}$  or  $\widehat{F}_\varepsilon = \widehat{F}_{\varepsilon, 1}$ . On the other hand, by definition we have the decomposition of the conditional quantile estimator as

$$\widehat{q}_\tau(z) - q_\tau(z) = \{\widehat{m}_h(z^\top \widehat{\gamma}; \widehat{\gamma}) - m(z^\top \gamma_0)\} + \{\widehat{q}_\varepsilon(\tau) - q_\varepsilon(\tau)\}.$$

Since the rate of  $\widehat{m}_h(z^\top \widehat{\gamma}; \widehat{\gamma}) - m(z^\top \gamma_0)$  is  $O_P(h^2 + 1/\sqrt{nh})$ , the rate of  $\widehat{q}_\tau(z) - q_\tau(z)$  is given by the nonparametric estimator of the location function. By standard nonparametric regression theory, choosing the bandwidth sequence such that  $nh^5 \rightarrow 0$ , we get

$$\sqrt{nh} \{\widehat{m}_h(z^\top \widehat{\gamma}; \widehat{\gamma}) - m(z^\top \gamma_0)\} \Rightarrow \mathcal{N}(0, R(k)V(z)),$$

with

$$V(z) = \frac{\text{Var}(Y \mid Z^\top \gamma_0 = z^\top \gamma_0)}{f(z^\top \gamma_0; \gamma_0)} \quad \text{and} \quad R(k) = \int k^2(u) du.$$

Since  $\widehat{q}_\varepsilon(\tau) - q_\varepsilon(\tau) = o_P(1/\sqrt{nh})$ , the  $\sqrt{nh}$ -asymptotic normality of the estimators  $\widehat{q}_\tau(z)$ , obtained with the empirical or smoothed estimators of  $F_\varepsilon$ , then follows.

Here, the asymptotic normality of  $\widehat{q}_\tau(z)$  is investigated by means of our simulation experiment. Figure 2.4 provides a visual representation of the distribution of  $\sqrt{nh}\{\widehat{q}_\tau(z) - q_\tau(z)\}$  for several sample sizes and values of  $\tau$ , with the bandwidths selected according to (2.24). We observe that the distribution of  $\widehat{q}_\tau(z)$  tends to a Gaussian distribution. The slopes of the QQ-lines are stable when  $n$  increases, confirming the rate of convergence  $O_P(1/\sqrt{nh})$  of  $\widehat{q}_\tau(z)$ ; here this rate is  $O_P(n^{-5/14})$  obtained with  $h \sim n^{-2/7}$ . Furthermore, we note that as the quantile level decreases, the variance of  $\sqrt{nh}\{\widehat{q}_\tau(z) - q_\tau(z)\}$  increases.

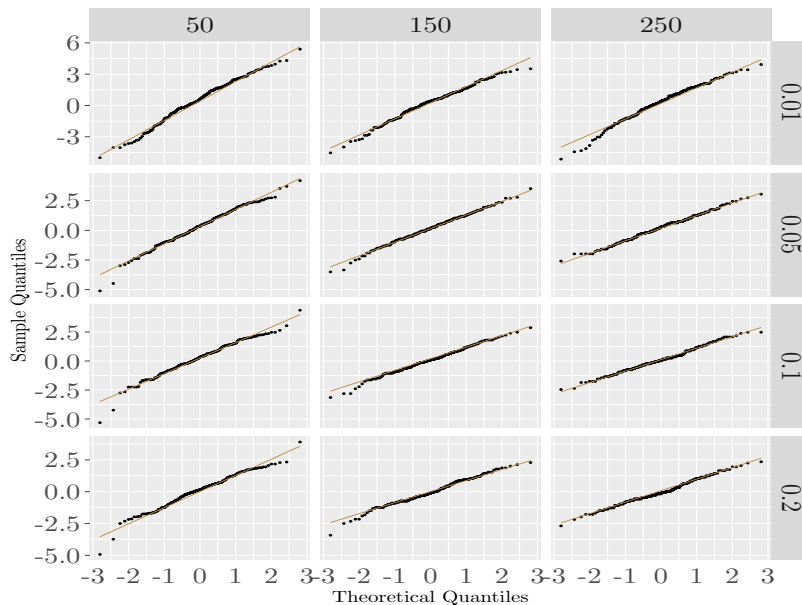


Figure 2.4: QQ-plots of  $\sqrt{nh}\{\hat{q}_\tau(z^{(1)}) - q_\tau(z^{(1)})\}$  in the homoscedastic semiparametric single-index regression setup. Here,  $\rho = 0.3$  is assumed, the distribution function estimator is based on  $\hat{F}_{\varepsilon,1}$  in (2.10), the bandwidth is  $h = Cn^{-2/7}$  according to (2.24).

### 2.3.2 Nonparametric location-scale regression

For the applications we keep in mind, there is a particular interest in the case of heteroscedasticity. Indeed in our use case the financial loss of a banking institution may exhibit high volatility depending on the economical and financial situation. We therefore aim extending the semiparametric model in Section 2.3.1 to the case where  $\sigma$  depends on covariates. As a transition towards this extension, let us recall some facts on the nonparametric location-scale model, as can be found in the literature.

Hence, we now consider the following nonparametric location-scale model

$$Y = m(X) + \sigma(X)\varepsilon, \quad \varepsilon \perp X, \quad X \in \mathbb{R}^{d_X},$$

which is a simplified variant of model (2.6) where  $X$  coincides with  $Z$ . This model was studied, for example, by Fan and Gijbels (1996), Akritas and Van Keilegom (2001), Neumeyer and Van Keilegom (2010) and Racine and Van Keilegom (2020). In the two latter references, the authors consider the case where  $X$  is a vector.

To estimate the conditional quantile function of order  $\tau \in (0, 1)$

$$q_\tau(x) = \inf_y \left\{ y : F_\varepsilon \left( \frac{y - m(x)}{\sigma(x)} \right) \geq \tau \right\},$$

we first have to estimate the functions  $m(\cdot)$  and  $\sigma(\cdot)$ . For this purpose we use again a

local linear smoother. For simplicity, we focus on the case of  $X \in \mathbb{R}$  (one covariate<sup>7</sup> only). More precisely, for  $x$  such that  $f_X(x) > 0$ , where  $f_X$  stands for the density of  $X$ , we consider

$$\widehat{m}(x) = \widehat{m}_h(x) = \arg \min_a \sum_{(a,b) \in \mathbb{R}^2} \sum_{i=1}^n \{Y_i - a - b(X_i - x)\}^2 k\left(\frac{X_i - x}{h}\right),$$

where  $k(\cdot)$  is a second order symmetric kernel and  $h$  is the bandwidth.

For the variance function, we notice that

$$\sigma^2(x) = \frac{1}{\text{Var}(\varepsilon)} \mathbb{E} [\{Y - m(X)\}^2 \mid X = x].$$

In general, in location-scale models  $\text{Var}(\varepsilon)$  is set equal to 1. In our context, the variance of  $Y$  is given<sup>8</sup> We then propose, say,

$$\text{Var}(\varepsilon) = \frac{\text{Var}(Y)}{2} =: \frac{\sigma_Y^2}{2}. \quad (2.26)$$

With such identification condition, we can estimate  $\sigma^2(x)$  using again a local linear type estimator, that is, using residuals instead of regression errors terms, for  $x$  such that  $f_X(x) > 0$ , we define

$$\widehat{\sigma}^2(x) = \widehat{\sigma}_h^2(x) = \times \arg \min_a \sum_{(a,b) \in \mathbb{R}^2} \sum_{i=1}^n \{[\text{Var}(\varepsilon)]^{-1} \widehat{r}_i^2 - a - b(X_i - x)\}^2 k\left(\frac{X_i - x}{h}\right)$$

with  $\widehat{r}_i = Y_i - \widehat{m}(X_i)$ .

See also Fan and Yao (1998). For notation simplicity, we used the same bandwidth for  $\widehat{m}(x)$  and  $\widehat{\sigma}^2(x)$ . However, for more refined results, different bandwidths are needed. See also the discussion in Section 2.2.1 on the necessity to reduce the bias of nonparametric estimator of the mean in order to get a better estimate of the variance function.

Finally, the distribution function  $F_\varepsilon$  can be estimated by any of the methods described in Section 2.2.2 (empirical and smoothed empirical with different types of kernels), using the standardized residuals

$$\widehat{\varepsilon}_i = \frac{Y_i - \widehat{m}(X_i)}{\widehat{\sigma}(X_i)}, \quad 1 \leq i \leq n.$$

Surprisingly, under some technical assumptions, it can be shown that the process  $\sqrt{n}(\widehat{F}_\varepsilon - F_\varepsilon)$  converges to a zero mean Gaussian process. See Akritas and Van Keilegom (2001) and

---

<sup>7</sup>Let us note that the simulation method discussed in Section 2.2.3 is not adapted to the case where  $m(\cdot)$  and  $\sigma(\cdot)$  depend on a 1-dimensional  $X$ , and  $Y \sim \mathcal{N}(0, 1)$ . We therefore omit simulations in the nonparametric location-scale setup.

<sup>8</sup>Recall that we are interested in the case where  $Y \sim \mathcal{N}(0, 1)$ .

Neumeyer and Van Keilegom (2010) for the case where  $\widehat{F}_\varepsilon$  is the empirical distribution of the residuals, and see Racine and Van Keilegom (2020) for the case of the smooth empirical distribution function with a symmetric kernel. Neumeyer and Van Keilegom (2010) and Racine and Van Keilegom (2020) allow  $X$  to be a vector.

### 2.3.3 Location-scale single-index modeling

In order to avoid the curse of dimensionality due to the nonparametric estimation of the multivariate location and scale functions  $m(\cdot)$  and  $\sigma(\cdot)$ , we here consider a separate semiparametric single-index modeling of these functions. The model (2.6) becomes

$$Y = m(Z^\top \gamma) + \sigma(X^\top \beta)\varepsilon, \quad \text{with } \varepsilon \perp (Z, X) \quad \text{and} \quad \mathbb{E}(\varepsilon) = 0, \quad (2.27)$$

where  $\gamma \in \mathbb{R}^{d_z}$  and  $\beta \in \mathbb{R}^{d_x}$  are unknown vectors and  $m(\cdot)$  and  $\sigma(\cdot)$  are now univariate functions to be estimated. This model was considered, for example, by Zhu et al. (2013). They propose semiparametric estimators using suitable estimating equations. However, one of the technical conditions, see their condition (C1), is infeasible. Yin et al. (2010) considered the mean function is zero, and use an alternative characterization of  $\beta$  which holds true in that context. By construction and the independence assumption in (2.27), we have

$$\mathbb{E}[Y \mid X, Z] = \mathbb{E}[Y \mid Z] = \mathbb{E}[Y \mid Z^\top \gamma], \quad (2.28)$$

and

$$\begin{aligned} \mathbb{E}\{[Y - \mathbb{E}[Y \mid X, Z]]^2 \mid X, Z\} &= \mathbb{E}\{[Y - m(Z^\top \gamma)]^2 \mid X, Z\} \\ &= \mathbb{E}\{[Y - m(Z^\top \gamma)]^2 \mid X\} = \mathbb{E}\{[Y - m(Z^\top \gamma)]^2 \mid X^\top \beta\}. \end{aligned} \quad (2.29)$$

Moreover,

$$\sigma^2(u) = \frac{1}{\text{Var}(\varepsilon)} \mathbb{E} \left[ \{Y - m(Z^\top \gamma)\}^2 \mid X^\top \beta = u \right].$$

For identification purposes, we impose

$$\gamma = (1, \tilde{\gamma}^\top)^\top \quad \text{with} \quad \tilde{\gamma} = (\gamma_2, \dots, \gamma_{d_z})^\top \in \tilde{\Gamma} \subset \mathbb{R}^{d_z-1}, \quad (2.30)$$

and

$$\beta = (1, \tilde{\beta}^\top)^\top \quad \text{with} \quad \tilde{\beta} = (\beta_2, \dots, \beta_{d_x})^\top \in \tilde{B} \subset \mathbb{R}^{d_x-1}, \quad (2.31)$$

Next, given the values  $\gamma \in \{1\} \times \tilde{\Gamma}$ ,  $\beta \in \{1\} \times \tilde{B}$ , we consider the regression functions

$$m(t; \gamma) = \mathbb{E}[Y \mid Z^\top \gamma = t], \quad t \in \mathbb{R},$$

and

$$\sigma^2(u; \gamma, \beta) = \frac{1}{\text{Var}(\varepsilon)} \mathbb{E}\{[Y - m(Z^\top \gamma; \gamma)]^2 \mid X^\top \beta = u\}, \quad u \in \mathbb{R}.$$

Then the single-index conditions (2.28) and (2.29) become

$$\mathbb{E}[Y \mid X, Z] = m(Z^\top \gamma; \gamma),$$

and

$$\mathbb{E}[\{Y - \mathbb{E}[Y \mid X, Z]\}^2 \mid X, Z] = \sigma^2(X^\top \beta; \gamma, \beta) \times \text{Var}(\varepsilon).$$

The functions  $m(\cdot; \gamma)$  and  $\sigma^2(\cdot; \gamma, \beta)$  are estimated using the local linear smoother. More precisely,  $m(\cdot; \gamma)$  is estimated by  $\widehat{m}_h(t; \gamma)$  defined in (2.20), for some second order symmetric kernel  $k(\cdot)$  and bandwidth  $h$ . With the convention (2.26) for  $\text{Var}(\varepsilon)$ , the variance function  $\sigma^2(\cdot; \gamma, \beta)$  is estimated by

$$\widehat{\sigma}_h^2(u; \beta, \gamma) = \arg \min_a \sum_{(a,b) \in \mathbb{R}^2} \sum_{i=1}^n \left[ \frac{2}{\sigma_Y^2} \{Y_i - \widehat{m}(Z_i^\top \gamma; \gamma)\}^2 - a - b(X_i^\top \beta - u) \right]^2 k\left(\frac{X_i^\top \beta - u}{h}\right). \quad (2.32)$$

For simplicity, we use the kernel and bandwidth notation from (2.20) in (2.32) too, but they can be different.

We recall in Section 2.3.1 that when the functions  $m(\cdot)$  and  $\sigma(\cdot)$  are given,  $\gamma$  and  $\beta$  can be easily estimated by least squares. Therefore, given the nonparametric estimators of the location and scale functions, we estimate the parameters  $\gamma$  and  $\beta$  by semiparametric least squares. More precisely, we define

$$\widehat{\gamma} = \arg \min_{\gamma \in \{1\} \times \bar{\Gamma}} \sum_{i=1}^n \{Y_i - \widehat{m}_h(Z_i^\top \gamma; \gamma)\}^2 \tau_n(Z_i), \quad (2.33)$$

and

$$\widehat{\beta} = \arg \min_{\beta \in \{1\} \times \widehat{B}} \sum_{i=1}^n \left[ (2/\sigma_Y^2) \{Y_i - \widehat{m}_h(Z_i^\top \widehat{\gamma}; \widehat{\gamma})\}^2 - \widehat{\sigma}_h^2(X_i^\top \beta; \beta, \widehat{\gamma}) \right]^2 \tau_n(Z_i) \zeta_n(X_i). \quad (2.34)$$

Here,  $\tau_n(\cdot)$  and  $\zeta_n(\cdot)$  are trimming functions avoiding observations in areas of low design. See also the discussion in Section 2.3.1.

Finally, the distribution function  $F_\varepsilon$  can be estimated by any of the methods described in Section 2.2.2 (empirical and smoothed empirical with different types of kernels), using the standardized residuals

$$\widehat{\varepsilon}_i = \frac{Y_i - \widehat{m}_h(Z_i^\top \widehat{\gamma}; \widehat{\gamma})}{\widehat{\sigma}_h(X_i^\top \widehat{\beta}; \widehat{\beta}, \widehat{\gamma})}, \quad 1 \leq i \leq n. \quad (2.35)$$

The bandwidth choice is a delicate matter. For notation simplicity, we use the same symbol  $h$  for the bandwidths in (2.33), (2.34) and (2.35). However, these should have different rates.

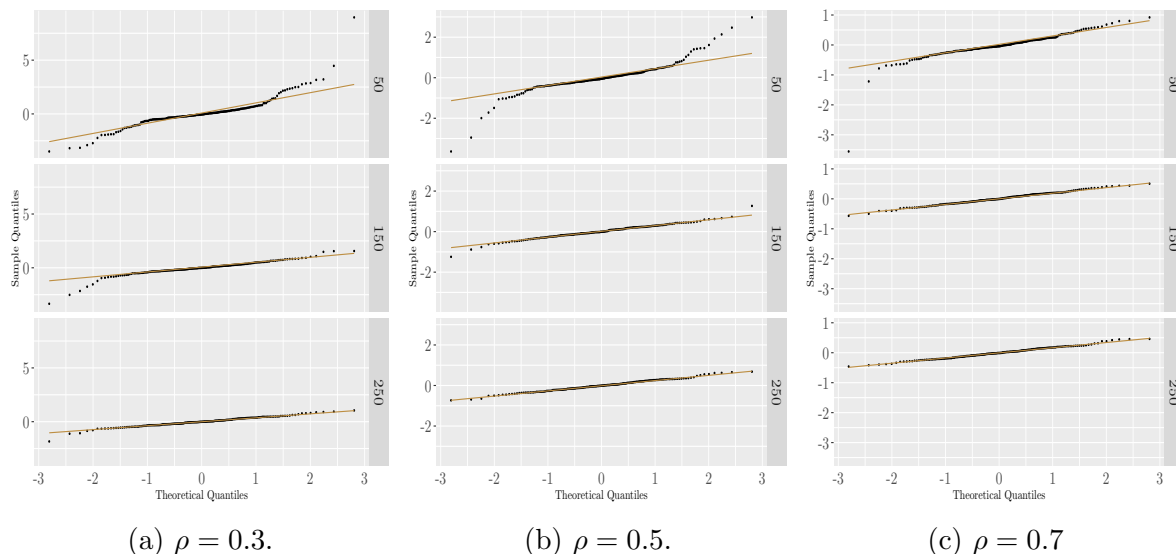


Figure 2.5: QQ-plots of  $\sqrt{n}(\hat{\gamma}_2 - \gamma_{02})$  for different signal-to-noise ratios in the location-scale single-index setup. The estimator  $\hat{\gamma}$  is obtained as in (2.33). Sample size  $n \in \{50, 150, 250\}$ .

We now present the results of a simulation experiment in the context of the semiparametric location-scale model (2.27). Similarly to Section 2.3.3, we consider  $\rho \in \{0.3, 0.5, 0.7\}$ ,  $\gamma_0 = (\gamma_{01}, \gamma_{02})^\top = (1, -0.4)^\top$  and  $\beta_0 = (\beta_{01}, \beta_{02})^\top = (1, -0.8)^\top$ . Moreover, the quantile level  $\tau$  is set equal to 0.01, 0.05, 0.1 and 0.2, the sample sizes are 50, 150 and 250, and each setup was replicated 200 times. We set

$$m(t) = \{(t + 2)/2\}^3 \quad \text{and} \quad \sigma(t) = \exp(t). \quad (2.36)$$

The standard deviation function<sup>9</sup> satisfies from the additional constraint of our simulation setup that requires this function to be a bijection from  $\mathbb{R}$  to  $(0, \infty)$ . The bandwidth rules are similar to the rules presented in Section 2.3.3. Additionally, the bandwidth for estimating the conditional standard deviation function  $\sigma(\cdot)$  is selected as in (2.24), the constant is now the empirical standard deviation of  $X_i^\top \hat{\beta}$ .

Figure 2.5 illustrates the estimating error of the single index parameter  $\gamma_{02}$  which decreases as the signal to noise ratio increases. Furthermore, Figure 2.5 illustrates that  $\sqrt{n}(\hat{\gamma}_{02} - \gamma_{02})$  has a Gaussian limit, as the theory indicates. In Figure 2.6 we present the results for  $\sqrt{n}(\hat{\beta}_{02} - \beta_{02})$ , which also has limit Gaussian behavior.

In this setup one still expects  $\rho$  to influence the accuracy of the estimator of  $\gamma$  and it is confirmed by Figure 2.5: smaller  $\rho$  leads to larger variance for  $\hat{\gamma}_{02}$ . Interestingly, in Figure 2.6 the accuracy seems to be equivalent for all signal-to-noise ratios, that is the value of  $\rho$  influences less the variance of  $\hat{\beta}_{02}$ . This may be because on the one hand, smaller  $\rho$  induces more precise estimates of the location function, and thus less reliable residuals.

<sup>9</sup>We could also have assumed a standard deviation function of the form  $\exp(at)$  with some  $a > 0$ , however it would only change the scale of the vector  $X$  by a factor  $a^{-1}$ ; see (2.13).

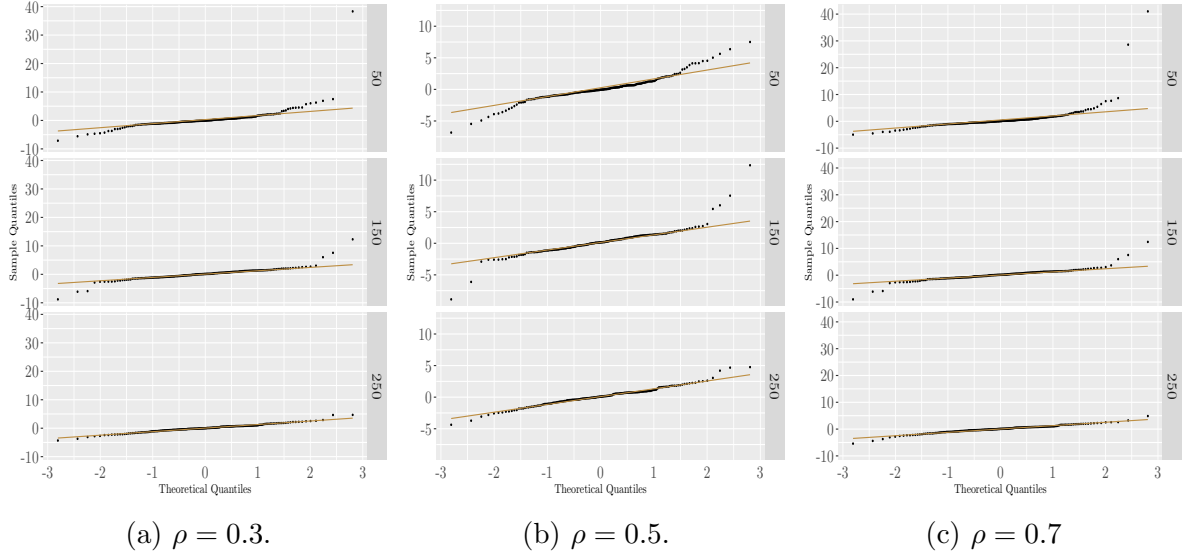


Figure 2.6: QQ-plots of  $\sqrt{n}(\hat{\beta}_2 - \beta_{02})$  for different signal-to-noise ratios in the location-scale single-index setup. The estimator  $\hat{\beta}$  is obtained as in (2.34). Sample size  $n \in \{50, 150, 250\}$ .

However, on the other hand, smaller  $\rho$  means larger  $\sqrt{1-\rho}$  which in turn means more information for estimating the conditional variance of the noise. The two antagonistic effects seem to cancel each other out in our experiment when the sample size increase.

In all setups, the distributions of  $\sqrt{n}(\hat{\beta}_{02} - \beta_{02})$  seems to converge to a Gaussian limit. However, large estimation errors are more frequent when estimating  $\beta$ , especially in the case of a low sample size. The reason could be that the initial estimates of the mean function introduce additional noise that significantly affects the estimation of  $\beta$  in small samples.

We next estimate the conditional quantiles of  $Y$  given  $W = (x, z)$ , for  $\tau \in \{0.05, 0.1, 0.2\}$ . In the heteroscedastic context, the conditional quantile of order  $\tau \in (0, 1)$ , given  $W = (x, z)$ , is obtained as

$$q_\tau(x, z) = \inf \left\{ y : F_\varepsilon \left( \frac{y - m(z^\top \gamma)}{\sigma(x^\top \beta)} \right) \geq \tau \right\}. \quad (2.37)$$

The conditional quantile is then estimated by

$$\hat{q}_\tau(x, z) = \inf \left\{ y : \hat{F}_\varepsilon \left( \frac{y - \hat{m}_h(z^\top \hat{\gamma}; \hat{\gamma})}{\hat{\sigma}_h(x^\top \hat{\beta}, \hat{\beta}; \hat{\gamma})} \right) \geq \tau \right\} = \hat{m}_h(z^\top \hat{\gamma}; \hat{\gamma}) + \hat{\sigma}_h(x^\top \hat{\beta}, \hat{\beta}; \hat{\gamma}) \hat{q}_\varepsilon(\tau),$$

with  $\hat{q}_\varepsilon(\tau) = \inf \{t : \hat{F}_\varepsilon(t) \geq \tau\}$ . For assessing the behavior of our conditional quantile estimator, we computed the logarithms of the squared errors for fixed covariates values: the values  $z^{(j)}, x^{(j)} \in \mathbb{R}^2$ ,  $1 \leq j \leq 3$ , are selected such that each of their component is equal to the estimated 0.25, 0.5 and 0.75-th quantile of the corresponding covariate component,

## CHAPTER 2. A SEMIPARAMETRIC LOCATION-SCALE MODEL WITH APPLICATION TO CREDIT RISK

respectively. The distribution function is estimated using the smoothing-based estimator  $\widehat{F}_{\varepsilon,1}$  in (2.10). The results presented in Figure 2.7 indicate that the single-index location-scale model provide quite accurate estimates, even with small to moderate sample sizes, and this regardless of the covariate values considered.

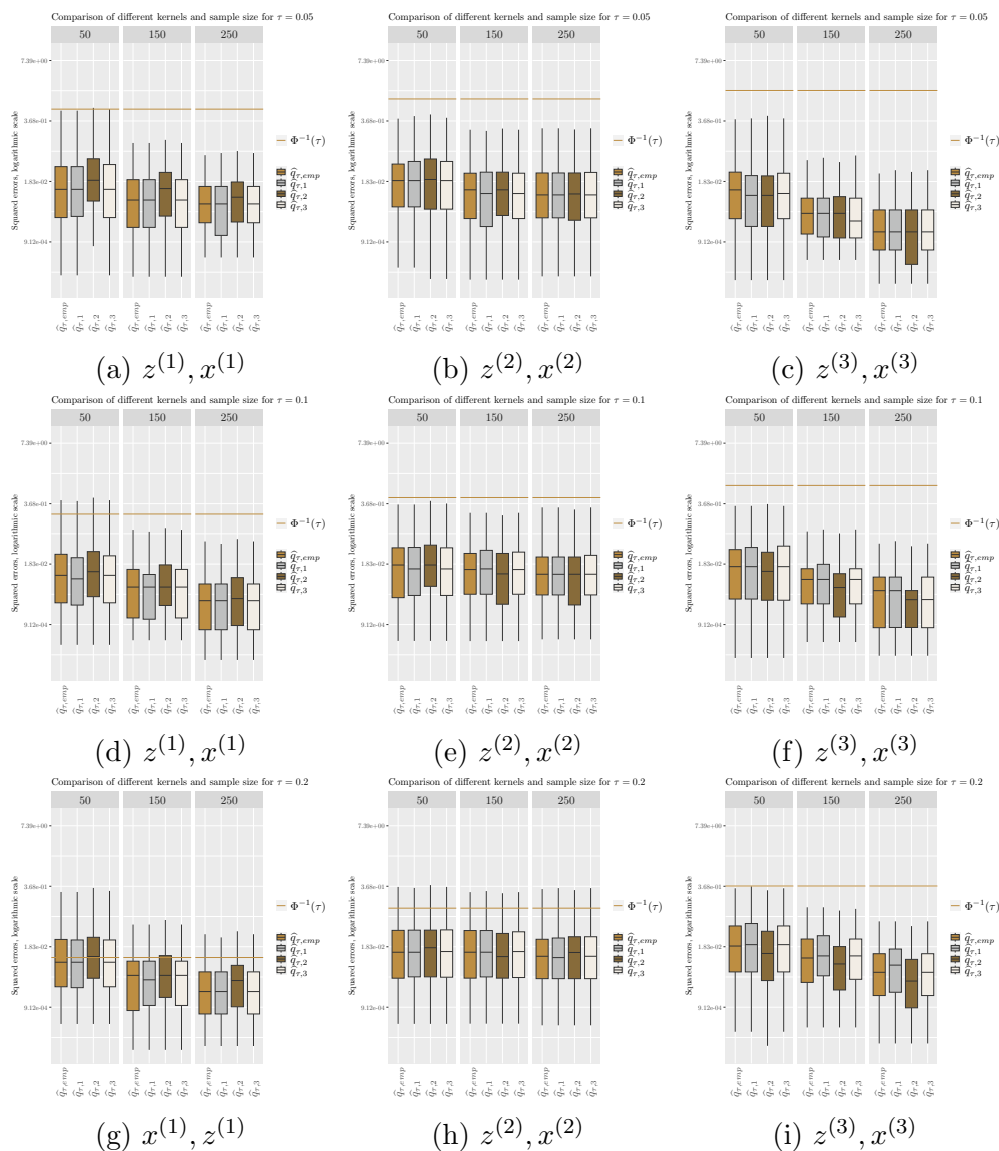


Figure 2.7: Estimated conditional quantiles in the semiparametric location-scale model (2.27). From the left to the right panels, the box-plots of  $\ln |\widehat{q}_\tau(x^{(j)}, z^{(j)}) - q_\tau(x^{(j)}, z^{(j)})|^2$ ,  $j \in \{1, 2, 3\}$ , are plotted. Each box-plot corresponds to one  $\widehat{F}_\varepsilon$  estimator presented in Section 2.2.2. The quantile levels are  $\tau = 0.05$  (top row of panels) middle,  $\tau = 0.1$  (middle row) and  $\tau = 0.2$  (bottom row), respectively. Here  $\rho = 0.7$ ,  $n \in \{50, 150, 250\}$ , and the results are obtained from 200 replications. For both  $\widehat{m}_h$  and  $\widehat{\sigma}_h$ , the bandwidth is of rate  $n^{-2/7}$  with the constant selected according to (2.24).

Next, Figure 2.8 provides a visual representation of  $\sqrt{nh}\{\widehat{q}_\tau(x^{(1)}, z^{(1)}) - q_\tau(x^{(1)}, z^{(1)})\}$ ,

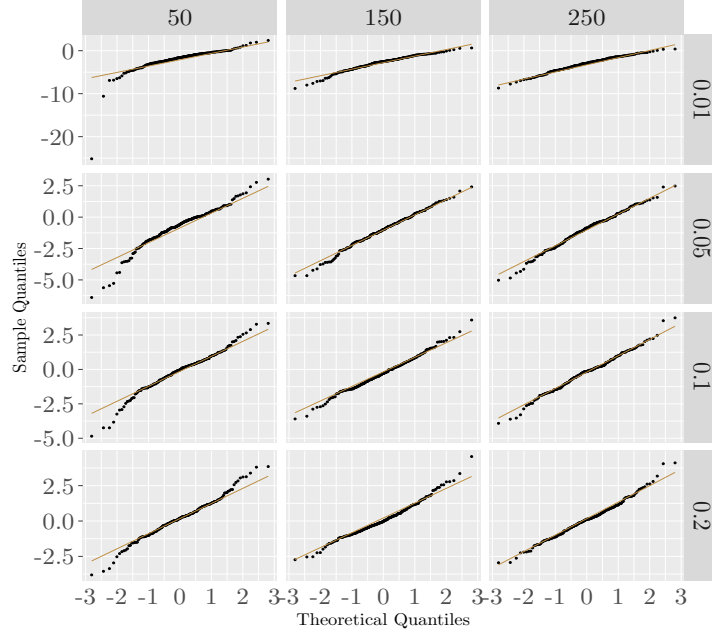


Figure 2.8: QQ-plots for  $\sqrt{nh}\{\hat{q}_\tau(x^{(1)}, z^{(1)}) - q_\tau(x^{(1)}, z^{(1)})\}$  in the location-scale single index setup. Here,  $\rho = 0.3$ ,  $n \in \{50, 150, 250\}$ ,  $\tau \in \{0.01, 0.05, 0.1, 0.2\}$  and the distribution function estimator is based on  $\hat{F}_{\varepsilon,1}$  in (2.10). The bandwidth  $h$  is of rate  $n^{-2/7}$  with the constant corresponding to  $\hat{\sigma}_h$  selected according to (2.24).

which seems to have Gaussian limit. However, when compared with the homoscedastic case, the Gaussian asymptotic approximation is significantly less accurate, especially in the case of the low sample size  $n = 50$  and when considering the  $\tau = 0.01$  quantile. However, as the size of the sample increases, the distribution of the estimator gets closer to a Gaussian distribution. Moreover, we observe that the estimated conditional quantile is biased and tends to underestimate the true value especially when considering  $\tau = 0.01$ . For the results in Figure 2.8, the multiplying factor  $\sqrt{nh}$  is taken with the bandwidth from  $\hat{\sigma}_h$ .

### 2.3.4 Heteroscedastic partially linear single-index model

In Section 2.3.3 we considered a semiparametric location-scale model using the single index paradigm for both the mean (location) and the variance (scale). The next step is to allow for more flexibility, beyond the assumption that a single linear combination of the covariates captures the full information on the location function, and another linear combination achieves the same property for the scale function. We now investigate the partially linear single-index model with a scale function modeled using the single-index assumption. More precisely, let  $Z \in \mathbb{R}^{d_Z}$ ,  $V \in \mathbb{R}^{d_V}$  and  $X \in \mathbb{R}^{d_X}$  be subvectors of  $W \in \mathbb{R}^{d_W}$ , with possibly common components (in which case  $d_Z + d_X + d_V > d_W$ ), and

$$Y = V^\top \omega + m(Z^\top \gamma) + \sigma(X^\top \beta) \varepsilon, \quad \text{with } \varepsilon \perp (Z, V, X) \quad \text{and} \quad \mathbb{E}(\varepsilon) = 0. \quad (2.38)$$

CHAPTER 2. A SEMIPARAMETRIC LOCATION-SCALE MODEL WITH APPLICATION TO CREDIT RISK

---

Here,  $\omega, \gamma$  and  $\beta$  are unknown parameter vectors, and  $m(\cdot), \sigma(\cdot)$  are unknown functions. Let  $\omega_0, \gamma_0$  and  $\beta_0$  denote the true values of the parameters. The case of constant  $\sigma(\cdot)$  was the studied by Carroll et al. (1997), as an extension of the partially linear model introduced by Robinson (1988) where the function  $m(\cdot)$  depends on the full vector  $Z$ .

The parameters  $\gamma$  and  $\beta$  can be identified using the conditions (2.30) and (2.31). The parameter  $\omega$  can be taken in the interior of a set  $\Omega \subset \mathbb{R}^{d_V}$ . However, in this context an additional condition is necessary in order to identify the linear effect of  $V$  and the nonlinear effect of  $Z$ . To understand what would be necessary in this context, let us rewrite the equation of the model (2.38) under the form

$$Y - \mathbb{E}(Y \mid Z^\top \gamma) = \{V - \mathbb{E}(V \mid Z^\top \gamma)\}^\top \omega + \sigma(X^\top \beta) \varepsilon.$$

We deduce that a sufficient condition for identifying  $\omega \in \Omega$  is

$$\inf_{\gamma \in \{1\} \times \tilde{\Gamma}} \lambda_{\min}(\text{Var}(V - \mathbb{E}(V \mid Z^\top \gamma))) > 0. \quad (2.39)$$

Here, for a positive semi-definite matrix  $A$ ,  $\lambda_{\min}(A)$  denotes the smallest eigenvalue of  $A$ . Condition (2.39) implies

$$\text{Var}(\{V^\top \omega - \mathbb{E}(V^\top \omega \mid Z^\top \gamma)\}) > 0, \quad \forall \omega \neq \mathbf{0}, \gamma \in \{1\} \times \tilde{\Gamma}.$$

This means that  $V$  and  $Z$  may share some covariates, but the non redundant ones should not have a zero coefficient in the linear combinations.

The estimation of the parameters of the location function  $(v, z) \mapsto v^\top \omega + m(z^\top \gamma)$  in model (2.38) can follow the procedure of Carroll et al. (1997). For simplicity, we consider a least squares criterion. Given  $\omega$  and  $\gamma$ , for some second order symmetric kernel  $k(\cdot)$  and bandwidth  $h$ , let

$$\hat{m}_h(t; \gamma, \omega) = \underset{a}{\operatorname{argmax}} \sum_{i=1}^n [\{Y_i - V_i^\top \omega\} - a - b(Z_i^\top \gamma - t)]^2 k\left(\frac{Z_i^\top \gamma - t}{h}\right). \quad (2.40)$$

Then, with some suitable trimming function and given pilot estimates  $\tilde{\omega}$  and  $\tilde{\gamma}$ , the estimators of  $\omega$  and  $\gamma$  can be obtained by semiparametric least squares :

$$(\hat{\omega}, \hat{\gamma}) = \underset{\omega \in \Omega, \gamma \in \{1\} \times \tilde{\Gamma}}{\operatorname{argmax}} \sum_{i=1}^n \{Y_i - V_i^\top \omega - \hat{m}_h(Z_i^\top \gamma; \tilde{\gamma}, \tilde{\omega})\}^2 \tau_n(Z_i). \quad (2.41)$$

Given a stopping rule, the final estimators of  $\omega$  and  $\gamma$  are obtained by iterating between (2.40) and (2.41). To initialize the iterations, Carroll et al. (1997) propose to use a parametric fit, that is to choose some location function and proceed with nonlinear least squares to estimate the parameters.

The idea used in (2.34) and (2.32) can be easily adjusted to estimate the scale function  $\sigma(\cdot)$ . It suffices to replace  $\hat{m}_h(Z_i^\top \hat{\gamma}; \hat{\gamma})$  by  $V_i^\top \hat{\omega} + \hat{m}_h(Z_i^\top \hat{\gamma}; \hat{\gamma}, \hat{\omega})$  with  $\hat{m}_h(\cdot; \cdot, \cdot)$  defined

according to (2.40), to get the scale function estimator, say,  $\widehat{\sigma}_h(\cdot; \cdot, \widehat{\gamma}, \widehat{\omega})$ . Finally, the standardized residuals are obtained by an obvious adjustment of (2.35); see also (2.45).

Let us discuss on how to adapt the simulation method presented in Section 2.2.3 in order to generate synthetic data according to the model (2.38), such that  $Y \sim \mathcal{N}(0, 1)$ . With one-to-one functions  $m(\cdot)$  and  $\sigma(\cdot)$ , for some  $\rho \in (0, 1)$ , by Box-Mueller's method, we have

$$Y = \underbrace{\sqrt{\rho}\sqrt{-2\ln(U_1)}\sin(2\pi U_2)}_{V^\top\omega + m(Z^\top\gamma)} + \underbrace{\sqrt{1-\rho}\sqrt{-2\ln(U_1)}\cos(2\pi U_2)}_{\sigma(X^\top\beta)} \underbrace{\varepsilon}_{\varepsilon}. \quad (2.42)$$

With  $X = (X_1, \tilde{X})$  and  $Z = (Z_1, \tilde{Z})$ , we can draw copies of the vector  $(V^\top, \tilde{X}^\top, \tilde{Z}^\top)^\top$  from a centered multivariate Gaussian distribution with some covariance matrix  $\Sigma$ . The two remaining covariate components are determined by the following relationships :

$$\begin{aligned} Z_1 &= m^{-1}\left(\sqrt{\rho}\sqrt{-2\ln(U_1)}\sin(2\pi U_2) - V^\top\omega\right) - \tilde{Z}^\top\tilde{\gamma}, \\ X_1 &= \sigma^{-1}\left(\sqrt{1-\rho}\sqrt{-2\ln(U_1)}\right) - \tilde{X}^\top\tilde{\beta}. \end{aligned} \quad (2.43)$$

Now, let us illustrate this model by simulation studies, for the quantiles level 0.01, 0.05, 0.1 and 0.2 and the sample sizes 50, 150 and 250. . Similarly to Sections 2.3.3 and 2.3.4, we consider  $\rho \in \{0.3, 0.5, 0.7\}$ ,  $\gamma = (1, -0.4)$  and  $\beta = (1, -0.8)$ , and the location and scale functions as in Section 2.3.3. Moreover, we set  $\omega = 0.1$ . For simplicity, the variance  $\Sigma$  of the vector  $(V^\top, \tilde{X}^\top, \tilde{Z}^\top)^\top$  is taken to be the identity matrix. The bandwidth selection methods for smoothing  $m(\cdot)$  and  $\sigma(\cdot)$  are the same as in Section 2.3.3.

In Figure 2.9, the estimation of  $\gamma$  is more accurate as the signal-to-noise ratio increases which is similar to what was noticed in the previous sections, see Figures 2.1 and 2.5. Furthermore, a similar pattern for the estimator of  $\omega$  can be noted in Figure 2.10, the accuracy of the estimator of  $\widehat{\omega}$  improves with larger  $\rho$ . The asymptotic normality of  $\sqrt{n}(\widehat{\omega} - \omega_0)$  and  $\sqrt{n}(\widehat{\gamma} - \gamma_0)$  has been established by Carroll et al. (1997) or Xia and Härdle (2006), and this property appears to be experimentally confirmed in our simulation setup; see Figures 2.9 and 2.10.

Regarding the conditional quantiles, in the case of the model (2.38), the conditional quantile given  $W = (z, v, x)$  can be adapted from (2.37) to

$$q_\tau(z, v, x) = \inf \left\{ y : F_\varepsilon \left( \frac{y - v^\top\omega - m(z^\top\gamma)}{\sigma(x^\top\beta)} \right) \geq \tau \right\}. \quad (2.44)$$

Then, the estimated conditional quantile can be expressed as

$$\widehat{q}_\tau(z, v, x) = V_i^\top\widehat{\omega} + \widehat{m}_h(Z_i^\top\widehat{\gamma}; \widehat{\gamma}, \widehat{\omega}) + \widehat{\sigma}_h(x^\top\widehat{\beta}, \widehat{\beta}; \widehat{\gamma}, \widehat{\omega})\widehat{q}_\varepsilon(\tau) \quad (2.45)$$

We evaluated the performance of the conditional quantiles estimator with  $z^{(j)} \in \mathbb{R}^2$ ,  $j \in \{1, 2, 3\}$ , representing estimates of the quantile of  $Z$  of order 0.25, 0.50 and 0.75, respectively, and  $x^{(j)} \in \mathbb{R}^2$ ,  $v^{(j)} \in \mathbb{R}$ ,  $j \in \{1, 2, 3\}$ , defined similarly for  $X$  and  $V$ .

CHAPTER 2. A SEMIPARAMETRIC LOCATION-SCALE MODEL WITH APPLICATION TO CREDIT RISK

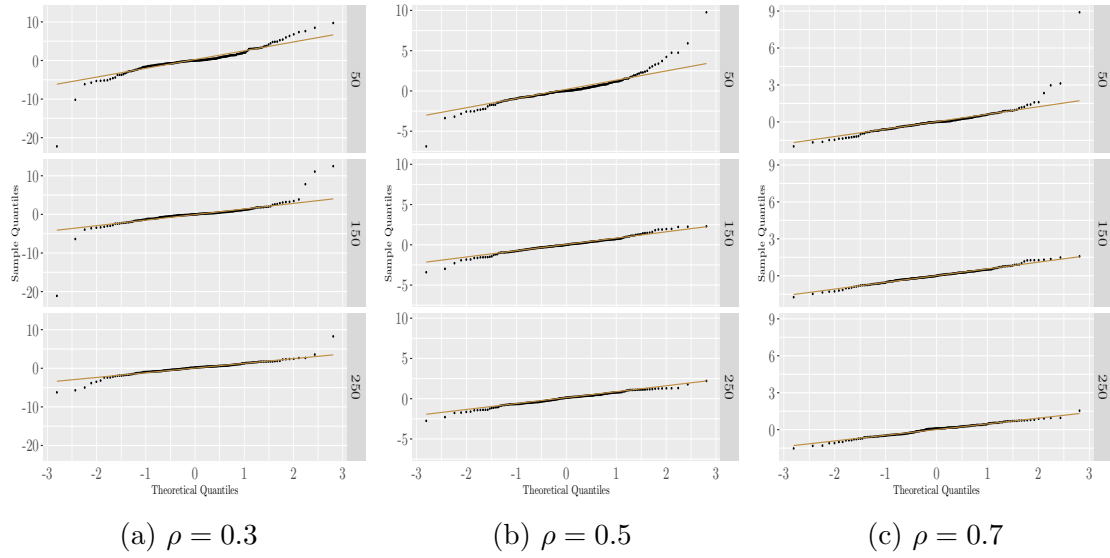


Figure 2.9: QQ-plots of  $\sqrt{n}(\hat{\gamma}_2 - \gamma_{02})$  in the heteroscedastic partially linear single index setup for different signal to noise ratios.

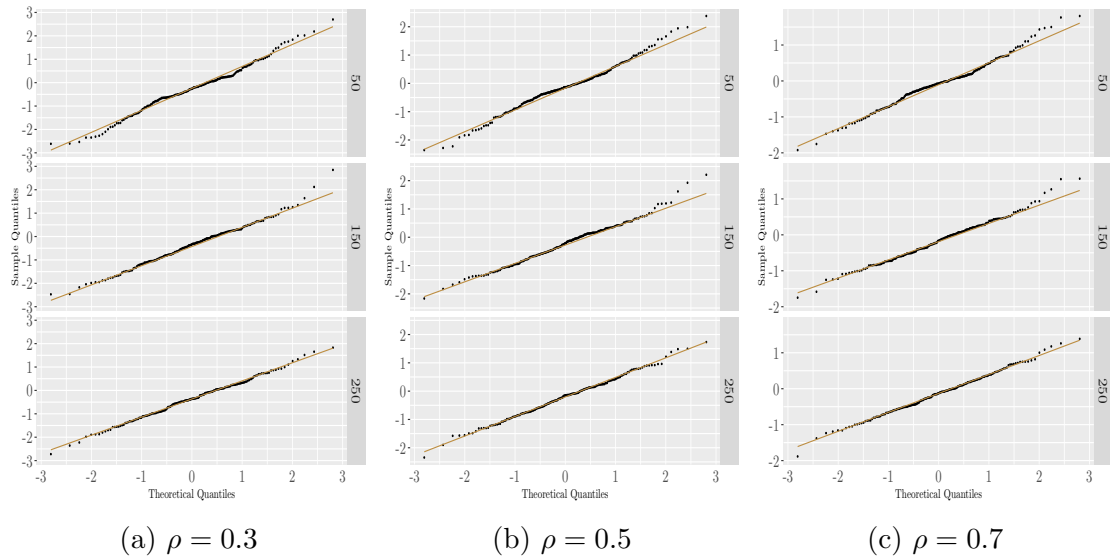


Figure 2.10: QQ-plots of  $\sqrt{n}(\hat{\omega} - \omega)$  in the heteroscedastic partially linear single index setup for different signal to noise ratios.

CHAPTER 2. A SEMIPARAMETRIC LOCATION-SCALE MODEL WITH APPLICATION TO CREDIT RISK

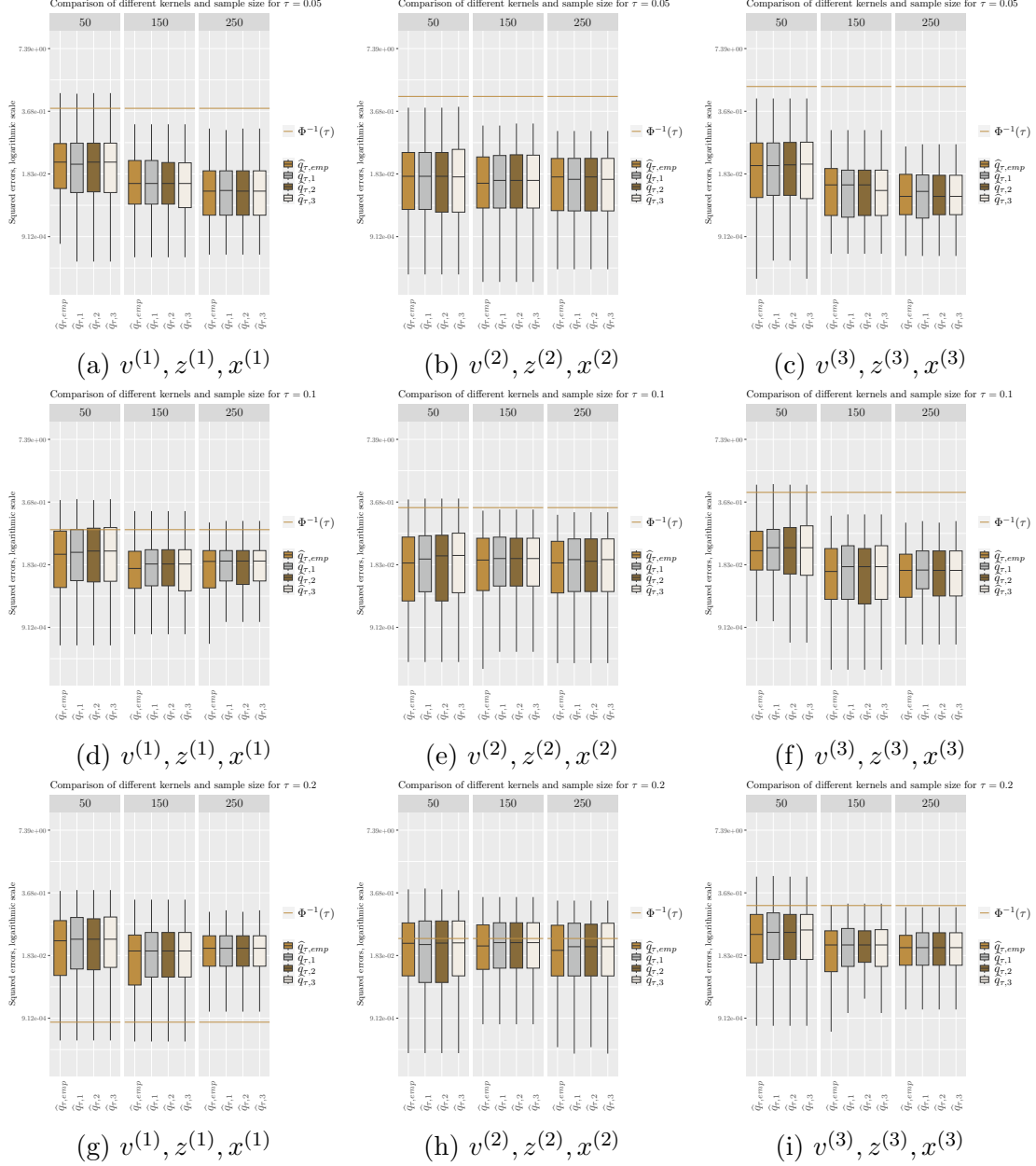


Figure 2.11: Estimated conditional quantiles in the heteroscedastic partially linear single-index model (2.38). From the left to the right panels, the box-plots of  $\ln |\hat{q}_\tau(z^{(j)}, v^{(j)}, x^{(j)}) - q_\tau(z^{(j)}, v^{(j)}, x^{(j)})|^2$ ,  $j \in \{1, 2, 3\}$ , are plotted. Each box-plot corresponds to one  $\hat{F}_\varepsilon$  estimator presented in Section 2.2.2. The quantile levels are  $\tau = 0.05$  (top row of panels) middle,  $\tau = 0.1$  (middle row) and  $\tau = 0.2$  (bottom row), respectively. Here  $\rho = 0.7$ ,  $n \in \{50, 150, 250\}$ , and the results are obtained from 200 replications. For both  $\hat{m}_h$  and  $\hat{\sigma}_h$ , the bandwidth is of rate  $n^{-2/7}$  with the constant selected according to (2.24).

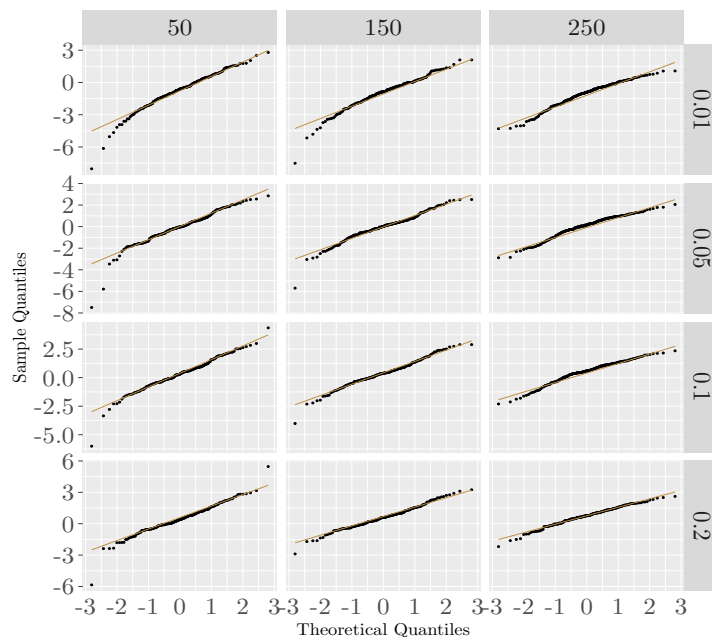


Figure 2.12: QQ-plots for  $\sqrt{nh}\{\hat{q}_\tau(z^{(1)}, v^{(1)}, x^{(1)}) - q_\tau(z^{(1)}, v^{(1)}, x^{(1)})\}$  in the model (2.38). Here,  $\rho = 0.3$ ,  $n \in \{50, 150, 250\}$ ,  $\tau \in \{0.01, 0.05, 0.1, 0.2\}$  and the distribution function estimator is based on  $\hat{F}_{\varepsilon,1}$  in (2.10). The bandwidth  $h$  is of rate  $n^{-2/7}$  with the constant corresponding to  $\hat{\sigma}_h$  selected according to (2.24).

We present the squared errors distribution using box-plots on a logarithmic scale, for  $\tau \in \{0.05, 0.1, 0.2\}$ . The error df  $F_\varepsilon$  is estimated using the smooth estimators presented in Section 2.2.2. The results in Figure 2.11 show good performance for the conditional quantile estimator in the heteroscedastic partially linear single-index model. The estimator of the error df has limited impact. This is explained by the fact that the rate of convergence of the conditional quantile estimator is given by the nonparametric estimators of the location and scale functions. Next, Figure 2.12 confirms the expected asymptotic normality property of  $\sqrt{nh}(\hat{q}_\tau(z^{(1)}, v^{(1)}, x^{(1)}) - q_\tau(z^{(1)}, v^{(1)}, x^{(1)}))$ , in the case where  $\rho = 0.3$ .

### 2.3.5 Semiparametric location-scale models with dependent data

The semiparametric models we discussed in the previous sections, and the associated estimation approaches, can be considered with a stationary time series  $\{(Y_i, W_i^\top)^\top\}$  (or equivalently  $\{(\varepsilon_i, W_i^\top)^\top\}$ ). Asymptotic results could be derived using limit theorems and uniform convergence results adapted to dependent data, for example imposing strong mixing conditions. See, for example, B. E. Hansen (2008). This possible extension is left for future investigation.

In the context of dependent data, we are particularly interested in the situation of an

autoregressive model for the common factor. We therefore want to include lagged values of  $Y$  as predictors. For simplicity, we focus on one lag only. More precisely, we consider that the times series  $\{Y_i\}$  is the stationary solution of the equation

$$Y_i = \mathbf{a}Y_{i-1} + m(Z_i^\top \gamma) + \sigma(X_i^\top \beta)\varepsilon_i, \quad 1 \leq i \leq n, \quad |\mathbf{a}| < 1, \quad \mathbb{E}(\varepsilon_i) = 0, \quad (2.46)$$

a variant of the equation (2.38) with  $V$  equal to the lag-1 value of  $Y$ . In this context,  $\mathbf{a}$  is the autoregressive coefficient to be estimated, and  $(\varepsilon_i, Z_i^\top, X_i^\top)^\top$ ,  $1 \leq i \leq n$ , are independent copies of  $(\varepsilon, Z^\top, X^\top)^\top$  with  $Z, X$  subvectors of  $W$  and  $\varepsilon \perp W$ .

Let us note that in the case of stationary series  $\{Y_i\}$  with  $\mathbb{E}(Y_i) = 0$ , we necessarily have  $\mathbb{E}[m(Z^\top \gamma)] = 0$ . This means that  $\{Y_i\}$  is the stationary solution of the AR(1) model

$$Y_i = \mathbf{a}Y_{i-1} + e_i, \quad \text{with} \quad e_i = m(Z_i^\top \gamma) + \sigma(X_i^\top \beta)\varepsilon_i, \quad i \geq 1,$$

and

$\{e_i\}$  is an independent white noise with  $\mathbb{E}(e_i^2) = \tau^2 := \mathbb{E}[m^2(Z^\top \gamma)] + \mathbb{E}[\sigma^2(X^\top \beta)] \text{Var}(\varepsilon)$ .

The model (2.46) can easily be extended to include several lags of the  $Y$  variable. More precisely, we can consider a zero-mean causal (or stable) AR( $p$ ) model<sup>10</sup> for some  $p \geq 1$ . The autoregressive coefficients can still be estimated by least squares, and the single-index functions are estimated as described in the previous sections using the residuals of the AR( $p$ ) fitted model. In the following and in our simulation experiments we simply take  $p = 1$ .

The estimator of  $\mathbf{a}$  in the AR(1) model is the standard least squares estimator, that is

$$\hat{\mathbf{a}} = \frac{\sum_{i=2}^n Y_i Y_{i-1}}{\sum_{i=2}^n Y_{i-1}^2} = a + \frac{(n-1)^{-1} \sum_{i=2}^n e_i Y_{i-1}}{(n-1)^{-1} \sum_{i=2}^n Y_{i-1}^2}.$$

With at hand  $\hat{\mathbf{a}}$ , we can estimate the single-index location and scale functions following the steps in Section 2.3.3, with the responses  $Y_i$  replaced by the AR(1) residuals  $Y_i - \hat{\mathbf{a}}Y_{i-1}$ .

The generation of synthetic time series can be adapted from the method presented in Section 2.3.4, starting from a standard Gaussian initial draw  $Y_0$  and modifying (2.42) as

$$Y_t = \mathbf{a}Y_{t-1} + \sqrt{1 - \mathbf{a}^2} \left( \sqrt{\rho} \sqrt{-2 \ln(U_1)} \sin(2\pi U_2) + \sqrt{1 - \rho} \sqrt{-2 \ln(U_1)} \cos(2\pi U_2) \right).$$

The vectors  $Z, X$  are generated as before, with the first components determined like in (2.43).

Let us investigate the estimators for the model (2.46) via some numerical experiments. We take  $\rho \in \{0.3, 0.5, 0.7\}$ . The functions  $m(\cdot)$  and  $\sigma(\cdot)$  and the index parameters  $\gamma, \beta$

---

<sup>10</sup>A causal, or stable, AR( $p$ ) model has all the roots of the characteristic polynomial outside the unit circle.

## CHAPTER 2. A SEMIPARAMETRIC LOCATION-SCALE MODEL WITH APPLICATION TO CREDIT RISK

are as in Section 2.3.3. See also (2.36) and the preceding paragraph. Moreover, we take  $a \in \{0.1, 0.2\}$ . If  $a$  is too large, the model will be overly influenced by past data, thereby diminishing the usefulness of the location-scale single-index modeling. Meanwhile, if  $a$  is too small, the practitioners may view it as unrealistic.

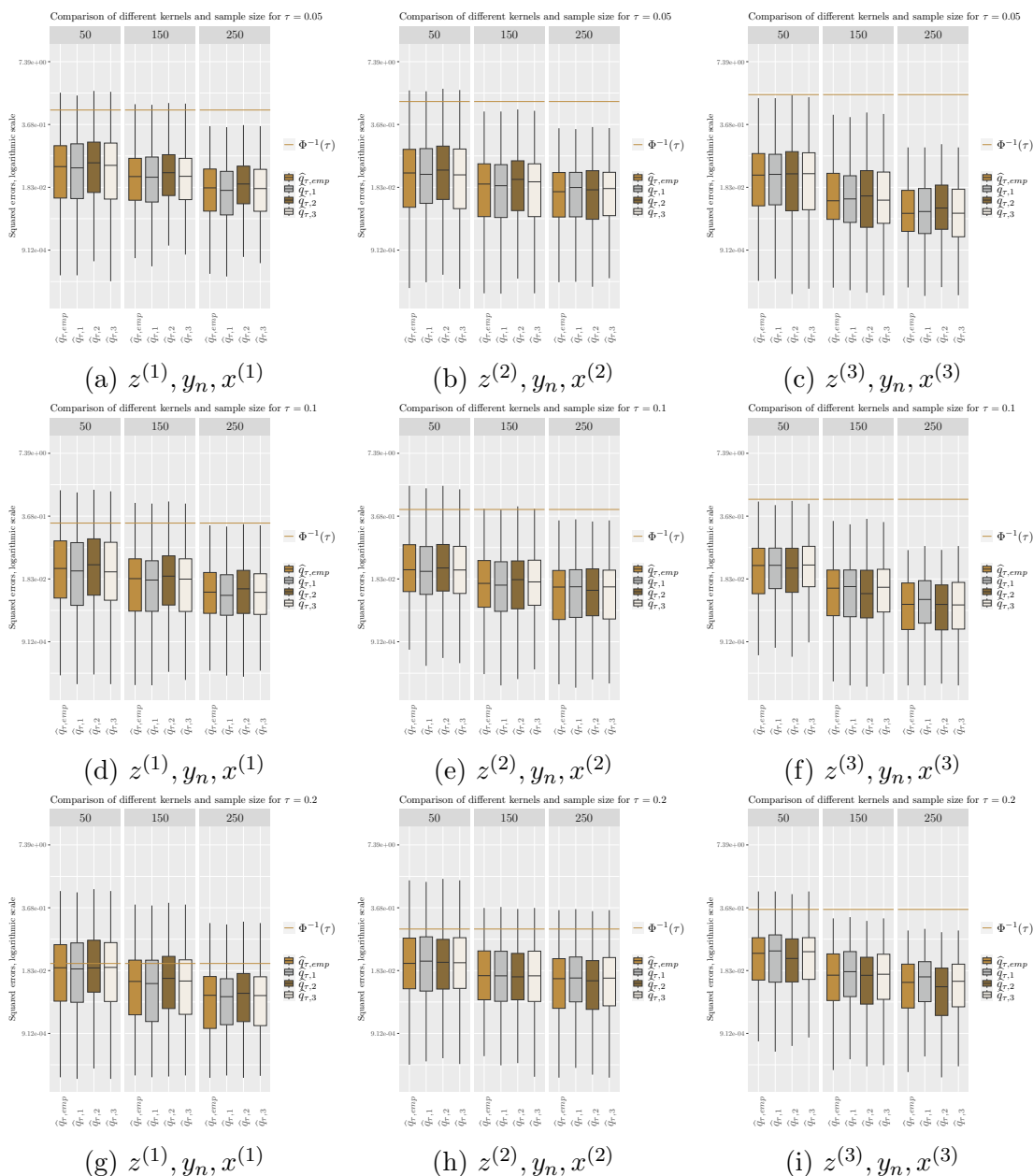


Figure 2.13: Estimated conditional quantiles in the autoregressive location-scale model (2.46). From the left to the right panels, the box-plots of  $\ln |\hat{q}_{\tau}(z^{(j)}, y_n, x^{(j)}) - q_{\tau}(z^{(j)}, y_n, x^{(j)})|^2$ ,  $j \in \{1, 2, 3\}$ , are plotted. They are obtained using the methodology presented Section 2.3.5. Here,  $\rho = 0.7$ . All the other elements of the setup are like in Figures 2.7 and 2.11.

The predicted values are still computed for  $z^{(1)}, z^{(2)}$  and  $z^{(3)}$  and  $x^{(1)}, x^{(2)}, x^{(3)}$  which

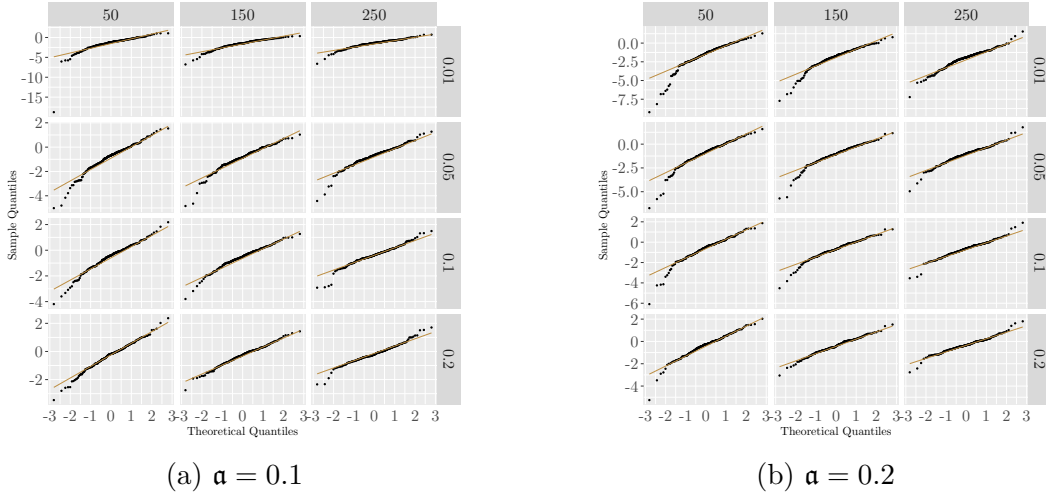


Figure 2.14: QQ-plots for  $\sqrt{nh}\{\hat{q}_\tau(z^{(1)}, y_n, x^{(1)}) - q_\tau(z^{(1)}, y_n, x^{(1)})\}$ . The estimators are obtained using the methods presented Section 2.3.5. Here,  $\rho = 0.5$ . All the other elements of the setup are like in Figures 2.8 and 2.12.

represents the estimated sample quantiles at level 0.25, 0.5 and 0.75 respectively for  $Z$  and  $X$ . Furthermore, we compute the predictions using the last observed common factor value, denoted  $y_n$ , mimicking what practitioners would do with real data. The conditional quantile given  $W = (z, y, x)$  can be directly adapted from (2.44). Figures 2.13 and 2.14 represent the logarithm of the squared errors and the distribution of  $\sqrt{nh}\{\hat{q}_\tau(z^{(j)}, y_n, x^{(j)}) - q_\tau(z^{(j)}, y_n, x^{(j)})\}$ , for different  $\tau$  and  $j \in \{1, 2, 3\}$ . Again, the asymptotic normality is confirmed by the simulation results.

It is worth noting the the performance of the conditional quantile estimator does not improve when the sample size increases in the case where the covariates are fixed to the values  $z^{(2)}$  and  $x^{(2)}$ . The the middle column in Figure 2.13. In the following we provide an explanation for this unexpected situation, in the context of the autoregressive model considered in this section.

The simulation method for the function value  $m(Z^\top \gamma_0)$  involves the choice of  $Z$  such that  $m(\gamma_0^\top Z) = \sqrt{1 - \alpha^2} \sqrt{\rho} \sqrt{-2 \ln(U_1)} \sin(2\pi U_2)$ . Figure 2.15a illustrates this function and the value taken at points  $-1.5$  and  $-2.5$ . In our simulation setup, the variable  $\sqrt{1 - \alpha^2} \sqrt{\rho} \sqrt{-2 \ln(U_1)} \sin(2\pi U_2)$  follows a Gaussian random variable with zero mean and variance  $(1 - \alpha^2)\rho$ . Therefore the probability to draw an observation in the segment  $[-2.5, -1.5]$  is the probability that a Gaussian random variable  $\mathcal{N}(0, (1 - \alpha^2)\rho)$  to take values in  $m([-2.5, -1.5]) = [-0.015625, 0.015625]$ . With  $\alpha = 0.1$  (resp.  $\alpha = 0.2$ ) this probability is approximately 0.01498 (resp. 0.01520). Therefore, even if, strictly speaking, the function  $m(\cdot)$  is invertible, in practice, with the sample sizes considered in our numerical experiments, very few observations will fall in this segment  $[-2.5, -1.5]$ . This is illustrated by Figure 2.15b which represents one simulated sample using the method presented in Section 2.2.3. It happens that the for the value  $z^{(2)}$  we considered,  $(1, z^{(2)})^\top \gamma_0 \approx -1.8$

CHAPTER 2. A SEMIPARAMETRIC LOCATION-SCALE MODEL WITH APPLICATION TO CREDIT RISK

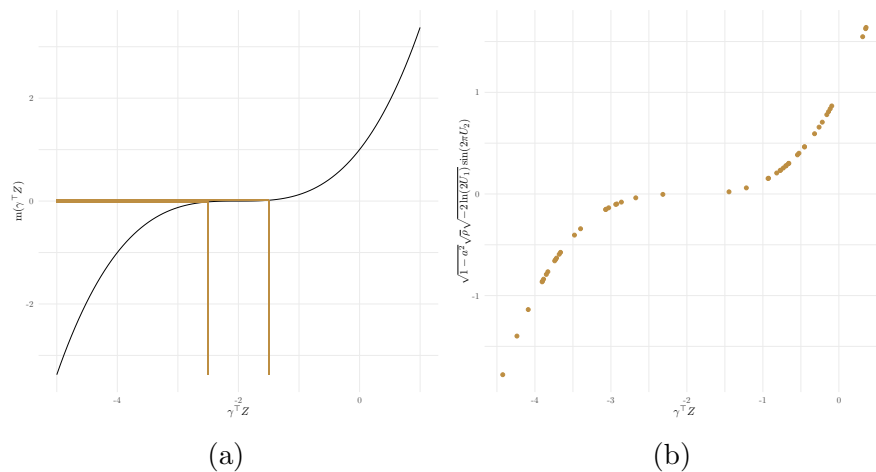


Figure 2.15: The left panel illustrates the function  $m(t) = \{(t + 2)/2\}^3$  and the values taken at point  $-2.5$  and  $-1.5$ . The right panel represents a simulated sample of size 50 from  $m(Z_i^\top \gamma_0)$ .

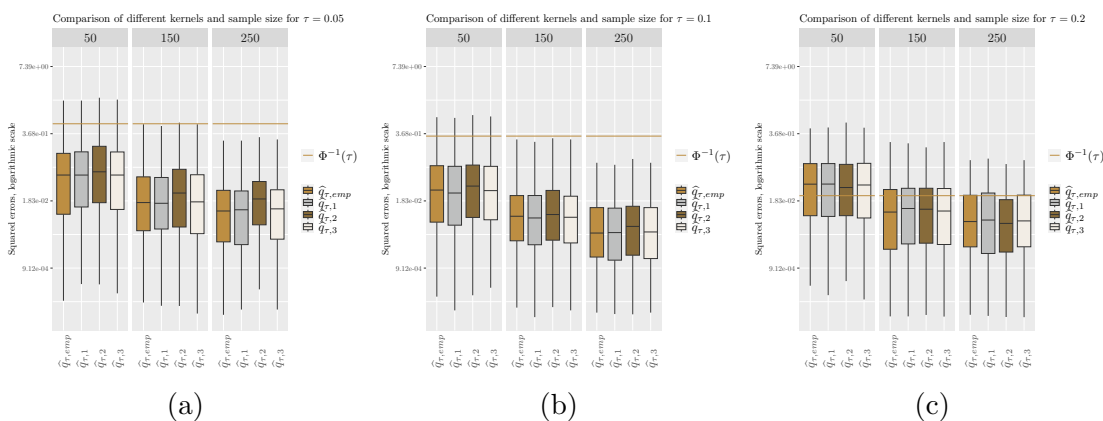


Figure 2.16: Changing the location function: the same box-plots as in the middle column of Figure 2.13, here  $m(x) = 3 + x + x^2/2 + x^3/10$ .

belongs to that segment, and it has very few or no neighbors in the interval defined by the bandwidth used for kernel smoothing. This configuration problem explains the poor behavior of the squared errors on the middle column of Figure 2.13.

The problem arises thus because our choice for the function  $m(\cdot)$  has a null derivative at  $t = 2$ . The problem can be avoided by considering a function  $m(\cdot)$  with positive derivative, such for example  $m(t) = 3 + t + t^2/2 + t^3/10$ . The results obtained with this alternative function are presented in Figure 2.16, where can be noticed that the box-plots of squared error are decreasing.

## 2.4 Theoretical grounds

In the previous sections, we presented several location-scale models which can be used for building flexible conditional quantile estimates of the common factor  $Y$  given covariates, or given covariates and lagged values of  $Y$ . We also presented several approaches for the estimation of the distribution of the standardized errors. In this section, we provide theoretical grounds for these methods.

Let us first discuss two major aspects underlying the theoretical justification of the semi-parametric single-index estimators we presented above. On the one hand, the semiparametric estimators proposed in the previous sections are obtained by the optimization of a criterion function depending on the parameter of interest. These optimization criteria, here derived by the least squares approach, are sums of terms depending also on unknown functions, and these functions are estimated by kernel smoothing. The kernel estimators are linear estimators and therefore, putting everything together, brings the optimization criteria under the form of multiple sums. In other words, by the construction of the criteria we will investigate below and the kernel estimators, we get  $U$ -statistics depending on the parameters we want to estimate. We then have to handle so-called  $U$ -processes indexed by the parameters of the model. Fortunately, under mild conditions, such processes fall into the class of  $U$ -processes indexed by *Euclidean families*<sup>11</sup> of functions with squared integrable envelopes, for which the uniform rates of convergence of the  $U$ -processes are available, see Sherman (1994b). (See also Nolan & Pollard, 1987, Lemma 22).

On the other hand, under mild assumption, using the Taylor expansion, the optimization criteria admit a quadratic approximation, from which the rate of convergence and the asymptotic distribution of the optima can be derived. The uniform rates of the  $U$ -processes, and the Law of Large Numbers, serve to show that the error of the quadratic approximation is negligible. Finally, the Central Limit Theorem, applied to sums of terms calculated with the true values of the parameters, provides the asymptotic normality of the semiparametric estimators.

Showing that the quadratic approximation error is negligible using the uniform rates of  $U$ -processes is now standard and appears in many published articles cited below. In the following, we therefore focus on deriving the expression for the quadratic approximation of the approximation criteria. In addition, we provide two lemmas of their own interest which formalize the quadratic approximation method for estimators obtained from general optimization criteria.

### 2.4.1 Technical lemmas

The following lemmas provide a general method for establishing the rate of convergence and the asymptotic distribution for the minimizer of a real-valued function. These lemmas

---

<sup>11</sup>In statistical textbooks, such sets of functions are also called classes of polynomial entropy number, or of polynomial complexity, or VC classes of functions.

are versions of results proposed by Sherman (1993), Sherman (1994b), Sherman (1994a).

The first lemma provides a general tool for proving the rates of convergence for consistent estimators.

**Lemma 2.4.1.** *Let  $S \subset \mathbb{R}^d$  and suppose  $\alpha_0$  minimizes a real-valued function  $G(\alpha)$  defined on  $S$ . Suppose further that a random counterpart,  $G_n(\alpha)$ , is minimized at a point  $\alpha_n$ , and  $\alpha_n$  converges in probability to  $\alpha_0$ . Let  $\{\delta_n\}$  be a sequence of non-negative real numbers such that  $\delta_n \rightarrow 0$  as  $n$  tends to infinity. Assume that :*

1. *there exists a neighborhood  $\mathcal{V}$  of  $\alpha_0$  and a constant  $\kappa > 0$  for which*

$$G(\alpha) - G(\alpha_0) \geq \kappa \|\alpha - \alpha_0\|^2, \quad \forall \alpha \in \mathcal{V}; \quad (2.47)$$

2. *uniformly over  $o_P(1)$  neighborhoods of  $\alpha_0$ ,*

$$G_n(\alpha) - G_n(\alpha_0) = G(\alpha) - G(\alpha_0) + U_n \|\alpha - \alpha_0\| + o_P(\|\alpha - \alpha_0\|^2) + O_P(\delta_n),$$

*with  $\beta > 0$  and  $\{U_n\}$  a sequence of random variables of order  $O_P(n^{-\beta})$ .*

*Then*

$$\|\alpha_n - \alpha_0\| = O_P(\max\{\delta_n^{1/2}, n^{-\beta}\}).$$

Let us point out that in the statement of the previous result, the sequence  $\{U_n\}$  can depend on  $\alpha$ , and then the order  $O_P(n^{-\beta})$  is supposed to hold uniformly over  $o_P(1)$  neighborhoods of  $\alpha_0$ .

*Proof of Lemma 2.4.1.* Since  $\alpha_n$  minimizes  $G_n(\alpha)$ ,

$$G_n(\alpha_n) - G_n(\alpha_0) \leq 0.$$

Since  $\alpha_n$  is consistent, it lies in a sequence of  $o_P(1)$  neighborhoods of  $\alpha_0$ . Moreover, with probability tending to 1 as  $n$  increases to infinity, we have  $\alpha_n \in \mathcal{V}$ . We then deduce that

$$\begin{aligned} 0 &\geq \kappa \|\alpha_n - \alpha_0\|^2 + U_n \|\alpha_n - \alpha_0\| + o_P(\|\alpha_n - \alpha_0\|^2) + O_P(\delta_n) \\ &\geq (\kappa/2) \|\alpha_n - \alpha_0\|^2 + U_n \|\alpha_n - \alpha_0\| + O_P(\delta_n), \end{aligned}$$

which<sup>12</sup> can be rewritten

$$(\kappa/2) \|\alpha_n - \alpha_0\|^2 + U_n \|\alpha_n - \alpha_0\| \leq O_P(\delta_n).$$

---

<sup>12</sup>Let us note that, by the definition of  $O_P(\cdot)$ , if a sequence of random variables, say  $\{Z_n\}$ , is  $O_P(\delta_n)$ , then  $\{-Z_n\}$  is also  $O_P(\delta_n)$ .

Adding  $U_n^2/2\kappa$  on both sides of the last inequality, we obtain

$$0 \leq (\kappa/2) (\|\alpha_n - \alpha_0\| + U_n/\kappa)^2 \leq O_P(\delta_n) + U_n^2/2\kappa = O_P(\delta_n + n^{-2\beta}).$$

Taking the square root both sides, we deduce

$$\|\alpha_n - \alpha_0\| \leq -U_n/\kappa + O_P(\delta_n^{1/2} + n^{-\beta}) = O_P(\delta_n^{1/2} + n^{-\beta}),$$

and the result follows. □

The second lemma provides sufficient conditions for deriving the asymptotic distribution of an estimator for which the rate of convergence has been established. Like for Lemma 2.4.1, the sequence  $\{W_n\}$  below can depend on  $\alpha$ , and then its order  $O_P(1)$  and the convergence in distribution should be understood uniform over  $O_P(1)$  neighborhoods of  $\alpha_0$ .

**Lemma 2.4.2.** *Suppose that  $\alpha_n$  defined in Lemma 2.4.1 satisfies  $\|\alpha_n - \alpha_0\| = O_P(n^{-\beta})$ , for some  $\beta > 0$ , and  $\alpha_0$  is an interior point of  $S \subset \mathbb{R}^d$ . Suppose also that uniformly over  $O_P(n^{-\beta})$  neighborhoods of  $\alpha_0$ ,*

$$G_n(\alpha) - G_n(\alpha_0) = \frac{1}{2}(\alpha - \alpha_0)^\top \Omega(\alpha - \alpha_0) + n^{-\beta}(\alpha - \alpha_0)^\top W_n + o_P(n^{-2\beta}),$$

where  $\Omega$  is a positive definite matrix and  $\{W_n\}$  is a sequence of random vectors of order  $O_P(1)$ . Then

$$n^\beta(\alpha_n - \alpha_0) = -\Omega^{-1}W_n + o_P(1).$$

In particular, if  $W_n$  converges in distribution to a non-degenerated multivariate Gaussian limit  $N_d(\mathbf{0}, \Delta)$  random vector, then<sup>13</sup>

$$n^\beta(\alpha_n - \alpha_0) \Rightarrow N_d(\mathbf{0}, \Omega^{-1}\Delta\Omega^{-1}).$$

*Proof of Lemma 2.4.2.* Because  $\alpha_0$  is an interior point of  $S$ ,  $\alpha_0 - n^{-\beta}\Omega^{-1}W_n$  lies in  $S$  with probability tending to one. When this happens, by the definition of  $\alpha_n$ , the identity (2.4.2) applied twice and little algebra, we get

$$\begin{aligned} 0 &\geq G_n(\alpha_n) - G_n(\alpha_0 - n^{-\beta}\Omega^{-1}W_n) \\ &= n^{-2\beta} \left\{ \frac{1}{2} [n^\beta(\alpha_n - \alpha_0) + \Omega^{-1}W_n]^\top \Omega [n^\beta(\alpha_n - \alpha_0) + \Omega^{-1}W_n] + o_P(1) \right\}. \end{aligned}$$

The  $o_P(1)$  can be assumed to absorb the cases where  $\alpha_0 - n^{-\beta}\Omega^{-1}W_n$  does not lie in  $S$ . Deduce then that  $n^\beta(\alpha_n - \alpha_0) + \Omega^{-1}W_n = o_P(1)$ . □

---

<sup>13</sup>Here,  $\Rightarrow$  denotes the convergence in distribution.

To explain how this type of technical results can help the asymptotic analysis, let us consider the classical setup of nonlinear least squares (NLS) in the framework of parametric models. In this context, we have

$$Y = m(X; \alpha) + e, \quad \mathbb{E}(e | X) = 0,$$

and the conditional variance of  $e$  given  $X$  is a bounded function. The function  $m$  is now known to belong to a given family of parametric functions, and there exists a unique parameter  $\alpha_0$  such that  $\mathbb{E}(Y | X) = m(X; \alpha_0)$ . Given an i.i.d. sample  $(Y_i, X_i^\top)^\top$ ,  $1 \leq i \leq n$ , the parameter  $\alpha_0$  can be estimated by (weighted) least squares, that is

$$\hat{\alpha}_{\text{NLS}} = \arg \min_{\alpha} G_n(\alpha) \quad \text{where} \quad G_n(\alpha) = \frac{1}{n} \sum_{i=1}^n \{Y_i - m(X_i; \alpha)\}^2 w(X_i).$$

The function  $G(\alpha)$  is simply the expectation of  $G_n(\alpha)$ . Here,  $w(\cdot) \geq 0$  is a given weight function. In the nonlinear, parametric regression context,  $w(\cdot)$  can be used to achieve efficiency, construct robustified versions of the estimator *etc.* Then, using  $\mathbb{E}(e | X) = 0$ , we have

$$G(\alpha) = \mathbb{E} [\{Y - m(X; \alpha)\}^2 w(X)] = \mathbb{E}(e^2 w(X)) + \mathbb{E} [\{m(X; \alpha) - m(X; \alpha_0)\}^2 w(X)], \quad (2.50)$$

and the unique minimizer of  $G(\cdot)$  is clearly  $\alpha_0$ , provided the identification condition

$$\{m(X; \alpha) - m(X; \alpha_0)\}^2 w(X) = 0 \quad \text{implies} \quad \alpha = \alpha_0,$$

holds true. Assuming the necessary regularity conditions for the regressions functions, and recalling that  $e = Y - m(X; \alpha_0)$  and  $\mathbb{E}(e | X) = 0$ , by Taylor expansion we can write

$$\begin{aligned} G(\alpha) &= G(\alpha_0) - 2(\alpha - \alpha_0)^\top \underbrace{\mathbb{E} [e \nabla_{\alpha} m(X; \alpha_0) w(X)]}_{=0} \\ &\quad + (\alpha - \alpha_0)^\top \mathbb{E} [\nabla_{\alpha} m(X; \alpha_0) \nabla_{\alpha} m(X; \alpha_0)^\top w(X)] (\alpha - \alpha_0) \\ &\quad + (\alpha - \alpha_0)^\top \underbrace{\mathbb{E} [e \nabla_{\alpha, \alpha}^2 m(X; \alpha_0) w(X)]}_{=0} (\alpha - \alpha_0) + o(\|\alpha - \alpha_0\|^2) \\ &= G(\alpha_0) + (\alpha - \alpha_0)^\top \mathbb{E} [\nabla_{\alpha} m(X; \alpha_0) \nabla_{\alpha} m(X; \alpha_0)^\top w(X)] (\alpha - \alpha_0) \\ &\quad + o(\|\alpha - \alpha_0\|^2). \end{aligned}$$

Here,  $\nabla_{\alpha}$  is gradient operator, that is column matrix-valued map of the first order derivatives with respect to  $\alpha$ , while  $\nabla_{\alpha, \alpha}^2$  is the Hessian operator, that is the squared matrix-valued operator of the second order derivatives with respect to  $\alpha$ .

The condition (2.47) is guaranteed by requiring that the smallest eigenvalue of the symmetric matrix  $\mathbb{E} [\nabla_{\alpha} m(X; \alpha_0) \nabla_{\alpha} m(X; \alpha_0)^\top w(X)]$ , which plays the role of  $2\Omega$  in Lemma 2.4.2, is positive. The  $\sqrt{n}$ -rate of convergence then follows from Lemma 2.4.1 applied

with  $\beta = 1/2$  and  $\delta_n = n^{-1}$ . With  $\beta = 1/2$ , we next apply Lemma 2.4.2. The random vector  $W_n$  is

$$W_n = -\frac{2}{\sqrt{n}} \sum_{i=1}^n e_i \nabla_{\alpha} m(X_i; \alpha_0) w(X_i),$$

and its asymptotic normality is guaranteed by the Central Limit Theorem, under mild assumptions. The matrix  $\Delta$  is then,

$$\Delta = 4 \text{Var} [e \nabla_{\alpha} m(X; \alpha_0) w(X)] = 4 \mathbb{E} [\sigma^2(X) \nabla_{\alpha} m(X; \alpha_0) \nabla_{\alpha} m(X; \alpha_0)^{\top} w(X)].$$

Finally, we can deduce that

$$\sqrt{n} (\hat{\alpha}_{\text{NLS}} - \alpha_0) \Rightarrow N_d(\mathbf{0}, V(\alpha_0)),$$

where

$$\begin{aligned} V(\alpha_0) &= \mathbb{E} [\nabla_{\alpha} m(X; \alpha_0) \nabla_{\alpha} m(X; \alpha_0)^{\top} w(X)]^{-1} \\ &\quad \times \mathbb{E} [\sigma^2(X) \nabla_{\alpha} m(X; \alpha_0) \nabla_{\alpha} m(X; \alpha_0)^{\top} w(X)] \\ &\quad \times \mathbb{E} [\nabla_{\alpha} m(X; \alpha_0) \nabla_{\alpha} m(X; \alpha_0)^{\top} w(X)]^{-1}. \end{aligned}$$

In the homoscedastic case, the asymptotic variance of the NLS estimator simplifies to

$$V(\alpha_0) = \sigma^2 \mathbb{E} [\nabla_{\alpha} m(X; \alpha_0) \nabla_{\alpha} m(X; \alpha_0)^{\top} w(X)]^{-1}.$$

The classical theory of the NLS estimators has been established by Jennrich (1969). See also van der Vaart (1998), Pollard and Radchenko (2006) for more refined results.

## 2.4.2 From nonlinear to semiparametric single-index regression

The asymptotic theory for single-index models can be derived along the lines of the nonlinear regression, with few major differences. With the notation from the Section 2.3.1, recall that, given a value  $\gamma$  satisfying the identification condition (2.18), we define

$$m(t; \gamma) = \mathbb{E}[Y \mid Z^{\top} \gamma = t], \quad t \in \mathbb{R}. \quad (2.51)$$

The single-index condition (2.17) then implies that there exists  $\gamma_0 \in \Gamma$  such that

$$\mathbb{E}(Y \mid Z) = m(Z^{\top} \gamma_0; \gamma_0), \quad \gamma_0 = (1, \tilde{\gamma}_0^{\top})^{\top} \in \Gamma = \{1\} \times \tilde{\Gamma}. \quad (2.52)$$

Thus, if the form of the function  $m(\cdot; \gamma)$  defined in (2.52) is given, as it is assumed in a parametric model, then, in the case of the single-index estimation, the parametric regression  $x \mapsto m(x; \alpha)$  is replaced by  $z \mapsto m(z^{\top} \gamma; \gamma)$ .

A key difference between parametric and semiparametric single-index models comes from the expression of the gradient of the regression function with respect to the index parameter. Indeed, in the single-index framework, where the link function is unknown, the index

parameter has two occurrences and the derivative is thus different. The following lemma provides the expression of this gradient for the function  $\gamma \mapsto m(z^\top \gamma_0; \gamma_0)$ , with respect to both occurrences of  $\gamma$ . See also Ichimura (1993), Härdle et al. (1993), Newey (1994), Dominitz and Sherman (2005).

**Lemma 2.4.3.** *Assume that  $\mathbb{E}(|Y|) < \infty$  and the support of  $Z$  is bounded. Moreover,  $\tilde{\gamma}_0$  is an interior point of the compact set  $\tilde{\Gamma} \subset \mathbb{R}^{d_Z-1}$ , and the function  $t \mapsto m(t; \gamma_0)$  defined in (2.51) is continuously differentiable, and let  $m'(t; \gamma_0)$  denote its derivative. Let  $\nabla_{\tilde{\gamma}}$  denote the vector (column matrix) of the first order derivatives with respect to the last  $d_Z - 1$  components of  $\gamma$ , while  $\tilde{Z}$  is the subvector of the last  $d_Z - 1$  components of  $Z$ . Then*

$$\nabla_{\tilde{\gamma}} m(Z^\top \gamma_0; \gamma_0) = m'(Z^\top \gamma_0; \gamma_0) \left[ \tilde{Z} - \mathbb{E}(\tilde{Z} \mid Z^\top \gamma_0) \right].$$

*Proof of Lemma 2.4.3.* The derivative of  $m(Z^\top \gamma; \gamma)$  with respect to  $\gamma$  at the point  $\gamma_0$  is given by the two partial derivatives of the map  $(t, \gamma) \mapsto m(t; \gamma)$  and the chain rule. The derivative with respect to  $t$  is  $m'(t; \gamma_0)$ . By the chain rule, its contribution to the derivative we want to get is  $m'(Z^\top \gamma_0; \gamma_0)Z$ . For the derivative with respect to the second occurrence of  $\gamma$  in  $m(Z^\top \gamma; \gamma)$ , let  $m(Z^\top \gamma_0) = \mathbb{E}[Y \mid Z] = \mathbb{E}[Y \mid Z^\top \gamma_0]$ , where the second equality is given by the single-index assumption. We can then write

$$m(t; \gamma) = \mathbb{E}(Y \mid Z^\top \gamma = t) = \mathbb{E}(m(Z^\top \gamma_0) \mid Z^\top \gamma = t) = \mathbb{E}(m(Z^\top \gamma_0 - Z^\top \gamma + t) \mid Z^\top \gamma = t).$$

On the other hand, by the regularity of the map<sup>14</sup>  $s \mapsto m(s)$ , and using the fact that  $Z$  is bounded, we have

$$m(Z^\top \gamma_0 - Z^\top \gamma + t) - m(t) = m'(t)Z^\top(\gamma - \gamma_0) + o(\|\alpha - \alpha_0\|).$$

Taking the derivative inside the expectation, the second contribution to the derivative we want to get is  $-m'(Z^\top \gamma_0; \gamma_0)\mathbb{E}(\tilde{Z} \mid Z^\top \gamma_0)$ . Gathering facts, and considering only the partial derivative with respect to the last  $d_Z - 1$  components, the result follows.  $\square$

To complete the analogy between the nonlinear parametric regression and the semiparametric single-index model, it remains to clarify the effect of not knowing the  $m(\cdot; \gamma)$ . If this function was given, one can immediately define a nonlinear estimator as in (2.19), that is solution of the problem

$$\min_{\gamma \in \{1\} \times \tilde{\Gamma}} \mathbb{E} \left[ \{Y - m(Z^\top \gamma; \gamma)\}^2 \tau(Z) \right]. \quad (2.53)$$

As mentioned above, here the role of the function  $\tau(\cdot)$  is somehow more complex than that of the weight in (2.50). In addition to the improvement of the asymptotic efficiency and/

---

<sup>14</sup>The regularity is guaranteed by the regularity of  $t \mapsto m(t; \gamma_0)$  and the fact that, by definition,  $m(\cdot) = m(\cdot; \gamma_0)$

or robustness, as mentioned after the definition (2.21), the function  $\tau(\cdot)$  has a technical role, that is to keep the nonparametric estimates of the nuisance functions under control in the areas of low design where such estimators behave erratically. For this purpose, it acts as a *trimming function*. For simplicity, here we will consider  $\tau(\cdot)$  fixed and given. It is widely recognized in the literature, and confirmed by our extensive numerical studies, that  $\tau(\cdot)$  has little influence in practice, and can even be omitted in practice.

The asymptotics of the estimator defined by the finite sample counterpart of (2.53) would follow from the general theory of the NLS, with the adjustment of the gradient of the regression as revealed by Lemma 2.4.3. The question would be whether the fact that the *nuisance function*  $m(\cdot; \gamma)$  is unknown matters? Can one hope to get the same asymptotic behavior<sup>15</sup> for the NLS if the true  $m(\cdot; \gamma)$  is replaced by some nonparametric estimator? In the context of the single-index models, the answer is positive. In fact, Newey (1994) constructs a general asymptotic theory for semiparametric estimators and shows in particular that the lowest asymptotic variance for a semiparametric estimator of  $\gamma$  one should expect is not affected by the fact that  $m(\cdot; \gamma)$  is unknown.

Despite this positive perspective, the remaining work is far from being obvious. One still has to build a suitable estimate of the nuisance function  $m(\cdot; \gamma)$  and to show that the error of estimating it is indeed negligible for the asymptotic variance.

### 2.4.3 Asymptotics for semiparametric single-index regression

Let us sketch the steps of the asymptotics for the single-index model (SIM) that we recall it in the following in a more general setting. Let  $Y \in \mathbb{R}$  and  $Z \in \mathbb{R}^{d_z}$ , and assume  $\mathbb{E}(Y^2) < \infty$ . Let us assume the model

$$Y = m(Z^\top \gamma) + \varepsilon, \quad \text{with} \quad \mathbb{E}(\varepsilon | Z) = 0, \quad \mathbb{E}(\varepsilon^2 | Z) \leq C < \infty,$$

where  $m(\cdot)$  is some function,

$$\gamma = (1, \tilde{\gamma}^\top)^\top \in \Gamma = \{1\} \times \tilde{\Gamma} \subset \mathbb{R}^{d_z},$$

and  $C$  is a constant. Let  $\gamma_0$  denote the true value of the parameter  $\gamma$ . Let  $(Y_i, Z_i^\top)^\top$ ,  $1 \leq i \leq n$ , be a random sample from  $(Y, Z^\top)^\top$ .

Let

$$G_n(\gamma) = G_n(\gamma; h) = \frac{1}{n} \sum_{i=1}^n \{Y_i - \hat{m}_h(Z_i^\top \gamma; \gamma)\}^2 \tau(Z_i), \quad \gamma \in \{1\} \times \tilde{\Gamma}, \quad (2.54)$$

where  $\hat{m}_h(\cdot; \gamma)$  is a nonparametric kernel estimator of the function  $m(\cdot; \gamma)$ , such as the Nadaraya-Watson estimator or the local linear estimator defined in (2.20). Following Delecroix et al. (2006), let us consider

$$\tau(Z_i) = \tau(Z_i; \gamma_0, c) = \mathbb{1}_{\{f(Z_i^\top \gamma_0; \gamma_0) \geq c > 0\}}, \quad \text{for some } c > 0. \quad (2.55)$$

---

<sup>15</sup>Here, the asymptotic behavior depends on the gradient, for the single-index model is the one given in Lemma 2.4.3.

CHAPTER 2. A SEMIPARAMETRIC LOCATION-SCALE MODEL WITH APPLICATION TO CREDIT RISK

---

Here,  $f(\cdot; \gamma_0)$  denotes the density of the index  $Z^\top \gamma_0$ , assumed to exist. This theoretical trimming, allows for milder conditions on the nonparametric estimation of the nuisance function  $m(\cdot; \gamma)$ . The full justification of this trimming requires a pilot consistent estimator  $\gamma_n$  and an approximated trimming  $\tau_n(Z_i) = \mathbb{1}_{\{\hat{f}(Z_i^\top \gamma_n; \gamma_n) \geq c > 0\}}$ , where  $\hat{f}(\cdot; \gamma)$  is a nonparametric estimator of the density of  $Z^\top \gamma$  (assumed to exist, at least for  $\gamma$  is a neighborhood of  $\gamma_0$ ). The pilot estimator can be obtained using a fixed trimming, not depending on the parameter  $\gamma$ . Under suitable conditions it can be shown that the estimator of  $\gamma_0$  defined with the approximated trimming  $\tau_n(Z_i)$  is asymptotically equivalent to the infeasible trimming  $\tau(Z_i)$  defined in (2.55). The details are provided in Delecroix et al. (2006) and are thus omitted. For simplicity, in the following we consider the theoretical trimming (2.55).

By the Taylor expansion,

$$\begin{aligned}
 G_n(\gamma) &= \underbrace{\frac{1}{n} \sum_{i=1}^n \{Y_i - \hat{m}_h(Z_i^\top \gamma_0; \gamma_0)\}^2 \tau(Z_i)}_{G_n(\gamma_0)} - (\tilde{\gamma} - \tilde{\gamma}_0)^\top \underbrace{\frac{2}{n} \sum_{i=1}^n \varepsilon_i \nabla_{\tilde{\gamma}} m(Z_i^\top \gamma_0; \gamma_0) \tau(Z_i)}_{2W_n/\sqrt{n}} \\
 &+ (\tilde{\gamma} - \tilde{\gamma}_0)^\top \underbrace{\left[ \frac{1}{n} \sum_{i=1}^n \nabla_{\tilde{\gamma}} m(Z_i^\top \gamma_0; \gamma_0) \nabla_{\tilde{\gamma}} m(Z_i^\top \gamma_0; \gamma_0)^\top \tau(Z_i) \right]}_{\Omega_n/2} (\tilde{\gamma} - \tilde{\gamma}_0) + R(\gamma; h). \quad (2.56)
 \end{aligned}$$

Recall that  $\nabla_{\tilde{\gamma}}$  denotes the vector (column matrix) of the first order derivatives with respect to the last  $d_Z - 1$  components of  $\gamma$ . The remainder  $R(\gamma; h)$  is composed of several parts. More precisely, we have

$$R(\gamma; h) = -2(\tilde{\gamma} - \tilde{\gamma}_0)^\top R_1(\gamma; h) + (\tilde{\gamma} - \tilde{\gamma}_0)^\top R_2(\gamma; h)(\tilde{\gamma} - \tilde{\gamma}_0) + R_3(\gamma; h),$$

where

$$\begin{aligned}
 R_1(\gamma; h) &= \underbrace{\frac{1}{n} \sum_{i=1}^n \{\hat{m}_h(Z_i^\top \gamma; \gamma) - m(Z_i^\top \gamma_0; \gamma_0)\} \nabla_{\tilde{\gamma}} m(Z_i^\top \gamma_0; \gamma_0) \tau(Z_i)}_{R_{11}(\gamma; h)} \\
 &+ \underbrace{\frac{1}{n} \sum_{i=1}^n \varepsilon_i \{\nabla_{\tilde{\gamma}} \hat{m}_h(Z_i^\top \gamma_0; \gamma_0) - \nabla_{\tilde{\gamma}} m(Z_i^\top \gamma_0; \gamma_0)\} \tau(Z_i)}_{R_{12}(\gamma; h)} \\
 &+ \underbrace{\frac{1}{n} \sum_{i=1}^n \{\hat{m}_h(Z_i^\top \gamma; \gamma) - m(Z_i^\top \gamma; \gamma)\} \{\nabla_{\tilde{\gamma}} \hat{m}_h(Z_i^\top \gamma_0; \gamma_0) - \nabla_{\tilde{\gamma}} m(Z_i^\top \gamma_0; \gamma_0)\} \tau(Z_i)}_{R_{13}(\gamma; h)},
 \end{aligned}$$

$$\begin{aligned}
 R_2(\gamma; h) &= \underbrace{\frac{1}{n} \sum_{i=1}^n \varepsilon_i \nabla_{\tilde{\gamma}, \tilde{\gamma}}^2 \widehat{m}_h(Z_i^\top \gamma_0; \gamma_0) \tau(Z_i)}_{R_{21}(\gamma; h)} \\
 &\quad - \underbrace{\frac{1}{n} \sum_{i=1}^n \{ \widehat{m}_h(Z_i^\top \gamma; \gamma) - m(Z_i^\top \gamma_0; \gamma_0) \} \nabla_{\tilde{\gamma}, \tilde{\gamma}}^2 \widehat{m}_h(Z_i^\top \gamma_0; \gamma_0) \tau(Z_i)}_{R_{22}(\gamma; h)} \\
 &\quad + \underbrace{\left[ \frac{1}{n} \sum_{i=1}^n \{ \nabla_{\tilde{\gamma}} \widehat{m}_h(Z_i^\top \gamma; \gamma) \nabla_{\tilde{\gamma}} \widehat{m}_h(Z_i^\top \gamma; \gamma)^\top - \nabla_{\tilde{\gamma}} m(Z_i^\top \gamma_0; \gamma_0) \nabla_{\tilde{\gamma}} m(Z_i^\top \gamma_0; \gamma_0)^\top \} \tau(Z_i) \right]}_{R_{23}(\gamma; h)},
 \end{aligned}$$

and  $R_3(\gamma; h)$  is given by the remainder of the second order Taylor expansion and, under some regularity conditions, is negligible.

Let  $\mathcal{H}_n$  be a suitable bandwidth range decreasing to zero. The asymptotic results<sup>16</sup> (rate of convergence and asymptotic distribution) for the semiparametric least squares estimator in SIM can be derived by using Lemmas 2.4.1 and 2.4.2, the uniform rates of convergence for the  $U$ -processes indexed by classes of functions with polynomial complexity (see Sherman, 1994b), and showing the following properties<sup>17</sup> :

$$\sup_{h \in \mathcal{H}_n} \|R_1(\gamma; h)\| \leq o_P(n^{-1/2}) + \|\gamma - \gamma_0\| \times o_P(1), \quad (2.57)$$

$$\sup_{h \in \mathcal{H}_n} \sup_{\gamma \in \Gamma} \|R_2(\gamma; h)\| = o_P(1),$$

and

$$\sup_{h \in \mathcal{H}_n} \|R_3(\gamma; h)\| \leq \|\gamma - \gamma_0\|^2 \times o_P(1),$$

with the  $o_P(n^{-1/2})$  and  $o_P(1)$  terms not depending on  $\gamma$ . See Delecroix et al. (2006). It is worth noting that Lemma 2.4.3 and the form of the trimming (2.55) guarantee that

$$\mathbb{E} [\nabla_{\tilde{\gamma}} m(Z^\top \gamma_0; \gamma_0) \tau(Z)] = \mathbb{E} \left[ m'(Z^\top \gamma_0; \gamma_0) \left\{ \tilde{Z} - \mathbb{E}(\tilde{Z} \mid Z^\top \gamma_0) \right\} \mathbb{1}_{\{f(Z_i^\top \gamma_0; \gamma_0) \geq c > 0\}} \right] = 0,$$

and this serves for the uniform rate in (2.57).

By Lemmas 2.4.1 and 2.4.2, under the assumptions guaranteeing the conditions of the lemmas, we get the following i.i.d., representation of the minimum of  $G_n(\gamma)$  defined in (2.54) :

$$\tilde{\gamma} = \{1 + o_P(1)\} \frac{1}{n} \sum_{i=1}^n \varepsilon_i \Omega(\gamma_0)^{-1} \nabla_{\tilde{\gamma}} m(Z_i^\top \gamma_0; \gamma_0) \tau(Z_i), \quad (2.58)$$

---

<sup>16</sup>The consistency of the semiparametric least squares estimator  $\hat{\gamma}$  is usually obtained from the uniform law of large numbers for  $U$ -statistics. For simplicity, the discussion on the consistency step is omitted. See for example Ichimura (1993) and Delecroix et al. (2006) for the details.

<sup>17</sup>Here, for a matrix  $A$ ,  $\|A\|$  denotes the Frobenius norm, which is the Euclidean norm when  $A$  is a column or row matrix. Note that for any  $\alpha \in \Gamma$ , by the definition of  $\Gamma$ , we have  $\|\gamma - \gamma_0\| = \|\tilde{\gamma} - \tilde{\gamma}_0\|$ .

where

$$\Omega(\gamma_0) = \mathbb{E} \left[ m'(Z^\top \gamma_0; \gamma_0)^2 \left\{ \tilde{Z} - \mathbb{E}(\tilde{Z} \mid Z^\top \gamma_0) \right\} \left\{ \tilde{Z} - \mathbb{E}(\tilde{Z} \mid Z^\top \gamma_0) \right\}^\top \tau(Z) \right]. \quad (2.59)$$

As a consequence of the i.i.d. representation, we get the convergence in distribution :

$$\sqrt{n} \left( \tilde{\gamma} - \tilde{\gamma}_0 \right) \Rightarrow \mathcal{N}_{d_Z-1} \left( \mathbf{0}, \Omega(\gamma_0)^{-1} \mathcal{V}(\gamma_0) \Omega(\gamma_0)^{-1} \right), \quad (2.60)$$

where

$$\mathcal{V}(\gamma_0) = \mathbb{E} \left[ \text{Var}(Y \mid Z) m'(Z^\top \gamma_0; \gamma_0)^2 \left\{ \tilde{Z} - \mathbb{E}(\tilde{Z} \mid Z^\top \gamma_0) \right\} \left\{ \tilde{Z} - \mathbb{E}(\tilde{Z} \mid Z^\top \gamma_0) \right\}^\top \tau(Z) \right].$$

In the case of homoscedastic errors, that is when  $\text{Var}(\varepsilon \mid Z) = \sigma^2$ , the asymptotic variance of  $\tilde{\gamma}$  simplifies to  $\sigma^2 \Omega_\gamma^{-1}$ .

By the arguments presented by Delecroix et al. (2006), the semiparametric single-index estimator defined in (2.21) with a feasible trimming tending to  $\tau(\cdot)$  is asymptotically equivalent to the  $\hat{\gamma}$  defined as the minimum of  $G_n(\gamma)$  defined in (2.54). As a consequence, the feasible estimator in (2.21) has the limit distribution derived in (2.60).

Let us now gather the facts and formally state results on the asymptotics of the estimator  $\hat{\gamma}$  and the resulting semiparametric regression functions estimator  $\hat{m}_h(z^\top \hat{\gamma}; \hat{\gamma})$ .

**Assumption 1.** *(The data generating process and the model)*

1. The observations  $(Y_i, Z_i^\top)^\top$ ,  $1 \leq i \leq n$ , are independent copies of  $(Y, Z^\top)^\top \in \mathbb{R} \times \mathbb{R}^{d_Z}$ . The vector  $Z$  is bounded,  $\text{Var}(Z)$  is positive definite, and  $\mathbb{E}(Y^4) < \infty$ . Moreover,  $\mathbb{E}(Y \mid Z)$  is a bounded function.
2. The parameter set  $\Gamma = \{1\} \times \tilde{\Gamma}$  is compact, with the elements  $\gamma = (1, \tilde{\gamma}^\top)^\top$ ,  $\tilde{\gamma} = (\gamma_2, \dots, \gamma_{d_Z})^\top \in \tilde{\Gamma} \subset \mathbb{R}^{d_Z-1}$ . There exists a unique value  $\gamma_0 = (1, \tilde{\gamma}_0^\top)^\top \in \Gamma$  such that  $\mathbb{E}(Y \mid Z) = \mathbb{E}(Y \mid Z^\top \gamma_0)$ . Moreover,  $\tilde{\gamma}_0$  is an interior point of  $\tilde{\Gamma}$ .
3. For each  $\gamma \in \Gamma$ , the random variable  $Z^\top \gamma$  admits a density  $t \mapsto f(t; \gamma)$ .
4. A constant  $c_0 > 0$  exists such that  $\mathbb{P}(f(Z^\top \gamma_0; \gamma_0) = c) = 0$ ,  $\forall 0 < c \leq c_0$ .
5. The  $(d_Z - 1) \times (d_Z - 1)$ -matrix  $\Omega(\gamma_0)$  defined in (2.59) is positive definite.

**Assumption 2.** *(Kernel smoothing)*

1. The kernel function  $k(\cdot)$  is differentiable, symmetric, positive and compactly supported. Moreover,  $k(\cdot)$  and the derivative  $k'(\cdot)$  are of bounded variation.
2. The bandwidth set is  $\mathcal{H}_n = [\underline{c}n^{-\nu}, \bar{c}n^{-\nu}]$ , for some exponent  $\nu \in (1/8, 1/4)$  and constants  $\underline{c}, \bar{c} > 0$ .

**Assumption 3.** (*Regularity conditions*)

1. The density of  $Z^\top \gamma$  is Hölder continuous as function of  $(t, \gamma)$  with  $\gamma \in \Gamma$ .
2. The functions  $(t, \gamma) \mapsto m(t; \gamma) = \mathbb{E}(Y \mid Z^\top \gamma = t)$  and  $(t, \gamma) \mapsto \text{Var}(Y \mid Z^\top \gamma = t)$ , with  $\gamma \in \Gamma$ , are Hölder continuous.
3. For any  $\gamma \in \Gamma$ , the functions  $t \mapsto f(t; \gamma)$  and  $t \mapsto m(t; \gamma)f(t; \gamma)$  are twice differentiable and the second order derivatives are Lipschitz continuous with the Lipschitz constant independent of  $\gamma$ .
4. For any  $\gamma \in \Gamma$  and any component  $Z^{(j)}$ ,  $2 \leq j \leq d_Z$ , the functions  $t \mapsto \mathbb{E}[Z^{(j)} \mid Z^\top \gamma = t]$  and  $t \mapsto \mathbb{E}[YZ^{(j)} \mid Z^\top \gamma = t]$  are twice continuously differentiable and the second order derivative is Lipschitz continuous with the Lipschitz constant independent of  $\gamma$ .
5. The function  $\gamma \mapsto m(t; \gamma)$  is twice continuously differentiable and the partial derivatives are continuous as functions of  $t$ . Moreover, the vector-valued function of the first order derivatives with respect to  $\gamma$  is Lipschitz continuous as function of  $(t, \gamma)$ .

The condition of the range  $\mathcal{H}_n$  allows for the data-driven choice of the bandwidth by joint optimization with respect to  $\gamma$  and the bandwidth, provided  $\widehat{m}_h(\cdot; \gamma)$  is the leave-one-out version of the nonparametric smoother.

**Proposition 2.4.4.** *The Assumptions 1 to 3 hold true. Then, the estimator  $\widehat{\gamma}$  obtained as a minimum of  $G_n(\gamma)$  defined in (2.54), with the trimming in (2.55) and any sequence of bandwidths in  $\mathcal{H}_n$ , admits the i.i.d. representation in (2.58). In particular, its sub-vector of the last  $d_Z - 1$  components converges in distribution to the Gaussian limit in (2.60).*

Proposition 2.4.4 is a version of Theorem 4.1 in Delecroix et al. (2006), tailored for a data-driven choice of  $h$ . In the next result, the bandwidth sequence is fixed at a suitable rate for the estimation of the df  $F_\varepsilon$ .

**Proposition 2.4.5.** *Assume the conditions of Proposition 2.4.5 Assumptions 1 to 3 hold true, for a sequence of bandwidth in the range  $\mathcal{H}_n = [\underline{c}n^{-\nu}, \bar{c}n^{-\nu}]$ , for some exponent  $\nu \in (1/4, 1/3)$  and constants  $\underline{c}, \bar{c} > 0$ . Let  $\mathcal{Z}_c = \{z : f(z^\top \gamma_0; \gamma_0) \geq c > 0\}$ , for some (small)  $c > 0$ . Then,*

$$\begin{aligned} \sup_{z \in \mathcal{Z}_c} \left| \widehat{m}_h(z^\top \widehat{\gamma}; \widehat{\gamma}) - m(z^\top \gamma_0; \gamma_0) \right| &= O_P(\{\ln n\}^{1/2} \{nh\}^{-1/2}), \\ \sup_{z \in \mathcal{Z}_c} \left| \nabla_{\widehat{\gamma}} \widehat{m}_h(z^\top \widehat{\gamma}; \widehat{\gamma}) - \nabla_{\widehat{\gamma}} m(z^\top \gamma_0; \gamma_0) \right| &= O_P(\{\ln n\}^{1/2} h^{-1} \{nh\}^{-1/2}), \\ \sup_{z \in \mathcal{Z}_c} \left| \nabla_{\widehat{\gamma}, \widehat{\gamma}}^2 \widehat{m}_h(z^\top \widehat{\gamma}; \widehat{\gamma}) - \nabla_{\widehat{\gamma}, \widehat{\gamma}}^2 m(z^\top \gamma_0; \gamma_0) \right| &= O_P(\{\ln n\}^{1/2} h^{-2} \{nh\}^{-1/2}). \end{aligned}$$

Proposition 2.4.5 is a simplified version<sup>18</sup> of Proposition 3.7 in Lopez et al. (2013). It can be used to show that, with probability tending to 1, the single-index regression functions is eventually in a certain Donsker class of functions (Lopez et al., 2013, See Assumption 10). The fact that the location and scale functions estimators belong to a Donsker class is a key arguments for deriving the asymptotics for the estimators of the distribution function  $F_\varepsilon$ . See Neumeyer and Van Keilegom (2010), Koul et al. (2017), Racine and Van Keilegom (2020).

#### 2.4.4 Asymptotics for semiparametric location-scale regression

Let us now investigate the model defined in (2.27). Recall that when the functions  $m(\cdot)$  and  $\sigma(\cdot)$  are given,  $\gamma$  and  $\beta$  can be easily estimated by least squares. Therefore, given the nonparametric estimators of the location and scale functions, we estimate the parameters  $\gamma$  and  $\beta$  by semiparametric least squares. More precisely, we define

$$\hat{\gamma} = \arg \min_{\gamma \in \{1\} \times \tilde{\Gamma}} \sum_{i=1}^n \{Y_i - \hat{m}_h(Z_i^\top \gamma; \gamma)\}^2 \tau(Z_i), \quad (2.61)$$

with  $\tau(Z_i) = \tau(Z_i; \gamma_0, c)$  a trimming<sup>19</sup> as in (2.55) with some  $c > 0$ , and

$$\hat{\beta} = \arg \min_{\beta \in \{1\} \times \tilde{B}} G_n(\beta; \hat{\gamma}, \hat{\gamma}). \quad (2.62)$$

Here,

$$\begin{aligned} G_n(\beta; \gamma^{(m)}, \gamma^{(\sigma)}) &= G_n(\beta; \gamma^{(m)}, \gamma^{(\sigma)}; h) \\ &= \sum_{i=1}^n \left[ (2/\sigma_Y^2) \{Y_i - \hat{m}_h(Z_i^\top \gamma^{(m)}; \gamma^{(m)})\}^2 - \hat{\sigma}_h^2(X_i^\top \beta; \beta, \gamma^{(\sigma)}) \right]^2 \tau_n(Z_i) \zeta(X_i), \end{aligned}$$

and

$$\begin{aligned} \hat{\sigma}_h^2(u; \beta, \gamma) &= \arg \min_{a, (a,b) \in \mathbb{R}^2} \sum_{\ell=1}^n \left[ \frac{2}{\sigma_Y^2} \{Y_\ell - \hat{m}_h(Z_\ell^\top \gamma; \gamma)\}^2 \tau_n(Z_\ell) - a - b(X_\ell^\top \beta - u) \right]^2 \\ &\quad \times k \left( \frac{X_\ell^\top \beta - u}{h} \right). \end{aligned}$$

Here, for simplicity, we consider a same bandwidth for smoothing the location and scale functions. Moreover, in this section  $\tau_n(\cdot)$  and  $\zeta(\cdot)$  are the infeasible trimming functions

$$\tau_n(Z_i) = \tau(Z_i; \gamma_0, c_n) = \mathbb{1}_{\{f(Z_i^\top \gamma_0; \gamma_0) \geq c_n\}} \quad \text{and} \quad \zeta(X_i) = \mathbb{1}_{\{f(X_i^\top \beta; \beta) \geq c'\}}, \quad (2.63)$$

---

<sup>18</sup>In (Lopez et al., 2013, Proposition 3.7) the responses where randomly right censored, and the result is stated uniformly over the parameter space  $\Gamma$ .

<sup>19</sup>Thus, here  $\tau(Z_i; \gamma_0, c)$  is the indicator of the event  $\{f(Z_i^\top \gamma_0; \gamma_0) > c\}$ .

avoiding observations in areas of low design. Here,  $c_n$  is a sequence of positive numbers decreasing to zero at a suitable rate, while  $c'$  can be a fixed small constant. See also the discussion in Section 2.3.3. It is worth noting that the trimming for defining  $\hat{\gamma}$  in (2.61) needs not to tend to the indicator of the whole support of  $Z$ , in theory we only need a trimming which guarantees that  $\hat{\gamma}$  is  $\sqrt{n}$ -asymptotically normal.

The asymptotic arguments based on the minimization of  $G_n(\beta; \hat{\gamma})$  follow the lines of Section 2.4.3. However, the estimation of  $\gamma_0$  affects the asymptotic variance of  $\hat{\beta}$ . Assuming the required regularity conditions, let us decompose

$$G_n(\beta; \hat{\gamma}, \hat{\gamma}) = G_n(\beta; \gamma_0, \gamma_0) + \nabla_1 G_n(\beta; \gamma_0, \gamma_0)^\top (\tilde{\gamma} - \tilde{\gamma}_0) + \nabla_2 G_n(\beta; \gamma_0, \gamma_0)^\top (\tilde{\gamma} - \tilde{\gamma}_0) + O_P(\|\tilde{\gamma} - \tilde{\gamma}_0\|^2). \quad (2.64)$$

Here,  $\nabla_1$  and  $\nabla_2$  are the  $(d_Z - 1) \times 1$  column matrix-valued operator of the first order derivatives with respect to (the last  $d_Z - 1$  components of)  $\gamma^{(m)}$  and  $\gamma^{(\sigma)}$ , respectively. By the results from Section 2.4.3, we have

$$\tilde{\gamma} - \tilde{\gamma}_0 = \{1 + o_P(1)\} \frac{1}{\sqrt{n}} \sum_{i=1}^n \Psi_i, \quad (2.65)$$

where

$$\Psi_i = \Psi(\varepsilon_i, Z_i, X_i) = \sigma(X_i^\top \beta_0) \varepsilon_i \times \Omega(\gamma_0)^{-1} \nabla_{\tilde{\gamma}} m(Z_i^\top \gamma_0; \gamma_0) \tau(Z_i; \gamma_0, c),$$

with

$$\Omega(\gamma_0) = \mathbb{E} \left[ m'(Z^\top \gamma_0; \gamma_0)^2 \left\{ \tilde{Z} - \mathbb{E}(\tilde{Z} \mid Z^\top \gamma_0) \right\} \left\{ \tilde{Z} - \mathbb{E}(\tilde{Z} \mid Z^\top \gamma_0) \right\}^\top \tau(Z; \gamma_0, c) \right]. \quad (2.66)$$

The quantity  $G_n(\beta; \gamma_0, \gamma_0)$  can be decomposed exactly like in (2.56), that is, with the trimming functions defined in (2.63), we can write

$$\begin{aligned} G_n(\beta; \gamma_0, \gamma_0) &= \underbrace{\frac{1}{n} \sum_{i=1}^n \left[ (2/\sigma_Y^2) \left\{ Y_i - \hat{m}_h(Z_i^\top \gamma_0; \gamma_0) \right\}^2 - \hat{\sigma}_h^2(X_i^\top \beta_0; \beta_0, \gamma_0) \right]^2 \tau_n(Z_i) \zeta(X_i)}_{G_n(\beta_0; \gamma_0, \gamma_0)} \\ &\quad - \underbrace{(\tilde{\beta} - \tilde{\beta}_0)^\top \frac{2}{n} \sum_{i=1}^n \eta_i \nabla_{\tilde{\beta}} \sigma^2(X_i^\top \beta_0; \beta_0) \tau_n(Z_i) \zeta(X_i)}_{2W_{0n}/\sqrt{n}} \\ &\quad + \underbrace{(\tilde{\beta} - \tilde{\beta}_0)^\top \left[ \frac{1}{n} \sum_{i=1}^n \nabla_{\tilde{\beta}} \sigma^2(X_i^\top \beta_0; \beta_0) \nabla_{\tilde{\beta}} \sigma^2(X_i^\top \beta_0; \beta_0)^\top \tau_n(Z_i) \zeta(X_i) \right]}_{\Omega_{0n}/2} (\tilde{\beta} - \tilde{\beta}_0) \\ &\quad + \text{negligible terms,} \quad (2.67) \end{aligned}$$

CHAPTER 2. A SEMIPARAMETRIC LOCATION-SCALE MODEL WITH APPLICATION TO CREDIT RISK

---

where  $\nabla_{\tilde{\beta}}$  is the  $(d_X - 1) \times 1$  column matrix-valued operator of the first order derivatives with respect to the last  $d_X - 1$  components of  $\beta$ , and

$$\begin{aligned} \eta_i &= (2/\sigma_Y^2)\{Y_i - \mathbb{E}[Y_i | X_i, Z_i]\}^2 - \sigma^2(X_i^\top \beta_0; \beta_0, \gamma_0) \\ &= (2/\sigma_Y^2)\{Y_i - m(Z_i^\top \gamma_0; \gamma_0)\}^2 - \sigma^2(X_i^\top \beta_0; \beta_0, \gamma_0). \end{aligned}$$

Note that

$$\mathbb{E}(\eta_i | Z_i, X_i) = 0 \quad \text{and} \quad \mathbb{E}(\eta_i^2 | Z_i, X_i) = \sigma^4(X_i^\top \beta_0; \beta_0, \gamma_0) \mathbb{E} \left[ \{(2/\sigma_Y^2)\varepsilon^2 - 1\}^2 \right].$$

On the other hand, by a Taylor expansion with respect to  $\beta$ , we get

$$\nabla_1 G_n(\beta; \gamma_0, \gamma_0)^\top = \nabla_1 G_n(\beta_0; \gamma_0, \gamma_0)^\top + (\tilde{\beta} - \tilde{\beta}_0)^\top \nabla_{\tilde{\beta},1}^2 G_n(\beta_0; \gamma_0, \gamma_0) + \text{negligible terms},$$

where  $\nabla_{\tilde{\beta},1}^2$  denotes the  $(d_X - 1) \times (d_Z - 1)$  matrix-valued operator of second order cross derivatives with respect to the last  $(d_X - 1)$  and  $(d_Z - 1)$  components of  $\beta$  and  $\gamma$ , respectively. Similarly, we get

$$\nabla_2 G_n(\beta; \gamma_0, \gamma_0)^\top = \nabla_2 G_n(\beta_0; \gamma_0, \gamma_0)^\top + (\tilde{\beta} - \tilde{\beta}_0)^\top \nabla_{\tilde{\beta},2}^2 G_n(\beta_0; \gamma_0, \gamma_0) + \text{negligible terms},$$

Under suitable conditions guaranteeing that the effect of the nonparametric smoothing vanishes asymptotically, the two quantities  $\nabla_{\tilde{\beta},1}^2 G_n(\beta; \gamma_0, \gamma_0)$  and  $\nabla_{\tilde{\beta},2}^2 G_n(\beta; \gamma_0, \gamma_0)$  are expected to concentrate around their expectation calculated with the true functions  $m(\cdot; \gamma)$  and  $\sigma^2(\cdot; \beta, \gamma)$ , respectively. Moreover, using the uniform rates of convergence for the  $U$ -processes indexed by families of functions with polynomial complexity<sup>20</sup>, the convergence of  $\nabla_{\tilde{\beta},1}^2 G_n(\beta; \gamma_0, \gamma_0)$  and  $\nabla_{\tilde{\beta},2}^2 G_n(\beta; \gamma_0, \gamma_0)$  holds uniformly with respect to  $\beta$ . More precisely, under suitable, mild conditions, we get<sup>21</sup>

$$\nabla_{\tilde{\beta},1}^2 G_n(\beta_0; \gamma_0, \gamma_0) = \{1 + o_P(1)\} \mathbf{G}_1 \quad \text{and} \quad \nabla_{\tilde{\beta},2}^2 G_n(\beta_0; \gamma_0, \gamma_0) = \{1 + o_P(1)\} \mathbf{G}_2,$$

with  $\mathbf{G}_1 = G_1(\beta_0, \gamma_0, \gamma_0)$  and  $\mathbf{G}_2 = G_2(\beta_0, \gamma_0, \gamma_0)$  some  $(d_X - 1) \times (d_Z - 1)$  limit matrices.

Now, combining (2.64), (2.65) and (2.67), we get

$$\begin{aligned} G_n(\beta; \hat{\gamma}, \hat{\gamma}) &= G_n(\beta_0; \gamma_0, \gamma_0) + (\tilde{\beta} - \tilde{\beta}_0)^\top \underbrace{\frac{1}{n} \sum_{i=1}^n \Upsilon_i}_{W_n/\sqrt{n}} + \frac{1}{2} (\tilde{\beta} - \tilde{\beta}_0)^\top \Omega_\sigma (\tilde{\beta} - \tilde{\beta}_0) \\ &\quad + \text{negligible terms}, \end{aligned}$$

---

<sup>20</sup>The complexity is expressed in terms of covering numbers. See Sherman (1994b).

<sup>21</sup>We put forward the argument that, in fact, the statement in the display remains true when  $\beta - \beta_0$  tends to zero. More precisely, we claim that, repeatedly using the rates of uniform convergence for  $U$ -processes indexed by Euclidean family of functions, under suitable conditions on the bandwidths rate,  $\nabla_{\tilde{\beta},q}^2 G_n(\beta_0; \gamma_0, \gamma_0) = \{1 + o_P(1)\} \mathbf{G}_q$  uniformly over  $o_P(1)$  neighborhoods of  $\beta_0$ ,  $q \in \{1, 2\}$ .

where<sup>22</sup>

$$\Omega_\sigma = 2\mathbb{E} \left[ \nabla_{\tilde{\beta}} \sigma^2(X^\top \beta_0; \beta_0) \nabla_{\tilde{\beta}} \sigma^2(X^\top \beta_0; \beta_0)^\top \zeta(X) \right],$$

and

$$\Upsilon_i = (\mathbf{G}_1 + \mathbf{G}_2) \Psi_i - 2\eta_i \nabla_{\tilde{\beta}} \sigma^2(X_i^\top \beta_0; \beta_0) \zeta(X_i) \in \mathbb{R}^{d_X-1}.$$

The asymptotic normality of the estimator  $\hat{\beta}$  defined in (2.62) follows by the Central Limit Theorem applied for the average of the i.i.d. zero-mean vectors  $\Upsilon_i$  and Lemma 2.4.2 :

$$\sqrt{n} \left( \hat{\beta} - \tilde{\beta}_0 \right) \Rightarrow \mathcal{N}_{d_X-1} \left( \mathbf{0}, \Omega_\sigma^{-1} \text{Var}(\Upsilon_i) \Omega_\sigma^{-1} \right).$$

### 2.4.5 Asymptotics for semiparametric autoregression

Consider that the times series  $\{Y_i\}$  is the stationary solution of the equation

$$Y_i = \mathbf{a}Y_{i-1} + m(Z_i^\top \gamma) + \sigma(X_i^\top \beta) \varepsilon_i, \quad 1 \leq i \leq n, \quad |\mathbf{a}| < 1, \quad (2.68)$$

where  $(\varepsilon_i, Z_i^\top, X_i^\top)^\top$ ,  $1 \leq i \leq n$ , are independent copies of  $(\varepsilon, Z^\top, X^\top)^\top$  with  $Z, X$  sub-vectors of  $W$  and  $\varepsilon \perp W$ ,  $\mathbb{E}(\varepsilon) = 0$ . Let  $\mathbf{a}_0 \in (-1, 1)$  denote the true value of the autoregressive parameter. We can rewrite (2.68) under the form of a standard AR(1) model

$$Y_i = \mathbf{a}Y_{i-1} + e_i, \quad \text{with } e_i = m(Z_i^\top \gamma) + \sigma(X_i^\top \beta) \varepsilon_i, \quad i \geq 1,$$

with  $\{e_i\}$  an independent white noise with

$$\mathbb{E}(e_i^2) = \tau^2 := \mathbb{E}[m^2(Z^\top \gamma)] + \mathbb{E}[\sigma^2(X^\top \beta)] \text{Var}(\varepsilon). \quad (2.69)$$

The estimator of  $a$  in the AR(1) model is the standard least squares estimator, that is

$$\hat{\mathbf{a}} = \frac{\sum_{i=2}^n Y_i Y_{i-1}}{\sum_{i=2}^n Y_{i-1}^2} = \mathbf{a}_0 + \frac{(n-1)^{-1} \sum_{i=2}^n e_i Y_{i-1}}{(n-1)^{-1} \sum_{i=2}^n Y_{i-1}^2}. \quad (2.70)$$

**Lemma 2.4.6.** *The estimator  $\hat{\mathbf{a}}$  is consistent and*

$$\sqrt{n} (\hat{\mathbf{a}} - \mathbf{a}_0) \Rightarrow \mathcal{N}(0, 1 - \mathbf{a}_0^2).$$

*Proof of Lemma 2.4.6.* The result follows by standard asymptotic results, the Law of Large Numbers (LLN) and the Central Limit Theorem (CLT), for weakly dependent stationary time series (see Chapter 7, Hamilton, 1994). By the LLN, the sequence  $\{Y_i\}$ , that is the stationary solution (2.68), has the property

$$\mathbb{E}(Y_i^2) = \frac{\tau^2}{1 - \mathbf{a}_0^2} \quad \text{and} \quad (n-1)^{-1} \sum_{i=2}^n Y_{i-1}^2 = \frac{\tau^2}{1 - \mathbf{a}_0^2} + o_P(1).$$

---

<sup>22</sup>The trimming  $\tau_n(Z)$  no longer appears in the definition of the limit  $\Omega$  because of the sequence  $c_n$  decreasing to 0.

On the other hand,  $\mathbb{E}[e_i Y_{i-1}] = \mathbb{E}[\mathbb{E}(e_i | Y_{i-1}) Y_{i-1}] = 0$  and, by the CLT, we have

$$\sqrt{n} \frac{1}{n-1} \sum_{i=2}^n e_i Y_{i-1} \Rightarrow \mathcal{N}(0, V(\mathbf{a}_0)),$$

where

$$V(\mathbf{a}_0) = \mathbb{E}[e_i^2 Y_{i-1}^2] = \mathbb{E}[\mathbb{E}(e_i^2 | Y_{i-1}) Y_{i-1}^2] = \tau^2 \times \frac{\tau^2}{1 - \mathbf{a}_0^2}.$$

Using these fact in (2.70), the result follows by Slutsky's Theorem.  $\square$

Note that by the definition of the model (2.68),

$$e_i = e_i(\gamma_0, \beta_0) = m(Z_i^\top \gamma_0) + \sigma(X_i^\top \beta_0) \varepsilon_i, \quad 1 \leq i \leq n,$$

with i.i.d.  $(e_i, Z_i^\top, X_i^\top)^\top$ , and

$$\mathbb{E}[e_i | Z_i, X_i] = m(Z_i^\top \gamma_0).$$

We can therefore reconsider the least squares estimator of the parameter  $\gamma_0$ , as defined in Section 2.4.3, with the responses  $Y_i$  replaced by the  $e_i$ . However, the  $e_i$  are not available, only the  $\widehat{e}_i$  can be used, where

$$\widehat{e}_i = Y_i - \widehat{\mathbf{a}} Y_{i-1} = \{Y_i - \mathbf{a}_0 Y_{i-1}\} - (\widehat{\mathbf{a}} - \mathbf{a}_0) Y_{i-1} = e_i - (\widehat{\mathbf{a}} - \mathbf{a}_0) Y_{i-1}.$$

We then redefine

$$G_n(\gamma) = G_n(\gamma; h) = \frac{1}{n} \sum_{i=1}^n \{\widehat{e}_i - \widehat{m}_h(Z_i^\top \gamma; \gamma)\}^2 \tau(Z_i), \quad \gamma \in \{1\} \times \widetilde{\Gamma},$$

where

$$\widehat{m}_h(u; \gamma) = \sum_{\ell=2}^n \mathcal{W}_{n,\ell}(u; h, \gamma) \{e_\ell + (\widehat{\mathbf{a}} - \mathbf{a}_0) Y_{\ell-1}\}, \quad \text{with} \quad \sum_{\ell=2}^n \mathcal{W}_{n,\ell}(h; \gamma) = 1.$$

The expression of  $\widehat{m}_h(u; \gamma)$ , more convenient for the current discussion, is an alternative writing of local linear or local constant (Nadaraya-Watson) estimators, which are both so-called *linear smoothers*. For example, in the case of the Nadaraya-Watson estimator,

$$\mathcal{W}_{n,\ell}(u; h, \gamma) = \frac{k((u - Z_\ell^\top \gamma)/h)}{\sum_{j=2}^n k((u - Z_j^\top \gamma))},$$

with the rule  $0/0 = 0$ . Recall that  $k(\cdot)$  is a second order kernel and  $h$  is the bandwidth. Let us note that both  $\{e_i\}$  and  $\{Y_{i-1}\}$  are zero-mean sequences, the first is i.i.d, the second is a stationary AR(1) series. This suggests that

$$\sum_{\ell=1}^n \mathcal{W}_{n,\ell}(u; h, \gamma) e_\ell \quad \text{and} \quad \sum_{\ell=1}^n \mathcal{W}_{n,\ell}(u; h, \gamma) Y_{\ell-1},$$

have a similar behavior (rate). Since  $\widehat{\mathbf{a}} - \mathbf{a}_0 = O_P(1/\sqrt{n})$ , we can confidently consider that the estimation of  $\mathbf{a}_0$  does not affect the kernel estimator of the location function  $m(\cdot; \gamma)$ , uniformly with respect to  $\gamma$ . More precisely, we can consider

$$\widehat{m}_h(u; \gamma) = \sum_{\ell=2}^n \mathcal{W}_{n,\ell}(u; h, \gamma) e_\ell, \quad \text{with} \quad \sum_{\ell=2}^n \mathcal{W}_{n,\ell}(h; \gamma) = 1.$$

Next, following (2.56), and noting that

$$\mathbb{E} [Y_{i-1} \nabla_{\tilde{\gamma}} m(Z_i^\top \gamma_0; \gamma_0)] = 0,$$

by the Taylor expansion, we can write

$$\begin{aligned} G_n(\gamma) &= \underbrace{\frac{1}{n} \sum_{i=1}^n \{\widehat{e}_i - \widehat{m}_h(Z_i^\top \gamma_0; \gamma_0)\}^2 \tau(Z_i)}_{G_n(\gamma_0)} \\ &\quad - (\tilde{\gamma} - \tilde{\gamma}_0)^\top \frac{2}{n} \sum_{i=1}^n e_i \nabla_{\tilde{\gamma}} m(Z_i^\top \gamma_0; \gamma_0) \tau(Z_i) + (\tilde{\gamma} - \tilde{\gamma}_0)^\top \Omega(\gamma_0) (\tilde{\gamma} - \tilde{\gamma}_0) \\ &\quad + \underbrace{(\widehat{\mathbf{a}} - \mathbf{a}_0) (\tilde{\gamma} - \tilde{\gamma}_0)^\top \frac{2}{n} \sum_{i=1}^n Y_{i-1} \nabla_{\tilde{\gamma}} m(Z_i^\top \gamma_0; \gamma_0) \tau(Z_i)}_{\text{negligible terms}} + \text{remainder terms}, \end{aligned}$$

with  $\Omega(\gamma_0)$  defined in (2.66). Gathering facts, we deduce

$$\sqrt{n} (\tilde{\gamma} - \tilde{\gamma}_0) \Rightarrow \mathcal{N}_{d_Z-1}(\mathbf{0}, \Omega(\gamma_0)^{-1} \mathcal{V}(\gamma_0) \Omega(\gamma_0)^{-1}), \quad (2.71)$$

where

$$\mathcal{V}(\gamma_0, \mathbf{a}_0) = \mathbb{E} \left[ \text{Var}(e_i | Z_i) m'(Z_i^\top \gamma_0; \gamma_0)^2 \left\{ \tilde{Z}_i - \mathbb{E}(\tilde{Z}_i | Z_i^\top \gamma_0) \right\} \left\{ \tilde{Z}_i - \mathbb{E}(\tilde{Z}_i | Z_i^\top \gamma_0) \right\}^\top \tau(Z_i) \right].$$

In this context,  $\text{Var}(e_i | Z_i) = \mathbb{E}[\sigma^2(X_i^\top \beta) | Z_i] \text{Var}(\varepsilon)$

**Remark 1.** *Let us note that the asymptotic variance in (2.71) is expressed in terms of the conditional variance of the  $e_i$  variable, which is different from the conditional variance of the  $Y_i$  given  $Z_i$ . In the context of the i.i.d. observations as considered in Section 2.4.3, we have the equality  $\text{Var}(\varepsilon | Z) = \text{Var}(Y | Z)$ . It is therefore interesting to express  $\mathcal{V}(\gamma_0, \mathbf{a}_0)$  using  $\text{Var}(Y_i | Z_i)$  instead of  $\text{Var}(e_i | Z_i)$ . By the definition of the model, we get*

$$\text{Var}(Y_i | Z_i) = \mathbb{E}[\mathbf{a}_0^2 Y_{i-1}^2] + \mathbb{E}[\sigma^2(X_i^\top \beta) | Z_i] \text{Var}(\varepsilon) = \frac{\mathbf{a}_0^2 \tau^2}{1 - \mathbf{a}_0^2} + \mathbb{E}[\sigma^2(X_i^\top \beta) | Z_i] \text{Var}(\varepsilon),$$

with  $\tau^2$  defined in (2.69). In particular, this reveals the large variance of the estimator of  $\gamma_0$  when the time series  $\{Y_i\}$  is close to a unit root series, that is  $\mathbf{a}_0 < 1$  is close to 1.

Let us now discuss the case where the function  $m(\cdot)$  is null, that is the times series  $\{Y_i\}$  is the stationary solution of the equation

$$Y_i = \mathbf{a}Y_{i-1} + \sigma(X_i^\top \beta)\varepsilon_i, \quad 1 \leq i \leq n, \quad |\mathbf{a}| < 1.$$

In this case, like in Section 2.3.3, we define

$$\sigma^2(u; \beta) = \frac{1}{\text{Var}(\varepsilon)} \mathbb{E}[\{Y_i - \mathbf{a}Y_{i-1}\}^2 \mid X_i^\top \beta = u] = \frac{1}{\text{Var}(\varepsilon)} \mathbb{E}[\{e_i - (\mathbf{a} - \mathbf{a}_0)Y_{i-1}\}^2 \mid X_i^\top \beta = u],$$

and estimate this function by

$$\widehat{\sigma}_h^2(u; \beta, \widehat{\mathbf{a}}) = \frac{2}{\sigma_Y^2} \sum_{\ell=2}^n \mathcal{W}_{n,\ell}(u; h, \gamma) \{e_\ell + (\widehat{\mathbf{a}} - \mathbf{a}_0)Y_{\ell-1}\}^2 \quad \text{with} \quad \sum_{\ell=2}^n \mathcal{W}_{n,\ell}(h; \gamma) = 1, \quad (2.72)$$

for some weights  $\mathcal{W}_{n,\ell}$ . Again, since  $\widehat{\mathbf{a}}$  is  $\sqrt{n}$ -consistent, it is clear that the feasible non-parametric estimator in (2.72) defined using the estimated squared  $e_\ell$ , is asymptotically equivalent to  $\widehat{\sigma}_h^2(u; \beta, \mathbf{a}_0)$ , that is the one defined with the true squared  $e_\ell$  instead. As a consequence, the feasible semiparametric estimator

$$\widehat{\beta} = \arg \min_{\beta \in \{1\} \times \widetilde{B}} \sum_{i=1}^n \left[ (2/\sigma_Y^2) \{Y_i - \widehat{\mathbf{a}}Y_{i-1}\}^2 - \widehat{\sigma}_h^2(X_i^\top \beta; \beta, \widehat{\mathbf{a}}) \right]^2 \zeta_n(X_i),$$

with  $\zeta_n(X_i)$  the indicator of the set  $\{\widehat{f}(X_i^\top \widetilde{\beta}; \widetilde{\beta}) \geq c'\}$ , for some  $c' > 0$ , asymptotically behaves like the infeasible one defined with  $\mathbf{a}_0$  instead of  $\widehat{\mathbf{a}}$ . The asymptotic distribution of  $\widehat{\beta}$  can be deduced from that of semiparametric single-index parameter estimator, as discussed<sup>23</sup> in Section 2.4.3.

Gathering facts, we deduce that the after computing  $\widehat{\mathbf{a}}$ , the estimators of  $\gamma_0$  and  $\beta_0$  in the model (2.68) can be done by obvious modifications of the estimators discussed in Section 2.4.4. Moreover, their asymptotic distributions are not affected by the estimation of  $\mathbf{a}_0$ , they behave as in the case where  $\mathbf{a}_0$  is given. The details are omitted.

## 2.4.6 Error distribution estimation

Let us first consider the homoscedastic single-index regression context with known<sup>24</sup> error variance  $\sigma_0^2$ . This is the case considered by Koul et al. (2017). In this case, the empirical distribution function estimator

$$\widehat{F}_{\varepsilon, \text{emp}}(t) = \frac{1}{N(\gamma_0; c)} \sum_{i=1}^n \tau(Z_i) \mathbb{1}_{\{\widehat{\varepsilon}_i \leq t\}}, \quad \text{with} \quad N(\gamma_0; c) = \sum_{i=1}^n \tau(Z_i), \quad t \in \mathbb{R},$$

---

<sup>23</sup>In the spirit of the sections above, the asymptotic investigation is considered for the estimator  $\widehat{\beta}$  defined with the infeasible trimming based on the event  $\{f(X_i^\top \beta_0; \beta_0) \geq c'\}$ . The difference between the two estimators is negligible, see Delecroix et al. (2006).

<sup>24</sup>Without loss of generality, we can consider  $\sigma_0 = 1$ .

where, for some (small) constant  $c > 0$ , and  $\tau(Z_i)$  denotes the indicator of the event  $\{f(Z_i^\top \gamma_0; \gamma_0) \geq c\}$ . Moreover,

$$\widehat{\varepsilon}_i = Y_i - \widehat{m}_h(Z_i^\top \widehat{\gamma}; \widehat{\gamma}), \quad 1 \leq i \leq n.$$

Let

$$p(\gamma_0; c) = \mathbb{E}[\tau(Z)] = \mathbb{P}(f(Z^\top \gamma_0; \gamma_0) > c),$$

and note that  $\mathbb{E}[N(\gamma_0; c)] = np(\gamma_0; c)$  and  $N(\gamma_0; c) - np(\gamma_0; c) = O_P(1/\sqrt{n})$ .

For simplicity, we consider the infeasible trimming  $\tau(Z_i)$ . The difference between  $\widehat{F}_{\varepsilon, \text{emp}}$  defined as above and that with a feasible pilot estimator and a kernel estimation of the density of the index, that is  $\tau_n(Z_i) = \mathbb{1}_{\{\widehat{f}(Z_i^\top \gamma_n; \gamma_n) \geq c > 0\}}$ , can be handled with the arguments of Delecroix et al. (2006); see also Koul et al. (2017).

Let  $f_\varepsilon$  denote the density of  $\varepsilon$ , that is  $F'_\varepsilon = f_\varepsilon$ , and let

$$\phi_1(e, y) = f_\varepsilon(y)e.$$

**Proposition 2.4.7.** *Assume the conditions of Proposition 2.4.5 hold true, for a sequence of bandwidths in the range  $\mathcal{H}_n = [\underline{c}n^{-\nu}, \bar{c}n^{-\nu}]$ , for some exponent  $\nu \in (1/4, 1/3)$  and constants  $\underline{c}, \bar{c} > 0$ . Moreover, the error variable  $\varepsilon$  has a finite third moment,  $F_\varepsilon$  is twice continuously differentiable, and  $\mathbb{E}[\{f'_\varepsilon(\varepsilon)/f_\varepsilon(\varepsilon)\}^2] < \infty$ . Then,*

$$\widehat{F}_{\varepsilon, \text{emp}}(y) - F_\varepsilon(y) = \frac{1}{np(\gamma_0; c)} \sum_{i=1}^n \tau(Z_i) [\mathbb{1}_{\{\varepsilon_i \leq y\}} - F_\varepsilon(y) + \phi_1(\varepsilon_i, y)] + R_{1,n}(y),$$

and

$$\sup_{y \in \mathbb{R}} |R_{1,n}(y)| = o_P(1/\sqrt{n}).$$

Proposition 2.4.7 is a version of Theorem 1 in Koul et al. (2017). The conditions we impose here are stronger than the ones they use. The result above implies the weak convergence of the empirical process indexed by the family of functions

$$\mathcal{F}_1 = \{(e, z) \mapsto \tau(z) [\mathbb{1}_{\{e \leq y\}} - F_\varepsilon(y) + \phi_1(e, y)] : e, y \in \mathbb{R}\}.$$

The limit in distribution of this process is given in the following corollary.

**Corollary 2.4.8.** *Assume the conditions of Theorem 2.4.7 hold true. Then, the process  $\sqrt{n}(\widehat{F}_{\varepsilon, \text{emp}} - F_\varepsilon)$  converges weakly to a zero-mean Gaussian process  $\mathbb{Z}$  with the covariance*

$$\begin{aligned} \text{Cov}(\mathbb{Z}(y_1), \mathbb{Z}(y_2)) &= \frac{1}{p(\gamma_0; c)} \mathbb{E} \left[ \{ \mathbb{1}_{\{\varepsilon \leq y_1\}} - F_\varepsilon(y_1) + f_\varepsilon(y_1)\varepsilon \} \right. \\ &\quad \left. \times \{ \mathbb{1}_{\{\varepsilon \leq y_2\}} - F_\varepsilon(y_2) + f_\varepsilon(y_2)\varepsilon \} \right]. \end{aligned}$$

*Proof of Corollary 2.4.8.* The weak convergence of the process  $\sqrt{n}(\widehat{F}_{\varepsilon, \text{emp}} - F_\varepsilon)$  is a consequence of the fact that  $\mathcal{F}_1$  is a Donsker class (Chapter 2.10, van der Vaart & Wellner, 2023) and the CLT (Chapter 2.5, van der Vaart & Wellner, 2023). The calculation of covariance function is straightforward, after noting that  $\mathbb{E}[\tau^2(Z)] = p(\gamma_0; c)$ .  $\square$

Let us now discuss the case where the variance of the error term is also estimated. This is the case considered by Neumeyer and Van Keilegom (2010). To get some insight on what changes when the variance is estimated, let us consider the simpler situation where the variance is unknown but independent of the covariates (*i.e.*, the variance  $\sigma_0^2$  is constant). For simpler writings, let us consider  $\text{Var}(\varepsilon) = 1$ . As discussed in Section 2.3.1, the standardization of the residuals is unnecessary in the case of homoscedastic errors, can be absorbed by the empirical distribution function estimator. However, to get insight on the effect of the variance estimation on the asymptotic representation of the empirical distribution function estimator, below we consider the standardized residuals. In this case, the standardized residuals<sup>25</sup> are

$$\widehat{\varepsilon}_i = \frac{Y_i - \widehat{m}_h(Z_i^\top \widehat{\gamma}; \widehat{\gamma})}{\widehat{\sigma}}, \quad 1 \leq i \leq n, \quad (2.73)$$

where

$$\widehat{\sigma}^2 = \frac{1}{N(\gamma_0; c)} \sum_{i=1}^n \{Y_i - \widehat{m}_h(Z_i^\top \widehat{\gamma}; \widehat{\gamma})\}^2 \tau(Z_i), \quad \text{with} \quad N(\gamma_0; c) = \sum_{i=1}^n \tau(Z_i).$$

Next, under mild conditions,

$$\begin{aligned} \widehat{\sigma}^2 - \sigma_0^2 &= \frac{1 + o_P(1)}{np(\gamma_0; c)} \sum_{i=1}^n \left[ \{Y_i - \widehat{m}_h(Z_i^\top \widehat{\gamma}; \widehat{\gamma})\}^2 - \sigma_0^2 \mathbb{E}(\varepsilon_i^2) \right] \tau(Z_i) \\ &= \frac{1}{np(\gamma_0; c)} \sum_{i=1}^n \left[ \{Y_i - m(Z_i^\top \gamma_0; \gamma_0)\}^2 - \sigma_0^2 \mathbb{E}(\varepsilon_i^2) \right] \tau(Z_i) + \text{negligible terms} \\ &= \frac{\sigma_0^2}{np(\gamma_0; c)} \sum_{i=1}^n \{\varepsilon_i^2 - 1\} \tau(Z_i) + \text{negligible terms}. \end{aligned}$$

The trimming is necessary because in general  $\widehat{m}_h(Z_i^\top \widehat{\gamma}; \widehat{\gamma}) - m(Z_i^\top \gamma_0; \gamma_0)$  is not negligible uniformly with respect to  $Z_i$ .

Since, by elementary algebra,  $\widehat{\sigma}^2 - \sigma_0^2 = (\widehat{\sigma} - \sigma_0)(2\sigma_0 + O_P(1/\sqrt{n}))$ , we next deduce

$$\frac{\widehat{\sigma} - \sigma_0}{\sigma_0} = \frac{1}{np(\gamma_0; c)} \sum_{i=1}^n \frac{1}{2} \{\varepsilon_i^2 - 1\} \tau(Z_i) + \text{negligible terms} = O_P(1/\sqrt{n}). \quad (2.74)$$

---

<sup>25</sup>Note that for the final purpose of computing conditional quantiles, here is unnecessary to fix the variance of the standardized error, as we did in (2.26) and the following section where the variance was not considered constant.

Gathering facts, and the property  $\hat{\gamma} - \gamma_0 = O_P(1/\sqrt{n})$ , Lemma 2.4.3 and a suitable bandwidth, we get

$$\begin{aligned}\tau(Z_i)\hat{\varepsilon}_i &= \frac{Y_i - \hat{m}_h(Z_i^\top \hat{\gamma}; \hat{\gamma})}{\hat{\sigma}} \tau(Z_i) \\ &= \frac{Y_i - \hat{m}_h(Z_i^\top \gamma_0; \gamma_0)}{\hat{\sigma}} \tau(Z_i) + \text{negligible terms} \\ &= \tau(Z_i)\varepsilon_i \times \frac{\sigma_0}{\hat{\sigma}} - \frac{\hat{m}_h(Z_i^\top \gamma_0; \gamma_0) - m(Z_i^\top \gamma_0; \gamma_0)}{\hat{\sigma}} \tau(Z_i) \\ &\quad + \text{negligible terms}, \quad 1 \leq i \leq n.\end{aligned}$$

The negligible terms are  $o_P(1/\sqrt{n})$  uniformly with respect to  $Z_i$  with  $\tau(Z_i) = 1$ . With the standardized residuals in (2.73), let

$$\hat{F}_{\varepsilon, \text{emp}}(y) = \frac{1}{N(\gamma_0; c)} \sum_{i=1}^n \tau(Z_i) \mathbb{1}_{\{\hat{\varepsilon}_i \leq y\}} = \frac{1}{N(\gamma_0; c)} \sum_{i=1}^n \tau(Z_i) \mathbb{1}_{\{\hat{\varepsilon}_i \leq y, \tau(Z_i)=1\}}, \quad t \in \mathbb{R}.$$

Following the lines of the proof of Theorem 2.1 in Neumeyer and Van Keilegom (2010), using their Lemma A.3 and their first statement in Lemma A.2 (that results require a simple modification due to the presence of the trimming), and the i.i.d. representation (2.74) above, by a Taylor expansion we write

$$\begin{aligned}\hat{F}_{\varepsilon, \text{emp}}(y) - F_\varepsilon(y) &= \frac{1}{N(\gamma_0; c)} \sum_{i=1}^n \tau(Z_i) [\mathbb{1}_{\{\hat{\varepsilon}_i \leq y\}} - F_\varepsilon(y)] \\ &\quad + \mathbb{E}_n \left\{ \left[ F_\varepsilon \left( y \frac{\hat{\sigma}}{\sigma_0} + \frac{\hat{m}_h(Z^\top \gamma_0; \gamma_0) - m(Z^\top \gamma_0; \gamma_0)}{\hat{\sigma}} \right) - F_\varepsilon(y) \right] \tau(Z) \right\} + o_P(1/\sqrt{n}) \\ &= \frac{1}{np(\gamma_0; c)} \sum_{i=1}^n \tau(Z_i) [\mathbb{1}_{\{\varepsilon_i \leq y\}} - F_\varepsilon(y)] \\ &\quad + f_\varepsilon(y) \mathbb{E}_n \left\{ \frac{1}{np(\gamma_0; c)} \sum_{i=1}^n \left[ \frac{y}{2} \{\varepsilon_i^2 - 1\} + \varepsilon_i \right] \tau(Z_i) \right\} + o_P(1/\sqrt{n}).\end{aligned}$$

Here,  $\mathbb{E}_n\{\cdot\}$  denotes the conditional expectation given the sample. Since the quantity inside the last expectation  $\mathbb{E}_n\{\cdot\}$  is a function of the sample, we derive the i.i.d. representation of the empirical distribution estimator. We formalize this representation in the following proposition, a version of Theorem 2.1 of Neumeyer and Van Keilegom (2010). Let

$$\phi_2(e, y) = f_\varepsilon(y)[e + y(e^2 - 1)/2].$$

**Proposition 2.4.9.** *Assume the conditions of Proposition 2.4.7 hold true, and the error variable  $\varepsilon$  has a finite moment of order four. Then,*

$$\hat{F}_{\varepsilon, \text{emp}}(y) - F_\varepsilon(y) = \frac{1}{np(\gamma_0; c)} \sum_{i=1}^n \tau(Z_i) [\mathbb{1}_{\{\varepsilon_i \leq y\}} - F_\varepsilon(y) + \phi_2(\varepsilon_i, y)] + R_{2,n}(y),$$

with

$$\sup_{y \in \mathbb{R}} |R_{2,n}(y)| = o_P(1/\sqrt{n}).$$

**Corollary 2.4.10.** *Assume the conditions of Theorem 2.4.7 hold true. Then, the process  $\sqrt{n}(\widehat{F}_{\varepsilon, \text{emp}}(y) - F_\varepsilon)$  converges weakly to a zero-mean Gaussian process  $\mathbb{Z}$  with the covariance*

$$\begin{aligned} \text{Cov}(\mathbb{Z}(y_1), \mathbb{Z}(y_2)) &= \frac{1}{p(\gamma_0; c)} \mathbb{E} \left[ \{\mathbb{1}_{\{\varepsilon \leq y_1\}} - F_\varepsilon(y_1) + \phi_2(\varepsilon, y_1)\} \right. \\ &\quad \left. \times \{\mathbb{1}_{\{\varepsilon \leq y_2\}} - F_\varepsilon(y_2) + \phi_2(\varepsilon, y_2)\} \right]. \end{aligned}$$

Proposition 2.4.9 can be extended to the case of non-constant variance  $\sigma^2$ . In the case of a nonparametric conditional variance, the results can be derived from those of Neumeyer and Van Keilegom (2010). Under more technical conditions, the representation in Proposition 2.4.9 is expected to remain valid. With a single-index modeling the i.i.d. representation of  $\widehat{F}_{\varepsilon, \text{emp}}(y)$  becomes even more technical because of the fact that the trimming  $\tau$  should tend to the constant function equal to 1, and a trimming based on the index  $X^\top \beta_0$  should be considered. We leave this investigation for future work.

## 2.4.7 Quantile estimation

Let us now consider the map that assigns to each cumulative distribution function  $F$  its  $\tau$ -th quantile  $q(\tau) := F^{-1}(\tau) = \inf\{y : F(y) \geq \tau\}$ . It is well known that this map is Hadamard differentiable at every distribution function  $F$  that is differentiable at  $q(\tau) = F^{-1}(\tau)$  with strictly positive derivative  $f(q(\tau))$ , tangentially to the set of functions that are continuous at  $q(\tau)$ . See van der Vaart and Wellner (Section 3.10.5.2, 2023).

Let  $q_\varepsilon(\tau) := F_\varepsilon^{-1}(\tau)$  and

$$\widehat{q}_\varepsilon(\tau) = \inf\{y : \widehat{F}_{\varepsilon, \text{emp}}(y) \geq \tau\},$$

where  $\widehat{F}_{\varepsilon, \text{emp}}$  is an empirical distribution function estimation for which an i.i.d. representation like in Propositions 2.4.7 and 2.4.9 holds true. By the delta-method and the expression of the derivative of the inverse map, we get the following result.

**Corollary 2.4.11.** *Assume that  $F_\varepsilon$  admits a continuous density  $f_\varepsilon$ , and  $\widehat{F}_{\varepsilon, \text{emp}}$  admits an i.i.d. representation like in Propositions 2.4.7 and 2.4.9 for some function  $\phi(\varepsilon, y)$ . Then, for any  $\tau \in (0, 1)$  such that  $f_\varepsilon(q_\varepsilon(\tau)) > 0$ ,*

$$\sqrt{n} \{\widehat{q}_\varepsilon(\tau) - q_\varepsilon(\tau)\} \Rightarrow \frac{\mathbb{Z}(q_\varepsilon(\tau))}{f_\varepsilon(q_\varepsilon(\tau))},$$

that is the limit is a centered normal distribution with variance

$$\frac{1}{f_\varepsilon^2(q_\varepsilon(\tau))p(\gamma_0; c)} \text{Var} \left[ \mathbb{1}_{\{\varepsilon \leq q_\varepsilon(\tau)\}} - \tau + \phi(\varepsilon, q_\varepsilon(\tau)) \right].$$

*Proof of Corollary 2.4.11.* The result is a direct consequence of the weak convergence of the empirical distribution function estimator  $\widehat{F}_{\varepsilon, \text{emp}}$  and (Theorem 3.10.4 and Lemma 3.10.21, van der Vaart & Wellner, 2023).  $\square$

Let us recall the purpose of our semiparametric location-scale modeling. We want to get estimates for

$$q_{\tau}(w) = \inf \left\{ y : F_{\varepsilon} \left( \frac{y - m(z)}{\sigma(x)} \right) \geq \tau \right\},$$

where  $w, z, x$  are values in the supports of  $W, Z, X$ , where  $W$  is a vector of covariates and  $Z, X$  are subvectors of  $W$ . The functions  $m(\cdot)$  and  $\sigma(\cdot)$  being unknown, we proposed different models which combines the appealing features of parametric and nonparametric paradigms. With quantile notation introduced above, we can write

$$q_{\tau}(w) = m(z) + \sigma(x)q_{\varepsilon}(\tau).$$

Let  $\widehat{m}(\cdot)$  and  $\widehat{\sigma}(\cdot)$  be generic estimates of the unknown functions, that could be that constructed in any of the models discussed above (single index, partially linear single-index...). Then the natural estimator of  $q_{\tau}(w)$  is

$$\widehat{q}_{\tau}(w) = \widehat{m}(z) + \widehat{\sigma}(x)\widehat{q}_{\varepsilon}(\tau).$$

By simple algebra, we get

$$\widehat{q}_{\tau}(w) - q_{\tau}(w) = \{\widehat{m}(z) - m(z)\} + \{\widehat{\sigma}(x) - \sigma(x)\}q_{\varepsilon}(\tau) + \sigma(x)\{\widehat{q}_{\varepsilon}(\tau) - q_{\varepsilon}(\tau)\}. \quad (2.75)$$

**Remark 2.** *The relationship (2.75) reveals several aspects. First, the rate of convergence and the asymptotic behavior of the conditional quantile estimator  $\widehat{q}_{\tau}(w)$  will be determined by the behavior of the estimators of the location and scale functions. This because we showed that  $\widehat{q}_{\varepsilon}(\tau) - q_{\varepsilon}(\tau) = O_P(1/\sqrt{n})$ , while  $\widehat{m}(z) - m(z)$  and  $\widehat{\sigma}(x) - \sigma(x)$  converge at slower rates as soon as the estimators  $\widehat{m}(z)$  and  $\widehat{\sigma}(x)$  involve functions to be estimated nonparametrically. It is therefore important to have flexible estimators for  $m(\cdot)$  and  $\sigma(\cdot)$  which avoid the curse of dimensionality due to several covariates. The different methods including a dimension reduction idea that we proposed in the previous sections are therefore better choices than the fully nonparametric estimators for the location and/or scale function. They are also better than the purely parametric models because they are less exposed to misspecification. Second, under usual conditions,  $\widehat{q}_{\tau}(w) - q_{\tau}(w)$  is asymptotically normal distributed. This is a consequence of the joint asymptotic normality of  $(\widehat{m}(z), \widehat{\sigma}(x))$  which can be guaranteed under suitable conditions in the different models discussed above. Third, there are several ‘optimal’ bandwidths to be considered in the single-index paradigm for the location and scale functions: the one for indices estimators  $\widehat{\alpha}$  and  $\widehat{\beta}$  with the rate  $n^{-1/5}$ , the one for computing the standardized residuals used to estimate  $F_{\varepsilon}$  with faster rate, say,  $n^{-2/7}$ , and, finally, the one for calculating the estimator of the conditional quantile using (2.75) which is again with the rate  $n^{-1/5}$ . The conclusion is that we are facing a quite challenging situation where, on the one hand, the theory provides some guidance and, on the other hand, the practitioner is reticent to multiple and somehow complex choices of the tuning parameters in the applications.*

## 2.5 Conclusion

In Chapter 2 we propose a single index location scale model and show how practitioners could apply this model to credit risk to estimated unexpected losses. Instead of using the marginal distribution of the common factor we propose to estimate its conditional distribution, given a set of covariates, we can therefore compute the probability of default given this conditional quantile with the Merton-Vašíček's model. In this chapter, we propose several variants: homoscedastic, heteroscedastic, a partially linear model and a model with dependent data. We propose the use of nonparametric estimates to account for potential nonlinearities. Additionally, we introduce a single index model as a dimension reduction technique to mitigate the curse of dimensionality.

For each variant, we perform numerical studies to assess the performance of each one against the regulatory model. In most cases, the proposed model outperforms the regulatory model sometimes of several order of magnitudes. Additionally, some theoretical properties are derived, most notably, it is showed that the estimator of the conditional quantile converges at the same rate as the nonparametric estimators of the location and scale functions. Futhermore implementation on private banking data will be presented in Chapter 3

## References

- Akritis, M. G., & Van Keilegom, I. (2001). Non-parametric estimation of the residual distribution. *Scand. J. Stat.*, 28(3), 549–567.
- Azzalini, A. (1981). A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika*, 68(1), 326–328.
- Box, G. E., & Muller, M. E. (1958). A note on the generation of random normal deviates. *Ann. Math. Statistics*, 29(2), 610–611.
- Cai, T. T., Levine, M., & Wang, L. (2009). Variance function estimation in multivariate nonparametric regression with fixed design. *J. Multivariate Anal.*, 100(1), 126–136.
- Carroll, R. J., Fan, J., Gijbels, I., & Wand, M. P. (1997). Generalized partially linear single-index models. *J. Amer. Statist. Assoc.*, 92(438), 477–489.
- Delecroix, M., Hristache, M., & Patilea, V. (2006). On semiparametric  $M$ -estimation in single-index regression. *J. Statist. Plann. Inference*, 136(3), 730–769.
- Dominitz, J., & Sherman, R. P. (2005). Some convergence theory for iterative estimation procedures with an application to semiparametric estimation. *Econometric Theory*, 21(4), 838–863.
- Fan, J., & Gijbels, I. (1996). *Local polynomial modelling and its applications* (Vol. 66). Chapman & Hall, London.
- Fan, J., & Yao, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika*, 85(3), 645–660.
- Hall, P., & Carroll, R. J. (1989). Variance function estimation in regression: The effect of estimating the mean. *J. Roy. Statist. Soc. Ser. B*, 51(1), 3–14.
- Hamilton, J. D. (1994). *Time series analysis*. Princeton University Press, Princeton, NJ.
- Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, 24(3), 726–748.
- Hardle, W., Hall, P., & Ichimura, H. (1993). Optimal smoothing in single-index models. *Ann. Statist.*, 21(1), 157–178.
- Härdle, W., Hall, P., & Ichimura, H. (1993). Optimal smoothing in single-index models. *Ann. Statist.*, 21(1), 157–178.
- Härdle, W., Müller, M., Sperlich, S., & Werwatz, A. (2004). *Nonparametric and semi-parametric models*. Springer-Verlag, New York.
- Härdle, W., & Tsybakov, A. (1997). Local polynomial estimators of the volatility function in nonparametric autoregression. *J. Econom.*, 81(1), 223–242.
- Horowitz, J. L. (2009). *Semiparametric and nonparametric methods in econometrics*. Springer, New York.
- Hristache, M., Juditsky, A., & Spokoiny, V. (2001). Direct estimation of the index coefficient in a single-index model. *Ann. Statist.*, 595–623.
- Ichimura, H. (1993). Semiparametric least squares (sls) and weighted sls estimation of single-index models. *J. Econom.*, 58(1-2), 71–120.
- Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist.*, 40, 633–643.

- Koul, H. L., Müller, U. U., & Schick, A. (2017). Estimating the error distribution in a single-index model. *From Statistics to Mathematical Finance: Festschrift in Honour of Winfried Stute*, 209–233.
- Liitiäinen, E., Corona, F., & Lendasse, A. (2010). Residual variance estimation using a nearest neighbor statistic. *J. Multivariate Anal.*, 101(4), 811–823.
- Lopez, O., Patilea, V., & Van Keilegom, I. (2013). Single index regression models in the presence of censoring depending on the covariates. *Bernoulli*, 19(3), 721–747.
- Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *J. Finance*, 29(2), 449–470.
- Muller, H.-G., & Stadtmüller, U. (1987). Estimation of heteroscedasticity in regression analysis. *Ann. Statist.*, 610–625.
- Neumeyer, N., & Van Keilegom, I. (2010). Estimating the error distribution in nonparametric multiple regression with applications to model testing. *J. Multivariate Anal.*, 101(5), 1067–1078.
- Newey, W. K. (1994). The asymptotic variance of semiparametric estimators. *Econometrica*, 62(6), 1349–1382.
- Newey, W. K., & McFadden, D. (1994). Large sample estimation and hypothesis testing. Elsevier.
- Nolan, D., & Pollard, D. (1987).  $U$ -processes: Rates of convergence. *Ann. Statist.*, 15(2), 780–799.
- Pollard, D., & Radchenko, P. (2006). Nonlinear least-squares estimation. *J. Multivariate Anal.*, 97(2), 548–562.
- Powell, J. L., Stock, J. H., & Stoker, T. M. (1989). Semiparametric estimation of index coefficients. *Econometrica*, 57(6).
- Racine, J. S., & Van Keilegom, I. (2020). A smooth nonparametric, multivariate, mixed-data location-scale test. *J. Bus. Econ. Stat.*, 38(4), 784–795.
- Robinson, P. M. (1988). Root- $n$ -consistent semiparametric regression. *Econometrica*, 931–954.
- Roncalli, T. (2020, April). *Handbook of financial risk management* [Book available at: <https://www.routledge.com/Handbook-of-Financial-Risk-Management/Roncalli/p/book/9781138501874>]. Chapman & Hall/CRC Financial Mathematics Series.
- Ruppert, D., Wand, M. P., & Carroll, R. J. (2003). *Semiparametric regression* (Vol. 12). Cambridge University Press, Cambridge.
- Shen, Y., Gao, C., Witten, D., & Han, F. (2020). Optimal estimation of variance in nonparametric regression with random design. *Ann. Statist.*, 48(6), 3589–3618.
- Sherman, R. P. (1993). The limiting distribution of the maximum rank correlation estimator. *Econometrica*, 61(1), 123–137. <https://doi.org/10.2307/2951780>
- Sherman, R. P. (1994a).  $U$ -processes in the analysis of a generalized semiparametric regression estimator. *Econometric Theory*, 10(2), 372–395.
- Sherman, R. P. (1994b). Maximal inequalities for degenerate  $U$ -processes with applications to optimization estimators. *Ann. Statist.*, 22(1), 439–459.
- van der Vaart, A. W. (1998). *Asymptotic statistics* (Vol. 3). Cambridge University Press, Cambridge.

- van der Vaart, A. W., & Wellner, J. A. (2023). *Weak convergence and empirical processes— with applications to statistics* (Second). Springer, Cham. <https://doi.org/10.1007/978-3-031-29040-4>
- Vašíček, O. (2002). The distribution of loan portfolio value. *Risk*, *15*(12), 160–162.
- Wang, L., Brown, L. D., Cai, T. T., & Levine, M. (2008). Effect of mean on variance function estimation in nonparametric regression. *Ann. Statist.*, *36*(2), 646–664.
- Xia, Y., & Härdle, W. (2006). Semi-parametric estimation of partially linear single-index models. *J. Multivariate Anal.*, *97*(5), 1162–1184.
- Yin, J., Geng, Z., Li, R., & Wang, H. (2010). Nonparametric covariance model. *Statist. Sinica*, *20*(1), 469–479.
- Zhu, L., Dong, Y., & Li, R. (2013). Semiparametric estimation of conditional heteroscedasticity via single-index modeling. *Statist. Sinica*, *23*(3), 1235–1255.

## Chapter 3

# Density estimation based on quantile regressions applied to credit risk

**Abstract :** This chapter introduces a method for estimating the conditional density of a random variable using a flexible parametric model, where the parameter vector can depend on predictors. The parameter is adjusted by matching conditional quantiles. The concept is based on a seminal paper by Adrian et al. (2019), which proposes a two-step estimation approach, and applies the idea to GDP growth. In this chapter we propose to introduce this methodology, present a package publicly available on CRAN which reproduce the results of Adrian et al. (2019) and apply this methodology for estimating extreme losses on a credit portfolio on real data. This chapter is extracted from an article under review.

## 3.1 Introduction

In Chapter 2 we presented a methodology to estimate the distribution function of a random variable of interest. We proposed a semi parametric location scale model where both the mean and covariance functions were estimated non parametrically. This non-parametric hypothesis may be complicated to apply in practice due to its limited ability to forecast observations outside the range of the sampled data. This limitation is particularly problematic in stress-testing scenarios, where covariates need to be pessimistic and plausible, and thus often fall outside the observed sample range. For instance, the 2023 scenario proposed by the ESRB<sup>1</sup> involves a recession between  $-3$  and  $-5\%$  per year, an inflation jumping to  $10\%$  in 2024 and declining slowly to approximately  $5\%$  by 2026 and a record long-term interest rate of  $10\%$  by 2024. To the best of my knowledge a scenario of  $10\%$  inflation rate and long term interest and a  $5\%$  recession is unprecedented and therefore the local linear model may be unreliable.

Moreover, the interpretation of such a model may be complicated not only for the risk management team but also for validation team and the regulator who supervise the validity of the models used by banks. The complexity arises from several factors. First, the lack of closed-form expressions for the estimated functions can make it difficult to communicate and understand the model's behavior and, the challenges in forecasting extreme or out-of-sample scenarios, as mentioned earlier, can lead to uncertainty and skepticism regarding the model's reliability.

Therefore, this chapter presents an alternative approach similar to that of Adrian et al. (2019), that address the interpretation and practical implementation challenges discussed earlier. This approach introduces several additional assumptions regarding the distribution, notably that the relationship between a quantile of the distribution and the covariates can be described by a linear relationship. By incorporating these assumptions, we aim to enhance the model's interpretability and facilitate its application in practice, particularly in stress-testing scenarios.

In this seminal paper, Adrian et al. (2019) proposed a two-step estimation approach. The first step involves estimating the distribution of the target using standard quantile regression for a finite set of quantiles. The second step involves fitting a parametric distribution by interpolating between the previously obtained quantiles and the theoretical ones, specifically in the case of Adrian et al. (2019) a skew-student distribution is used. This estimating strategy can be viewed as the construction of multiple functions linking the covariates and the parameters of the chosen distribution in the second step. Finally, given these estimated conditional distributions, we compute time-varying forecasts for any distribution-based risk measure, such as VaR, expected shortfall (ES, thereafter). In practice, VaR and ES are the most often used conditional risk measures, which is why the term at-Risk is often used.

---

<sup>1</sup>All scenario are available in this documentation : [https://www.bankingsupervision.europa.eu/ecb/pub/pdf/ssm.Report\\_2023\\_Stress\\_Test~96bb5a3af8.en.pdf](https://www.bankingsupervision.europa.eu/ecb/pub/pdf/ssm.Report_2023_Stress_Test~96bb5a3af8.en.pdf)

Another advantage of the approach proposed by Adrian et al. (2019) is that, although the model is recent, it has already been adopted by regulators and financial institutions across many jurisdictions. For instance, the IMF used the Growth at Risk model to design a macroprudential stress-test scenario in south America<sup>2</sup>. Therefore, from a financial institution's perspective, implementing such models in risk management activities poses fewer risks. Since regulators are already familiar with the model, it will require less justification for its use.

This chapter will be organized as follows, first we will present the estimation strategy, then we will present results on real data obtained from a French bank. The goal of this Chapter is not to justify the performance of such a model but rather present how can a risk management team may implement this method and what are the results that can be obtained. Moreover our goal is to compare the output of the model presented in Chapter 2 and this other model.

## 3.2 Presentation of the methodology

When one comes to forecasting a variable  $Y_{t+h}$  given the information set available at time  $t$ ,  $X_t$ , the most complete information a forecaster can provide on the distribution of  $Y_{t+h}$  is summarized by the pdf of said conditional distribution, denoted by  $f_{Y_{t+h}|X_t}(y)$ . When this conditional probability distribution depends on a set of parameters  $\theta$ , the corresponding parametric pdf is denoted by  $f_{Y_{t+h}|X_t}(y; \theta)$ . Within this context, density forecasting simply consists of plotting the pdf  $f_{Y_{t+h}|X_t}(y; \theta)$  for any value  $y \in \Omega$  and time  $t$ , where  $\Omega$  is the set of admissible values for the variable. Then, the information contained in  $f_{Y_{t+h}|X_t}(y; \theta)$  may be used to compute confidence intervals, high-density regions, point forecasts, etc.

This conditional density may be used to forecast any downside or upside risk measure at time  $t + h$ . The VaR or the ES most often used in finance to measure the risk of a credit portfolio. The VaR is a metric that quantifies the potential loss of a portfolio over a given time horizon for a coverage rate  $\tau \in ]0, 1[$ . The concept of VaR can be extended to any stationary economic variable. Adrian et al. (2019) proposed to use this risk measure as a metric for macroeconomic risk by introducing the concept of Growth-at-Risk, i.e., the VaR of the GDP growth rate. Similarly, we can define inflation-at-Risk (IMF (2022)), unemployment-at-Risk, real-estate-at-Risk, etc. Whatever the economic variable considered, the VaR is formally defined as the  $\tau$ -quantile of the conditional distribution.

$$\text{VaR}_{Y_{t+h}|X_t}(\tau) = F_{Y_{t+h}|X_t}^{-1}(\tau; \theta),$$

where  $F_{Y_{t+h}|X_t}(y; \theta)$  is the cdf associated with the conditional distribution, we assume that the conditional cumulative distribution function is continuous. By definition, the conditional VaR satisfies the following property:

$$\mathbb{P}(Y_{t+h} \leq \text{VaR}_{Y_{t+h}|X_t}(\tau) | X_t) = \tau.$$

---

<sup>2</sup>For more details, see : <https://www.imf.org/en/Publications/Departmental-Papers-Policy-Papers/Issues/2020/01/31/Stress-Testing-at-the-IMF-48825>

The interpretation is obvious: if this distribution holds, there is a probability  $\tau$  chances that the considered variable (e.g., growth rate, inflation, etc.) would be smaller than the value  $\text{VaR}_{Y_{t+h}|X_t}(\tau)$  at horizon  $t+h$  given the information available at time  $t$ . Two remarks should be made here. First, contrary to what is generally done in finance, we do not define the VaR as the *opposite* of the quantile.<sup>3</sup> Second, we do not index the VaR by the confidence level  $1-\tau$  but by the risk  $\tau$ .<sup>4</sup> Hence, we do not follow the usual conventions because the risk can be in either the right tail (for example, for PD or the inflation rate) or the left tail (for example, for the GDP growth rate) of the distribution, depending on the economic variable being considered. The VaR is indexed by the risk of being “lesser than” a certain value. This will avoid confusion, because a 99%-VaR means that there is a 99% chance of getting a realization smaller than this value.

The second risk measure is the ES. The ES at level  $\tau$  is defined as the expected value of  $y_{t+h}$  conditional on being smaller than the VaR defined at the confidence level  $\tau$ .

$$\text{ES}_{Y_{t+h}|X_t}(\tau) = \mathbb{E}(Y_{t+h} | Y_{t+h} \leq \text{VaR}_{Y_{t+h}|X_t}(\tau)) = \frac{1}{\tau} \int_0^\tau \text{VaR}_{Y_{t+h}|X_t}(\gamma) d\gamma$$

Thus, the ES can be interpreted as the average value obtained in the worst  $\tau$  of cases, where  $\tau$  is the confidence level. The expected shortfall can significantly differ from the VaR if the distribution has very heavy left tail. Contrary to the VaR, the ES is a coherent risk measure (Artzner et al. (1999)), as it satisfies the subadditivity property. Furthermore, this measure is more sensitive than the VaR to a heavy tailed loss distribution. In 2013, the Basel Committee proposed shifting from the VaR to the ES at 2.5% for the internal models approach, which is now part of the new standards for the minimum capital requirements for market risk. As for the VaR, this risk measure can be extended to any stationary economic variable. For instance, we can define a GDP growth-expected shortfall or an inflation-expected shortfall.

While density forecasting is attractive in theory, it is complex to implement in practice. Indeed, it implies relating a time-varying pdf, i.e.,  $f_{Y_{t+h}|X_t}(y; \theta)$ , to a set of factors  $X_t$  at each time  $t$ . As we have only one realization of both  $Y_{t+h}$  and  $X_t$  for each period  $t$ , the identification and estimation of  $f_{Y_{t+h}|X_t}(y; \theta)$  is impossible without parametric or non-parametric assumptions. In this context, the aR methodology offers a fairly simple technical solution that allows the conditional density of a macroeconomic variable to be fully characterized as a parametric function of a set of factors. Then, in a second step, this density can be used to derive risk measure forecasts, especially VaR forecasts. The aR

---

<sup>3</sup>In finance, when we consider returns or profit and loss distributions, the  $\tau$ -quantile is generally a negative value when  $\tau$  is small (e.g., 1% or 5%) as it corresponds to an extreme loss. Then the VaR is usually defined as the opposite of the quantile to avoid to report a minus sign. In contrast, for a macroeconomic variable, the quantile for small or high probability values, may be positive or negative in this case.

<sup>4</sup>Usually in finance, the VaR is indexed by the confidence level, i.e.,  $\text{VaR}_{Y_{t+h}|X_t}(1-\tau) = F_{Y_{t+h}|X_t}^{-1}(\tau; \theta)$  with  $\mathbb{P}(Y_{t+h} \leq \text{VaR}_{Y_{t+h}|X_t}(1-\tau)|X_t) = \tau$ .

approach was initially proposed by Adrian et al. (2019) to model the full distribution of the GDP growth rate as a function of financial and economic conditions and to show that lower quantiles of the distribution are more sensitive to financial conditions than upper quantiles.<sup>5</sup> Their results show that the mean of GDP growth is negatively correlated with measures of financial volatility and that periods of low volatility are often followed by negative growth outcomes. Their results have important implications for the relationship between financial stability and macroeconomic performance. They indicate that GDP vulnerability changes frequently, which is inconsistent with the literature on disaster risk and economic growth but consistent with evidence from the term structure of asset prices across multiple markets. They also find that GDP growth volatility is largely driven by the left side of the conditional distribution and attribute the decline in volatility during the period before the financial crisis to a decline in downside risk to GDP growth.

### 3.3 aR approach

#### 3.3.1 A general framework

The aR approach introduced by Adrian et al. (2019) is a two step semi-parametric methodology that allows (i) forecasting the conditional density of a macroeconomic variable at horizon  $h$  given some factors  $x$  known at time  $t$ , and (ii) forecasting the associated time-varying downside or upside risk measures, typically conditional VaR and ES. Adrian et al. (2019) focused on the Growth-at-Risk, namely, the VaR defined on the GDP growth rate. Here, we generalize their approach by considering any stationary economic variable  $y_t$  and any risk measure (VaR, ES, expectiles, downside entropy, etc.).

1. **Quantile regressions:** consider a dataset  $\{Y_{j+h}, X_j\}_{j=1, \dots, t}$  and a set of  $k$  probabilities  $\tau_1, \dots, \tau_k$ , e.g., 5%, 25%, 75%, and 95% for  $k = 4$ . For each probability, we run a quantile regression to estimate a linear relationship between the  $\tau_j$ -quantile of  $Y_{t+h}$  and a set of explicative variables  $X_t$ . Denote the corresponding *empirical* quantiles by  $\hat{Q}_{Y_{t+h}|X_t}(\tau_j)$  for  $j = 1, \dots, k$ .
2. **Parametric assumption:** Consider a parametric distribution  $f_{Y_{t+h}|X_t}(y; \theta)$  associated with a set of  $m$  parameters  $\theta$ . This pdf should be sufficiently flexible to capture the main stylized facts of the empirical distribution of  $y$ , e.g., asymmetry or kurtosis. In the sequel, we consider a skewed t-distribution (Azzalini and Capitanio (2003)). From this parametric function, we can deduce the *theoretical* quantile for any probability  $\tau$ , denoted as  $Q_{Y_{t+h}|X_t}(\tau; \theta)$ .
3. **Quantile smoothing:** The parameters  $\theta$  are estimated by equalizing the empirical quantiles to their theoretical counterparts. Formally, we search for parameters  $\theta$

---

<sup>5</sup>The financial conditions are represented by the NFCI indicator of Brave and Butters (2018). This indicator is based on the receiver operating characteristic (ROC) curve associated with predictive models of financial imbalances and future financial stress.

that minimize the distance between  $\hat{Q}_{Y_{t+h}|X_t}(\tau_j)$  and  $Q_{Y_{t+h}|X_t}(\tau_j; \theta)$  for  $j = 1, \dots, k$ . For  $k = m$ , the system is just identified and the estimation method is similar to a method of classical moments (L. P. Hansen (1982)). Denote the corresponding estimate by  $\hat{\theta}_t$ .

4. **Density forecasts and risk measures:** The pdf  $f_{Y_{t+h}|X_t}(y; \hat{\theta}_t)$  entirely represents the conditional probability distribution of  $Y_{t+h}$  and its dependence on the factors  $X_t$ . The density forecast of  $Y_{t+h}$  is simply given in closed form by the function  $f_{Y_{t+h}|X_t}(y; \hat{\theta}_t)$ . From this function, we can deduce the forecasts of any downside risk measure, particularly the  $\text{VaR}_{Y_{t+h}|X_t}(\tau)$  and  $\text{ES}_{Y_{t+h}|X_t}(\tau)$  for any coverage rate  $\tau$ . Note that the coverage rate  $\tau$  can be different from the probabilities  $\tau_1, \dots, \tau_k$  initially selected to calibrate the distributional parameters.
5. **Time-varying risk measures:** By increasing the dataset to  $\{Y_{j+h}, X_j\}_{j=1, \dots, t+1}$ , we can compute the conditional pdf for  $Y_{t+h+1}$  given the information at time  $t+1$  and the corresponding risk measures, and so on. By repeating steps 1 to 4, we obtain a sequence of time-varying density forecasts and risk measures for an horizon of  $h$  periods at time  $t = 1, \dots, T$ .

Several remarks can be made here. First, optimizing the parameters of a parametric cumulative distribution function to fit the quantiles is central in the aR methodology. It is well known that the accuracy of quantile regressions may be poor, especially when one considers extreme quantiles, i.e., very high or very low coverage rates. Then, the VaR or ES estimates issued from the quantile regressions (step 1) may be not sufficient to well represent the tail relationship between  $y$  and  $x$ . In particular, we could observe the crossing of empirical quantile regression curves.<sup>6</sup> Then, a natural solution consists of estimating only a few quantiles ( $k = 4$  in Adrian et al. (2019)) and then, linking them through a function. This function is simply defined as a parametric density function. Then, by design, the quantiles and risk measures can be computed, in particular, there is no crossing of the quantile curve and  $F^{-1}$  is well defined.

Second, when the number of parameters to be estimated  $m$  is equal to the number  $k$  of conditions on the quantiles, the estimation method of parameters  $\theta$  can be interpreted as a method of classical moments. Each condition consists of equalizing the empirical  $\tau_j$ -quantile, estimated by a quantile regression, to the theoretical  $\tau_j$ -quantile. An extension consists of using more conditions than parameters to estimate, i.e.,  $k > m$ .

Third, the aR approach can be used to produce ES or VaR for very low/high coverage rates (extreme risks), e.g., 0.001 or 0.999. This allows us to represent the dynamics of extreme risks that could occur on  $y$ , which is consistent with a stress-test exercise. The main advantage of the aR methodology is that it does not require the estimation of these

---

<sup>6</sup>The crossing quantile curves issue occurs when the estimated quantile curves for different quantile levels cross each other. This is a violation of the monotonicity property of quantile functions, which states that the quantile function should be non-decreasing in the quantile level.

extreme value quantiles directly through quantile regressions. In contrast, the latter are generally considered only for medium coverage rates (Adrian et al. (2019)).

Several improvements of the aR methodology have been proposed since the seminal paper of Adrian et al. (2019). Among many others, we can cite Chavleishvili et al. (2021) who propose a structural quantile vector autoregressive model to quantify future risks to economic activity caused by elevated levels of financial stress as well as by economic vulnerability to shocks. In a recent paper, Lhuissier (2022) extends the Growth-at-Risk (GaR) model to study the financial conditions and macroeconomic downside risks in the euro area. His extension is based on a regime-switching skew-normal model with time-varying transition probabilities. While Adrian et al. (2019) consider a two-step quantile approach that requires fitting a skew-t distribution to the estimated quantiles, Lhuissier (2022) directly estimates the time-varying parameters of a skew-normal distribution within a Markov-switching framework. This direct approach allows great flexibility and tractability in modelling time variation in risk measures as a function of economic and financial conditions.

### 3.3.2 Quantile regressions

The first step of the aR approach is to perform  $k$  successive quantile regressions. Developed by Koenker and Bassett (1978), the quantile regression consists of estimating the conditional  $\tau$ -quantile of the dependent variable  $y$  as a linear function of a set of explicative variables  $x$ . The quantile regression differs from a standard linear regression model as it does not focus on the mean but on the quantiles associated with the conditional distribution of the dependent variable. In other words, a quantile regression allows us to estimate the conditional quantile function of the dependent variable  $y$ . Consider the same framework as in Section 3.1 and denote  $G_{y_{t+h}|x_t}(y)$  to be the cdf of the true (unknown) conditional distribution of the macroeconomic variable  $y_{t+h}$ , the corresponding conditional quantile function is given by:

$$Q_{Y_{t+h}|X_t}^G(\tau) = G_{Y_{t+h}|X_t}^{-1}(\tau) = \inf(Y_{t+h} \in \Omega : G_{Y_{t+h}|X_t}(Y_{t+h}) \geq \tau) \quad \forall \tau \in [0, 1].$$

The true quantile function is unobserved. However, each quantile can be estimated by a quantile regression that assumes a linear relationship between the  $\tau$ -quantile and a linear index of explanatory variables  $X_t$ , such that:

$$Q_{Y_{t+h}|X_t}(\tau) = X_t^\top \beta_\tau.$$

Note that the parameters  $\beta_\tau$  are specific to the coverage rate, i.e., there is one quantile regression for each probability  $\tau$ . For an aR-type model, we run  $k$  quantile regressions for  $\tau_1, \dots, \tau_k$ . The more  $\tau$  is chosen close to 0 or 1, the more the estimation will require a larger dataset to be precise. Thus, we generally avoid to considering extreme probabilities. Furthermore, as the aim of the aR model is to reveal the full conditional distribution of the dependent variable, it is recommended to consider several probabilities that cover the

whole space between 0 and 1. Typically, Adrian et al. (2019) consider  $k = 4$  coverage rates, namely, 5%, 25%, 75%, and 95%.

Consider a dataset  $\{y_{j+h}, x_j\}_{j=1, \dots, t}$ . Then, the conditional quantile regression slope  $\beta_\tau$  is chosen to minimize the quantile weighted absolute value of errors:

$$\hat{\beta}_\tau = \arg \min_{\beta \in \mathbb{R}^q} \sum_{j=1}^{t-h} \left( \tau \mathbb{1}_{(Y_{j+h} \geq X_j^\top \beta)} |Y_{j+h} - X_j^\top \beta| + (1 - \tau) \mathbb{1}_{(Y_{j+h} < X_j^\top \beta)} |Y_{j+h} - X_j^\top \beta| \right).$$

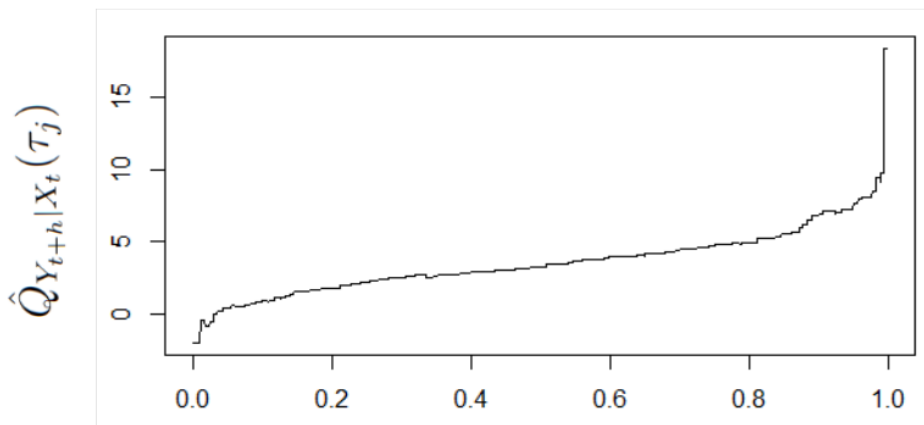
where  $q$  is the number of explicative variables and  $\mathbb{1}_{(\cdot)}$  is an indicator function equal to 1 if the condition in parentheses is valid and 0 otherwise. The output of the  $\tau$ -quantile regression is the *empirical* conditional quantile of  $y_{t+h}$  defined as:

$$\hat{Q}_{Y_{t+h}|X_t}(\tau) = X_t^\top \hat{\beta}_\tau.$$

It is important to note that the quantile regression for a medium coverage rate generally requires approximately 50 to 80 observations, i.e., 12 to 15 years of quarterly data or more than 4 years of monthly data. However, in IMF (2022), the authors propose an alternative estimation method for small and noisy samples, typically with less than 30 points. This method is based on (i) the Theil-Sen robust estimator (Sen (1968) and Theil (1950)), which improves the OLS estimator to make it more accurate for estimation on small samples and is robust to outliers, and (ii) the ‘‘Firth Logit’’ model (Firth (1993)). The latter is a logistic regression model with penalized likelihood. The Firth model estimates the binary probability of an event, which is similar to the classical logistic model. In the case of the aR model, this event is thus coded as being a dependent variable higher than a given value.

One of the main advantages of the aR model is that it requires only a few quantile regressions to reveal the full probability distribution of the dependent variable. Indeed, to determine the probability distribution function, we would alternatively consider performing quantile regressions for an infinite (or very large, in practice) number of quantile levels  $\tau_j \in [0; 1]$ . However, such an approach would be impossible in practice: the quantile estimates would be too difficult to map into a probability distribution function because of approximation error and estimation noise. Indeed, for two quantiles associated with  $\tau_1$  and  $\tau_2$ , with  $\tau_1 = \tau_2 + \delta$  and  $\delta$  in a neighbourhood of 0, the estimated probability distribution function might not be increasing and monotonic. To illustrate this quantile crossing issue, we consider the same dataset and the same specification as Adrian et al. (2019) for the estimation of the GaR for the US between 1973 and 2018. The dependent variable is the GDP quarterly growth rate. The quantile regression includes a constant and the Chicago Fed’s National Financial Conditions Index (NFCI) with a one lag as explanatory variables. We run 1,000 quantile regressions for probabilities between 0 and 1 with a step size of 0.1%. The corresponding estimated quantiles are reported in Figure 3.1. We observe a quantile curve crossing: the estimated quantile function is not increasing and monotonic. This result illustrates the usefulness of the smoothing considered in the aR approach.

Figure 3.1: Estimated quantile function without smoothing



Note: this figure displays the estimated quantiles obtained with quantile regressions of the GDP growth rate on the lagged NFCI (US, 1973-2018). Quantile are reported on the x-axis, and the estimated quantiles are reported on the y-axis. We estimated the quantiles for  $\tau_j = \{0.01, 0.02, \dots, 0.99\}$ ,  $h$  corresponds to one quarter and  $t$  corresponds to the first quarter of 1975.

The second main step of the aR methodology is the parametric smoothing of the quantiles, which allows us to solve the quantile crossing issue and to produce coherent estimates of the conditional distribution of the dependent variable. The intuition of the smoothing procedure is simple and appealing. Consider a parametric probability density function sufficiently flexible to capture the main stylized facts of the distribution of the dependent variable. Then, we estimate its parameters by matching the corresponding theoretical quantiles to the empirical quantiles issued from the quantile regressions. The idea here is similar to the well known indirect inference method (Gourieroux et al. (1993)). Once the parameters have been estimated, we obtain a parametric pdf that, by design, can be used to compute any quantile without any crossing issue, including at the extreme tails of the distribution.

Depending on the series considered, several density functions can be used as parametric smoothing functions, e.g., GED distribution, skew normal distribution, etc. Adrian et al. (2019) consider a skewed t-distribution (Azzalini and Capitanio (2003)) that allows us to capture both asymmetry and kurtosis in the distribution. The corresponding pdf is given by:

$$f(y; \mu, \sigma, \alpha, \nu) = \frac{2}{\sigma} t\left(\frac{y - \mu}{\sigma}; \nu\right) T\left(\alpha \frac{y - \mu}{\sigma} \sqrt{\frac{\nu + 1}{\nu + \left(\frac{y - \mu}{\sigma}\right)^2}}; \nu + 1\right),$$

where  $t(\cdot)$  and  $T(\cdot)$  correspond to the pdf and cdf of the t-Student distribution respectively. The four parameters of the distribution are the mean  $\mu$ , the standard deviation  $\sigma$ , the degrees of freedom  $\nu$ , and the skewness parameter  $\alpha$ .

More generally, consider a parametric conditional distribution denoted by  $f_{Y_{t+h}|X_t}(y; \theta)$ , where  $\theta \in \Theta \subset \mathbb{R}^m$  is the vector of parameters. For a given value of  $\theta$ , it is possible to derive a closed form or a numerical approximation of the *theoretical* quantile function, defined as:

$$Q_{Y_{t+h}|X_t}(\tau; \theta) = F_{Y_{t+h}|X_t}^{-1}(\tau; \theta) = \inf(Y_{t+h} \in \Omega : F_{Y_{t+h}|X_t}(Y_{t+h}; \theta) \geq \tau) \quad \forall \tau \in [0, 1],$$

where  $F_{Y_{t+h}|X_t}(Y_{t+h}; \theta)$  is the cdf associated with the conditional distribution. As soon as  $\theta$  is known, we fully characterize the conditional distribution of  $Y_{t+h}$ , and we can compute the corresponding risk measures, i.e., VaR and ES.

How to estimate  $\theta$ , the parameter vector? As previously mentioned, Adrian et al. (2019) propose a method that consists of matching the empirical quantiles  $\hat{Q}_{Y_{t+h}|X_t}(\tau_j)$  issued from the quantile regressions, to the theoretical quantiles  $Q_{Y_{t+h}|X_t}(\tau_j; \theta)$ , for a set of  $k$  coverage rates  $\tau_1, \dots, \tau_k$ . If  $k = m$ , the method implies solving the following program:

$$\hat{\theta}_t = \arg \min_{\theta} \sum_{j=1}^k \left( \hat{Q}_{Y_{t+h}|X_t}(\tau_j) - Q_{Y_{t+h}|X_t}(\tau_j; \theta) \right)^2.$$

Note that this estimation method gives equal weight to the  $k$  quantiles conditions. When the number of quantiles  $k$  is larger than the number of parameters  $m$ , we can use a GMM estimator (L. P. Hansen (1982)), with:

$$\hat{\theta}_t = \arg \min_{\theta} M(\tau; \theta)' (\mathbb{V}(M(\tau; \theta)))^{-1} M(\tau; \theta),$$

with  $M(\tau; \theta) = (m(\tau_1; \theta), \dots, m(\tau_k; \theta))'$  and  $m(\tau_j; \theta) = \hat{Q}_{Y_{t+h}|X_t}(\tau_j) - Q_{Y_{t+h}|X_t}(\tau_j; \theta)$ .

In their seminal application for the GaR model, Adrian et al. (2019) consider  $m = 4$  parameters  $(\mu_t; \sigma_t; \alpha_t; \nu_t)$  and  $k = 4$  coverage rates, namely 5%, 25%, 75%, and 95%. For each quarter, the 4 parameters of the skewed t-distribution are chosen to minimize the following function:

$$\{\hat{\mu}_t, \hat{\sigma}_t, \hat{\alpha}_t, \hat{\nu}_t\} = \arg \min_{\mu, \sigma, \alpha, \nu} \sum_{j=1}^4 \left( \hat{Q}_{Y_{t+h}|X_t}(\tau_j) - F^{-1}(\tau_j; \mu, \sigma, \alpha, \nu) \right)^2,$$

where  $F(y; \mu, \sigma, \alpha, \nu)$  is the cdf of the skewed t-distribution.

In any consideration of the estimation method used, the conditional distribution of the forecasts  $Y_{t+h}$  given the information available at time  $t$  is summarized by the function  $f_{Y_{t+h}|X_t}(\cdot; \hat{\theta}_t)$ . This estimated pdf can be used to produce coherent VaR or ES forecasts without crossing quantiles for coverage rates that are different from those initially considered for the quantiles regressions. To illustrate this idea, we report the estimated density associated with the US GDP growth rate at a given date on Figure 3.2. The vertical dotted lines represent the 4 empirical quantiles at 5%, 25%, 75%, and 95% issued from

the quantile regressions (first step). The vertical yellow lines represent the theoretical quantiles deduced from the skewed t-distribution with the optimal parameters  $\hat{\theta}_t$ . By definition, the empirical and theoretical quantiles are confused. Moreover, we report the quantile at 1% that can be deduced from the density forecast  $f_{Y_{t+h}|X_t}(Y_{t+h}; \hat{\theta}_t)$  on the right panel of Figure 3.2. This extreme quantile has not been estimated by a quantile regression but is deduced from the smoothing procedure. This clearly illustrates the usefulness of the aR methodology.

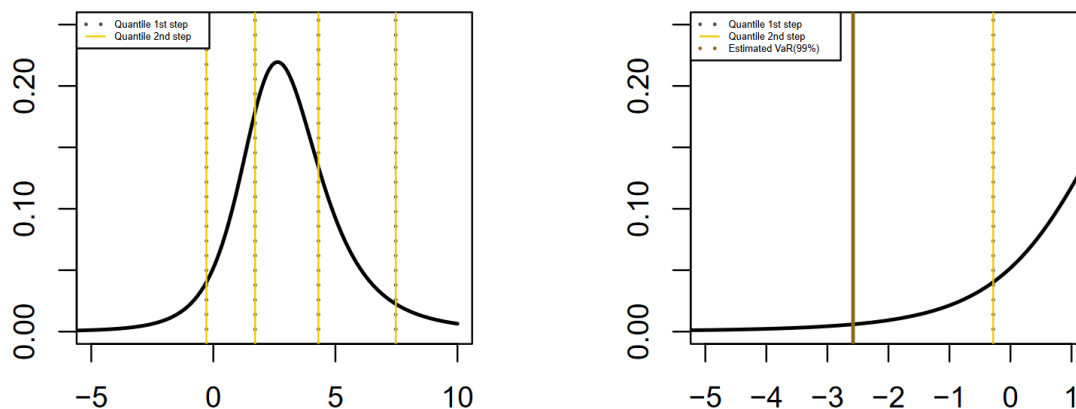


Figure 3.2: This figure compares the quantiles obtained after the first and second step for the model presented by Adrian et al. (2019) we consider the last observed period. We note that these quantiles closely match. Also, this figure represents the obtained conditional distribution for the last observation of the sample and a VaR is also computed.

### 3.4 Application on real data

In the context of credit risk management, we defined the expected loss as the amount of loan losses that can be reasonably forecasted based on historical data, current economic conditions, and other quantifiable factors. Expected loss is typically calculated after computing the probability of default (PD) the expected loss given default (LGD) and the Exposure at Default (EAD). In contrast, we defined the unexpected loss as the amount of losses that exceed expectations due to either unforeseen events or changes in market conditions or because of statistical model uncertainty. Therefore unexpected losses are inherently difficult to predict and require credit risk managers to maintain a buffer against potential losses, ensuring the bank’s financial stability and resilience in the face of unexpected events.

Let us recall some financial vocabulary. First, the through-the-cycle (TTC) PD represents the average probability of default of a borrower over a long period of time, smoothing out the effects of economic cycle, for a precise definition of this concept, and the implication of the Basel agreement, see Aguais (2008). This metric is ideal for estimating EL. Unex-

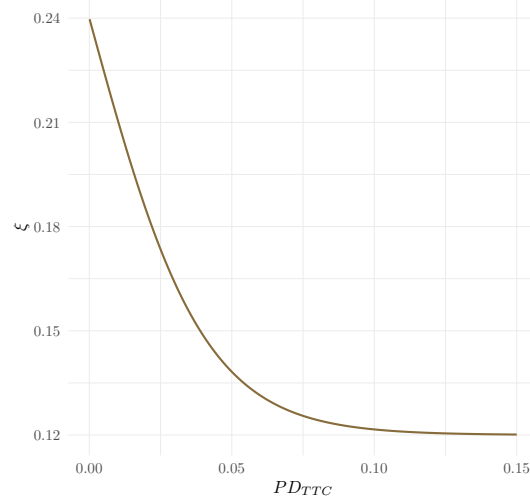


Figure 3.3: Illustration of (3.1) for  $PD_{TTC}$  varying between 0 and 0.15.

pected losses, however, focus on potential deviations from the average due to unforeseen circumstances. The point in time PD reflects the borrower's default risk at a specific moment, incorporating the current (or future) economic climate. By using PIT PD, banks can account for the increased risk associated with economic downturns, leading to a more accurate picture of potential losses exceeding expectations. In essence, TTC PD can be used for computing EL, while PIT PD may be more suitable when modeling unexpected losses.

Specifically, the point in time probability of default  $PD_{PIT}$  given economic condition  $y$ , and the trough the cycle probability of default  $PD_{TTC}$  are linked as follows :

$$PD_{PIT}(y) = \pi(y), \quad PD_{TTC} = \mathbb{E}_Y[\pi(y)].$$

The goal of this section is to derive and estimate the function  $\pi$ . In the case of the regulatory model, this function is :

$$PD_{PIT}(y) = \Phi \left( \frac{\Phi^{-1}(PD_{TTC}) - \sqrt{\xi} \cdot y}{\sqrt{1 - \xi}} \right).$$

This model allows to define a pure PIT probability of default from a TTC probability of default given the realization of some macroeconomic scenarios, indeed, under the Basel III regulatory formula, the correlation factor can be defined as a function of the TTC Probability of default :

$$\xi = \frac{0.12 \times (1 - \exp(-50 \times PD_{TTC}))}{1 - \exp(-50PD_{TTC})} + \frac{0.24 \times [1 - (1 - \exp(-50 \times PD_{TTC}))]}{1 + \exp(-50 \times PD_{TTC})}. \quad (3.1)$$

The underlying rationale is that larger companies, with inherently lower TTC PDs due to the diversification of their markets, tend to be more susceptible to broad economic

downturns because of this diversification. This translates to a stronger correlation between their default risk and the common factor. Conversely, smaller companies, often with higher TTC PDs due to less diversification, might operate in specific markets less affected by the economic condition. This limited exposure weakens the correlation between their default risk and the common factor, Figure 3.3 represents this regulatory formula.

Therefore, on a given portfolio we can estimate the correlation factor and the TTC probability of default. The former is given by (3.1), the regulatory formula, and the latter can be estimated by an empirical estimator :

$$\widehat{PD}_{TTC} = \frac{1}{T} \sum_{t=1}^T PD_{PIT,t}.$$

Where  $T$  is the number of observed periods of time. Therefore in model the only unknown random variable is the common factor  $y$ . The goal of the following sections will be to estimate the conditional distribution of the common factor using the methodologies presented in Sections 2.3.3 and 3.2.

### 3.4.1 Application on S&P500 data

We will apply our method on the ratings published by S&P500 on publicly traded and non publicly traded company. They rate 5718 unique companies over the last decade. We observe 78816 ratings over the span of this time horizon allocated across more than 350 categories. However most of them are included in the following rates :  $AAA$ ,  $AA$ ,  $A$ ,  $BBB$ ,  $BB$ ,  $B$ ,  $C$  and  $D$  the latter meaning that the company defaulted on some or all of its debt. Traditionnally, the rates  $AAA$  to  $BBB$  indicates investment grades company, that are seen as less risky and the company rated from  $BB$  to  $CCC$  are seen as risky and known as speculative grades. The analyses of different companies occurs at different times.

The most recent ratings has been made in july 2022 and the oldest was recorded in 1987. However, we observe only a few ratings each year until june 2010 where it become necessary for rating company to publish there ratings. From june 2010 to july 2022 we observe between 147 and 1199 ratings. The maximum behind attained during the covid crisis. Therefore we are able to estimate a migration matrix every month (or less) for 12 years. Our goal is to obtain the migration matrix where the term  $p_{G,g,t}$  represents the true probability during period  $t$  to go from rating  $G \in \{AAA, AA, A, BBB, BB, B, C, D\}$  to rating  $g \in \{AAA, AA, A, BBB, BB, B, C, D\}$ , the period is a quarter in this case. Moreover the grades are an ordered set where  $g + 1$  means one higher rating than  $g$ . For instance, if  $g = AA$  then  $g + 1 = AAA$ . Also note that we will assume there are no possible transition from the default state, if a company defaulted on its debt it can not operate again which is not consistent with the data (see Table 3.1) but is consistent with the Merton-Vašíček model. In our dataset we observe  $n_{g,t}^G \in \mathbb{N}$ , transitions from rating  $G$  to  $g$  and  $n_t = \sum_{G,g} n_{g,t}^G$ , transitions during period  $t$ . Given our dataset we can estimate

CHAPTER 3. DENSITY ESTIMATION BASED ON QUANTILE REGRESSIONS  
APPLIED TO CREDIT RISK

	<i>A</i>	<i>AA</i>	<i>AAA</i>	<i>BBB</i>	<i>D</i>	<i>BB</i>	<i>B</i>	<i>C</i>	Other	NR	n_obs
<i>A</i>	0.7117	0.0256	0.0303	0.1076	0.0010	0.0219	0.0569	0.0042	0.0010	0.0397	1915
<i>AA</i>	0.1461	0.5959	0.0258	0.0831	0.0029	0.0544	0.0401	0.0090	NA	0.0429	349
<i>AAA</i>	0.1440	0.0194	0.7535	0.0360	NA	NA	0.0194	NA	NA	0.0277	361
<i>B</i>	0.0347	0.0076	0.0019	0.0698	0.0025	0.0912	0.6567	0.0289	0.0005	0.1045	2106
<i>BB</i>	0.0344	0.0029	0.0004	0.0993	0.0011	0.6822	0.1004	0.0098	0.0023	0.0364	1743
<i>BBB</i>	0.0865	0.0092	0.0048	0.7098	0.0013	0.0656	0.0647	0.0056	0.0022	0.0506	2288
<i>C</i>	0.0276	0.0075	NA	0.0498	0.0398	0.0647	0.1162	0.6244	NA	0.0697	402
<i>D</i>	0.0462	0.0154	NA	0.0462	0.2154	0.0769	0.1538	NA	NA	0.3846	65
NR	0.0998	0.0187	0.0031	0.1482	0.0151	0.1389	0.2871	0.0421	0.0005	0.2418	641
Other	NA	NA	NA	0.4211	NA	0.1579	0.1053	NA	0.2632	0.0526	19

Table 3.1: Transition Matrix

probability  $p_{G,g,t}$  with:

$$\hat{p}_{G,g,t} = \frac{1}{n_t} \sum_{n_t} n_{g,t}^G.$$

Following Belkin et al. (1998) we define the probability of transition from rating  $G$  to rating  $g$  in the Merton-Vasicek model.

$$\Delta(x_g^{G+1}, x_g^G, y_t) = \Phi\left(\frac{x_g^{G+1} - \sqrt{\xi} y_t}{\sqrt{1 - \xi}}\right) - \Phi\left(\frac{x_g^G - \sqrt{\xi} y_t}{\sqrt{1 - \xi}}\right).$$

Where  $x_g^{G+1}$  represents barrier between different ratings. They are defined as in Belkin et al. (1998). Intuitively, these barriers are defined on the through-the-cycle migration matrix,  $x_g^{G+1}$  is defined based on the TTC transition probability between ratings  $g$  and  $G + 1$ . The goal is then to find the common factor by minimizing the following problem :

$$\min_{y_t} \sum_G \sum_g \frac{n_{t,g} [\hat{p}_t(G, g) - \Delta(x_g^{G+1}, x_g^G, y_t)]^2}{\Delta(x_g^{G+1}, x_g^G, y_t) [1 - \Delta(x_g^{G+1}, x_g^G, y_t)]}. \quad (3.2)$$

The problem (3.2) can be summarized as a weighted least square problem where we minimize the distance between the transition probability given by the Merton-Vašiček model (e.g,  $\Delta(x_g^{G+1}, x_g^G, Y_t)$ ) and the observed transition probability (e.g.  $\hat{p}_t(G, g)$ ), where  $\xi$  is calibrated using the regulatory formula 3.1. On the previously described data, the obtained common factor is presented in Figure 3.4.

On the obtained data we can fit the two models presented in previous chapters. However, we first need to choose the vector of covariates. We made an arbitrary selection for this vector. Our sample is composed of ratings of large corporates therefore we consider the corporate yield of *BBB* rated companies and the unemployment rate observed in the US. We consider this unemployment rate as most rated companies by S&P are either american

## CHAPTER 3. DENSITY ESTIMATION BASED ON QUANTILE REGRESSIONS APPLIED TO CREDIT RISK

---

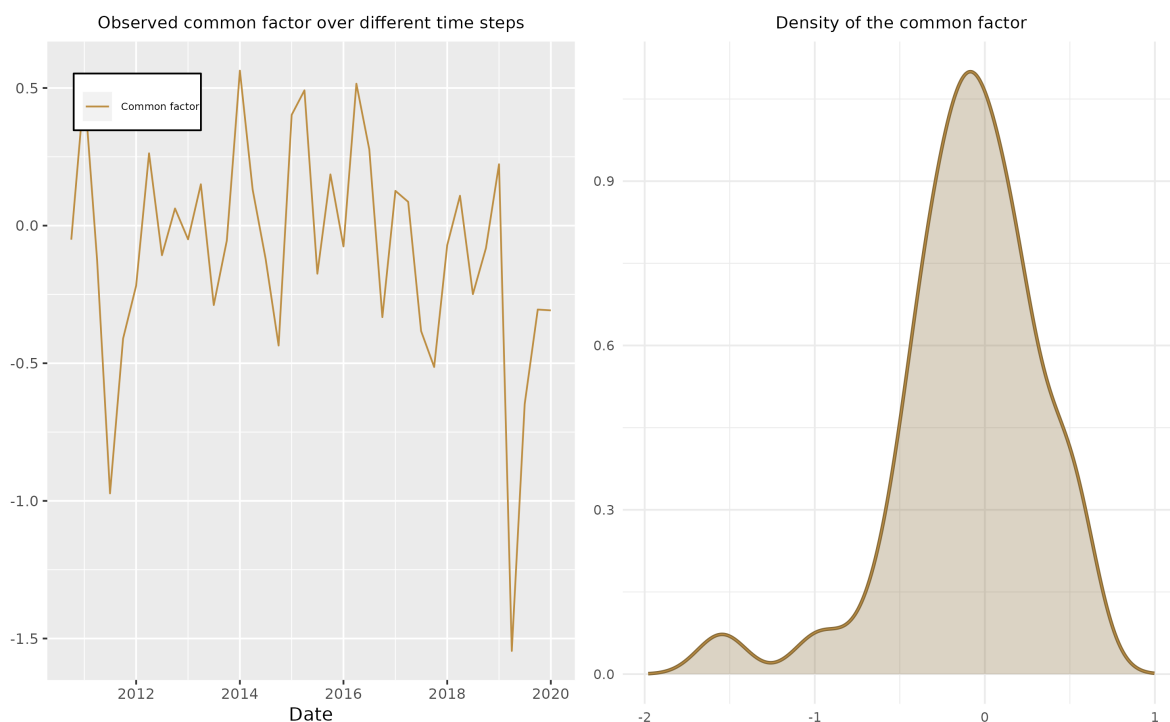


Figure 3.4: Observed common factor and its density obtained from the S&P data.

company or possess a large business unit in this country. Moreover it seems reasonable to think that an economic crisis in the US will also propagate to other countries. Also bond yields of companies should be a good indicator of credit risks. The higher the yield the higher the credit risk.

On the data illustrated by Figure 3.4, using the at-risk methodology presented in Chapter 3, the obtained coefficients of the multiple quantile regressions are presented in Table 3.2. We see that in most cases if either the Unemployment rate or the corporate bond yield increases the common factor tends to decrease this is not surprising as both covariates should be correlated to corporate risk of default. The variables selection process has not been discussed in this manuscript and is of high interest for practitioner. Indeed one covariate could be significant at one quantile but not for one other, in this case it is not entirely clear if that covariate should be included in the model. One other method would be to select a model that minimizes an information criterion, such as the BIC. However, the purpose of this section is not to construct a comprehensive probability of default model. Instead, our goal is to demonstrate the output of the presented methodologies and compare them with the regulatory model in two simple cases.

Given the quantile regressions presented in Table 3.2 we can perform the smoothing step, presented in 3.3.1 we computed the in-sample conditional distribution of the common factor. The results are presented in Figure 3.5.

quantile	0.05	0.25	0.75	0.95
Intercept	<b>0.99</b> (0.59)	<b>1.17</b> (1.03)	<b>1.39</b> (1.00)	<b>1.81</b> (0.89)
Unemployment rate	<b>-0.08</b> (0.05)	<b>0.01</b> (0.05)	<b>-0.02</b> (0.05)	<b>-0.02</b> (0.04)
BBB corporate yield	<b>-0.25</b> (0.17)	<b>-0.37</b> (0.26)	<b>-0.26</b> (0.24)	<b>-0.30</b> (0.19)

Table 3.2: Coefficients of the quantile regression, the standard deviation are reported between brackets.



Figure 3.5: Conditional distribution, the estimation is based on the methodology presented in Chapter 3. The quantile regressions obtained after the first step are presented in Table 3.2. The second step is based on a skewed-t distribution.

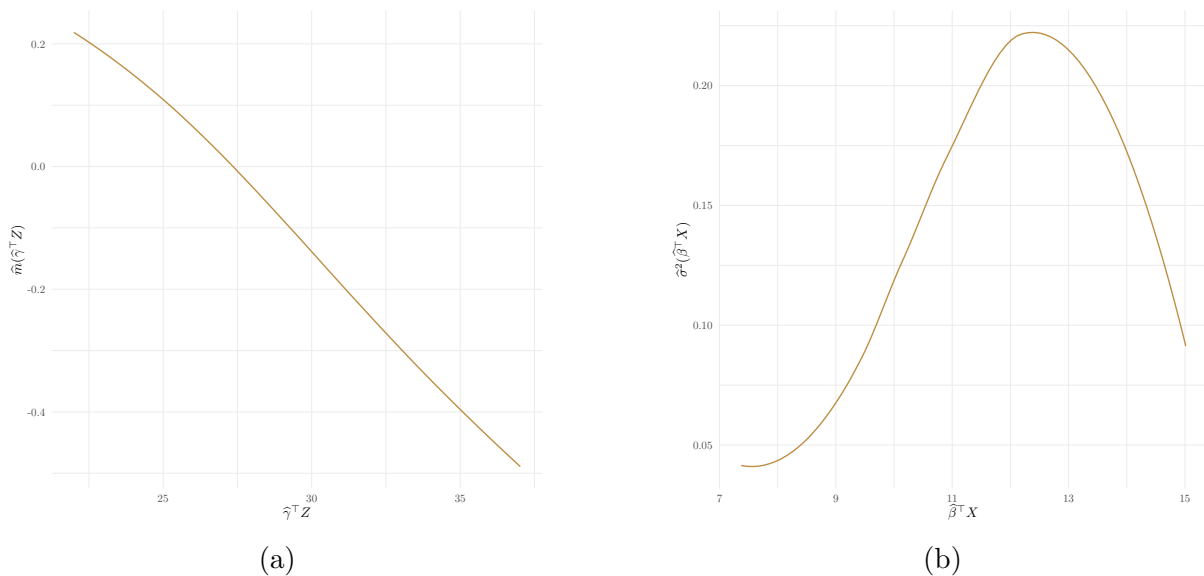


Figure 3.6: On the left panel is represented an estimator of the mean function,  $\hat{m}$ , the right panel displays an estimates of the conditional variance function,  $\hat{\sigma}^2$ .

Moreover, it is important to note that the model introduced in this chapter, as well as the one described in Chapter 2, can be applied to the same dataset to estimate the conditional distribution of the target variable. Each model employs a distinct methodology: the aR approach assumes a linear relationship between the quantiles of the conditional distribution and the covariates, whereas the location-scale single-index model allows for more general and potentially nonlinear relationships. However, since both methodologies are applied to the same dataset, their outputs should, at a minimum, exhibit comparable patterns or trends in the estimated conditional distribution.

Therefore, in the case of the location-scale model, we strictly followed the steps given in 2.3.2. We obtain :  $\hat{\gamma} = (1, 5.60)$ . Hence, an increase of the corporate bond yield tends to increase the default rates of the companies as the mean function is decreasing, moreover and more interestingly, this function is found to be almost linear, as illustrated by Figure 3.6. The ratio of  $\hat{\gamma}_{02}/\hat{\gamma}_{01}$  is similar to the ratio of the coefficients found in the quantile regressions models therefore an increase of any of the covariates should have similar impact on the mean estimated probability of default regardless of the methodology used.

We considered the same set of covariates to fit the conditional variance function, we obtained,  $\hat{\beta} = (1, 1.17)$  however the conditional variance has a non linear shape, the Figure 3.6 illustrates that an increase of either the unemployment rate of the corporate bond yield tends to increase the conditional variance up to a certain threshold.

We could estimate the quantile of the common factor on these set of covariates given economic projection. Let assume that the unemployment rate and the bond yield are both equal to 10% then, if we denote by  $\hat{q}_{\tau,1}(w)$  and  $\hat{q}_{\tau,2}(w)$  the quantile estimated by

respectively the at-risk methodology and the local linear model, then we obtain :

$$\hat{q}_{\tau,1}(w) = -6.08, \quad \hat{q}_{\tau,2}(w) = -2.24.$$

However, the hypothesized scenario is quite extreme, as no corporate yields in our observed sample exceeded 5% and the highest unemployment rate is 9.1%. Therefore the supposed scenario is very extreme and the quantile obtained by the at-risk methodology seems more credible. This is not surprising, given that no economic crisis occurred during the period of our observations and the non parametric estimator of the mean function and the conditional variance are expected to behave poorly on out of sample data.

Also, we do not easily have access to projection scenario therefore in the following section we will apply both methodologies on a private dataset and project both models based on the macroeconomic scenarios assumed by the bank in their internal stress-test.

### 3.4.2 Application on private banking data

In this section we present the results obtained on the private bank data. We have quarterly default rates for various homogeneous portfolios, observed from the last quarter of 2009 to the last quarter of 2022, resulting in 53 observations. The data have been anonymized to prevent the disclosure of sensitive information. This process ensures that the common factor remains a Gaussian random variable, and the methodology and interpretation are unaffected by the anonymization. Unlike the S&P data, this dataset includes corporate data from smaller companies that are not frequently rated. Consequently, the bank measures default rates for each time period. This scenario can be considered a special case where only two grades are possible: defaulted and non-defaulted.

The goal of this section is to present how to adapt the previously presented methodologies when estimated extreme financial losses due to credit risk. Assume that we observe economic scenarios provided by regulators, public institutions, academia, or other services within the bank. The aim is to determine the financial losses based on these scenarios. Under the IFRS9 accounting framework, it is advised to consider multiple scenarios that represent a range of likely outcomes, from a severe recession to an optimistic growth scenario. A standard practice is to include at least a central or baseline scenario with moderate growth and an adverse scenario with a crisis. Additionally, some banks also consider an optimistic growth scenario. While other banks may use more than these three scenarios, in this section, we will present the projected results for a baseline and an adverse scenarios.

Therefore, we apply the same procedure to estimate the common factor and obtain, Figure 3.7 illustrates the observed default rates in this portfolio, and the distribution of the common factor. We notice that in this dataset the common factor is centered and has an estimated variance of approximately 0.84. This variance is slightly, which could be explained by the choice of  $\xi$  and the economic cycle during which the observations were

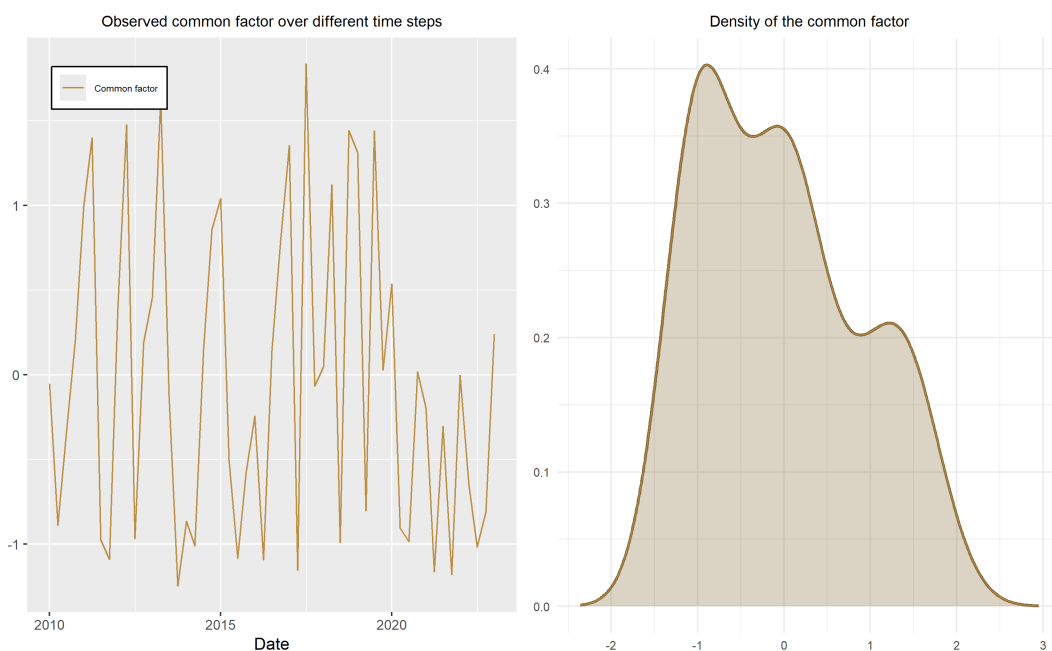


Figure 3.7: On the left panel we observe the historic common factor for each time period and the corresponding probability of default. On the right panel an estimation of the distribution of the common factor is displayed.

gathered. Indeed, no major economic crisis (at least in the seems of an increased default rates) seems to have occurred in the portfolio during this decade.

On these historic data we fit both model presented in Sections 2.3.3 and 3.2 on macroeconomic variables easily obtainable from public institutions. These macroeconomic data contains unemployment rates, long term interest rates, housing price index, sectoral value added, economic growth, inflation etc. For sensitivity reason we will not disclose, the exact scenarios, covariates used, the obtained probability of default nor the value of  $\xi$ . However note that the baseline scenario is standard value taken by economic variables, for instance the economic growth is close to 2%, unemployment rate is close to the one observed in recent years, etc. The adverse scenario is a crisis scenario with a recession, increased unemployment rate, and higher inflation. Figure 3.8 presents the projected common factor on both scenarios using the methodology presented in Section 3.2. Figure 3.9 presents the results using methodology presented in Section 2.3.3.

Given the value of  $\xi$  and a quantile of interest arbitrarily fixed by practitioners, an estimation of the conditional distribution of the common factor is sufficient to compute the conditional probability of default, as given by Equation (2.4). We see that for both methodology the common factor is lower in the adverse scenario compared to the baseline. The probability of default is therefore higher in the adverse scenario.

We notice similar behavior as Adrian et al. (2019), the variance of the conditional distri-

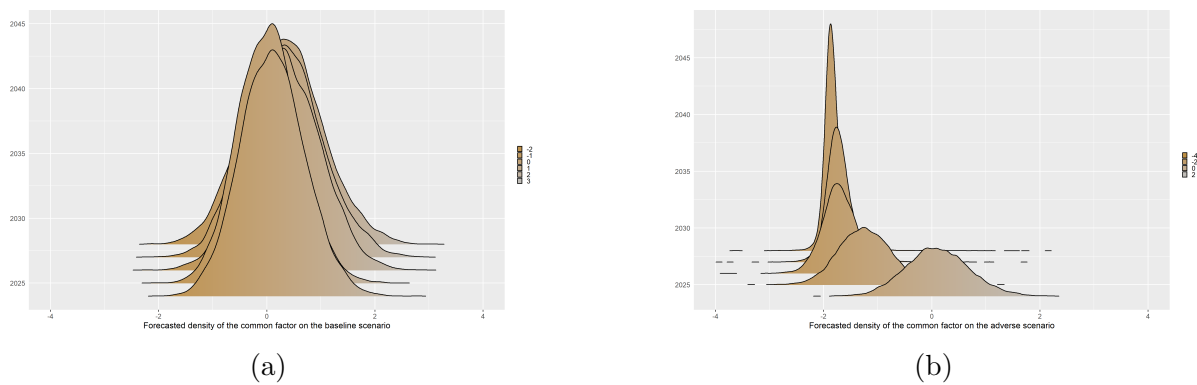


Figure 3.8: Projected common factor on the baseline and adverse scenario on respectively the left and right panel. Both the estimates and the figures have been generated using the CRAN package Lajaunie et al. (2023)

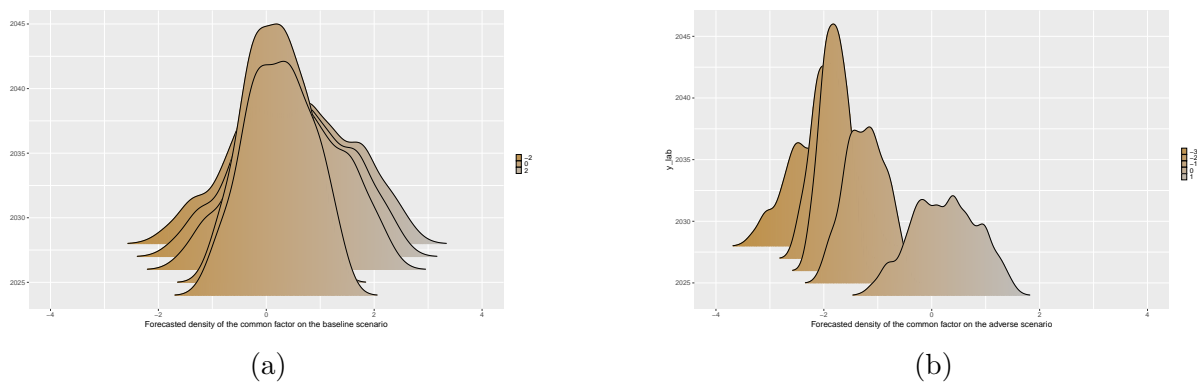


Figure 3.9: Projected common factor on the baseline and adverse scenario on respectively the left and right panel.

	Merton- Vašíček	At-Risk, Section 3.2	Local linear single index, Section 2.3.3
Observed	-3.71	-0.57	-0.64
Adverse	-3.71	-3.84	-3.70
Baseline	-3.71	-2.18	-2.53

Table 3.3: Comparison of models for different scenarios, note that only the last year for each scenario is presented, we considered the quantile  $\alpha = 0.001$ .

bution is very dependent on the covariates, though in our case the higher variance is found on the baseline scenario. Moreover, we note that the obtained quantiles using the regulatory models vary quite significantly from the one obtained using the presented methods. Indeed, in the case of the baseline scenario, the quantiles are much higher leading to a lower probability of default, however in the adverse scenario, the conditional quantiles are lower than the one obtained using the marginal distribution of the common factor, see Table 3.3 that summarize the obtained results using all methodologies.

For both models we considered the same covariates and only considered a model with 2 covariates which may be seen as too simplistic, however our data sample is small and the estimating strategies can fit very general relationships. We used the same bandwidths as described in the empirical Section 2.3.3 and show the results for the smooth estimator of the distribution function, (2.10). Considering the methodology presented in Chapter 3 we considered the same hypothesis as presented in Section 3.3.1 and considered a skew-t distribution for the second step.

As expected and as showed by Figure 3.8 and 3.9 we see that a decline in economic growth tends to decrease the value taken by the common factor. Indeed the distribution in the adverse scenario have lower mean when compared to the baseline scenario. Also, more interestingly, we see that the relationship between the covariates and the mean of the conditional distribution of the common factor is linear, see Figure 3.10. Therefore, the probability of default increases non linearly.

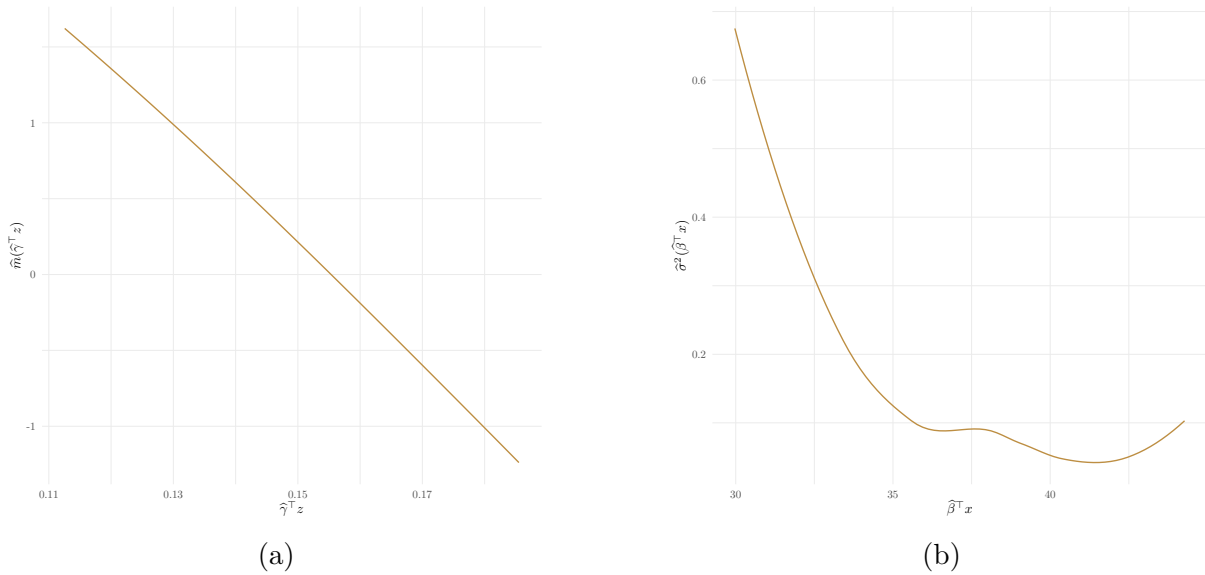


Figure 3.10: On the left panel is represented an estimator of the mean function,  $\hat{m}$ , the right panel displays an estimates of the conditional variance function,  $\hat{\sigma}^2$ .

### 3.5 Conclusion

The aR approach is a semi-parametric methodology that allows forecasting the full conditional distribution of a variable at an horizon  $h$  as a function of factors observed at time  $t$ . An aR model can be simply estimated from historical data (i) by running a few of quantile regressions for a set of coverage rates, and then (ii) by using a method of moments to estimate the parameters of a parametric distribution that is sufficiently flexible to capture the main features of the true distribution revealed by the empirical quantiles. The last step can be viewed as a smoothing procedure of the empirical quantiles. This two-step approach ensures consistency of the quantiles produced by the estimated parametric distribution. In particular, it avoids the well-known issue of quantile crossing. As a complement to this article, we propose an R package that automatically estimates an aR-type model and produces the forecasts for both density and risk measures (VaR and ES) with a variety of specification options, the package is available on the CRAN, see Lajaunie et al. (2023).

Several extensions of the aR methodology can be considered to improve its adequacy to the bank stress-test framework. First, we plan to extend the estimation procedure to an over-identified case (i.e., with more moment conditions than parameters to estimate) to consider the variance of the quantile estimates and to weight their relative importance. Second, we propose a bivariate aR-type model that allows us to model the conditional joint distribution of a couple of risk parameter and macroeconomic variables. Considering the joint distribution will allow banks to derive new relevant risk measures (Benoit et al., 2017), such as the conditional Value-at-Risk (CoVaR) or the Marginal Expected Shortfall.

## References

- Adrian, T., Boyarchenko, N., & Giannone, D. (2019). Vulnerable growth. *Am. Econ. Rev.*, *109*(4), 1263–1289.
- Aguais, S. (2008, January). Designing and implementing a Basel II compliant pit-ttc ratings framework [Article available at: <https://mpra.ub.uni-muenchen.de/7004/>].
- Artzner, P., Delbaen, F., Eber, J.-M., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, *9*(3), 203–228.
- Azzalini, A., & Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *J. R. Stat. Soc.*, *65*(2), 367–389.
- Belkin, B., Suchower, S., & Forest, L. (1998). A one-parameter representation of credit risk and transition matrices [Article available at: <https://www.z-riskengine.com/media/hqtnwlm/a-one-parameter-representation-of-credit-risk-and-transition-matrices.pdf>]. *CreditMetrics monitor*, *1*(3), 46–56.
- Benoit, S., Colliard, J.-E., Hurlin, C., & Pérignon, C. (2017). Where the risks lie: A survey on systemic risk. *Rev. Finance*, *21*(1), 109–152.
- Brave, S., & Butters, R. A. (2018). Diagnosing the financial system: Financial conditions and financial stress. *29th issue (June 2012) of the International Journal of Central Banking*.
- Chavleishvili, S., Engle, R. F., Fahr, S. A., Kremer, M., Manganelli, S., & Schwaab, B. (2021). The risk management approach to macro-prudential policy [Article available at: <https://www.ecb.europa.eu/pub/pdf/scpwps/ecb.wp2565~e7b85e3982.en.pdf>]. *ECB working paper*, *2565*.
- Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, *80*(1), 27–38.
- Gourieroux, C., Monfort, A., & Renault, E. (1993). Indirect inference. *J. Appl. Econom.*, *8*, S85–S118.
- Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, *50*(4), 1029–1054.
- IMF. (2022). West african economic and monetary union: Financial sector assessment program—financial system stability assessment [Report available at: <https://www.imf.org/en/Publications/CR/Issues/2022/05/11/West-African-Economic-and-Monetary-Union-Financial-Sector-Assessment-ProgramFinancial-517823>]. *Country Report*, *136*.
- Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, *46*(1), 33–50.
- Lajaunie, Q., Flament, G., Hurlin, C., Kazemi, S., & Lajaunie, M. Q. (2023). Package ‘atrisk’ [Package available on the CRAN at: <https://cran.r-project.org/web/packages/atRisk/index.html>].
- Lhuissier, S. (2022). Financial conditions and macroeconomic downside risks in the euro area. *Eur. Econ. Rev.*, *143*, 104046.
- Sen, P.-A. (1968). Estimates of the regression coefficient based on kendall’s tau. *J. Am. Stat. Assoc.*, *63*(324), 1379–1389.

CHAPTER 3. DENSITY ESTIMATION BASED ON QUANTILE REGRESSIONS  
APPLIED TO CREDIT RISK

---

Theil, H. (1950). A rank-invariant method of linear and polynomial regression analysis.  
*Nederl. Akad. Wetensch., Proc.*, 53, 386–392.

## Chapter 4

# Macroeconomic scenarios for climate risk assessment

**Abstract :** In this chapter, we propose to build a statistical model for factor productivity and illustrate our predictions on a common toy model : DICE. Data show that the variation of primary *exergy*, or simply called *exergy* is a good predictor of the TFP variation. That remark allows us to design new economic scenarios that respect the Paris Agreement. These scenarios are of particular interest for policy makers. Moreover, our model is capable to reconcile different projections by the introduction of environmental constraints. On one hand, if *exergy* increases, we have similar projections with standard economic models. On the other hand, if *exergy* decreases, we obtain projections closer to other Integrated Assessment Models, where economic output declines Capellán-Pérez et al. (2020) and Meadows et al. (1972). This chapter has been published in Structural Change and Economic Dynamics, for more details see Flament (2023).

## 4.1 From regulatory stress-testing to climate stress-testing

### 4.1.1 Some definitions and regulatory constraints

For a long time, banks insurance and financial institutions did not place Environmental, Social, and Governance (ESG) criteria at the heart of their strategy. Regarding environmental criteria, banking activity was considered intrinsically low-polluting: unlike other sectors, banks do not emit toxic fumes, do not use chemicals, and do not consume fossil fuels in their direct production processes. This view is now outdated, and banks must now be able to justify that they are full-fledged actors in sustainable finance. They are expected to rise to the expectations placed upon them to become the drivers of the transition towards a more sustainable development model. To achieve this, they must be able to meet ESG requirements, which form the triad of a more sustainable financial system, in line with society's aspirations:

- A financial system that supports the energy transition towards a low-carbon economy, promotes low-ecological-impact activities, and discourages polluting activities.
- A financial system that not only benchmark short term financial criteria but also long term impacts of its activities.
- A financial system that is governed in a more transparent manner, with greater diversity in its members and in its thought patterns.

In this transition towards sustainable finance, the aspect of supporting the transition to a low carbon economy presents a crucial challenge. On one hand, climate change generates large-scale and more frequent weather phenomena that could destabilize local economic and lead to the erosion of asset values. On the other hand, policies aimed at combating climate change could also weaken the financial system if the transition to a low-carbon economy were too sudden or too late, thereby degrading the value of financial assets. This is why supervisors and central banks are calling on financial actors to quickly implement the necessary internal measures to monitor and mitigate this risk. Consequently, financial actors, including banks, need to measure three types of risks:

- **Physical risk**, the risks resulting from damage directly caused by weather and climate phenomena. Physical risk measures the direct impact of climate change on people and assets (droughts, floods, etc.). There are two types of climate risk: acute and chronic. Acute physical risk refers to sudden and extreme weather events due to climate change, like floods or wildfires, that can cause immediate damage and financial losses, while chronic physical risk involves long-term, gradual climate changes, such as rising sea levels or increased average temperatures, that pose ongoing threats to assets and investments in the banking industry.

- **Transition risk** refers to the financial risks that arise from the process of adjusting to a low-carbon economy, including changes in policy, technology, and market sentiment that can lead to stranded assets, reduced profitability, or increased costs for high-carbon industries and their stakeholders in the banking sector. For instance, most modern industrial processes rely on fossil fuels that emit carbon dioxide. Replacing these energy sources involves significant uncertainty, which may cause disruptions to certain activities and, consequently, to financial institutions.
- **Liability risk** refers to the potential financial exposure of banks due to legal claims or regulatory penalties stemming from their role in financing activities that contribute to climate change, or their failure to adequately manage climate-related risks.

Until very recently, non-financial risks were not considered, most efforts were focused on modeling financial losses stemming from financial crises, such as market turmoil, liquidity crunches, and high probabilities of default in the retail market. However, the underlying causes of these crises were not addressed nor mentioned as the scenario only mentioned the severity of the recession or the loss of values in some financial market. However, in 2015, Mark Carney, then Governor of the Bank of England, delivered a speech titled *Breaking the tragedy of the horizon - climate change and financial stability*. This speech initiated a series of studies on non-financial risks across many jurisdictions. On one hand, it initiated a series of so called stress-testing exercises focused on climate risks. On the other hand, regulators began to mandate disclosures of climate-related risks by financial institutions. For example, the CRR3<sup>1</sup> not only emphasizes the importance of quantitative exercises but also intensifies communication requirements regarding environmentally related financial risks. Furthermore, it requires financial institutions to disclose a transition plan outlining their strategies for monitoring both short-term and long-term economic transitions towards a low-carbon economy.

The focus of the previous chapters were on modeling the probability of default with any given vector of covariates. However this vector of covariates and the assumed model used to generate the stress-testing scenario was not discussed. If one consider irrelevant scenario or not stringent enough ones then the stress-testing exercise itself becomes irrelevant even if the best credit risk modeling techniques are used. Describing a scenario where only financial covariates are projected is already a difficult task and well discussed in the literature, see Adrian et al. (2019) or Parlato and Philippon (2022) for instance. However in the case of climate change one also have to integrate climate related factors, their impact on financial covariates and project them on very long term horizon. The risk of designing scenarios that does not truly reflect either physical or transition risk is very high which may ultimately undermine the credibility and effectiveness of the climate stress-testing exercise, leading to an underestimation of climate-related financial risks.

---

<sup>1</sup>For additional information on CRR3/CRD6, see the note from the banque de France : [https://acpr.banque-france.fr/sites/default/files/medias/documents/20240325\\_revue\\_acpr\\_paquet\\_bancaire.pdf](https://acpr.banque-france.fr/sites/default/files/medias/documents/20240325_revue_acpr_paquet_bancaire.pdf)

### 4.1.2 Differences between climate and traditional stress-tests

Unlike traditional exercises where the assumption is that a financial crisis has occurred, climate stress-tests involve the regulator first assessing the impact of global warming and the transition to a clean economy on broader economic indicators (GDP, unemployment, inflation, asset values...). The risk managing team of each financial institution then evaluates the impact of these economic factors on extreme losses. This approach is fundamentally different from traditional stress-tests, which do not consider underlying driver for the crisis.

For instance, the latest stress-test conducted by the ECB assumed a scenario of high inflation, a severe recession, and high interest rates, but the underlying cause of that crisis is not described<sup>2</sup>, the difference is striking when compared to the 2023 climate stress-test conducted by the ACPR<sup>3</sup> where the transition scenario, including a carbon tax and possible costs due to physical risks, is discussed and translated into macroeconomic data. This additional layer of complexity introduces significant uncertainty into the final results because, on top of the uncertainty inherent to statistical models discussed in preceding chapters, climate stress-tests also incorporate the uncertainty of modeling the effects of global warming (or lack thereof) on macroeconomic indicators.

Moreover, climate stress-tests bear one other major uncertainty. Indeed, the time horizon of the stress is at least of several decades if not a century, whereas traditional stress-tests typically limit themselves to a maximum of three years. This additional assumption is necessary to consider physical risks that will materialize over a longer period. However, this poses a problem for banks that have products on their balance sheets with maturities ranging from a few days (such as those in a trading book) to at most two decades, in the case of long-term corporate bonds, real estate mortgages or states bonds.

Many assets held by financial institutions may not exist in a few years, making it difficult to accurately assess the long-term impact of climate risks on their portfolios. Additionally, the mismatch between the short-term nature of many financial instruments and the long-term horizon of climate stress-tests introduces significant complexity and uncertainty into the analysis.

To address this complexity, regulators have introduced the concept of dynamic balance sheet. This requires all participants in the climate stress-test to replace matured products in their portfolios with new ones. The replacement strategy must align with the plan communicated to the regulator, ensuring that the bank's risk profile remains consistent with the strategic plan communicated to the regulator.

However, there is considerable uncertainty as to whether future management teams will

---

<sup>2</sup>See the result of the 2023 Euro area banks stress-test : [https://www.bankingsupervision.europa.eu/ecb/pub/pdf/ssm.Report\\_2023\\_Stress\\_Test~96bb5a3af8.en.pdf](https://www.bankingsupervision.europa.eu/ecb/pub/pdf/ssm.Report_2023_Stress_Test~96bb5a3af8.en.pdf)

<sup>3</sup>See the technical assumptions : [https://acpr.banque-france.fr/sites/default/files/medias/documents/2023\\_main\\_assumptions\\_and\\_scenarios\\_of\\_the\\_acpr\\_climate\\_exercise.pdf](https://acpr.banque-france.fr/sites/default/files/medias/documents/2023_main_assumptions_and_scenarios_of_the_acpr_climate_exercise.pdf)

adhere to these plans, and it is also uncertain whether the projects these banks plan to finance will indeed materialize<sup>4</sup>. As outlined by the previously mentioned publications of the regulator, it was assumed that the balance sheet was dynamic after 2025 therefore even if every model used is a perfect representation of the reality it would still be dependent on the evolution of the banks balance sheet that could be very far from the real strategy that will be adopted by financial institutions a decade latter.

However, even though the uncertainty is high and the modelling challenges unprecedented climate stress-tests were conducted in many jurisdictions. It started in [the netherlands in 2018](#)<sup>5</sup>, then followed by a similar exercise in [France in 2020](#)<sup>6</sup>. In 2021, both [the Bank of Canada](#)<sup>7</sup> and the [Bank of England](#)<sup>8</sup> conducted their own climate stress-tests. Finally, the European Central Bank launched a Europe-wide climate stress-test in 2022, followed by an exercise focused on insurance companies in 2023.

### 4.1.3 Current practices for integrating climate change into stress-testing

In most of the literature, see Battiston and Monasterolo (2020) for instance, it is assumed that climate impacts should first be assessed using macroeconomic models and then such macroeconomic data are integrated to credit risk models as covariates. This methodology is nowadays the current practice for conducting climate stress-testing (Cartellier (2022)). The NGFS uses three models. The REMIND-MAgPIE model (Luderer et al. (2015)), MESSAGEix-GLOBIOM (Krey et al. (2020)) and GCAM (Calvin et al. (2019)). In previous chapter, we focused on the second modeling step, developing a new methodology that financial management teams in a bank could use to assess credit risk. On one hand the previous methodologies could be employed in a climate stress-test if the economic forecasts integrate such components. On the other hand it could also be employed in a more traditional setting.

In this chapter, we argue that the design of the model used to assess the impact of climate-related risks on the economy is crucial. If not done properly, it may lead to radically different conclusions when translated into financial variables. Therefore, the focus of this chapter is on the design of such a model.

These scenarios are designed by economists and consist of different pathways. A pathway

---

<sup>4</sup>See a technical note from the ECB on this matter : <https://www.bankingsupervision.europa.eu/ecb/pub/pdf/ssm.climateriskstresstest2021~a4de107198.en.pdf>

<sup>5</sup>Document available at this url : [https://www.dnb.nl/media/pdnpdalc/201810\\_nr-7-2018-\\_an\\_energy\\_transition\\_risk\\_stress\\_test\\_for\\_the\\_financial\\_system\\_of\\_the\\_netherlands.pdf](https://www.dnb.nl/media/pdnpdalc/201810_nr-7-2018-_an_energy_transition_risk_stress_test_for_the_financial_system_of_the_netherlands.pdf).

<sup>6</sup>Document available at this URL : <https://acpr.banque-france.fr/en/scenarios-and-main-assumptions-acpr-pilot-climate-exercise>

<sup>7</sup><https://www.bankofcanada.ca/wp-content/uploads/2021/11/BoC-OSFI-Using-Scenario-Analysis-to-Assess-Climate-Transition-Risk.pdf>

<sup>8</sup>Document available at this URL : <https://www.bankofengland.co.uk/stress-testing/2022/results-of-the-2021-climate-biennial-exploratory-scenario>

can be seen as a narrative where common assumptions are made. In this case, assumptions include population growth, the potential costs of clean energy and clean production technologies, fossil fuel reserves, technological development, and to a certain extent, GDP growth. Given a pathway, different policies are assumed for transitioning toward a low-carbon economy or not, and all relevant information is computed for each time period for each policy. In the case of the NGFS scenarios, the projected data are very general. For instance, one can find the quantity of coal consumed each year, the quantity of electricity and its price for each energy source and for every region of the world. For more details, one can refer to the technical document published by the NGFS, Bertram et al. (2021).

Moreover, climate scientists (see, for instance, Petit et al. (1999)) have demonstrated that the relationship between carbon dioxide concentration and temperature is nearly linear. In the macroeconomic scenario developed by the NGFS, the quantity of fossil fuel burned is known for each time period, allowing to estimate the temperature increase caused by human activities. This allows economists to integrate a feedback loop, known as the damage function. This function represents the impact of increased climate hazards on annual production.

Therefore, economists can estimate the impact on the economy of different set of policies. In the case of the NGFS scenarios, the studied policy consists of three broad categories. First is a scenario where a common agreement is made to develop green technologies with unlimited cooperation and all policies are implemented to respect the Paris agreement, this is the less costly scenario and is known as the Net Zero Scenario. The second scenario, known as the Delayed Transition Scenario, corresponds to a situation where the transition is delayed by several years. However, after this initial delay, the common effort is assumed to be the same as in the Net Zero Scenario. The third scenario is known as the Hot House World and assumes that no effort will be made to limit global warming below manageable thresholds. Therefore, in this case, the damages are the highest of all scenarios.

The climate stress-tests consists of a combination of a common pathway (Population growth, technological progress, ...) and a set of policies (Carbon tax, incentive to develop clean energy, ...) that is translated to a set of possible outcome (Energy Price, GDP, Unemployment, ...) and what would be called by statisticians, explanatory variables that will then be used for assessing extreme loss in a stress-test scenario.

However to the best of my knowledge, banks generally use what the NGFS (Bertram et al. (2021)) refers to semi endogenous variables (GDP, Inflation, Unemployment rate, ...). Some of these variables, especially GDP have to some extent a deterministic trend. This assumption is very problematic when assessing the impact of climate change onto credit risk. Therefore the following sections of this Chapter discuss a model that could be used to generate macroeconomic scenarios especially GDP growth.

## 4.2 Introduction to climate integrated macroeconomic models

Since 1988, the Intergovernmental Panel for Climate Change (IPCC) tries to evaluate the impact of human activities on the climate and has very recently launched yet another worrisome report IPCC (2021). Around the same time, W. D. Nordhaus (1991) published a pioneer work was, for the first time, a cost benefit analysis which discussed the optimal greenhouse gas emissions (GHG) abatement policy. Nordhaus then proposed the first iteration of the Dynamic Integrated model of Climate and Economy (DICE) W. D. Nordhaus et al. (1992) to take into account the damage of climate change and greenhouse gas mitigation on the economy. His work seems to have met with a huge success, as he has been granted the economic Nobel prize in 2018. Many economists have expressed criticisms about DICE N. Stern (2013). Stern states that economic growth should not be taken for granted and new models needs to be built to add-in possible recessions. In DICE, economic declines are impossible due to the exogenous growth of Total Factor Productivity (TFP). To build new models where economic decline is a possible scenario, it is crucial to change the exogeneity hypothesis of the TFP. This quantity is often interpreted as a proxy for technological progress, review from Isaksson et al. (2007) states that this progress is often led by education, health, infrastructure, and financial developments. In practice, it seems very difficult to quantify all these variables to build a meaningful model of TFP. Some other models rely on econometric relationships Alestra et al. (2022), using proxy variables such as relative price of energy or number of years in the schooling system. Some others rely on purely economic assumptions, such as the catch-up model developed by Aghion et al. (2001), Pradhan and Ghosh (2022), and Vandebussche et al. (2006).

Some authors tried to endogenize the TFP Alestra et al. (2022) by building an easily tractable and interpretable econometric model to generate long-term growth scenarios. The most striking difference between their contribution and the W. Nordhaus (2018) model, is the assumption of endogenousness of the TFP. However, the Alestra et al. (2022) model lacks crucial properties needed for relevant future Gross Domestic Product (GDP) forecasts. Firstly, the variables used to predict TFP seem to be very hard to project, which is a problem given the key role the TFP must play in the model. One may wonder to what extent the model of Alestra et al. (2022) is endogenous as the predictor variable can not be easily projected. Secondly, they assume energy consumption will be set by GDP and the energy mix will adapt according to the relative price of each energy source year by year. Unfortunately such an assumption is very optimistic, because the energy mix does not evolve greatly from a year to another. Moreover, regardless of energy prices, a nuclear, coal or gas power plant is built for several decades.

In this Chapter, we focus on the DICE model, however many other IAM are used in the literature. For instance, the NGFS uses GCAM Calvin et al. (2019), and REMIND Luderer et al. (2015). GCAM set exogenous path for labor productivity. REMIND has a different specification where it defines a production function that embeds final energy.

In a way, it may seem similar to our model, but they tune efficiency parameters to match exogenous GDP growth path Luderer et al. (2015):

«It assigns an efficiency parameter to each production factor in the various macroeconomic CES (Constant Elasticity of Substitution) functions. The changes of efficiency parameters over time are tuned such that baseline economic growth and energy intensity improvements match exogenous scenario specifications. »

Which makes GDP growth exogenous just as all other models. Thus, it means that the NGFS takes IPCC predictions into account with exogenous GDP growth paths. These assumptions are at best misleading and lead to systematic risk as financial institutions are unable to properly measure climate risk. We think these models are inherently dangerous in a stress-test framework hence we must build better models to reflect climate risk onto economic scenarios. We believe such risks can be taken into account by a Bayesian framework that naturally introduce credibility intervals. These intervals are well suited to understand tail risks of different scenarios. That framework also allows us to introduce endogenous variables that reflects climate change.

Our goal is to build a statistical model using observable variables to predict future values of TFP. This work distinguishes itself from Alestra et al. (2022) by introducing the *exergy* variable and building credibility intervals around the quantity of interest the total factor productivity. Moreover it also differs from Bercegol and Benisty (2022) as we propose long-term prediction and a different model. We then illustrate our projections with a toy model DICE. The main advantage of our method is that it uses one observable variable that seems to be a driver of TFP. It also has the appealing property of being very easily estimated from a climate trajectory. Using that model, then, we would like to show projections of GDP using our modified version of the DICE model W. Nordhaus (2018). Hence, such projections should be more reliable than previous ones.

The following Sections will be organized as follows. Firstly, we shall describe the economic assumptions used. We shall then, in the following section, describe all the assumptions made to develop our own version with a focus on TFP modeling. In the fourth section, we will discuss two ways to build an *exergy* scenario using exogenous assumptions and endogenous ones. In the fifth section, we will build the climate change module. The aim of this module is to translate climate change into economic damage. In the sixth section, we will derive our full model and compare it to some other propositions such as W. Nordhaus (2018).

## 4.3 Economic model assumptions

### 4.3.1 The DICE model

The DICE model introduced by W. D. Nordhaus et al. (1992) and recently reevaluated W. Nordhaus (2014, 2018), is an economic model that describes the interaction between the economy and the environment. In particular, such a model answers the following question: what would be the cost of the emissions reduction needed to meet the Paris Agreement ? To answer such a question, Nordhaus identifies two costs.

The first is the direct cost of climate change. Unfortunately, climate change increases the likelihood and the magnitude of a disastrous event. Recent events, such as the heat dome in Canada, wildfire in north Africa, southern Europe or the chronic drought in western USA, are only a few indicators of what is meant to come. Obviously environmental disasters are very costly for a society. Hurricane Katrina costed more than \$100 billion Shreve and Kelman (2014) and global warming is expected to kill millions if not billions of people Mora et al. (2017, 2018). Thus, Nordhaus, applying the findings of Tol (2009), estimated a damage function in order to quantify these damages. That function has been the center of many criticisms for being too optimistic in extreme scenarios, see for example N. Stern (2013) and Weitzman (2010).

The second cost is the adaptation to a warming climate and the investments required to lower the greenhouse gas emissions of our economy. These investments and decommissions may slow down or reverse growth. First, we need to decommission all internal combustion engines (planes, ships, cars, etc), all electric generators that burn fossil fuels, change all the current cement works and steel mills, on a global scale, by 2050. We then need to invest in a low carbon economy. But one may argue that, to date, green investments are not as cost effective as traditional investments, else they would already be financed and a net-zero economy would be reached without any public intervention. That difference, between traditional and low carbon investments, might introduce a cost of greening the economy. That cost is represented in the model by the  $\omega(t)$  function, which represents indirect damages Pradhan and Ghosh (2019).

Overall, Nordhaus considers a standard Cobb-Douglas production function multiplied by two other functions representing direct and indirect damage due to climate change:

$$Y(t) = (1 - \omega(t)) \frac{A(t)K(t)^\beta L(t)^{1-\beta}}{1 + \Omega(t)}. \quad (4.1)$$

In (4.1),  $Y(t)$  represents the output of the economy at time  $t$ . Many other production functions exist (see Thompson (2006)) but we shall not discuss them as we argue that the most important parameter is the TFP, regardless of the production function thus, we focus on building an endogenous TFP model with the simplest production function. In (4.1),  $K(t)$  and  $L(t)$  respectively represent capital and labor at time  $t$ , respectively  $\Omega(t)$  is the amount of damage caused by climate change at time  $t$ , while  $\omega(t)$  represents the

cost of emissions reduction and adaptation. We will then add a relationship between the ghg concentration in the atmosphere and climate change to counterbalance the cost of greening the economy with the lower cost of climate catastrophe. They have the following definitions:

$$\begin{aligned}\Omega(t) &= \alpha_1 T(t) + \alpha_2 T(t)^2 \stackrel{DICE}{\approx} 0.0026 \times T(t)^2, \\ \omega(t) &= \theta_1 \mu(t)^{\theta_2}.\end{aligned}\tag{4.2}$$

In (4.2),  $T(t)$  is the mean temperature increase of the earth at time  $t$  while  $\alpha_1$  and  $\alpha_2$  are regression parameters, Tol (2009) described one method to estimate them. In (4.2),  $\mu(t)$  is the pace of the reduction of greenhouse gases emissions and  $\theta_1, \theta_2$  are predefined constants, W. Nordhaus and Sztorc (2013) state that  $\theta_2 = 2.8$  and, between 2020 and 2050,  $\theta_1$  grows from 0.039 to 0.068. If one is interested in estimating economic output growth rate, then the most important driver is TFP  $A(t)$ . According to the specification of W. Nordhaus and Sztorc (2013), this variable is defined exogenously by:

$$\begin{aligned}A(t) &= A(t-1)(1 + g_A(t)), \\ g_A(t) &= g_A(t-1)(1 + \delta_A), \\ g_A(2015) &= 7.9\% \text{ per 5 years}, \\ \delta_A &= 0.6\% \text{ per 5 years}.\end{aligned}$$

This model specification leads to a very high economic growth, of about 1.9% per year for the coming century, while allowing this growth to be independent of any other variable. Growth is mainly due to TFP, as it grows by about 1.5% per year. The remaining growth comes from population increase and a small capital accumulation. Overall, it means that the cost of climate change is almost null. More and more contributions in the recent literature consider that this is at least, very optimistic, if not completely misleading Pindyck (2017).

## Notation

The notations used throughout this Chapter are summarized in 4.1. We use a subscript to specify the country while the time will be between parentheses. When there is no risk of confusion, the country subscript is omitted.

Symbole	Significance
$Y_c(t)$	Gross Domestic Product (GDP)
$A_c(t)$	Total Factor Productivity (TFP)
$K_c(t)$	Capital stock
$I_c(t)$	Investment
$L_c(t)$	Number of working hours
$\Omega_c(t)$	Direct damage from climate change
$\omega_c(t)$	Investments to transition toward a green economy
$T(t)$	Temperature variation since the Industrial Revolution
$E_c(t)$	<i>Exergy</i>
$TP_c(t)$	Total primary energy
$G(t)$	Greenhouse gas emissions
$SCO_2(t)$	Additional greenhouse gas in the atmosphere.
$\mathcal{C}$	List of countries (see Table 4.5)

Table 4.1: Notations

Therein, bold symbols designate vectors. For example,  $\mathbf{Y}(t)$  is a vector with the GDP for a set of countries. Moreover, we will note the logarithm of every quantity, using lowercase letters. For instance,  $a_c(t) = \log(A_c(t))$ .

### 4.3.2 GDP dynamic

As we stated in the introduction, we shall modify the DICE model to introduce endogenous TFP growth, using the standard two factors Cobb-Douglas function, following Alestra et al. (2022) and W. Nordhaus and Sztorc (2013) and, more precisely, with the notations in Table 4.1 :

$$Y_c(t) = A_c(t)K_c(t)^\beta L_c(t)^{1-\beta}.$$

Here  $Y_c(t)$  represents the GDP for country  $c$  at time  $t$  without any climate or transition damage (e.g,  $\omega$  and  $\Omega$ ). The TFP, capital and labor are represented respectively by  $A_c(t)$ ,  $L_c(t)$ ,  $K_c(t)$ . The parameter  $\beta$  represents the elasticity between capital and labor. We will set that value as W. Nordhaus and Sztorc (2013), that is  $\beta = 0.3$ . We keep the damage function introduced by Nordhaus ((4.2) or a variant), however, as we will discuss in the following sections, we will endogenize the indirect damage into the TFP growth model (see (4.2)), thus, we shall not use the indirect damage function introduced by Nordhaus. It will remove significant unknown as that function is nearly impossible to calibrate, yet, no observation exists to test its forecasts. Thus, the main goal of this Chapter is to build a statistical model with some exogenous variables to predict the variations of  $A_c(t)$ .

Following Solow (1956) or W. Nordhaus and Sztorc (2013), we use a classical model for

capital variation. Let  $\delta_K$  be the depreciation of the capital. The dynamic of  $K(t)$  is then:

$$K(t) = (1 - \delta_K)K(t - 1) + I(t).$$

Even if this is highly debatable, we will keep the same value of  $\delta_K$  as W. Nordhaus and Sztorc (2013) for comparability purpose. However, it would be interesting to modify this hypothesis as climate change will likely increase its value. For example, a very aggressive policy towards carbon emissions reduction would increase the depreciation rate of every industrial structure required by the fossil fuels industry. It is the stranded assets problem, see for example: Van der Ploeg and Rezai (2020).

### 4.3.3 TFP dynamic

The variable that leads to sustained growth in the aforementioned model, is TFP, as many researchers Ayres et al. (2003), Johansson et al. (2015), Keen et al. (2019), D. I. Stern (2011), and Warr and Ayres (2006) discussed, growth rate could be very dependant on an observed variable, *exergy*. To simplify the notation, we shall refer to *primary exergy* as *exergy*. Some others refer to that notion under the name *final energy* Bercegol and Benisty (2022). *Exergy* is a thermodynamic quantity which represents the amount of work a given system could do before reaching thermodynamical equilibrium. This quantity could be estimated using the amount of energy used and the efficiency of its transformation into useful work. We argue that useful work is a good predictor of TFP as productivity is driven by *exergy* intensive industrial process Cleveland et al. (1984). Hence, the higher the *exergy* the higher the number of produced goods. All other variables being equal, it should equate to more GDP. Thus, in the framework developed by Nordhaus, we argue that, the variations of *exergy* should be a good predictor of variations of the TFP.

Moreover, as explained by Keen et al. (2019), we do not want to add *exergy* as a substitute for labor or capital: *Exergy* can not be replaced by capital because capital needs it to transform flow of natural resources into produced goods. labor may to some extent replace a part of *exergy* as the human body can perform physical work. However, the human body is order of magnitude less powerful than other sources of energy. It may be the reason why humans historically replaced their body (and to some extent all living species) by external machines powered by fossil fuels Huber (2009). This is also an argument to dismiss the proposed production functions that directly include energy Van der Werf (2008) and an argument to include *exergy* into the error term. That development is also explored by Bercegol and Benisty (2022) who try to explain past TFP growth using *exergy*, though, they used a different functional form, and we propose two advantages over their method: first, our goal is to build forward looking models. Second, we try as much as possible to provide credibility intervals for all our scenarios.

Our modeling also echoes the idea of Santos et al. (2018, 2021) where they tried to model TFP with the efficiency of *exergy*. That efficiency represents the amount of real useful work done in various sectors. For instance the efficiency of internal combustion engines

that transform chemical energy onto kinetic energy or the efficiency of coal power plants that also transform chemical energy into electricity that will finally be used by machines and transformed into other form of energy (either thermal, kinetic or potential). While our modeling seems to be very close to their approach, it still has drawbacks.

Firstly, we want to be able to forecast long term GDP, but we do not know the exact amount of work needed in each sector over a long period. That makes the computation of *exergy* efficiency impossible in the long run. Moreover, that computation requires the efficiency of each transformation in each sector which is also unknown over the long run due to technical innovations. Although we could have a rough idea of such quantity, we do not want to introduce more unknown variables into our model in order to reduce long-term variance. That is the reason we wanted to use *exergy* variations. Even though that model might have a worst goodness of fit, we do not rely on unknown variables to project TFP which is a great feature for long term forecasting. However, we think such approaches might be very interesting to develop a sectoral approach. Primary *exergy* can be computed using primary energy and conversion factor (for further details : Brockway et al. (2014)).

## 4.4 Modeling the impact of *exergy* on the TFP

### 4.4.1 Data

Before diving into our proposed Bayesian models, let us describe the data we used. It consists of time series ranging from 1972 to 2018 for seven countries, gathered from the [IEA world energy balances](#)<sup>9</sup> and the Feenstra et al. (2015) data base. We studied the following countries: United States, United Kingdom, France, Germany, Japan, Canada and Italy. They are the largest ranked economies by GDP. We purposely excluded India, China and Brazil for which we could only gather *exergy* data from 1990 to 2018. We must also warn the reader, that our data consisted only of renewables sources which includes among other things, hydroelectricity, solar and wind. Unfortunately we wanted to have a more granular view of those sources thus we made the assumptions that hydro powers were already fully developed, in the studied countries. Thus, from 2000 all new renewables comes from wind and solar. We might overestimate a little the actual real amount of wind and solar, but anyway these assumptions does not introduce huge error as renewables only account for a fraction of a percent of final energy consumption in these countries.

### 4.4.2 Bayesian auto-regressive model

We focused our attention on the use of Bayesian models. We think these models offer two main advantages over traditional models, first, their output are the parameters of the posterior density of our value of interest (e.g TFP variation). As the output of

---

<sup>9</sup><https://www.iea.org/reports/world-energy-balances-overview/world>

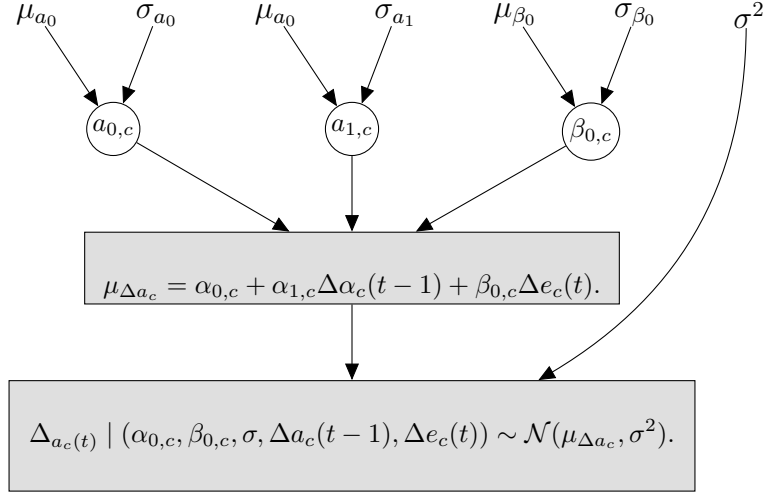


Figure 4.1: Proposed hierarchical Bayesian model, the priors being given in Table 4.2

these models are densities, they provide a natural framework to compute credibility intervals that are of special interest for long-term scenario generation. Second, they offer a mathematical framework to introduce prior knowledge on the data generating process. In this study, that property will be used to introduce the fact that western economies has similar dependencies upon *exergy*. These models are also widely used to model spatial dependencies, see for instance : McMillan et al. (2005).

The main goal of our model is to estimate the density of TFP using past values of observed TFP and current values of observed *exergy*. Let us assume that  $\mathcal{C}$  is the set of considered countries. For every country  $c \in \mathcal{C}$ , we could denote the logarithm of TFP  $a_c(t)$  and the logarithm of *exergy*  $e_c(t)$  for any  $t \in [2, \dots T]$ . Let us assume that the data generating process depends on 3 parameters :  $\alpha_{0,c}$ ,  $\alpha_{1,c}$  and  $\beta_{0,c}$ . Then, our first proposed model could be written :

$$\Delta a_c(t) = \alpha_{0,c} + \alpha_{1,c} \Delta a(t-1) + \beta_{0,c} \Delta e_c(t) + \epsilon_{c,t}. \quad (4.3)$$

Where  $\Delta$  is the difference operator, that is, for instance:  $\Delta a_c(t) = a_c(t) - a_c(t-1)$ . We could assume that any country has a true value similar to other developed economies. The underlying idea is that any western economy relies on energy to produce goods and the technology used in these countries might be to some extent similar. Thus, the *exergy* requirement to produce goods should be similar. Hence, we assume we have some priors on the distribution of the parameters and then knowing these priors, we could draw the regression coefficients of our model. Figure 4.1 illustrates our proposed hierarchical Bayesian model. Where  $\forall c \in \mathcal{C}$  :

$$\alpha_{0,c} | \mu_{\alpha_0}, \sigma_{\alpha_0}^2 \sim \mathcal{N}(\mu_{\alpha_0}, \sigma_{\alpha_0}^2), \quad \alpha_{1,c} | \mu_{\alpha_1}, \sigma_{\alpha_1}^2 \sim \mathcal{N}(\mu_{\alpha_1}, \sigma_{\alpha_1}^2) \quad \text{and} \quad \beta_{0,c} | \mu_{\beta_0}, \sigma_{\beta_0}^2 \sim \mathcal{N}(\mu_{\beta_0}, \sigma_{\beta_0}^2).$$

To check the convergence of our model and to ensure the validity of our assumption, we

CHAPTER 4. MACROECONOMIC SCENARIOS FOR CLIMATE RISK ASSESSMENT

---

estimated our model with different priors settings ranging from very informative to more uninformative ones. We summarized our setup in 4.2.

Parameters	Prior 1	Prior 2	Prior 3
$\mu_{\alpha_0}$	$\mathcal{N}(0, 1)$	$\mathcal{N}(0, 100)$	$\mathcal{N}(0, 0.25)$
$\mu_{\alpha_1}$	$\mathcal{N}(0, 1)$	$\mathcal{N}(0, 100)$	$\mathcal{N}(0.05, 0.25)$
$\mu_{\beta_0}$	$\mathcal{N}(0, 1)$	$\mathcal{N}(0, 100)$	$\mathcal{N}(0.1, 0.25)$
$\sigma^2$	$\mathcal{Ca}(1)$	$\mathcal{Ca}(10)$	$\mathcal{Ca}(0.1)$
$\sigma_{\alpha_0}^2$	$\mathcal{Ca}(10)$	$\mathcal{Ca}(10)$	$\mathcal{Ca}(1)$
$\sigma_{\alpha_1}^2$	$\mathcal{Ca}(10)$	$\mathcal{Ca}(10)$	$\mathcal{Ca}(1)$
$\sigma_{\beta_0}^2$	$\mathcal{Ca}(10)$	$\mathcal{Ca}(10)$	$\mathcal{Ca}(1)$
Tuning steps	5000	5000	5000
Draws	2000	2000	2000

Table 4.2: Parameters and priors for model (4.3)

Only one experiment, corresponding to Prior 2, will be shown below. The plots from all experiments described are provided in the following repository<sup>10</sup> and we used Salvatier et al. (2016) the PyMC3 python package. In Table 4.2,  $\mathcal{Ca}(\beta)$  corresponds to a half-Cauchy density of parameter  $\beta$ ,  $f_{\mathcal{Ca}(\beta)}$ , and  $\mathcal{N}(\mu, \sigma^2)$  to a Gaussian density,  $f_{\mathcal{N}(\mu, \sigma^2)}$ , of mean  $\mu$  and variance  $\sigma^2$ . These densities are defined as :

$$\forall x \in \mathbb{R} : f_{\mathcal{Ca}(\beta)}(x) = \frac{2}{\pi\beta \left[1 + \left(\frac{x}{\beta}\right)^2\right]} \mathbb{1}_{x>0} \quad \text{and} \quad f_{\mathcal{N}(\mu, \sigma^2)}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{x - \mu}{\sigma}\right)^2\right).$$

With the priors described in Table 4.3, the posterior densities for our desired parameters are shown in Figure 4.2.

The first striking result of Figure 4.2 is that the *exergy* variations coefficient is significantly greater than 0 for almost all countries (except France). The other interesting result is that every coefficient is similar for every country with a mean value near 0.2. It may indicate that every country has the same dependence upon *exergy*. The lag coefficient is not significant for all countries and the constant term which might account for exogenous technological improvement is greater than 0 for every country but two. In this model, exogenous improvement only accounts for a few tenths of a percent, thus, the exogenous TFP variation supposed by W. Nordhaus (2018) is explain in this model by two components: the first one is indeed an exogenous element that drives TFP upward. It may be interpreted as technological advancement. The second is *exergy* variation that may come from the necessity to increase *exergy* use to improve the productivity, in terms of produced goods, of already existing assets. 4.2 shows the predictions of our model on historic TFP variations.

---

<sup>10</sup>Code and data can be found at : <https://github.com/GFlament-Adway/BDICE>.

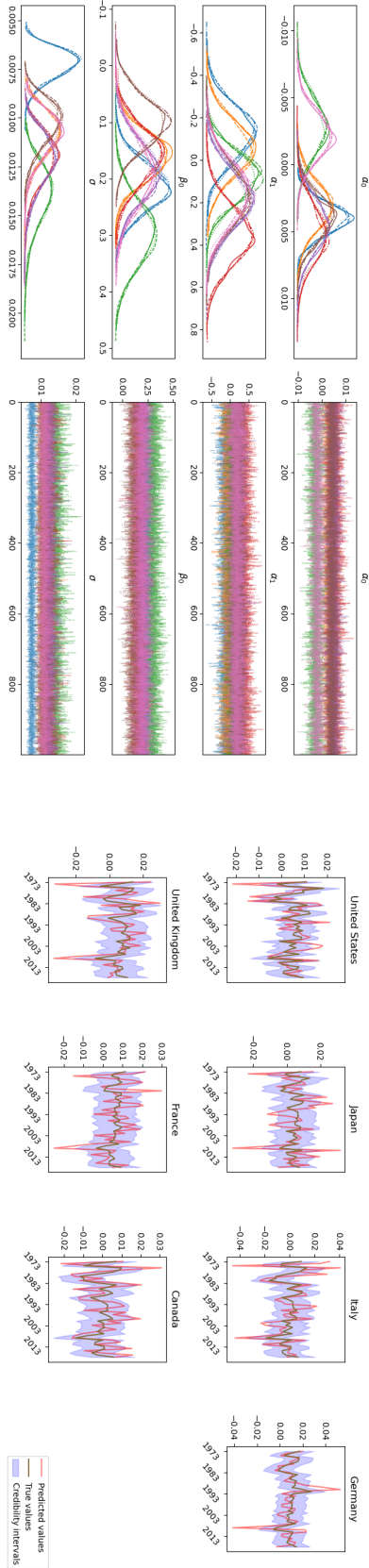


Figure 4.2: Left panel: Two estimates according to model (4.3) (dotted and plain lines) for each posterior density of:  $\alpha_0$ ,  $\alpha_1$ ,  $\beta_0$  and  $\sigma$ . Each color corresponds to one country (see Table 4.2 for an exhaustive list), represented by a plain and a dotted line. Middle panel: sampled values used to estimate posterior densities. Right panel: Posterior means, 0.1 and 0.9 quantiles for each country, according to model (4.3) The  $R^2$  is 0.31 and the mean squared error (mse) is  $10^{-3}$ .

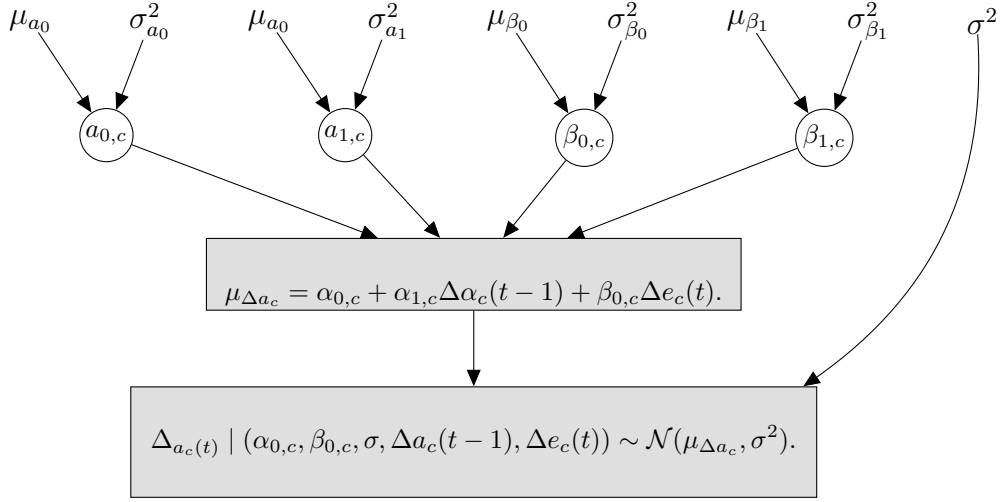


Figure 4.3: Proposed hierarchical Bayesian model, the priors being given in Table 4.3

According to this result, we observe huge differences with the trajectory proposed by W. Nordhaus (2018). For instance, in the DICE model, the growth rate is near 1.9% per year. That trajectory is compatible with our modeling if and only if we suppose an increase of *exergy* use by about 2% a year. Unfortunately, that growth seems to be in contradiction to the Paris Agreement as we shall explain in Section 4.5. On the other hand, if one tries to meet the Paris agreement and reduce greenhouse gas emissions then we have to reduce *exergy* in the short term. Hence we have to reduce economic output. It is the main difference between our model and the classical DICE, we add what may be interpreted as ecological constraint.

### 4.4.3 Bayesian model with an additional exogenous variable.

The preceding modeling still lacks one important feature: indeed, a country that imports a lot of manufactured goods and exports a lot of services might have a relatively low correlation simply because our data does not consider such cases. Let us simply add one other variables, which is the variation of world *exergy*. Precisely, this is the variation of the sum of *exergy* from our listed countries which could be considered as a proxy that accounts for imports to a given country. That proxy is unfortunately not perfect as it doesn't contains the information on what is the *exergy* contents of these trades. One further research perspective would be to try to find a relationship between a financial trade in a sector and the *exergy* embedded in that trade. Eventhough it has limitations, 4.3 presents you the proposed model.

$$\Delta a_c(t) = \alpha_{0,c} + \alpha_{1,c} \Delta a_c(t-1) + \beta_{0,c} \Delta e_c(t) + \beta_{1,c} \Delta \left( \sum_{c \in \mathcal{C}} e_c(t) \right) + \epsilon_{c,t}. \quad (4.4)$$

Parameters	Prior 4	Prior 5	Prior 6	Prior 7	Prior 8
$\mu_{\alpha_0}$	$\mathcal{N}(0, 1)$	$\mathcal{N}(0, 100)$	$\mathcal{N}(0, 1)$	$\mathcal{N}(0, 100)$	$\mathcal{N}(0, 1)$
$\mu_{\alpha_1}$	$\mathcal{N}(0.05, 1)$	$\mathcal{N}(0, 100)$	$\mathcal{N}(0, 1)$	$\mathcal{N}(0, 100)$	$\mathcal{N}(0, 1)$
$\mu_{\beta_0}$	$\mathcal{N}(0, 1)$	$\mathcal{N}(0, 100)$	$\mathcal{N}(0, 1)$	$\mathcal{N}(0, 100)$	$\mathcal{N}(0, 1)$
$\mu_{\beta_1}$	$\mathcal{N}(0.1, 0.25)$	$\mathcal{N}(0, 100)$	$\mathcal{N}(0, 1)$	$\mathcal{N}(0, 100)$	$\mathcal{N}(0, 1)$
$\sigma$	$\mathcal{Ca}(0.1)$	$\mathcal{Ca}(10)$	$\mathcal{Ca}(1)$	$\mathcal{Ca}(10)$	$\mathcal{Ca}(1)$
$\sigma_{\alpha_0}$	$\mathcal{Ca}(1)$	$\mathcal{Ca}(10)$	$\mathcal{Ca}(1)$	$\mathcal{Ca}(10)$	$\mathcal{Ca}(1)$
$\sigma_{\alpha_1}$	$\mathcal{Ca}(1)$	$\mathcal{Ca}(10)$	$\mathcal{Ca}(1)$	$\mathcal{Ca}(10)$	$\mathcal{Ca}(1)$
$\sigma_{\beta_0}$	$\mathcal{Ca}(1)$	$\mathcal{Ca}(10)$	$\mathcal{Ca}(1)$	$\mathcal{Ca}(10)$	$\mathcal{Ca}(1)$
$\sigma_{\beta_1}$	$\mathcal{Ca}(1)$	$\mathcal{Ca}(10)$	$\mathcal{Ca}(1)$	$\mathcal{Ca}(10)$	$\mathcal{Ca}(1)$
Tuning steps	5000	5000	5000	20000	20000
Draws	2000	2000	2000	2000	2000

Table 4.3: Tested priors of model (4.4).

Only one experiment (Prior 7) will be shown below but all plots from all described experiment could be checked<sup>11</sup>.

The constant, lag and the variance posterior densities, see Figure 4.4, are similar to the previous posterior densities, see Figure 4.2, hence our interpretation remains the same for these coefficients. However, the posterior densities for  $\beta_{0,c}$  seems to be centered around a positive value but a lower one than obtained in the previous model. It means that the variations of the *exergy* of a country  $c$  does not explain much of its TFP variation. However, the variation of the *exergy* of the world explains a lot of the variations of the TFP for other countries, especially for Italy. It may reflect interdependencies and the importance of imports for western economies. It would be of great interest to include China in the same model because they manufacture much of the goods consumed in western economies, but, unfortunately we lack their data. That model seems to have a better interpolation of the observed data and will be used for further analyses.

#### 4.4.4 Forecast values during the Covid-19 crisis and the recovery

We must check the relevance of our model to forecast real TFP values. We have been living through a very unique period in modern history. Indeed, during the Covid-19 crisis, it was the first time humanity reduced its greenhouse gas emissions at such rates. Of course, it didn't last long. However, it is a very unique chance for us to compare our model with the real data that are being published. As the greenhouse gas reduction only comes from a slowdown in the use of fossil fuel, and not a drastic change in the way we produce energy, it may directly be translated to *exergy* variations. That is, we assume that GHG variations are equal to *exergy* variations. That assumption represents

---

<sup>11</sup>Code and data can be found at : <https://github.com/GFlament-Adway/BDICE>

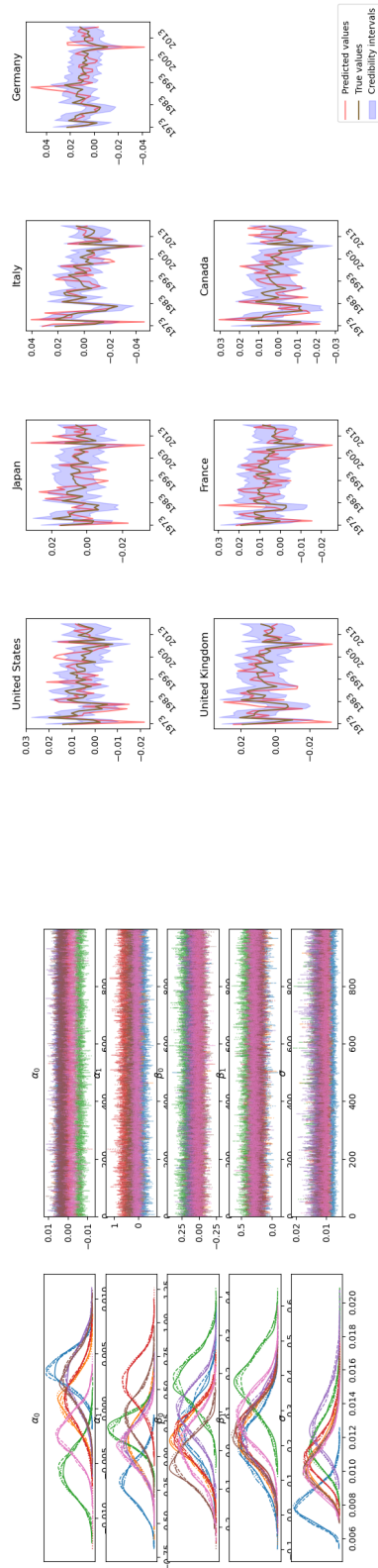


Figure 4.4: Left panel:  $\alpha_0$ ,  $\alpha_1$ ,  $\beta_0$ ,  $\beta_1$  and  $\sigma$  Posterior density. Each color corresponds to one country (see Table 4.2 for an exhaustive list), 2 chains have been generated represented by a plain and a dotted line. The sampled values are also shown. Right panel: Posterior means and 0.1 and 0.9 quantiles for each country. The  $R^2$  is 0.31 and the mean squared error (mse) is  $10^{-3}$ .

Country	Estimated <i>exergy</i> variation (%)	Model (4.3) (%)	Model (4.4) (%)	Observed GDP variation (%)
United States	-10.9	-2.0 (-2.7, -1.3)	-2.2(-3, -1.4)	-2.3
France	-14.3	-2.1 (-3.0, -1)	-2(-3, -0.8)	-4.2
Canada	-9.6	-2.5 (-3.2, -1.8)	-2.9(-3.9, -1.9)	-5.6
Italy	-10.7	-3.2 (-4.2, -2.2)	-5.8(-6.7, -4.8)	-5.9
Germany	-9.7	-1.7 (-2.7, -0.8)	-2.6(-3.8, -1.2)	-1.4
Japan	-6.5	-1 (-1.5, -0.4)	-2(-3.2, -0.1)	-1.7
United Kingdom	-11.2	-1.7 (-2.4, -1.1)	-3(-4.2, -1.8)	-4

Table 4.4: For each country, models (4.3) and (4.4) predictions during the Covid-19 crisis. The mean predictions are given. The 0.1 and 0.9 quantiles are shown between parenthesis.

an approximation of the true *exergy* variation and should be treated as such. Still, we can not provide more extreme test as, historically, no such variations have been observed. 4.4 presents the forecast TFP under the first waves of Covid-19, while 4.5 presents the forecast TFP after the 2021 recovery for different scenarios. Our model is able to explain, at least partially and for a wide range of countries, the huge variations seen during the Covid-19 crisis. The prediction for some countries seems to be too optimistic (France and Canada) that might be explained by specificities of their economy, for example tourism represents about 10% of France GDP. Our model probably does not properly take into account all the covid restrictions that limited international travel. Canada on the other hand rely heavily on the oil and gas and the mining sector, even though our model reflects properly the quantity drop it can not reflect the price drop of commodities seen during 2020. Besides that drawback, our model is able to make such predictions even though no such variations are observed in the training data. The recent ban on Russian fossil fuels should result in a drop of *exergy* for western economies and, according to our model, result in a drop of TFP.

## 4.5 Future *exergy* availability impact on real GDP

### 4.5.1 Model for future *exergy* availability

*Exergy* values are computed using the methodology presented by Brockway et al. (2014). Hence, we use the same energy sources as the International Energy Agency (IEA): Coal, Oil, Natural gas, nuclear, hydroelectricity, biomass, solar and wind. To produce *exergy* scenarios, we have to produce scenarios for each of these energy sources. We could consider an exogenous path, which could be interpreted as a government policy that set a track to be followed. Or we could consider a track according to the future relative price of each energy source. That second methodology has the appealing property of directly taking into account a carbon tax which is thought to be an efficient way to reduce carbon emissions (see for instance: Gillingham and Stock (2018)). Under that second idea, we could inverse the question. One may wonder what would be the ideal carbon tax in order to reach some level of greenhouse gas emission reduction. Answering that question in an economic

	Country	Prediction (%) for 5% <i>exergy</i> growth	Prediction (%) for 7% <i>exergy</i> growth
Model (4.3)	United States	1.5(1.2, 1.8)	1.9(1.5, 2.4)
	France	1.4(1.1, 1.8)	1.8(1.3, 2.2)
	Canada	0.9(0.6, 1.2)	1.4(1, 1.8)
	Italy	1.1(0.7, 1.6)	1.6(1.1, 2.3)
	Germany	2.3(1.8, 2.8)	2.8(2.2, 3.5)
	Japan	1.3(1, 1.6)	1.7(1.3, 2.1)
	United Kingdom	1.9(1.4, 2.3)	2.4(1.8, 2.9)
Model (4.4)	United States	1.7(1, 2.4)	2.2(1.4, 3.0)
	France	1.2(0.5, 1.8)	1.6(0.8, 2.3)
	Canada	1(0.4, 1.6)	1.6(0.8, 2.4)
	Italy	1.8(1.3, 2.3)	2.9(2.2, 3.5)
	Germany	1.2(0.5, 1.8)	2.5(1.7, 3.4)
	Japan	1.6(1, 2.1)	2.1(1.4, 2.8)
	United Kingdom	1.7(1, 2.4)	2.4(1.5, 3.2)

Table 4.5: For each country, models (4.3) and (4.4) predictions for 2021 under different *exergy* scenarios. The mean predictions are given. The 0.1 and 0.9 quantiles are shown between parenthesis.

model where *exergy* has its fair share, is key for policy makers. Unfortunately, the scenario developed by Nordhaus seems unsuitable from that perspective. The economic growth depends mainly on exogenous assumptions, hence, the economic impact of such taxes is very small.

If we are able to draw a path for that carbon tax, we could easily incorporate it into the relative price of each sources. That idea closely matches the approach of Alestra et al. (2022), based on elasticities between energy input estimated by D. I. Stern (2012) or Papageorgiou et al. (2017) one could compute scenarios of energy production. But as the main goal of this Chapter is to develop a TFP model, and for the sake of simplicity, we used exogenous path for all energy sources.

## 4.5.2 Exogenous trajectory

As we use the Brockway et al. (2014) methodology, we need the primary energy from each energy source to compute *exergy*. Future values of energy input could be set according to regulatory objectives, for example, limiting global warming to 2°C as proposed by the Paris Agreement. It states:

*Holding the increase in the global average temperature to well below 2°C above pre-industrial levels and pursuing efforts to limit the temperature increase to*

*1.5°C above pre-industrial levels, recognizing that this would significantly reduce the risks and impacts of climate change.*

To ensure we are on track to meet that Agreement, we have to lower our carbon emissions by approximately 4% per year from now on until we reach net zero. Indeed the annual global greenhouse gases emissions has to decrease by approximately 4% a year Chen et al. (2022). Thus, to illustrate our model, we will simply assume that every country reduces its emissions by that figure. Hence, we ensure that global decrease meets the Paris agreement criteria. Of course, the energy sector is not the only greenhouse gas emitter, but if we assume every economic sector (even agriculture) has to decrease at the same pace, then we know that coal, oil and natural gas consumption have to be reduced by 4% a year. Now if we assume an increase by 1% per year of nuclear and hydro-electricity and a 5% per year increase of new renewables, then we could compute a scenario for *exergy*. These scenarios are basic projections and one could study the impact of different path on the GDP. These projections are completely unrealistic as they completely ignore the difficulty of such industrial, political and social efforts. However, they are of great interest to see our model best case transition to a net zero economy while keeping global warming below 2°C.

As we stated earlier, we could design an *exergy* scenario to follow the growth trajectory of the traditional DICE model. Unfortunately such a scenario assumes an increase of greenhouse gas emissions by 2% per year which is not remotely close to meet the Paris Agreement. The most pessimistic scenario under the Paris Agreement scenario results in a net loss of TFP.

## 4.6 The climate module

One could use the standard damage function used by, see for example (4.2). It is equivalent to that used by Alestra et al. (2022). In the traditional settings an increase of 6°C would result in a damage of 8% to the GDP. We may use a different damage function such as that presented by Weitzman (2012). In this paper, Weitzman (2012) modified the damage function to measure a form of uncertainty. The science of climate change rely on models that are very accurate but still have a few unknowns that are taken into account by the IPCC. Weitzman (2012) illustrated it with the following example: a doubling of  $CO_2$  concentration might lead to an increase between 2° and 4.5° and the most likely temperature increase is 3°. This uncertainty will lead to variability in the output of the damage function (e.g the damage to the economy) and we will discuss in following sections how to integrate that variability into our model.

We use the standard model estimated by Hsiang and Kopp (2018), that states that the relation between global warming and greenhouse gas concentration is linear. Let  $T(t)$  be the temperature at time  $t$  and  $SCO_2(t)$  be the total amount of  $CO_2$  in the atmosphere, that quantity is very closely related to the quantity of fossil fuels burnt since the beginning

CHAPTER 4. MACROECONOMIC SCENARIOS FOR CLIMATE RISK ASSESSMENT

Probability \ $SCO_2(t^*)$	400 ( $\approx$ RCP 2.6)	450	500	550	600 ( $\approx$ RCP 4.5)	650	700 ( $\approx$ RCP 6.0)
$\mathbb{P}(T(t^*) > 2)$	27%	50%	66%	74%	79%	82%	84%
$\mathbb{P}(T(t^*) > 2.5)$	9%	33%	50%	62%	69%	74%	77%
$\mathbb{P}(T(t^*) > 3)$	2%	17%	34%	48%	57%	64%	69%
$\mathbb{P}(T(t^*) > 3.5)$	0.4%	7%	21%	34%	45%	53%	60%
$\mathbb{P}(T(t^*) > 4)$	0.05%	2%	11%	22%	33%	42%	49%

Table 4.6: Temperature increase probability under different greenhouse gas concentration scenarios.

of the industrial revolution. We assumed the same sinks as in W. Nordhaus (2018). Then, a rather simple and accurate modeling is:

$$T(t) = \tau SCO_2(t).$$

The usual value of  $\tau$  is 0.0008. But we would like to go a bit further. Unfortunately, given a stock of greenhouse gases in the atmosphere, the temperature increases can not be determined exactly. As IPCC (2007) stated, a doubling of greenhouse gas concentration will likely cause an increase of  $3^\circ C$ , but with low probability, it could result in a warming lower than  $2^\circ C$  or higher than  $4.5^\circ C$ . Weitzman (2010) tries to fit distribution according to some intervals given by the IPCC. If we use what he calls thin-tailed density (Gaussian density), we get tail probabilities described in 4.6, computed for a given time  $t^*$ . The time  $t^*$  represents the moment the greenhouse gas concentration reaches equilibrium. It is not necessary the time when fossil fuels stop being burnt as some feedback loops might be involved in this phenomenon hence prolonging the accumulation of greenhouse gases in the atmosphere.

Given the previous densities, we can estimate the probability of damage greater than some given threshold. Recall that  $\Omega_N(T)$  and  $\Omega_W(T)$  respectively represents the damage function of Nordhaus and Weitzman. We could consider 2 damage functions:

$$\Omega_N(T) = \frac{1}{1 + 0.0026.T(t)^2}, \tag{4.5}$$

$$\Omega_W(T) = \frac{1}{1 + 0.00239.T(t)^2 + 0.000156.T(t)^7}. \tag{4.6}$$

Where (4.5) corresponds to the damage function proposed by W. Nordhaus and Sztorc (2013), while (4.6) corresponds to that proposed by Weitzman (2010) with a slight modification for computational reasons. As the global warming now becomes a random variable, the damage also becomes a random variable. We could compute damage tail damage probabilities presented in Table 4.7.

CHAPTER 4. MACROECONOMIC SCENARIOS FOR CLIMATE RISK ASSESSMENT

$SCO_2(t^*)$	400 ( $\approx$ RCP 2.6)	450	500	550	600 ( $\approx$ RCP 4.5)	650	700 ( $\approx$ RCP 6.0)
$\mathbb{P}(1 - \Omega_N(T(t^*)) > 1\%)$	28%	53%	67%	75%	80%	83%	85%
$\mathbb{P}(1 - \Omega_W(T(t^*)) > 1\%)$	43%	65%	76%	81%	85%	87%	88%
$\mathbb{P}(1 - \Omega_N(T(t^*)) > 5\%)$	$10^{-3}\%$	0.6%	5%	13%	22%	31%	39%
$\mathbb{P}(1 - \Omega_W(T(t^*)) > 5\%)$	10%	33%	50%	62%	69%	74%	78%
$\mathbb{P}(1 - \Omega_N(T(t^*)) > 10\%)$	$10^{-9}\%$	$10^{-4}\%$	0.04%	0.5%	2%	5%	9%
$\mathbb{P}(1 - \Omega_W(T(t^*)) > 10\%)$	0.4%	20%	38%	51%	60%	67%	71%
$\mathbb{P}(1 - \Omega_N(T(t^*)) > 20\%)$	$10^{-25}\%$	$10^{-13}\%$	$10^{-7}\%$	$10^{-5}\%$	$10^{-3}\%$	0.02%	0.1%
$\mathbb{P}(1 - \Omega_W(T(t^*)) > 20\%)$	0.7%	9%	24%	38%	49%	56%	63%

Table 4.7: Damage probability under different greenhouse gas concentration scenarios.

For example, in Table 4.7 the probability of the damage being greater than 1% is 43% for the Weitzman damage function with a greenhouse gas concentration of 400ppm. As expected, large damages have very low probability to happen using the Nordhaus damage function. Nevertheless, as we have no observation it is very hard to calibrate such functions and the computed probabilities have to be interpreted as indications rather than precise estimates. A recent discussion about the construction of such function has been presented by Barnett et al. (2022).

## 4.7 Real GDP forecasts

We computed the real output (Figure 4.5) by combining several assumptions made in this Chapter. We tested both Nordhaus (4.5) and Weitzman (4.6) damage function. We used priors 3 and 4 as described in Table 4.2 and Table 4.3. The population growth matches the assumption of W. Nordhaus (2018) and the capital stock model is described in Section 4.3.2. We also computed a second simulation with a pessimistic case of climate warming (Figure 4.6). Moreover, Figures 4.5 and 4.6 show global economic output if one assumes that the TFP variation will follow the US model. That illustration tries to compare our findings with the assumptions of W. Nordhaus (2018). Of course, one could compute the same model for all the considered countries. We also decided to illustrate our findings on a 30 years time span, to compare our scenarios with the one of the NGFS<sup>12</sup> and DICE. Moreover our Bayesian framework makes the computation of credibility intervals easy. For instance, to generate one scenario we generated one sample according to our posterior densities for each year. We repeated (1000 times in the following figures) that process to obtain enough scenarios to compute credibility intervals.

Our projections are very different from those computed by DICE W. Nordhaus (2018). He obtains around 200 trillion \$2010 in 2050 that is roughly twice our projected GDP. That enormous differences comes from the integration of the impact of natural resources into the TFP that is a major driver of that GDP growth. Indeed W. Nordhaus and Sztorc (2013) assumes a roughly 1.9%/year growth of the GDP, that assumption is unsatisfactory

<sup>12</sup><https://www.ngfs.net/ngfs-scenarios-portal/>

CHAPTER 4. MACROECONOMIC SCENARIOS FOR CLIMATE RISK ASSESSMENT

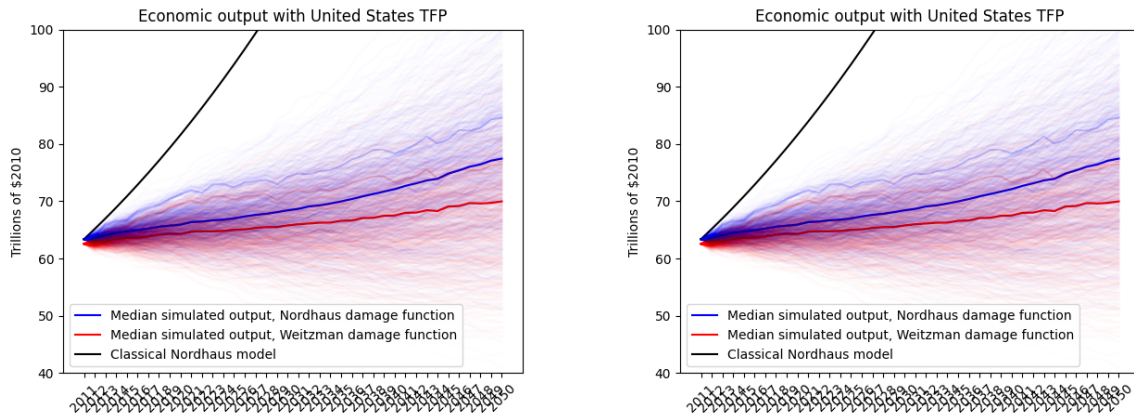


Figure 4.5: Left panel: Simulated output paths using priors 3 from (4.3) and Nordhaus damage function, Right panel: Simulated output paths using priors 4 from (4.4) and Nordhaus damage function.

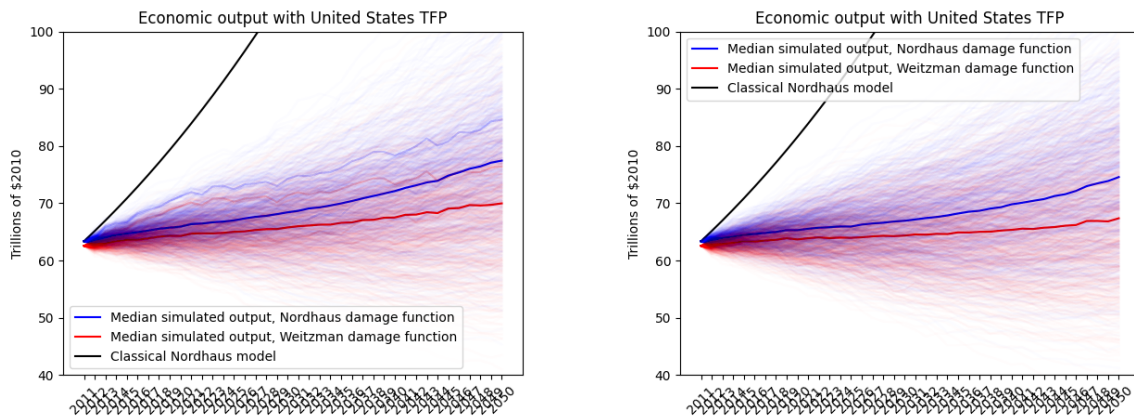


Figure 4.6: Left panel: Simulated output paths using priors 3 model (4.3) and damage function with a 0.1 quantile (4.6) temperature increase, Right panel: Simulated output paths using priors 4 from (4.4) and damage function with a 0.1 quantile (4.6).

and we rather modeled it thanks to *exergy*, that changes alone drives all the difference between our model and DICE. This changes drastically the consequences and the policy required to handle the social consequences of GDP growth slowdown. Each trajectory represent one draw from the posterior estimated by a model presented in 4.4.2 or 4.4.3. If we assume such predictions holds true for any country, worldwide, then it raises the question of welfare redistribution. Under a pessimistic scenario, where GDP declines or stays roughly constant, we can not ask for the same decline in western countries and in the Sub-Saharan countries. In the first case, it would require a mild comfort reduction, in the second case, it might result in a catastrophic humanitarian crisis.

We must stress that, even if one thinks climate change will not cause as much risk as transition towards a low carbon economy, the amount of fossil fuels on earth is a given constant quantity. One can not burn that resource recklessly for eternity. It must come to an end. A low carbon economy will happen regardless of political or social will. The only remaining question about that transition should be the timing and the pace.

Moreover, one must keep in mind that our scenarios are best case scenarios and undermines major technological uncertainties. For instance we assumed the transition from fossil fuels to electricity to be instantaneous for any industrial process. We assumed the physical property (availability, density, etc) of low carbon electricity sources to be identical to fossil fuel sources. That is obviously very optimistic. We also assumed climate change to have no impact on population growth and we assumed perfect and complete collaborations between all countries. These assumptions are far from satisfactory. However, we believe such scenarios are a first step in the right direction to consider some ecological constraints in the economic literature.

## 4.8 Conclusion

The main conclusion of this work is the importance of *exergy* variations to explain TFP variations. In almost all western economies, even through the Covid-19 and the 2008 financial crisis, our model seems to explain relatively well observed data. The *exergy* contribution to the TFP seems to be equivalent for all countries. We introduce a proxy for imports that improved our model. It underlines its importance in modern western economies. A secondary contribution of this work is the improvement of forward looking scenarios used in real world policy assessments. Confidence intervals are crucially important for these exercises, and our methodology introduces a natural framework to compute them.

This work contributes to a larger discussion where economists try to introduce energy as a main factor of economic development Ayres et al. (2003), Bercegol and Benisty (2022), and Santos et al. (2018, 2021). Our work also tries to help policy makers understand the effect of different energy policy. Directly integrating tax or grants into *exergy* scenarios seems to be a natural extension of this work. Further studies of the methodology are crucial to improve the usefulness of our model. Also, building a separated model with

## CHAPTER 4. MACROECONOMIC SCENARIOS FOR CLIMATE RISK ASSESSMENT

---

sectorized scenarios is a legitimate future research direction.

Finally, in this Chapter, we mainly discussed the relationship between Growth of Domestic Product and *exergy*. Unfortunately besides the greenhouse gas problem, a significant ecological puzzle has to be solved and maybe, access to clean energy would result in an ecological catastrophe. Even with clean energy, destroying forests to grow crops, over-fishing or dumping plastic into the ocean would still result in a biodiversity disaster. Considering other dimensions of the ecological problem is mandatory to correctly handle the problem but is still beyond the scope of this paper.

## References

- Adrian, T., Boyarchenko, N., & Giannone, D. (2019). Vulnerable growth. *Am. Econ. Rev.*, 109(4), 1263–1289.
- Aghion, P., Harris, C., Howitt, P., & Vickers, J. (2001). Competition, imitation and growth with step-by-step innovation. *Rev. Econ. Stud.*, 68(3), 467–492.
- Alestra, C., Cette, G., Chouard, V., & Lecat, R. (2022). Growth impact of climate change and response policies: The advanced climate change long-term (ACCL) model. *J. Policy Model.*, 44(1), 96–112.
- Ayres, R. U., Ayres, L. W., & Warr, B. (2003). Exergy, power and work in the us economy, 1900–1998. *Energy*, 28(3), 219–273.
- Barnett, M., Brock, W., & Hansen, L. P. (2022). Climate change uncertainty spillover in the macroeconomy. *NBER*, 36(1), 253–320.
- Battiston, S., & Monasterolo, I. (2020). 11. a science-based climate-stress testing framework to integrate forward-looking climate transition risk into existing supervisory. *Joint JRC-EBA workshop on Banking Regulation and Sustainability*, 41.
- Bercegol, H., & Benisty, H. (2022). An energy-based macroeconomic model validated by global historical series since 1820. *Ecol. Econ.*, 192, 107253.
- Bertram, C., Hilaire, J., Kriegler, E., Beck, T., Bresch, D., Clarke, L., Cui, R., Edmonds, J., Charles, M., Zhao, A., et al. (2021). Ngfs climate scenario database: Technical documentation v2. 2.
- Brockway, P. E., Barrett, J. R., Foxon, T. J., & Steinberger, J. K. (2014). Divergence of trends in us and uk aggregate exergy efficiencies 1960–2010. *Environ. Sci. Technol.*, 48(16), 9874–9881.
- Calvin, K., Patel, P., Clarke, L., Asrar, G., Bond-Lamberty, B., Cui, R. Y., Di Vittorio, A., Dorheim, K., Edmonds, J., Hartin, C., et al. (2019). Gcam v5. 1: Representing the linkages between energy, water, land, climate, and economic systems. *Geoscientific Model Development*, 12(2), 677–698.
- Capellán-Pérez, I., de Blas, I., Nieto, J., de Castro, C., Miguel, L. J., Carpintero, Ó., Mediavilla, M., Lobejón, L. F., Ferreras-Alonso, N., Rodrigo, P., et al. (2020). Medeas: A new modeling framework integrating global biophysical and socioeconomic constraints. *Energy Environ. Sci.*, 13(3), 986–1017.
- Cartellier, F. (2022). Climate stress testing, an answer to the challenge of assessing climate-related risks to the financial system? [Available at SSRN: <https://ssrn.com/abstract=4179311> or <http://dx.doi.org/10.2139/ssrn.4179311>].
- Chen, D., et al. (2022). Figure TS.4 in IPCC, 2021: Technical summary. the physical science basis. contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change.
- Cleveland, C. J., Costanza, R., Hall, C. A., & Kaufmann, R. (1984). Energy and the us economy: A biophysical perspective. *Science*, 225(4665), 890–897.
- Feenstra, R. C., Inklaar, R., & Timmer, M. P. (2015). The next generation of the penn world table. *Am. Econ. Rev.*, 105(10), 3150–82.

- Flament, G. (2023). Impact of the energy transition on long-term factor productivity. *Struct. Change Econ. Dyn.*, 66(100), 393–406.
- Gillingham, K., & Stock, J. H. (2018). The cost of reducing greenhouse gas emissions. *J. Econ. Perspect.*, 32(4), 53–72.
- Hsiang, S., & Kopp, R. E. (2018). An economist’s guide to climate change science. *J. Econ. Perspect.*, 32(4), 3–32.
- Huber, M. T. (2009). Energizing historical materialism: Fossil fuels, space and the capitalist mode of production. *Geoforum*, 40(1), 105–115.
- IPCC. (2007). The physical science basis. *Contribution of working group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, 996.
- IPCC. (2021). The sixth assessment report -the physical science basis [See the full report at : <https://www.ipcc.ch/report/ar6/wg1/>].
- Isaksson, A., et al. (2007). Determinants of total factor productivity: A literature review. *Research and Statistics Branch, UNIDO*, 1(101), 672.
- Johansson, D. J., Lucas, P. L., Weitzel, M., Ahlgren, E. O., Bazaz, A. B., Chen, W., den Elzen, M. G., Ghosh, J., Grahn, M., Liang, Q.-M., et al. (2015). Multi-model comparison of the economic and energy implications for china and india in an international climate regime. *Mitig. Adapt. Strateg. Glob. Chang.*, 20, 1335–1359.
- Keen, S., Ayres, R. U., & Standish, R. (2019). A note on the role of energy in production. *Ecol. Econ.*, 157, 40–46.
- Krey, V., Havlik, P., Kishimoto, P. N., Fricko, O., Zilliacus, J., Gidden, M., Strubegger, M., Kartasasmita, G., Ermolieva, T., Forsell, N., Gusti, M., Johnson, N., Kikstra, J., Kindermann, G., Kolp, P., Lovat, F., McCollum, D. L., Min, J., Pachauri, S., ... Riahi, K. (2020). Messageix-globiom documentation – 2020 release. <https://doi.org/10.22022/iacc/03-2021.17115>
- Luderer, G., Leimbach, M., Bauer, N., Kriegler, E., Baumstark, L., Bertram, C., Giannousakis, A., Hilaire, J., Klein, D., Levesque, A., Mouratiadou, I., Pehl, M., Pietzcker, R., Piontek, F., Roming, N., Schultes, A., Schwanitz, V. J., & Strefler, J. (2015, November). Description of the remind model (version 1.6) [Available at SSRN: <https://ssrn.com/abstract=2697070> or <http://dx.doi.org/10.2139/ssrn.2697070>].
- McMillan, N., Bortnick, S. M., Irwin, M. E., & Berliner, L. M. (2005). A hierarchical Bayesian model to estimate and forecast ozone through space and time. *Atmos. Environ.*, 39(8), 1373–1382.
- Meadows, D. H., Randers, J., & Meadows, D. L. (1972). The limits to growth (1972) [First edition of the book available at: <https://www.library.dartmouth.edu/digital/digital-collections/limits-growth>].
- Mora, C., Counsell, C. W., Bielecki, C. R., & Louis, L. V. (2017). Twenty-seven ways a heat wave can kill you: Deadly heat in the era of climate change. *Circulation: Cardiovascular Quality and Outcomes*, 10(11), e004233.
- Mora, C., Spirandelli, D., Franklin, E. C., Lynham, J., Kantar, M. B., Miles, W., Smith, C. Z., Freel, K., Moy, J., Louis, L. V., et al. (2018). Broad threat to humanity

- from cumulative climate hazards intensified by greenhouse gas emissions. *Nat. Clim. Change*, 8(12), 1062–1071.
- Nordhaus, W., & Sztorc, P. (2013). User’s manual for DICE-2013r [Code and user’s manual available at : <https://github.com/psztorc/DICE>].
- Nordhaus, W. (2014). Estimates of the social cost of carbon: Concepts and results from the dice-2013r model and alternative approaches. *J. Assoc. Environ. Resour. Econ.*, 1(1/2), 273–312.
- Nordhaus, W. (2018). Evolution of modeling of the economics of global warming: Changes in the DICE model, 1992–2017. *Climatic change*, 148(4), 623–640.
- Nordhaus, W. D. (1991). To slow or not to slow: The economics of the greenhouse effect. *Econ. J.*, 101(407), 920–937.
- Nordhaus, W. D., et al. (1992). *The “DICE” model: Background and structure of a dynamic integrated climate-economy model of the economics of global warming* (tech. rep.) (Available at : <https://elischolar.library.yale.edu/cgi/viewcontent.cgi?article=2251&context=cowles-discussion-paper-series>). Cowles Foundation for Research in Economics, Yale University.
- Papageorgiou, C., Saam, M., & Schulte, P. (2017). Substitution between clean and dirty energy inputs: A macroeconomic perspective. *Rev Econ. Stat.*, 99(2), 281–290.
- Parlatore, C., & Philippon, T. (2022, April). *Designing stress scenarios* (Working Paper No. 29901). National Bureau of Economic Research. <https://doi.org/10.3386/w29901>
- Petit, J.-R., Jouzel, J., Raynaud, D., Barkov, N. I., Barnola, J.-M., Basile, I., Bender, M., Chappellaz, J., Davis, M., Delaygue, G., et al. (1999). Climate and atmospheric history of the past 420,000 years from the vostok ice core, antarctica. *Nature*, 399(6735), 429–436.
- Pindyck, R. S. (2017). The use and misuse of models for climate policy. *Rev. Environ. Econ. Policy*, 11(1), 100–114.
- Pradhan, B. K., & Ghosh, J. (2019). Climate policy vs. agricultural productivity shocks in a dynamic computable general equilibrium (cge) modeling framework: The case of a developing economy. *Econ. Model.*, 77, 55–69.
- Pradhan, B. K., & Ghosh, J. (2022). A computable general equilibrium (cge) assessment of technological progress and carbon pricing in india’s green energy transition via furthering its renewable capacity. *Energy Econ.*, 106, 105788.
- Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. (2016). Probabilistic programming in python using pymc3 [Documentation available at : <https://www.pymc.io/projects/docs/en/stable/learn.html>]. *PeerJ Computer Science*, 2, e55.
- Santos, J., Borges, A. S., & Domingos, T. (2021). Exploring the links between total factor productivity and energy efficiency: Portugal, 1960–2014. *Energy Econ.*, 101, 105407.
- Santos, J., Domingos, T., Sousa, T., & Aubyn, M. S. (2018). Useful exergy is key in obtaining plausible aggregate production functions and recognizing the role of energy in economic growth: Portugal 1960–2009. *Ecol. Econ.*, 148, 103–120.

## CHAPTER 4. MACROECONOMIC SCENARIOS FOR CLIMATE RISK ASSESSMENT

---

- Shreve, C. M., & Kelman, I. (2014). Does mitigation save? reviewing cost-benefit analyses of disaster risk reduction. *Int. J. Disaster Risk Reduct.*, 10, 213–235.
- Solow, R. M. (1956). A contribution to the theory of economic growth. *The quarterly journal of economics*, 70(1), 65–94.
- Stern, D. I. (2011). The role of energy in economic growth. *Ann. N. Y. Acad. Sci.*, 1219(1), 26–51.
- Stern, D. I. (2012). Interfuel substitution: A meta-analysis. *J. Econ. Surv.*, 26(2), 307–331.
- Stern, N. (2013). The structure of economic modeling of the potential impacts of climate change: Grafting gross underestimation of risk onto already narrow science models. *J. Econ. Lit.*, 51(3), 838–59.
- Thompson, H. (2006). The applied theory of energy substitution in production. *Energy Econ.*, 28(4), 410–425.
- Tol, R. S. (2009). The economic effects of climate change. *J. Econ. Perspect.*, 23(2), 29–51.
- Van der Ploeg, F., & Rezai, A. (2020). Stranded assets in the transition to a carbon-free economy. *Annu. Rev. Resour. Econ.*, 12, 281–298.
- Van der Werf, E. (2008). Production functions for climate policy modeling: An empirical analysis. *Energy Econ.*, 30(6), 2964–2979.
- Vandenbussche, J., Aghion, P., & Meghir, C. (2006). Growth, distance to frontier and composition of human capital. *J. Econ. Growth*, 11(2), 97–127.
- Warr, B., & Ayres, R. (2006). REXS: A forecasting model for assessing the impact of natural resource consumption and technological change on economic growth. *Struct. Change Econ. Dyn.*, 17(3), 329–378.
- Weitzman, M. L. (2010). What is the “damages function” for global warming—and what difference might it make? *Clim. Change*, 1(01), 57–69.
- Weitzman, M. L. (2012). Ghg targets as insurance against catastrophic climate damages. *J. Public Econ. Theory*, 14(2), 221–244.

CHAPTER 4. MACROECONOMIC SCENARIOS FOR CLIMATE RISK  
ASSESSMENT

---

# General conclusion and perspectives

## Conclusion

In this manuscript we propose three main contributions to modeling the impact of climate risk on credit risk and bank solvency. First, we show that the inclusion of macroeconomic covariates is possible in the Merton-Vašíček framework by modeling the conditional distribution of the common factor. We therefore propose two types of models. The first one is based on a location-scale semiparametric single-index approach. The finite sample performance of the different variants of the location-scale model is studied using numerical experiments, and some theoretical asymptotic properties are provided. This modeling approach is flexible thanks to the semiparametric estimation of the mean and scale functions. Also, the curse of dimensionality is prevented by the single-index assumptions.

The semiparametric location-scale approach involves kernel smoothing for the mean and scale functions estimation. The nonparametric methods are known to have unreliable performance for out-of-sample values. This can be particularly problematic for stress-testing, where the extreme values of the covariates are also of interest. Therefore, we propose a second modeling approach based on a two-step procedure where, in the first step linear quantile regressions are performed, and in the second step the distance between a quantile function belonging to a flexible parametric family and the estimated linear quantiles is minimized. Using a parametric family paradigm makes this second modeling approach easier to interpret and monitor for both practitioner in a banking institution and the banking supervisor. The first location-scale single-index model is more flexible and may perform better in the case of highly nonlinear relationships. However, if the quantile regressions are either close to linear or the observed sample is small, the second two-step modeling approach may yield more meaningful results. As in other data modeling situations where the practitioners have to choose between the guarantee of a model-free approach that requires a large number of observations and the simplicity and interpretability of a restricted model, the former can be used as a benchmark, and if both approaches lead to similar results, the latter may ultimately be preferred in practice. Indeed, the internal models used by the banks are heavily monitored by the regulator and the at-Risk approach in Chapter 3 may be easier to interpret and validate in this context.

The simulation studies show that the location-scale single-index models produce very

different results from the regulatory model, sometimes by several orders of magnitude. Moreover, on a real scenario used internally by a French bank, the two conditional quantile methods proposed in Chapters 2 and 3 lead to similar results. For very extreme scenarios, the out-of-sample results of the location-scale single-index regressions differ from those of the at-Risk methodology. This discrepancy could be due to the nonparametric smoothers involved in the estimation of the mean and scale functions, which are expected to behave poorly on areas of low design. However, it should be noted that there are no real benchmarks for this case study and the estimated conditional quantiles obtained by both methods cannot yet be back-tested.

In this manuscript we are interested in incorporating climate risks into credit risk and bank solvency. Most methodologies with a similar objective, discussed by Cartellier (2022), consist of two parts. First, a climate model incorporates climate risks into macroeconomic variables and then some projections of these covariates are calculated. Based on these projections and a credit risk model, practitioners then calculate credit risk, which therefore includes climate risks through these macroeconomic variables. However, we argue that even if the credit risk model were perfect, there would still be some uncertainty due to the climate model. One of the most important factors used by internal credit risk models is GDP growth. We have found that the scenario used by the NGFS traditionally includes an exogenous path. Therefore, most of the projections of this growth that are then used in credit risk models have a growth trajectory that is similar to the historical one or, at worst, a trajectory based on the best guess. We argue that such assumptions are unsatisfactory and could lead to a dramatic underestimation of the financial loss associated with climate risks. We therefore propose to model the growth path of the Total Factor Productivity (TFP) with a hierarchical Bayesian model. The *exergy* was found to be related to this quantity; if the *exergy* increases, then the TFP tends to increase. We then propose projections of the *exergy* based on possible transition paths. It was found that the faster the transition, the lower the GDP growth but the lower the physical risk. Moreover, it was found that the variation of the global *exergy* was more important than the variation of each country's *exergy* in explaining the variation of the TFP for each country. We interpret this result as a consequence of the interconnected nature of the global market, where goods produced in one country can be consumed elsewhere in the world.

## Perspectives

The modeling ideas proposed in chapters 2 and 3 have been applied to real credit data in a French bank. The models implemented were of great interest to the institution, but some further adjustments can be considered.

In Chapter 2, we haven't discussed any variable selection method. This was of primary concern for practitioners as they have to justify the inclusion of each of one internally as well as to the regulator. Given that the asymptotic distributions of the single-index parameter can be derived, one can use this asymptotic distribution for testing coefficients

and thus select covariates. Other method exists, for instance Kong and Xia (2007) adapted cross-validation method for the specific case of the single-index model, the author called this method the separated cross-validation method. In the case of a parametric model leave- $m$ -out, the cross-validation can be consistent under mild assumptions. In the case of nonparametric model, the leave-one-out crossvalidation is consistent and often used in practice. However, in the case of the single-index model, if  $n$  is the number of observations and  $m$  the number of observations not used for estimating the single-index parameter, then Kong and Xia (2007) showed that the leave- $m$ -out cross-validation is consistent if  $m/n \rightarrow c \in [2/3, 1]$ . The main idea of the separated cross-validation method is to estimate both the single-index parameter and the regression function on a subset of the observed data and select the model that minimizes a cross-validation criterion. However it requires to estimate all quantities on each subset of size  $n - m$  which could be very computationally intensive. Furthermore, Peng and Huang (2011) discussed a penalized least squares approach when estimating the single-index parameter, and this could be a faster solution for variable selection. However none of these references considered a location-scale model as we considered above, which is expected to be more challenging. For example, the choice of a certain subset of covariates for estimating the mean function will certainly modify the estimates obtained for the scale function. Therefore, the adaptation of the cross-validation methods mentioned above may not be straightforward and remains to be investigated.

Our simulation studies reveal an asymptotically Gaussian behavior of the conditional quantiles we proposed. However, we did not derive confidence intervals or bands around these conditional quantile predictions. The confidence sets could be important for practitioners, for instance for justifying so called overlays or Margin of Conservatism (MoC). The overlays are often used to manually modify statistical models output to introduce novel risks that are not easily quantified by traditional models. (See for example the BCE's calculations during the 2020-2022 period<sup>13</sup>.) Therefore deriving confidence band around the predicted quantiles could be used retrospectively to justify the appropriateness and magnitude of the overlays applied, providing a quantitative justification. Additionally, MoC<sup>14</sup> is a concept used in the Internal Ratings-Based Approach of the Basel III accord to adjust for shortcomings in quantitative risk estimates. It aims to address issues with data quality and model risk. Institutions are required to identify and rectify deficiencies in their data and models, and apply MoC to account for any remaining uncertainties. By deriving confidence bands around the predicted quantiles, practitioners can better quantify the uncertainties and justify the MoC applied for model risk. This ensures that the risk parameters (PD in this case) are adequately conservative, helping to mitigate model risk and maintain the robustness of the risk management framework.

---

<sup>13</sup>This discussion is available at: <https://www.bankingsupervision.europa.eu/press/blog/2023/html/ssm.blog230526~29af0452d6.en.html>

<sup>14</sup>See a report from the EBA on this matter : <https://www.eba.europa.eu/sites/default/files/documents/10180/1720738/0212ecde-426d-4e18-84f8-04b036dcce00/EBA%20Report%20on%20IRB%20modelling%20practices.pdf>

The previous comments also apply to the at-Risk approach. In fact, the variable selection methods have never been discussed. Moreover, in contrast to the single index-location scale models, the asymptotic distribution of the parameters has not been investigated at all. In our opinion, while the asymptotics for the at-Risk approach seem feasible, the variable selection problem is not obvious in the case of the at-Risk approach, because for some quantiles some covariates may be effective while for other quantiles they may not be. These issues merit further investigation.

The model in Chapter 3 proposes a linear relationship between the covariates and the conditional quantile of interest. In the spirit of the semiparametric approach studied in Chapter 2, we can consider a single-index quantile regression instead. See, e.g., Wu et al. (2010). More precisely, in Chapter 3 we considered

$$Q_{Y|X}(\tau) = X^\top \beta_\tau, \quad \beta_\tau \in \mathbb{R}^d,$$

where  $Q_{Y|X}(\tau)$  is the  $\tau$ -th conditional quantile of  $Y$  given  $X$ . More generally, one can consider

$$Q_{Y|X}(\tau) = g(X^\top \beta_\tau),$$

where  $g(\cdot)$  is some unknown function to be estimated nonparametrically. When  $g(\cdot)$  is given, the parameter  $\beta_\tau$  can be estimated by minimizing a sample version of the expected risk given by the loss function  $\rho_\tau(u) = |u| + 2(\tau - 1)u$ ,  $u \in \mathbb{R}$ , usually called the *check function*. Indeed, it holds that when  $g(\cdot)$  is given, the true value of the parameter  $\beta_\tau$  is solution of the problem

$$\min_{\beta} \mathbb{E} [\rho_\tau(Y - g(X^\top \beta_\tau))].$$

When  $g(\cdot)$  is unknown, the natural idea is then to proceed like in Chapter 2, that is replace the unknown function by a nonparametric kernel estimator defined again using an expected loss of the check function when  $\beta_\tau$  is fixed. See Wu et al. (2010) for the details. This remains to be considered in future studies. However, even without searching more flexible conditional quantiles to be matched in the framework of the at-Risk approach, significant unfolded theoretical aspects persist regarding that type of approach. These methods gained considerable popularity in academic literature and among practitioners following the publication of Adrian et al. (2019). However, the definition of the key parameters, those of the conditional distribution obtained after the second step, remains unclear. Consequently, to the best of my knowledge, no theoretical guarantees have been established in the literature concerning these parameters. Additionally, the error introduced by the second step has yet to be quantified, for instance by more intensive numerical studies.

In the article presented in Chapter 4, I only derived the GDP growth rate based on some transition scenarios. However, banking institutions do not compute credit risk models solely based on the GDP growth rate, whether for long-term or short-term scenarios. Therefore one research perspective is to also model other macroeconomic variable : Inflation, Unemployment rate, .... Moreover, the central bank recently mentioned an effort

## CHAPTER 4. MACROECONOMIC SCENARIOS FOR CLIMATE RISK ASSESSMENT

---

to develop short-term scenarios for financial institutions Allen et al. (2023) and NGFS (2023). Most scenarios proposed by the NGFS calculate outputs yearly up to 2100. However, for certain risks, financial institutions may require shorter-term scenarios. For instance, in liquidity risk management, banks need to compute metrics such as the NSFR and LCR, the latter being the ratio between high liquid assets held by the bank and total liquid outflows over 30 days. Other risks, such as market risks, have even shorter time horizons. Therefore, if regulators introduce climate risk into areas like liquidity risk, they must derive scenarios with monthly frequency. Academic work is ongoing in this area. I believe that integrating short-term variations of *exergy* to model the short-term variations of GDP growth, inflation, or unemployment rates using econometric models, such as those developed in 2 and Chapter 3, could be of great interest. Finally, practitioners often model the probability of default based on sectoral value added, some of these covariates was in fact used in Section 3.4. The macroeconomic variables modeled in Chapter 4 represents an aggregated GDP. However, an indicator with a higher granularity (for instance, the sectorials values added) would be of great interest for practitioners. Therefore, adapting input-output model (see for instance Adenot et al. (2022)), or DSGE model could be highly beneficial in this context.

## References

- Adenot, T., Briere, M., Counathe, P., Jouanneau, M., Le Berthe, T., & Le Guenedal, T. (2022). Cascading effects of carbon price through the value chain: Impact on firm's valuation [Available at : [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4043923](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4043923), or : <https://dx.doi.org/10.2139/ssrn.4043923>].
- Adrian, T., Boyarchenko, N., & Giannone, D. (2019). Vulnerable growth. *Am. Econ. Rev.*, *109*(4), 1263–1289.
- Allen, T., Boullot, M., Dées, S., de Gaye, A., Lisack, N., Thubin, C., & Wegner, O. (2023, September). Using short-term scenarios to assess the macroeconomic impacts of climate transition [Available at : [https://publications.banque-france.fr/sites/default/files/medias/documents/wp922\\_0.pdf](https://publications.banque-france.fr/sites/default/files/medias/documents/wp922_0.pdf)].
- Cartellier, F. (2022). Climate stress testing, an answer to the challenge of assessing climate-related risks to the financial system? [Available at SSRN: <https://ssrn.com/abstract=4179311> or <http://dx.doi.org/10.2139/ssrn.4179311>].
- Kong, E., & Xia, Y. (2007). Variable selection for the single-index model. *Biometrika*, *94*(1), 217–229.
- NGFS. (2023, October). Conceptual note on short-term climate scenarios [Available at : <https://www.ngfs.net/sites/default/files/medias/documents/conceptual-note-on-short-term-climate-scenarios.pdf>].
- Peng, H., & Huang, T. (2011). Penalized least squares for single index models. *J. Statist. Plann. Inference*, *141*(4), 1362–1379.
- Wu, T. Z., Yu, K., & Yu, Y. (2010). Single-index quantile regression. *J. Multivariate Anal.*, *101*(7), 1607–1621.

**Titre:** Modélisation de l'impact du risque climatique sur le crédit et la solvabilité bancaire

**Mots clés :** Climate stress-testing, estimation non paramétrique, fonction quantile, modèle dit location-scale, modèle semi-paramétrique à direction révélatrice, risque de crédit.

**Résumé:** Cette thèse propose une adaptation du modèle de Merton-Vašíček. Ce modèle à facteur commun permet de calculer les pertes extrêmes liées à un portefeuille de crédits. Usuellement, ce facteur est considéré comme étant Gaussien et le modèle ne permet pas l'intégration de variables macroéconomiques. Dans ce manuscrit nous proposons de définir les pertes extrêmes à partir des quantiles du facteur commun conditionnels à des variables macroéconomiques. Nous proposons deux types de modèles semi-paramétriques pour estimer ces quantités. Nous montrons leur pertinence aussi bien sur des exercices numériques que sur des données réelles.

Cependant, intégrer le risque climatique dans le calcul de pertes extrêmes nécessite l'introduction de scénarios macroéconomiques qui serviront ensuite de variables explicatives dans ce modèle de Merton-Vašíček. Nous proposons donc d'intégrer des contraintes environnementales dans un IAM, le modèle DICE. En particulier, nous proposons d'intégrer l'*exergy* dans la modélisation de la productivité totale des facteurs avant d'intégrer cette modélisation pour générer des trajectoires de (dé-)croissance économique, qui pourrait ensuite être intégrée comme variable explicative aux modèles précédemment présentés.

**Title:** Modeling the impact of climate risk on credit and bank solvency

**Keywords :** Climate stress-testing, credit risk, location-scale regression, nonparametric estimation, quantile function, semi-parametric single index.

**Abstract:** This manuscript proposes an adaptation of the Merton-Vašíček model. This single-factor model allows for the calculation of financial losses related to credit. Typically, this factor is considered Gaussian, and the model does not allow for the integration of macroeconomic variables. In this manuscript, we propose to compute this loss based on the quantile of the common factor conditional on macroeconomic variables. We propose two semiparametric types of models to estimate these quantities. We demonstrate their relevance through both numerical exercises and real data.

However, integrating climate risk into the calculation of extreme losses requires the introduction of macroeconomic scenarios that will then serve as explanatory variables in this Merton-Vašíček model. We have therefore proposed to integrate environmental constraints into an IAM, the DICE model. Specifically, we propose to integrate *exergy* into the modeling of the Total Factor Productivity (TFP). Next, it is possible to incorporate the predictions of the TFP to generate trajectories of (de-)growth that could eventually be integrated as explanatory variables into the conditional quantile models.

