



HAL
open science

Infinite ARCH processes, dynamic betas, and financial applications

Julien Royer

► **To cite this version:**

Julien Royer. Infinite ARCH processes, dynamic betas, and financial applications. Statistics [math.ST]. Institut Polytechnique de Paris, 2022. English. NNT : 2022IPPAG012 . tel-03910879

HAL Id: tel-03910879

<https://theses.hal.science/tel-03910879v1>

Submitted on 22 Dec 2022

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.



INSTITUT
POLYTECHNIQUE
DE PARIS

NNT : 2022IPPAG012

Thèse de doctorat



Processus ARCH d'ordre infini, Bêtas dynamiques et applications financières

Thèse de doctorat de l'Institut Polytechnique de Paris
préparée à l'École nationale de la statistique et de l'administration économique

École doctorale n°574 École doctorale de mathématiques Hadamard (EDMH)
Spécialité de doctorat : Mathématiques appliquées

Thèse présentée et soutenue à Palaiseau, le 9 décembre 2022, par

JULIEN ROYER

Composition du Jury :

François Roueff Professeur, Télécom Paris et Institut Polytechnique de Paris	Président
Alessandra Luati Professeur, Imperial College London et Università di Bologna	Rapporteur
Olivier Scaillet Professeur, Université de Genève	Rapporteur
Liudas Giraitis Professeur, Queen Mary University of London	Examineur
Christian Francq Professeur, Université de Lille et ENSAE-CREST	Directeur de thèse
Jean-Michel Zakoïan Professeur, Université de Lille et ENSAE-CREST	Directeur de thèse
Mabrouk Chetouane Responsable de la stratégie de marché, Natixis IM	Invité

Remerciements

Le manuscrit qui s'ouvre sur ces quelques mots vient présenter les étapes de la construction du jeune chercheur que je suis. Si cette thèse souligne alors une construction individuelle, elle n'aura en rien été un processus solitaire et je tiens ici à remercier les nombreuses personnes qui ont contribué à ce travail initiatique.

Je souhaite tout d'abord remercier Christian Francq et Jean-Michel Zakoïan. Leur bienveillance, leur disponibilité, et leur rigueur m'auront permis de me lancer dans des projets que je ne me croyais pas capable de mener à bien. Leur passion pour la recherche est des plus contagieuse et je leur suis particulièrement reconnaissant de m'avoir guidé, tout en me laissant explorer mes propres idées. Avoir un tel duo de directeurs de thèse est un privilège et je garderai un souvenir heureux de nos discussions, qu'elles touchent à l'économétrie ou à l'art moderne (voir brut), ainsi que de nos balades dans les calanques, la campagne provençale ou sous l'infinité des arches de la Via di San Luca¹. J'associe à ces remerciements Christiane Francq pour une relecture bienvenue et des discussions passionnantes.

I am extremely grateful to Professors Alessandra Luati and Olivier Scaillet for reviewing this manuscript and for their insightful comments. Additionally, I warmly thank Professors Liudas Giraitis and François Roueff for accepting to take part in my jury.

Je suis extrêmement reconnaissant envers Mabrouk Chetouane, sans qui cette thèse n'aurait jamais pu voir le jour. Je le remercie pour sa confiance, sa curiosité scientifique, et le temps qu'il m'aura accordé pour mener à bien mes recherches. Mabrouk m'a permis de surmonter les écueils qui jonchent le parcours d'un doctorant CIFRE et je garderai de notre collaboration des souvenirs hauts en couleurs (au sens propre comme au sens figuré). Nos discussions sur la finance, le vin, ou la « sapologie » ne font, elles, que commencer.

Je tiens, par ailleurs, à remercier les membres du laboratoire de finance du CREST pour m'avoir accueilli, soutenu et encouragé au cours de ces trois années et quelques mois. En particulier, je souhaite remercier Christian Gourieroux, Christian Robert et Francesco Violante pour l'intérêt qu'ils auront porté à mes recherches, ainsi que Jean-David Fermanian, Caroline Hillairet et Peter Tankov pour nos échanges. Je remercie chaleureusement Fanda Traoré, sans qui ce département ne tournerait certainement pas rond (ou carré si on se rapporte à l'architecture de l'ENSAE). Je remercie enfin les doctorants, post-doctorants ou étudiants en visite qui auront partagé une partie de mon aventure : Alicia, Christopher, David, Fallou, Fanny, François, Frederik, Hugo, Ishak, Jalel, Laura, Marcos, Maria, Matar, Ophélie, Théo, Thomas et Yousra. Plus personnellement, je salue mes co-auteurs David et Thomas. Que serait un réveillon de la Saint Sylvestre sans *call for paper* ? Enfin, j'associe à ces remerciements Sébastien Laurent pour ses conseils, son intérêt pour ma recherche, ainsi qu'un séjour très agréable à Aix.

¹Tout le crédit de ce jeu de mot génial sur les ARCH(∞) revient à Christian.

En parallèle de ma vie étudiante, j'ai eu la chance de travailler au sein du groupe Amundi. Je tiens ainsi à remercier Gilles Guez et Laurent Gonon d'avoir accepté de m'accueillir en CIFRE au sein de BFT IM. Je remercie tout particulièrement les membres de l'équipe de recherche : Fabien, Romain et Béline (avec un B), sous l'égide de Mabrouk dont les envolées lyriques nous auront tout autant amusé que galvanisé. J'ai adoré notre temps ensemble, à discuter de processus, de convergence, de philosophie bayésienne (OU PAS), et de bien d'autres choses encore. Je salue aussi chaleureusement les amis de la Belle Époque : Julien, Félicien, Lucas, Jean-Marc, Elodie, Sophie, mais aussi tous les collaborateurs de BFT IM pour une expérience importante dans ma vie professionnelle. Je joins à ces remerciements Hassan, Killian, Marcel, Cindy et Zakaria pour leur amitié. Enfin, je remercie particulièrement Franck et son cluster Matlab sans qui mes nombreuses simulations n'auraient pu aboutir.

Mes balbutiements de chercheurs et ma curiosité scientifique doivent beaucoup à ma formation dauphinoise, et je souhaite remercier chaleureusement Philippe "PhB" Bernard pour avoir su inculquer à ses étudiants, promotion après promotion, que la finance quantitative se nourrit de la recherche académique. Je remercie aussi amicalement ses principaux apôtres, qui m'auront encouragé à rejoindre l'ENSAE et poursuivre mes études: Mabrouk Chetouane (encore lui), Florian Ielpo, Matthieu Garcin, Chafic Merhy, et Guillaume Simon.

Je n'aurais sûrement jamais commencé (et encore moins terminé) une thèse sans le soutien indéfectible de ma famille. En premier lieu, mes parents qui m'ont appris la persévérance et, surtout, m'ont donné la chance d'être curieux. Je les remercie pour la confiance qu'ils ont toujours apporté à ma sœur et moi, nous encourageant à aller au bout de nos projets. Les associés à la conclusion de mon doctorat restera une de mes plus grandes fiertés. Je remercie, bien sûr, ma sœur qui aura toujours été présente à mes côtés. Enfin, je remercie ma tante et ma grand-mère. Faire une thèse sur les processus à mémoire longue apparaît comme une cruelle ironie de la vie. Je remercie aussi tous les amis qui auront su me sortir de mes équations et faire de ces dernières années une période riche et heureuse: Alex C, Alex J, ~~Fatima~~Chahrazade, Darius, Edouard, Emilie, Estelle, Florent, Florian, Jérémy, Julien, Khalil, Laure, Léo, Louis, Luc, Mathilde, Nada, Romain, Zak. Les derniers mots de ce déjà bien trop long prologue vont affectueusement à Celia, ma première relectrice et mon soutien le plus attentionné.

Contents

1	Introduction	11
1.1	Résumé des travaux de thèse en français	11
1.1.1	Contexte	11
1.1.2	Motivations	15
1.1.3	Synthèse des principaux résultats	19
1.2	Summary of the thesis works	29
1.2.1	Context	29
1.2.2	Motivations	33
1.2.3	Main results	36
2	Conditional asymmetry in Power ARCH(∞) models	47
2.1	Introduction	48
2.2	Asymmetric Power ARCH(∞) model	49
2.3	Statistical inference of an APARCH(∞) process	52
2.4	Specification tests	56
2.4.1	Portmanteau goodness-of-fit test for APARCH(∞) models	57
2.4.2	Testing for linear constraints on the parameters	58
2.4.3	Testing for GARCH(1,1) specifications	59
2.5	Simulations	60
2.6	Application: Are GARCH(1,1)-type models suitable for peripheral markets?	62
2.7	Concluding remarks	64
2.8	Appendix A: Proofs and technical results	67
2.8.1	Existence of a stationary APARCH(∞) solution	67
2.8.2	Statistical inference of an APARCH(∞) process	68
2.8.3	Specification tests	86
2.9	Appendix B: GARCH(1,1)-type test when the parameter d is considered fixed	91
2.9.1	Asymptotics for the Wald statistic	91
2.9.2	Finite sample properties of the statistic	92
2.10	Appendix C: Additional Monte Carlo experiments	94
2.11	Appendix D: Additional application – Are GARCH(1,1)-type specifications suitable to model peripheral stocks?	95

3	A multivariate ARCH(∞) model with exogenous variables and dynamic conditional betas	97
3.1	Introduction	98
3.2	ARCH-X(∞) extension of the CHAR model	99
3.2.1	Stationarity conditions	100
3.2.2	Particular specifications	101
3.3	Statistical inference	104
3.4	Simulations	107
3.5	Application: Do semibetas matter when modeling conditional betas?	108
3.6	Conclusion	110
3.7	Appendix A: Proofs and technical results	112
3.8	Appendix B: Finite sample properties of the Wald statistics when testing for linear constraints	116
4	Empirical Asset Pricing with Score-Driven Conditional Betas	117
4.1	Introduction	118
4.2	Testing in the Score-Driven Betas framework	119
4.2.1	A Score-Driven Conditional Regression model	119
4.2.2	Asymptotics of standard test statistics	122
4.2.3	Residual Bootstrap approximation of the test statistics distributions	124
4.3	Conditional Risk Premia Estimation Procedure	126
4.4	Monte Carlo experiments	127
4.5	Application: Assessing a dynamic carbon-related risk premium	129
4.5.1	Construction of a carbon-related risk factor	129
4.5.2	Carbon factor significance testing	131
4.5.3	Risk premium estimation	132
4.6	Conclusion	133
4.7	Appendix A: Proofs and technical results	135
4.7.1	Proof of Theorem 4.1	135
4.7.2	Proof of Theorem 4.2	137
4.8	Appendix B: Details on empirical results	141
5	Conclusion and Perspectives	145
5.1	Conclusion	145
5.2	Perspectives	146

List of Figures

1.1	Daily returns of the CAC 40 Index from January 1995 to May 2020	30
1.2	Autocorrelations of the CAC 40 Index daily returns	31
1.3	Illustration of the conditional asymmetry on the CAC 40 Index daily returns . .	34
2.1	Effect of a shock on the conditional variance of a GARCH(1,1) process and an ARCH(∞) process	51
2.2	Empirical behavior of the symmetry test statistics	62
2.3	Kernel density estimators of the GARCH-type and goodness-of-fit test statistics .	63
2.4	Empirical behavior of the GARCH(1,1)-type test statistic when d is fixed	93
2.5	GJR-GARCH(1,1)-type test for portfolios formed on Size	95
3.1	Conditional betas between the Market portfolio and two industry portfolios (Agriculture and Paper)	110
3.2	Wald statistic empirical and theoretical asymptotic distributions	116
4.1	Empirical results of the two-step risk premia estimation procedure on Monte Carlo experiments	128
4.2	Performance of Long-Only portfolios constructed on scope 2 carbon intensity . .	131
4.3	Green Minus Brown conditional risk premium	133
4.4	Conditional risk premia of the Fama French and Green Minus Brown factors . .	143

List of Tables

2.1	Estimation results for simulations of an APARCH(∞) process	61
2.2	Tests and validation of VaR forecasts on different equity indices	65
2.3	Observed power of the GARCH(1,1)-type test when the model is misspecified . .	94
2.4	Estimation results for simulations of a symmetric ARCH(∞) process and a TARCH(∞) process	94
3.1	Estimation results for simulations of a Cholesky ARCH(∞) with semibetas . . .	108
3.2	Estimation results for simulations of a Cholesky GARCH with realized semibetas	108
3.3	p-values of the Wald test on the importance of the semibetas for each industry portfolios	111
3.4	Persistence parameters for each industry portfolios	111
4.1	Finite sample properties of the test statistics for factor significance	127
4.2	Proportion of industry portfolios with a significant beta to the Green Minus Brown factor	131
4.3	Annualized average risk premia of the Fama-French and Green Minus Brown factors	132
4.4	Average scope 2 carbon intensity by sector	141
4.5	Financial ratios summary statistics by scope 2 carbon intensity deciles	141
4.6	Proportion of missing values by sector for scope 2 carbon intensity	141
4.7	Comparison of the dynamic conditional betas with a 2-year rolling window estimators	142
4.8	Potential explanatory variables of the conditional carbon-related risk premium .	144

Chapter 1

Introduction

1.1 Résumé des travaux de thèse en français

L'objectif de cette thèse est d'étudier les modèles de volatilité conditionnelle. Plus précisément, cette thèse se concentrera sur deux domaines de recherche de la littérature en économétrie financière. Premièrement, la modélisation de séries temporelles présentant des volatilités fortement persistantes. Dans cette optique, de nouvelles extensions des modèles ARCH d'ordre infini, à la fois dans le cadre univarié et multivarié, seront introduites et les aspects probabilistes et statistiques de ces processus seront discutés. Deuxièmement, la modélisation des coefficients de régression dynamiques dans les modèles factoriels linéaires sera abordée. Bien que ces sujets de recherche semblent éloignés, nous verrons au long de ce manuscrit qu'ils sont tous deux liés à la modélisation de la (co-)variance conditionnelle. Les spécifications introduites dans les chapitres suivants seront illustrées par des applications sur des données financières réelles, fournissant des arguments en faveur des extensions proposées.

1.1.1 Contexte

L'étude des phénomènes dynamiques est au cœur de la statistique et de la théorie des probabilités depuis le siècle dernier. Dans son cadre le plus usuel, un processus stochastique à temps discret $(X_t)_{t \in \mathbb{Z}}$ est supposé suivre une dynamique de la forme

$$X_t = \varphi(X_{t-1}, X_{t-2}, \dots; \eta_t)$$

où φ décrit la dépendance temporelle de la variable aléatoire X_t et $(\eta_t)_{t \in \mathbb{Z}}$ est une suite de variables aléatoires indépendantes et identiquement distribuées (iid). L'analyse des séries temporelles s'est d'abord concentrée sur les fonctions φ linéaires, suite aux travaux précurseurs de Yule (1927), il y a près d'un siècle. Dans ce cas, X_t est appelé un processus autorégressif (AR). Les propriétés théoriques des processus AR ont été la source d'une riche littérature (voir par exemple le manuel de référence de Brockwell et Davis (2009)). Ces modèles de séries chronologiques linéaires ont été appliqués à de nombreux domaines, allant des variables économiques aux variables climatiques. Cependant, leur dynamique est peu propice à la modélisation des séries

temporelles financières.

Bien que faisant partie des données les plus étudiées, les séries temporelles financières sont complexes à modéliser. Cette complexité est due à l'existence de propriétés statistiques similaires, observées sur une majorité d'actifs indépendamment de leur type (obligations d'entreprises, actions, matières premières...). Le premier de ces *faits stylisés* est l'absence de dépendance temporelle linéaire des rendements financiers. Cette propriété exclut naturellement l'utilisation de spécifications AR pour modéliser les variations de prix ou les rendements et a été à l'origine de l'*Hypothèse de marche aléatoire* des séries de prix et de la définition de l'*efficience de marché* de Fama (1970; 1991). Cependant, bien qu'ils ne soient pas autocorrélés, les rendements financiers ne sont pas indépendants car les transformations non linéaires des séries de rendements présentent une dépendance temporelle. Cette observation a conduit au rejet de l'*Hypothèse de marche aléatoire* (voir par exemple Lo et MacKinlay (1988)) et à l'émergence d'un deuxième fait stylisé : la présence de *clusters* de volatilité. Les rendements qui s'écartent fortement de leur moyenne apparaissent en *paquets* donnant lieu à une alternance de périodes de forte et de faible volatilité. Ce phénomène est récurrent mais non périodique et peut être directement observé sur la trajectoire des séries temporelles comme le montre la figure 1.1 présentant les rendements quotidiens de l'indice CAC 40 de janvier 1995 à mai 2020.

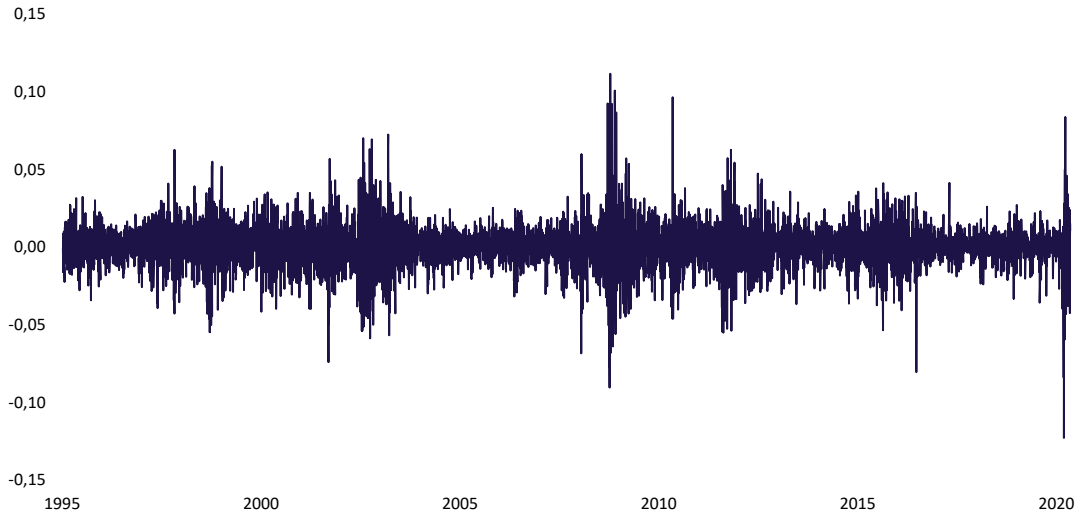


FIGURE 1.1 : Rendements quotidiens de l'indice CAC 40 de janvier 1995 à mai 2020

Ce fait empirique est difficilement compatible avec l'hypothèse de constance de la variance conditionnelle et a conduit à l'introduction de modèles conditionnellement hétéroscédastiques avec le modèle ARCH d'Engle 1982. Dans ce modèle non linéaire, la dynamique de X_t s'écrit sous la forme multiplicative

$$X_t = \sigma_t \eta_t \quad (1.1)$$

où σ_t est une fonction positive mesurable par rapport à la filtration $\mathcal{F}_{t-1} = \{X_u, u < t\}$. En particulier, dans le modèle ARCH(q) d'Engle, $0 < q < \infty$, la fonction σ_t est une fonction linéaire

des q premières valeurs passées de X_t^2 . Ce modèle a été étendu par Bollerslev 1986, donnant le célèbre modèle ARCH généralisé (GARCH), qui est sans doute devenu l'un des modèles les plus importants de la littérature en économétrie financière. Dans ce modèle, la variance conditionnelle de X_t est donnée par la dynamique suivante

$$\sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i X_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 \quad (1.2)$$

où ω est une constante strictement positive, $(\alpha_i)_{i=1,\dots,q}$ et $(\beta_j)_{j=1,\dots,p}$ sont des constantes positives, et (η_t) est une suite de variables aléatoires centrées, indépendantes et identiquement distribuées de variance unitaire.

Bien que les modèles GARCH soient bien adaptés à la modélisation des séries temporelles présentant à la fois une absence d'autocorrélation et des *clusters* de volatilité, la structure particulière de la variance conditionnelle impose une décroissance exponentielle de l'autocovariance des rendements au carré qui est en contradiction avec leur autocorrélogramme empirique. En effet, comme l'ont noté pour la première fois Ding et Granger (1996), le processus de volatilité présente empiriquement une forte persistance qui contredit l'utilisation des modèles GARCH classiques. La figure 1.2 présente l'autocorrélogramme des rendements, des rendements au carré et de la valeur absolue des rendements de l'indice CAC 40. On constate que même après une centaine de retards, les autocorrélations restent éloignées de 0 pour les séries de rendements transformées.

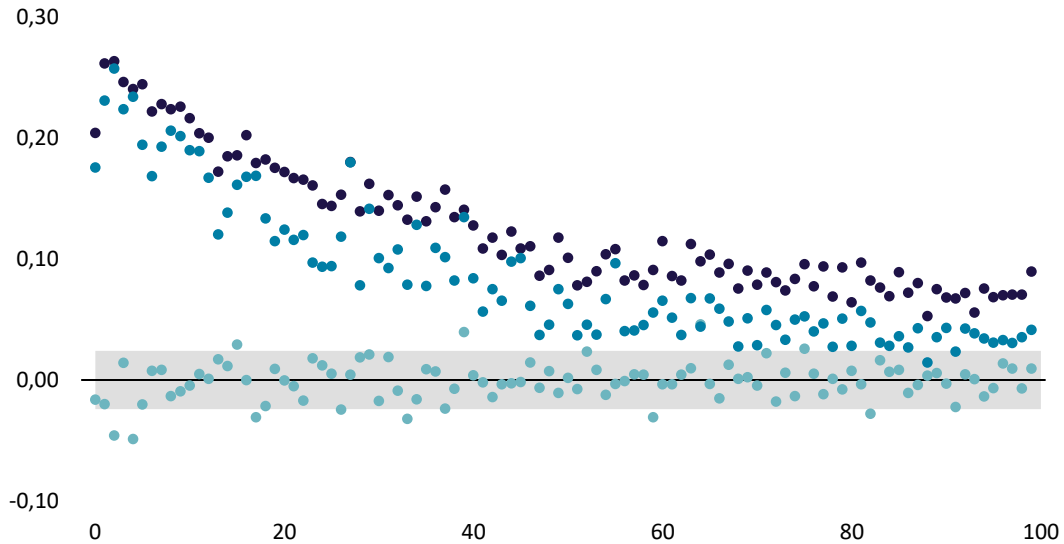


FIGURE 1.2 : Autocorrélations des rendements (bleu clair), rendements au carré (bleu) et valeur absolue des rendements (bleu foncé) de l'indice CAC 40

Afin de permettre une modélisation plus flexible de la décroissance des coefficients dans l'équation de la volatilité, Robinson (1991) a introduit le modèle autorégressif conditionnellement hétéroscédastique d'ordre infini (ARCH(∞)) comme une extension des modèles GARCH à *mé-*

moire courte. Sous cette spécification, la dynamique de la série temporelle est définie par

$$\begin{cases} X_t = \sigma_t \eta_t \\ \sigma_t^2 = \omega + \sum_{i=1}^{\infty} \alpha_i X_{t-i}^2 \end{cases} \quad (1.3)$$

avec $\omega > 0$, α_i , $i = 1, \dots$, une suite de constantes positives, et η_t une suite de variables aléatoires indépendantes et identiquement distribuées, centrées et de variance unitaire. En relâchant la décroissance exponentielle des autocorrélations des processus GARCH au carré, les ARCH(∞) conviennent aux séries à mémoire longue. De façon remarquable, en inversant le polynôme de retard dans l'équation (1.2), on peut montrer qu'un modèle GARCH(p, q) admet une représentation ARCH(∞). En particulier, l'équation de variance conditionnelle du GARCH(1,1) s'écrit comme suit :

$$\sigma_t^2 = \frac{\omega}{1 - \beta_1} + \sum_{i=1}^{\infty} \alpha_1 \beta_1^{i-1} X_{t-i}^2$$

ce qui met en évidence la décroissance exponentielle des coefficients imposée par la structure GARCH. Pour contourner cette décroissance exponentielle, une spécification ARCH(∞) populaire est le FIGARCH (fractionally integrated generalized autoregressive conditionally heteroskedastic) de Baillie et al. (1996). S'appuyant sur le modèle ARFIMA introduit par Granger et Joyeux (1980) et Hosking (1981)¹, le modèle FIGARCH(p, d, q) est défini comme suit

$$\begin{cases} X_t = \sigma_t \eta_t \\ \sigma_t^2 = \omega_0 + \left[1 - (1-L)^d \frac{\zeta(L)}{\psi(L)} \right] X_t^2 \end{cases} \quad (1.4)$$

où ψ et ζ sont respectivement des polynômes de degrés p et q , $d \in (0, 1)$ et $(1-L)^d$ représente l'opérateur de différence fractionnaire

$$(1-L)^d = \sum_{k=0}^{\infty} \pi_k(d) L^k$$

avec L l'opérateur retard usuel tel que $L^k X_t = X_{t-k}$ pour $k \in \mathbb{Z}$ et

$$\pi_k(d) = \frac{\Gamma(k-d)}{\Gamma(k+1)\Gamma(-d)}.$$

Pour illustrer la décroissance plus lente des coefficients dans le modèle FIGARCH, on peut considérer la spécification FIGARCH(0, d , 0) obtenue avec $\psi(L) = \zeta(L) = 1$. Le modèle s'écrit alors comme

$$\sigma_t^2 = \omega_0 - \sum_{k=1}^{\infty} \pi_k(d) X_{t-k}^2.$$

Ainsi, par identification des coefficients α_i dans l'équation (1.3), nous avons

$$\alpha_i = -\pi_i(d) = -\frac{\Gamma(i-d)}{\Gamma(i+1)\Gamma(-d)}.$$

¹Voir Giraitis et al. (2012) pour une revue des modèles linéaires à mémoire.

et la formule de Stirling permet d'obtenir $\alpha_i \sim i^{-d-1}$.

L'existence d'une solution strictement stationnaire et non anticipative au modèle ARCH(∞) a été prouvée par Giraitis et al. (2000), Kazakevičius et Leipus (2002), et Douc et al. (2008) sous la condition

$$A_s \mu_{2s} < 1 \quad (1.5)$$

pour un certain $s \in (0, 1]$, où $A_s = \sum_{i=1}^{\infty} \alpha_i^s$ et $\mu_{2s} = E|\eta_t|^{2s}$. Cette condition n'est cependant pas compatible avec le concept habituel de mémoire longue car elle implique la sommabilité des autocovariances de X_t^2 (voir Giraitis et al. (2000) et Zaffaroni (2004)). On peut noter que cette condition exclut également l'existence d'un modèle ARCH(∞) intégré stationnaire pour lequel $\sum_{i=1}^{\infty} \alpha_i = 1$. La question de l'existence d'une solution stationnaire avec un moment d'ordre quatre fini pour le modèle FIGARCH a fait l'objet d'une longue discussion académique qui a abouti au récent article de Giraitis et al. (2018) dans lequel l'existence d'une telle solution a été établie. Néanmoins, les modèles ARCH(∞) restent compatibles avec une décroissance lente de la fonction d'autocorrélation des carrés et sont parfois appelés modèles à *mémoire modérée*.

1.1.2 Motivations

Bien que les modèles à mémoire modérée soient des candidats naturels pour modéliser les séries chronologiques financières, leur utilisation par les praticiens a malheureusement été limitée. L'objectif de cette thèse est de renforcer l'utilisation des modèles de type ARCH(∞). A cet effet, nous proposerons plusieurs extensions du cadre ARCH(∞) usuel afin de concilier les extensions populaires de type GARCH et les modèles à mémoire modérée.

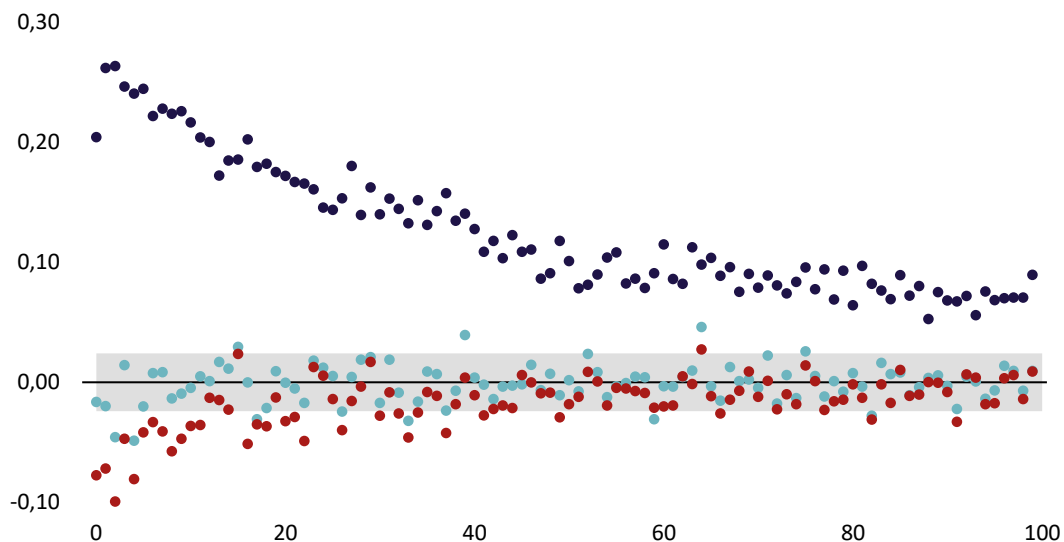


FIGURE 1.3 : Illustration de l'asymétrie conditionnelle sur les rendements de l'indice CAC 40

En particulier, le chapitre 2 considère une généralisation du modèle ARCH(∞) classique, per-

mettant la prise en compte de l'asymétrie conditionnelle, à savoir l'impact plus élevé sur la volatilité conditionnelle des observations négatives par rapport aux observations positives de même ampleur. La figure 1.3 illustre ce fait stylisé des séries financières². Outre l'absence d'autocorrélation des rendements (en bleu clair) et l'autocorrélation positive significative des valeurs absolues des rendements (en bleu foncé) déjà exposées dans la figure 1.2, elle présente la corrélation entre X_t^+ et X_{t-h} pour $h = 1, \dots, 100$ où X_t est le rendement quotidien de l'indice CAC 40 et $X_t^+ = \max(X_t, 0)$ (en rouge). En admettant que le processus X_t est stationnaire de second ordre et admet une dynamique comme l'équation (1.1), nous avons

$$\rho(X_t^+, X_{t-h}) = K \text{cov}(\sigma_t, X_{t-h}) = K[\text{cov}(\sigma_t, X_{t-h}^+) + \text{cov}(\sigma_t, X_{t-h}^-)]$$

où $\rho(Y, Z)$ désigne la corrélation entre les variables aléatoires Y et Z . Comme le montre la figure 1.3, pour plusieurs retards, la corrélation entre X_t^+ et X_{t-h} est clairement négative, ce qui signifie

$$\text{cov}(\sigma_t, X_{t-h}^+) < \text{cov}(\sigma_t, -X_{t-h}^-)$$

et souligne l'impact plus important des valeurs négatives passées de X_t sur la variance conditionnelle par rapport à l'impact des valeurs positives. Cette asymétrie a été observée pour la première fois par Black (1976) et Christie (1982) et est souvent appelée *effet de levier* car les auteurs suggèrent que ce phénomène est dû à la modification du ratio capitaux propres/dette de l'entreprise après une forte baisse du prix des actions, ce qui entraîne un risque plus élevé et donc une plus grande volatilité. Il est toutefois peu probable que ce seul mécanisme explique l'ampleur de l'asymétrie (voir par exemple Schwert (1989)). Une autre explication populaire est la "rétroaction de la volatilité", qui explique l'asymétrie par la variation dans le temps des primes de risque et des anticipations de la volatilité (par exemple, French et al. (1987), Campbell et Hentschel (1992) et Bekaert et Wu (2000)). Bien que de nombreuses extensions des modèles GARCH aient été introduites dans la littérature (les spécifications populaires incluent le modèle GJR-GARCH de Glosten et al. (1993) ou le modèle GARCH à seuil (*Threshold-GARCH*) de Zakoïan (1994)), à notre connaissance, les tentatives pour capturer à la fois l'asymétrie et les propriétés de mémoire des séries chronologiques financières ont été rares. Le chapitre 2 propose un nouveau modèle ARCH(∞) pour prendre en compte cet *effet de levier*.

Bien que les modèles de volatilités conditionnelles aient été initialement développées pour capturer l'hétéroscédasticité des séries temporelles univariées, de nombreuses extensions ont été proposées pour modéliser des systèmes multivariés. Ces modèles GARCH multivariés présentent un grand intérêt pour la gestion des risques ou la construction de portefeuilles, où la matrice de covariance conditionnelle est essentielle au processus de décision d'investissement. Dans ces modèles, un processus vectoriel à valeurs réelles \mathbf{X}_t de dimension m , $\mathbf{X}_t = (X_{1,t}, \dots, X_{m,t})'$ est défini comme suit

$$\mathbf{X}_t = \mathbf{H}_t^{1/2} \boldsymbol{\eta}_t$$

où $(\boldsymbol{\eta}_t)$ est une suite de processus iid centrés à valeurs dans \mathbb{R}^m ayant la matrice identité pour

²Cette illustration est inspirée du chapitre 10 de l'ouvrage de Francq et Zakoïan (2019).

covariance. Diverses spécifications de la matrice de covariance conditionnelle \mathbf{H}_t ont été proposées, la première étant le modèle diagonal de Engle et al. (1984). Ses extensions populaires comprennent notamment le GARCH à corrélation conditionnelle constante (CCC) de Bollerslev (1990) et l’extension de Jeantheau (1998), le modèle BEKK de Engle et Kroner (1995) ainsi que le modèle à corrélation conditionnelle dynamique (DCC) de Engle (2002) et Tse et Tsui (2002)³. Les conditions de stationnarité pour le modèle CCC et DCC ont été obtenues respectivement par Aue et al. (2009) et Fermanian et Malongo (2017) tandis que les résultats asymptotique pour l’estimation par quasi maximum de vraisemblance (QML) ont, entre autres, été dérivés par Jeantheau (1998), Ling et McAleer (2003), Comte et Lieberman (2003) et Francq et Zakoïan (2012).

Même si une importante littérature est consacrée à l’étude des modèles multivariés à mémoire courte, leurs homologues ARCH(∞) ont rarement été explorés. Les conditions de stationnarité d’un modèle ARCH(∞) multivarié ont néanmoins été discutées par Doukhan et al. (2006) tandis que Bardet et Wintemberger (2009) ont établi la convergence et la normalité asymptotique de l’estimateur QML (QMLE) pour une classe générale de processus causaux multidimensionnels. Cette dernière contribution ne vise cependant pas directement la modélisation de modèles à mémoire modérée multivariés et la condition de moments supérieurs pour le processus de génération des données peut être difficile à concilier avec les spécifications ARCH(∞). Le chapitre 3 introduit un nouveau modèle de volatilité multivarié dans lequel les covolatilités peuvent présenter une persistance plus élevée que l’équation GARCH multivariée usuelle. Les conditions d’existence d’une solution stationnaire unique sont établies ainsi que la convergence et la normalité asymptotique du QMLE. En particulier, nous nous appuyons sur le modèle Cholesky-GARCH de Darolles et al. (2018) qui permet de modéliser directement les bêtas conditionnels entre les actifs dans un système multivarié par une décomposition Cholesky de la matrice de covariance conditionnelle.

Par ailleurs, la modélisation d’une régression linéaire conditionnelle présente un grand intérêt en pratique. Considérons la régression conditionnelle de Y_t sur un ensemble de m variables exogènes $\mathbf{X}_t = (X_{1,t}, \dots, X_{m,t})'$

$$Y_t = \beta_t' \mathbf{X}_t + \eta_t = \sum_{k=1}^m \beta_{k,t} X_{k,t} + \eta_t$$

où β_t sont les coefficients de pente conditionnels et η_t sont des termes d’erreur qui sont orthogonaux à \mathbf{X}_t . Dans les modèles d’évaluation d’actifs, Y_t décrit généralement les rendements excédentaires d’un actif financier ou d’un portefeuille, et les variables exogènes \mathbf{X}_t sont des facteurs de risque systématiques ayant un impact sur la *cross-section* des actifs. Par exemple, le Capital Asset Pricing Model (CAPM) de Sharpe (1964) et Lintner (1965) utilise les rendements excédentaires du portefeuille de marché dans un modèle à facteur unique, tandis que Fama et French (1993) considèrent deux facteurs supplémentaires capturant les effets de taille et de valeur. Dans ces modèles factoriels, une hypothèse récurrente mais problématique est l’invariance temporelle des paramètres de pente (les fameux bêtas). Une telle hypothèse est difficile à jus-

³Pour une revue des modèles GARCH multivariés, voir Bauwens et al. (2006).

tifier empiriquement et est souvent rejetée dans les applications sur données réelles (voir par exemple Engle (2016)). À notre connaissance, Goldfeld et Quandt (1973) et Hamilton (1989) ont été les premiers à considérer des coefficients de régression dynamiques à travers des modèles à changement de régime. Cependant, dans ces modèles, les coefficients de pente sont considérés comme constants dans chaque état, ce qui peut être restrictif lorsque l'on considère des régimes de longue durée. Shanken (1990) et Ferson et Harvey (1991) ont proposé d'utiliser des variables instrumentales financières ou économiques pour récupérer des bêtas variant dans le temps tandis que plus récemment Gagliardini et al. (2016) ont établi des résultats théoriques sur l'estimation des primes de risque dynamiques dans ce cadre. Cette méthode nécessite toutefois la sélection de variables exogènes pour piloter l'ensemble de la dynamique des bêtas. En particulier, Ghysels (1998) a noté que les coefficients de pente obtenus peuvent être très sensibles au choix des variables instrumentales. S'appuyant sur les travaux de Bollerslev (1988), Engle (2016) a récemment proposé une approche différente pour estimer les bêtas conditionnels. En supposant que

$$\begin{pmatrix} Y_t \\ \mathbf{X}_t \end{pmatrix} \Big| \mathcal{F}_{t-1} \sim \mathcal{N} \left(\mathbf{0}, \mathbf{H}_t := \begin{bmatrix} H_{YY,t} & \mathbf{H}_{YX,t} \\ \mathbf{H}_{XY,t} & \mathbf{H}_{XX,t} \end{bmatrix} \right)$$

où \mathbf{H}_t est mesurable par rapport à la filtration \mathcal{F}_{t-1} , la distribution conditionnelle de Y sachant X est alors

$$Y_t | \mathbf{X}_t, \mathcal{F}_{t-1} \sim \mathcal{N}(\mathbf{H}_{YX,t} \mathbf{H}_{XX,t}^{-1} \mathbf{X}_t, H_{YY,t} - \mathbf{H}_{YX,t} \mathbf{H}_{XX,t}^{-1} \mathbf{H}_{XY,t})$$

et les bêtas dynamiques sont donnés par

$$\beta_t = \mathbf{H}_{XX,t}^{-1} \mathbf{H}_{XY,t}.$$

Il suffit donc de choisir une spécification de \mathbf{H}_t pour récupérer les coefficients de pente dynamiques. Bien que facile à mettre en œuvre, cette méthode souffre de nombreux inconvénients. Les paramètres dynamiques d'intérêt ne sont pas directement modélisés mais obtenus par une transformation de la matrice de covariance conditionnelle, ce qui rend difficile l'interprétation financière. Plus problématique encore, les bêtas obtenus s'avèrent souvent très volatils, comme le notent Grassi et Violante (2021). Pour contourner ces problèmes, Darolles et al. (2018) ont introduit le modèle Cholesky-GARCH permettant de modéliser directement les bêtas conditionnels dans un cadre GARCH multivarié. Le chapitre 3 présente une extension ARCH(∞) de ce modèle. Cependant, elle nécessite l'estimation du système multivarié complet, même si une seule équation présente un intérêt, ce qui peut s'avérer fastidieux dans un exercice d'évaluation d'actifs sur un grand panel. Pour pallier cette lacune du modèle Cholesky-GARCH, Grassi et Violante (2021) proposent un modèle Cholesky-GARCH par blocs, basé sur une décomposition Cholesky par blocs entre l'ensemble des facteurs et l'ensemble des actifs individuels. Bien que mieux adapté à un exercice empirique d'évaluation d'actifs, ce modèle impose la constance des primes de risque, ce qui peut être restrictif. Comme alternative, le chapitre 4 propose d'exploiter le modèle Autoregressive Conditional Beta (ACB) récemment introduit par Blasques et al. (2022a).

1.1.3 Synthèse des principaux résultats

Avant d'exposer les principaux résultats de la thèse, nous précisons la notation qui sera généralement utilisée dans la suite du manuscrit afin d'éviter les confusions avec les notations générales utilisées dans les premières sous-sections de l'introduction. Dorénavant, nous désignons par ε_t une série de rendements financiers centrés de processus de volatilité conditionnelle σ_t mesurable par rapport à la filtration $\mathcal{F}_{t-1} = \{\varepsilon_u, u < t\}$.

Le chapitre 2 se concentre sur les modèles de volatilité conditionnelle univariés. En particulier, il vise à étendre la littérature ARCH(∞) classique en permettant une réponse différente du processus de volatilité conditionnelle aux rendements négatifs ou positifs pour prendre en compte l'*effet de levier* mentionné précédemment. Le modèle proposé s'appuie sur la spécification APARCH(p, q) de Ding et al. (1993), où le processus de volatilité conditionnelle est le suivant

$$\sigma_t^\delta = \omega + \sum_{i=1}^q \alpha_i^+ |\varepsilon_{t-i}|^\delta \mathbf{1}_{\varepsilon_{t-i} \geq 0} + \alpha_i^- |\varepsilon_{t-i}|^\delta \mathbf{1}_{\varepsilon_{t-i} < 0} + \sum_{j=1}^p \beta_j \sigma_{t-j}^\delta$$

avec $\omega > 0$, où les coefficients sont des constantes positives et δ est une constante strictement positive. En particulier, on définit un processus *Asymmetric Power* ARCH(∞) comme suit

$$\begin{cases} \varepsilon_t = \sigma_t \eta_t \\ \sigma_t^\delta = \omega + \sum_{i=1}^{\infty} \alpha_i^+ |\varepsilon_{t-i}|^\delta \mathbf{1}_{\varepsilon_{t-i} \geq 0} + \alpha_i^- |\varepsilon_{t-i}|^\delta \mathbf{1}_{\varepsilon_{t-i} < 0} \end{cases}$$

où (η_t) est une suite de variables aléatoires iid centrées de variance unitaire, $\omega > 0$, $\delta > 0$, et où α_i^+ et α_i^- , $i = 1, \dots$, sont des suites de constantes positives. Il est intéressant de noter que les processus APARCH(∞) encapsulent de nombreux modèles à mémoire courte. Par exemple, le GARCH(1,1) symétrique usuel admet une représentation APARCH(∞) à partir de l'équation (1.3). Par des calculs analogues, on peut montrer que le Threshold-GARCH de Zakoïan (1994) ou le GJR-GARCH de Glosten et al. (1993) admettent la représentation APARCH(∞) suivante

$$\begin{cases} \varepsilon_t = \sigma_t \eta_t \\ \sigma_t^\delta = \frac{\omega}{1-\beta} + \sum_{i=1}^{\infty} \beta^{i-1} (\alpha^+ \mathbf{1}_{\varepsilon_{t-i} \geq 0} + \alpha^- \mathbf{1}_{\varepsilon_{t-i} < 0}) |\varepsilon_{t-i}|^\delta \end{cases}$$

où $\delta = 1$ et 2 respectivement. Comme pour le modèle GARCH, la décroissance exponentielle des suites de coefficients dans ces représentations implique une mémoire courte du processus de volatilité.

Le premier résultat théorique du chapitre 2 porte sur l'existence d'une solution stationnaire du modèle proposé. Pour tout $s > 0$, soit

$$A_s^+ = \sum_{i=1}^{\infty} (\alpha_i^+)^s, \quad A_s^- = \sum_{i=1}^{\infty} (\alpha_i^-)^s$$

et

$$\mu_{\delta s}^+ = \mathbb{E}|\mathbb{1}_{\eta_t \geq 0} \eta_t|^{\delta s}, \quad \mu_{\delta s}^- = \mathbb{E}|\mathbb{1}_{\eta_t < 0} \eta_t|^{\delta s}.$$

Le théorème 2.1 établit que s'il existe $s \in (0, 1]$ tel que $A_s^+ \mu_{\delta s}^+ + A_s^- \mu_{\delta s}^- < 1$, alors il existe une unique solution strictement stationnaire, ergodique et non anticipative de l'équation APARCH(∞) telle que $\mathbb{E}|\varepsilon_t|^{\delta s} < \infty$. La preuve est basée sur un développement de Volterra et, en ce sens, suit les travaux de Giraitis et al. (2000), Kazakevičius et Leipus (2002), et Douc et al. (2008). En particulier, la solution prend la forme de la série de Volterra suivante

$$\begin{cases} \varepsilon_t = \sigma_t \eta_t \\ \sigma_t^\delta = \omega + \omega \sum_{k=1}^{\infty} \sum_{i_1, \dots, i_k \geq 1} a_{i_1, t-i_1} \dots a_{i_k, t-i_1-\dots-i_k} |\eta_{t-i_1}|^\delta \dots |\eta_{t-i_1-\dots-i_k}|^\delta \end{cases}$$

avec $a_{i, t-j} = \alpha_i^+ \mathbb{1}_{\eta_{t-j} \geq 0} + \alpha_i^- \mathbb{1}_{\eta_{t-j} < 0}$. Il est intéressant de noter que la condition de stationnarité établie dans le théorème 2.1 se réduit à la condition de stationnarité (1.5) dans le cas ARCH(∞) où $\delta = 2$ et $A_s^+ = A_s^- = A_s$ puisque $\mu_{2s}^+ + \mu_{2s}^- = \mu_{2s}$.

Sur le plan statistique, ce chapitre traite de l'estimation QML du modèle. Les résultats asymptotiques du QMLE ont été établis pour la première fois dans le contexte des modèles ARCH(∞) par Robinson et Zaffaroni (2006). S'appuyant sur ces auteurs, le chapitre 2 introduit la forme paramétrique suivante

$$\begin{cases} \varepsilon_t = \sigma_t(\boldsymbol{\theta}_0) \eta_t \\ \sigma_t^\delta(\boldsymbol{\theta}_0) = \omega_0 + \sum_{i=1}^{\infty} \alpha_i^+(\boldsymbol{\phi}_0) |\varepsilon_{t-i}|^\delta \mathbb{1}_{\varepsilon_{t-i} \geq 0} + \alpha_i^-(\boldsymbol{\phi}_0) |\varepsilon_{t-i}|^\delta \mathbb{1}_{\varepsilon_{t-i} < 0} \end{cases}$$

où $\alpha_i^+(\cdot), \alpha_i^-(\cdot) : \Phi \rightarrow [0, \infty]$ sont des fonctions connues et $\boldsymbol{\phi}_0$ est un vecteur de paramètres inconnu de taille $r \times 1$, et ω_0 est une constante positive inconnue. On souhaite estimer $\boldsymbol{\theta}_0 = (\omega_0, \boldsymbol{\phi}_0)'$ sur un espace de paramètres Θ , sur la base de n observations $\varepsilon_1, \dots, \varepsilon_n$.

Le théorème 2.2 énonce la convergence du QMLE, noté $\tilde{\boldsymbol{\theta}}_n$. Pour établir la convergence de l'estimateur, les hypothèses présentées dans le chapitre 2 sont pour la plupart en accord avec celles proposées par Robinson et Zaffaroni (2006). Cependant, les hypothèses sur η_t sont moins restrictives car nous ne spécifions pas de contrainte particulière sur la distribution, contrairement à Robinson et Zaffaroni (2006) où la fonction de densité de probabilité de η_t est supposée bien se comporter au voisinage de 0. De plus, nos hypothèses sur α_i sont également moins fortes puisque nous permettons aux coefficients d'être égaux à 0 et n'imposons pas $\alpha_i(\boldsymbol{\phi}_0) \leq K \alpha_j(\boldsymbol{\phi}_0)$ pour $i \geq j \geq 1$.

La normalité asymptotique du QMLE est établie dans le théorème 2.3 qui stipule que

$$\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{L} \mathcal{N}(\mathbf{0}, (\kappa_\eta - 1)\mathbf{J}^{-1})$$

où

$$\kappa_\eta = \mathbb{E}(\eta_t^4) \quad \text{et} \quad \mathbf{J} = \frac{4}{\delta^2} \mathbb{E}_{\boldsymbol{\theta}_0} \left[\frac{1}{\sigma_t^{2\delta}(\boldsymbol{\theta}_0)} \frac{\partial \sigma_t^\delta(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \frac{\partial \sigma_t^\delta(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} \right]$$

est une matrice définie positive.

La structure de la preuve du théorème suit celle des résultats précédents dans le cas GARCH classique et est similaire à celle présentée dans l'ouvrage de Francq et Zakoïan (2019). Cependant, les démonstrations des résultats intermédiaires diffèrent et, notamment, l'impact des valeurs initiales requiert une attention particulière. La négligeabilité asymptotique des valeurs initiales est prouvée en utilisant un argument similaire à celui de Robinson et Zaffaroni (2006) et repose sur le taux de décroissance des fonctions α_i détaillé dans les hypothèses **A3(ii)** et **A8**. Une difficulté particulière dans le cadre de ce chapitre est que l'existence de moments pour le rapport $\sigma_t^\delta(\boldsymbol{\theta}_0)/\sigma_t^\delta(\boldsymbol{\theta})$ n'est pas assurée car les fonctions $\alpha_i^{+(-)}$ peuvent être égales à zéro pour certains i .

La distribution asymptotique de $\tilde{\boldsymbol{\theta}}_n$ a été obtenue en supposant que le vecteur des paramètres appartenait à l'intérieur de l'espace des paramètres $\boldsymbol{\Theta}$. Cette hypothèse est trop restrictive lorsque certains paramètres sont au bord de $\boldsymbol{\Theta}$. Le théorème 2.4 généralise les résultats du théorème 2.3 en établissant la distribution asymptotique de $\tilde{\boldsymbol{\theta}}_n$ lorsque $\boldsymbol{\theta}_0$ peut se trouver au bord de $\boldsymbol{\Theta}$. Dans ce cas, de manière similaire à Francq et Zakoïan (2007), la distribution asymptotique est la projection orthogonale d'une distribution vectorielle normale sur un cône convexe. Il convient de souligner que la distribution asymptotique du QMLE est obtenue sans aucune hypothèse supplémentaire sur les moments de ε_t .

En plus des résultats d'existence et d'inférence, le chapitre 2 présente quelques procédures de test. En particulier, la distribution asymptotique d'un test de qualité d'ajustement de type Portmanteau est établie dans le théorème 2.6. Depuis leur introduction par Box et Pierce (1970), les tests Portmanteau sont devenus largement utilisés pour tester l'adéquation des spécifications économétriques et ont été étendus aux modèles de volatilité par Li et Mak (1994). Les propriétés asymptotiques de ces tests ont été établies par Berkes et al. (2003a) pour le modèle GARCH(p, q) usuel et par Carbon et Francq (2011) dans le cas APARCH(p, q). A notre connaissance, ces résultats n'avaient pas encore été étendus à la littérature ARCH(∞). Considérons l'hypothèse nulle de qualité d'ajustement H_0^{GoF} qu'un processus (ε_t) satisfait le modèle APARCH(∞). L'autocovariance empirique au retard $h \leq n - 1$ des résidus au carré est donnée comme suit

$$\hat{r}_h = n^{-1} \sum_{t=h+1}^n (\hat{\eta}_t^2 - 1)(\hat{\eta}_{t-h}^2 - 1), \quad \text{avec} \quad \hat{\eta}_t^2 = \varepsilon_t^2 / \tilde{\sigma}_t^2(\tilde{\boldsymbol{\theta}}_n).$$

Soit $\hat{\mathbf{r}}_m = (\hat{r}_1, \dots, \hat{r}_m)$ pour tout $1 \leq m \leq n$, et $\hat{\mathbf{C}}_m$ la matrice de taille $m \times (r + 1)$ dont les éléments (h, k) sont donnés par

$$\hat{\mathbf{C}}_m(h, k) = -\frac{2}{\delta n} \sum_{t=h+1}^n (\hat{\eta}_{t-h}^2 - 1) \frac{1}{\tilde{\sigma}_t^\delta(\tilde{\boldsymbol{\theta}}_n)} \frac{\partial \tilde{\sigma}_t^\delta(\tilde{\boldsymbol{\theta}}_n)}{\partial \theta_k}.$$

En notant

$$\hat{\mathbf{J}}_n = \frac{4}{\delta^2} \frac{1}{n} \sum_{t=1}^n \frac{1}{\tilde{\sigma}_t^{2\delta}(\tilde{\boldsymbol{\theta}}_n)} \frac{\partial \tilde{\sigma}_t^\delta(\tilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}} \frac{\partial \tilde{\sigma}_t^\delta(\tilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}'} \quad \text{et} \quad \hat{\kappa}_\eta = \frac{1}{n} \sum_{t=1}^n \frac{\varepsilon_t^4}{\tilde{\sigma}_t^4(\tilde{\boldsymbol{\theta}}_n)}$$

des estimateurs convergents de \mathbf{J} et κ_η , la statistique du test Portmanteau est alors donnée par

$$n \hat{\mathbf{r}}_m' \hat{\mathbf{D}}^{-1} \hat{\mathbf{r}}_m \quad \text{avec} \quad \hat{\mathbf{D}} = (\hat{\kappa}_\eta - 1)^2 \mathbf{I}_m - (\hat{\kappa}_\eta - 1) \hat{\mathbf{C}}_m \hat{\mathbf{J}}_n^{-1} \hat{\mathbf{C}}_m'$$

et suit une loi χ_m^2 asymptotiquement comme le montre le théorème 2.6.

On peut remarquer que dans les modèles APARCH(∞), l'hypothèse de symétrie peut généralement s'exprimer comme une représentation contrainte particulière. Un test de significativité de l'asymétrie peut donc se construire en testant une restriction implicite sur le paramètre $\boldsymbol{\theta}_0$. Le chapitre 2 présente deux tests de contraintes linéaires. Considérons une hypothèse de la forme générale

$$H_0 : \mathbf{R} \boldsymbol{\theta}_0 = \mathbf{k}, \quad H_1 : \mathbf{R} \boldsymbol{\theta}_0 \neq \mathbf{k}$$

où \mathbf{R} est la matrice de contraintes et \mathbf{k} un vecteur constant. Soit c le rang de la matrice \mathbf{R} . La proposition 2.1 établit la distribution asymptotique du triptyque de statistiques usuelles : les statistiques de Wald, du score de Rao et du ratio de quasi-vraisemblances. Notons que ces distributions sont dérivées sans imposer que le vecteur de paramètres appartienne à l'intérieur de l'espace des paramètres. L'utilisation de représentations contraintes similaires est courante pour tester l'asymétrie dans le cadre de modèles (semi-)paramétriques, voir par exemple Nelson (1991).

Le second test de contrainte linéaire étudié dans le chapitre 2 présente une difficulté supplémentaire. Par simplicité, considérons la représentation ARCH(∞) suivante

$$\begin{cases} \varepsilon_t = \sigma_t(\boldsymbol{\theta}_0) \eta_t \\ \sigma_t^2(\boldsymbol{\theta}_0) = \frac{\omega_0}{1 - \beta_0} + \sum_{i=1}^{\infty} (\alpha_0 \beta_0^{i-1} + \gamma_0 i^{-(d_0+1)}) \varepsilon_{t-i}^2 \end{cases}$$

avec $\alpha_0 > 0$, $\beta_0 > 0$, $\gamma_0 \geq 0$, et $d_0 > 0$. D'après l'équation (1.4), cette dynamique est équivalente à celle du GARCH(1,1) si le paramètre γ_0 est nul. Tester la validité d'une représentation GARCH(1,1) peut donc s'effectuer en testant

$$H_0^{\text{GARCH}} : \gamma_0 = 0, \quad H_1^{\text{ARCH}(\infty)} : \gamma_0 > 0.$$

Bien que ce test puisse paraître trivial, il présente une difficulté majeure. En effet, le paramètre d_0 n'est pas identifié sous l'hypothèse nulle, on ne peut donc pas obtenir la loi asymptotique des statistiques de test comme présenté dans la proposition 2.1. La proposition 2.4 établit la distribution de la statistique de Wald lorsque le paramètre d_0 est supposé connu et fixé à une valeur $\bar{d} > 0.5$. A cette condition, la distribution asymptotique est un mélange entre une mesure de Dirac en 0 et une loi du χ^2 à un degré de liberté. De plus, une procédure de rééchantillonnage basé sur les résidus est introduite afin d'approximer la loi asymptotique de la statistique de

Wald sans avoir à fixer le paramètre d à une valeur arbitraire. La validité de cette approche est vérifiée à l'aide d'expériences de Monte Carlo.

Le chapitre 2 se termine par une application des tests proposés sur les rendements quotidiens de divers indices d'actions internationaux sur la période allant de janvier 1995 à décembre 2020. Sans surprise, la grande majorité des indices rejettent l'hypothèse de symétrie, ce qui est un résultat classique dans la littérature financière. Moins attendu est le fait que près de la moitié des trente indices rejettent l'hypothèse d'une spécification GARCH(1,1), et huit rejettent le modèle GJR-GARCH(1,1) ou TGARCH(1,1) au niveau de 5%. Il est remarquable de constater que tous les indices qui rejettent l'hypothèse de mémoire courte sont issus de marchés émergents. Cela suggère que le niveau de développement d'un marché financier a des implications sur les propriétés de persistance affichées par ses actifs. Une explication possible provient de la difficulté d'investir dans des marchés périphériques avec moins d'investisseurs et des instruments moins liquides, ce qui conduit à une intégration plus lente des chocs et finalement à une persistance plus élevée. De plus, un exercice hors échantillon illustre l'utilité du modèle APARCH(∞) lors du calcul de prévisions quotidiennes pour la *Value-at-Risk*.

La prise en compte de la persistance plus élevée du processus de volatilité conditionnelle s'est avérée utile dans le chapitre 2. Cependant, le cadre considéré ne permet pas de modéliser des systèmes multivariés. Les aspects probabilistes et statistiques des modèles ARCH(∞) multivariés ont rarement été étudiés dans la littérature académique. À cette fin, le chapitre 3 exploite le modèle Cholesky-GARCH de Darolles et al. (2018). Plus précisément, soit $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, \dots, \varepsilon_{m,t})'$ un vecteur de $m \geq 2$ séries de rendement vérifiant un modèle multivarié de volatilités

$$\boldsymbol{\varepsilon}_t = \boldsymbol{\Sigma}_t^{1/2} \boldsymbol{\eta}_t$$

où $(\boldsymbol{\eta}_t)$ est une suite de vecteurs aléatoires iid $(\mathbf{0}, \mathbf{I}_m)$. La décomposition de Cholesky de la matrice de variance-covariance conditionnelle considérée par Pourahmadi (1999) permet d'obtenir

$$\boldsymbol{\Sigma}_t = \mathbf{L}_t \mathbf{G}_t \mathbf{L}_t'$$

avec $\mathbf{G}_t = \text{diag}(\mathbf{g}_t)$ une matrice diagonale et \mathbf{L}_t une matrice unitriangulaire inférieure. Soit $\mathbf{v}_t = (v_{1,t}, \dots, v_{m,t})'$ avec $v_{k,t} = \sqrt{g_{k,t}} \eta_{k,t}$ pour $k = 1, \dots, m$, les facteurs orthogonaux obtenus récursivement à partir des séries observables $\boldsymbol{\varepsilon}_t$. En prenant $\boldsymbol{\Sigma}_t^{1/2} = \mathbf{L}_t \mathbf{G}_t^{1/2}$, on obtient

$$\boldsymbol{\varepsilon}_t = \boldsymbol{\Sigma}_t^{1/2} \boldsymbol{\eta}_t = \mathbf{L}_t \mathbf{v}_t$$

et donc $\mathbf{B}_t \boldsymbol{\varepsilon}_t = \mathbf{v}_t$ avec $\mathbf{B}_t = \mathbf{L}_t^{-1}$. Soit $\ell_{ij,t}$ (resp. $-\beta_{ij,t}$) l'élément de la ligne i et la colonne j de la matrice \mathbf{L}_t (respectivement \mathbf{B}_t). Alors, le modèle Cholesky-GARCH permet une modélisation jointe de la dynamique des bêtas conditionnels et de la variance conditionnelle des facteurs orthogonaux $v_{k,t}$. Dans ce modèle, les volatilités conditionnelles \mathbf{g}_t des facteurs orthogonaux sont souvent supposées suivre une équation GARCH multivariée, telle le modèle *Extended-CCC*

GARCH de Jeantheau (1998) où

$$\mathbf{g}_t = \boldsymbol{\omega} + \sum_{i=1}^q \mathbf{A}_i \mathbf{v}_{t-i}^2 + \sum_{j=1}^p \mathbf{b}_j \mathbf{g}_{t-j} \quad (1.6)$$

avec $\mathbf{v}_t^2 = (v_{1,t}^2, \dots, v_{m,t}^2)$, et où le vecteur $\boldsymbol{\omega} \in \mathbb{R}^m$ et les matrices \mathbf{A}_i et \mathbf{b}_j de taille $m \times m$ sont à entrées positives. Cependant, comme pour le modèle GARCH univarié, sous cette structure le modèle Cholesky-GARCH ne peut pas rendre compte de la forte persistance de certaines séries de rendements financiers. Pour pallier à la caractéristique de mémoire courte du modèle (1.6) nous considérons une représentation ARCH(∞) multivariée de la forme

$$\mathbf{g}_t = \boldsymbol{\omega} + \sum_{i=1}^{\infty} \mathbf{A}_i \mathbf{v}_{t-i}^2 + \sum_{i=1}^{\infty} \boldsymbol{\Pi}_i \mathbf{X}_{t-i} \quad (1.7)$$

où les matrices \mathbf{A}_i , de taille $m \times m$, et les matrices $\boldsymbol{\Pi}_i$, de taille $m \times P$, $i \in \{1, 2, \dots\}$ sont à entrées positives. En particulier, le vecteur \mathbf{X}_t comporte P variables exogènes positives telles que $(\boldsymbol{\eta}'_t, \mathbf{X}'_t)'$ est stationnaire et ergodique, avec $\boldsymbol{\eta}_t$ indépendant de $\{\mathbf{X}_{t-i}, i > 0\}$. L'addition de variables exogènes s'est montrée utile à la fois dans le cas univarié (voir par exemple Francq et Thieu (2019)) et le cas multivarié (voir par exemple Bollerslev et al. (2020b)), mais à notre connaissance, elle n'avait jamais été étudiée dans la littérature ARCH(∞). Notons que nous n'empêchons pas les variables exogènes d'être corrélées entre elles ainsi que corrélées avec le processus d'innovations. Ces dernières ne sont donc ni faiblement ni fortement exogènes au sens de Engle et al. (1983), mais exogènes dans le sens où leur dynamique n'est pas spécifiée par le modèle Cholesky-ARCH(∞).

Pour compléter la définition du modèle Cholesky-ARCH(∞), la dynamique des bêtas conditionnels doit être spécifiée et nous proposons un processus de la forme générale suivante

$$\boldsymbol{\beta}_t = \mathbf{c}(\mathbf{v}_{t-1}, \dots, \mathbf{v}_{t-q}, \mathbf{g}_{t-1}^{1/2}, \dots, \mathbf{g}_{t-q}^{1/2}, \mathbf{X}_{t-1}, \dots, \mathbf{X}_{t-q}) + \sum_{k=1}^p \mathbf{C}_k \boldsymbol{\beta}_{t-k}, \quad (1.8)$$

pour deux entiers p and q , avec \mathbf{c} une fonction mesurable de $\mathbb{R}^{(2m+P)q}$ dans $\mathbb{R}^{m(m-1)/2}$.

Les résultats théoriques du chapitre 3 commencent par les propriétés probabilistes du modèle. Le théorème 3.1 établit des conditions pour l'existence d'une solution unique strictement stationnaire, ergodique et non anticipative du modèle Cholesky-ARCH(∞), étendant les résultats de Doukhan et al. (2006). Une condition nécessaire supplémentaire pour l'existence des moments du processus des rendements ($\boldsymbol{\varepsilon}_t$) est fournie dans la proposition 3.1. Le chapitre 3 présente ensuite les résultats d'inférence statistique. Les propriétés asymptotiques du QMLE sont dérivées et présentées dans le théorème 3.2 montrant la convergence et la normalité asymptotique de l'estimateur.

Une caractéristique attrayante du modèle ARCH(∞) multivarié proposé, défini par le couple d'équations (1.7)-(1.8), est sa polyvalence. En particulier, en se basant sur le chapitre 2, une

spécification potentielle pour les variances conditionnelles individuelles est donnée par

$$g_{k,t} = \omega + \underbrace{\sum_{i=1}^{\infty} (\alpha_k \beta_k^{i-1} + \gamma_k i^{-d-1}) v_{k,t-i}^2}_{\text{effet ARCH}(\infty) \text{ spécifique}} + \underbrace{\sum_{j \neq k} \delta_{k,j} v_{j,t-1}^2}_{\text{contagion}} + \underbrace{\pi_k' \mathbf{X}_{t-1}}_{\text{exogénéité}}$$

qui encapsule le modèle Cholesky-GARCH ainsi que le modèle ARCH(∞) diagonal. Les variables populaires dans les équations de volatilité conditionnelle comprennent les volumes échangés, les mesures de volatilité implicite obtenues à partir des options cotées sur les marchés financiers, ainsi que les mesures de volatilité réalisée calculées à l'aide de données intrajournalières à haute fréquence (voir par exemple Engle et Patton (2001) et Francq et Thieu (2019)). La sélection des variables exogènes dans l'équation des bêtas conditionnels n'a cependant jamais été discutée. En se basant sur la littérature sur la volatilité réalisée, nous proposons d'utiliser des mesures de bêtas réalisés. Les bêtas réalisés, introduits pour la première fois par Barndorff-Nielsen et Shephard (2004) et Andersen et al. (2006), sont calculés à partir des rendements observés sur des intervalles de temps fixes. Soit $\boldsymbol{\varepsilon}_{w,t} = (\varepsilon_{w,1,t}, \dots, \varepsilon_{w,m,t})'$ le vecteur des rendements sur l'intervalle de temps w . La matrice de covariance réalisée $\mathbf{C}_t^{w,W}$ est alors définie comme la somme des W produits extérieurs des vecteurs de rendements. Par exemple, w peut être un jour et W un mois, ou w peut être un intervalle de 5 minutes et W un jour. Pour faciliter la notation, nous désignons par la matrice de covariance réalisée comme suit

$$\mathbf{C}_t := \mathbf{C}_t^{w,W} = \sum_{w=1}^W \boldsymbol{\varepsilon}_{w,t} \boldsymbol{\varepsilon}_{w,t}'.$$

De plus, pour $1 < k \leq m$, pour une matrice \mathbf{M} de taille $m \times m$, soient $\mathbf{M}_{[kk]}$ la sous-matrice carrée formée des $k-1$ premières lignes et colonnes, et $\mathbf{M}_{[k]}$ le sous-vecteur formé par les $k-1$ premières lignes et la k -ième colonne de \mathbf{M} . Pour $1 < i \leq m$, les bêtas réalisés entre l'actif i et les actifs $j = 1, \dots, i-1$ sont alors définis par

$$\bar{\boldsymbol{\beta}}_{i,t} = (\bar{\beta}_{ij,t})_{j=1,\dots,i-1} = (\mathbf{C}_{t[ii]})^{-1} \mathbf{C}_{t[i]}.$$

Dans l'esprit du modèle *Heterogeneous* AR de volatilité réalisée (HAR-RV) introduit par Corsi (2009), le bêta conditionnel quotidien pourrait dépendre des mesures de bêta réalisés calculées à des fréquences plus basses

$$\beta_{ij,t} = \varpi_{ij} + \pi_{ij}^{\mathcal{W}} \bar{\beta}_{ij,t-1}^{\mathcal{W}} + \pi_{ij}^{\mathcal{M}} \bar{\beta}_{ij,t-1}^{\mathcal{M}} + c_{ij} \beta_{ij,t-1}$$

où $\bar{\beta}_{ij,t}^{\mathcal{W}}$ désigne le bêta réalisé hebdomadaire en t calculé sur 5 jours ouvrés, et $\bar{\beta}_{ij,t}^{\mathcal{M}}$ le bêta réalisé mensuel en t calculé sur 21 jours ouvrés. Par ailleurs, les variables exogènes potentiellement utiles dans la dynamique des bêtas conditionnels pourraient être les semi-bêtas réalisés récemment introduits par Bollerslev et al. (2022). Considérons la décomposition de la matrice de covariance réalisée

$$\mathbf{C}_t = \mathcal{P}_t + \mathcal{N}_t + \mathcal{S}_t^+ + \mathcal{S}_t^-$$

en composantes de semi-covariance dérivées par Bollerslev et al. (2020a)

$$\begin{aligned}\mathcal{P}_t &= \sum_{w=1}^W \varepsilon_{w,t}^+ \varepsilon_{w,t}^{+'}, & \mathcal{S}_t^+ &= \sum_{w=1}^W \varepsilon_{w,t}^- \varepsilon_{w,t}^{+'}, \\ \mathcal{N}_t &= \sum_{w=1}^W \varepsilon_{w,t}^- \varepsilon_{w,t}^{-'}, & \mathcal{S}_t^- &= \sum_{w=1}^W \varepsilon_{w,t}^+ \varepsilon_{w,t}^{-'}\end{aligned}$$

où $\varepsilon_{w,t}^{+(-)}$ désigne la partie positive (respectivement négative) composantes par composantes de $\varepsilon_{w,t}$. Chaque composante de semi-covariance produit alors un semi-bêta réalisé

$$\begin{aligned}\bar{\beta}_{i,t}^{\mathcal{P}} &= (\mathbf{C}_{t[ii]})^{-1} \mathcal{P}_{t[i]}, & \bar{\beta}_{i,t}^{\mathcal{S}^+} &= -(\mathbf{C}_{t[ii]})^{-1} \mathcal{S}_{t[i]}^+, \\ \bar{\beta}_{i,t}^{\mathcal{N}} &= (\mathbf{C}_{t[ii]})^{-1} \mathcal{N}_{t[i]}, & \bar{\beta}_{i,t}^{\mathcal{S}^-} &= -(\mathbf{C}_{t[ii]})^{-1} \mathcal{S}_{t[i]}^-\end{aligned}$$

générant une décomposition du bêta réalisé usuel en quatre composantes

$$\bar{\beta}_{i,t} = \bar{\beta}_{i,t}^{\mathcal{P}} + \bar{\beta}_{i,t}^{\mathcal{N}} - \bar{\beta}_{i,t}^{\mathcal{S}^+} - \bar{\beta}_{i,t}^{\mathcal{S}^-}.$$

La dynamique du bêta conditionnel pourrait alors suivre

$$\beta_{ij,t} = \varpi_{ij} + \pi_{ij}^{\mathcal{P}} \bar{\beta}_{ij,t-1}^{\mathcal{P}} + \pi_{ij}^{\mathcal{N}} \bar{\beta}_{ij,t-1}^{\mathcal{N}} + \pi_{ij}^{\mathcal{S}^+} \bar{\beta}_{ij,t-1}^{\mathcal{S}^+} + \pi_{ij}^{\mathcal{S}^-} \bar{\beta}_{ij,t-1}^{\mathcal{S}^-} + c_{ij} \beta_{ij,t-1} \quad (1.9)$$

permettant une contribution plus précise des risques de baisse et de hausse.

Remarquablement, l'équation (1.9) se réduit à

$$\beta_{ij,t} = \varpi_{ij} + \pi_{ij}^{\mathcal{P}} \bar{\beta}_{ij,t} + c_{ij} \beta_{ij,t}$$

si $\pi_{ij}^{\mathcal{P}} = \pi_{ij}^{\mathcal{N}} = -\pi_{ij}^{\mathcal{S}^+} = -\pi_{ij}^{\mathcal{S}^-}$. Tester l'avantage informationnel de l'inclusion de semi-bêtas réalisés au lieu de bêtas réalisés classiques peut donc être réalisé en testant

$$\begin{cases} H_0 : \pi_{ij}^{\mathcal{P}} = \pi_{ij}^{\mathcal{N}} & \text{et} & \pi_{ij}^{\mathcal{P}} = -\pi_{ij}^{\mathcal{S}^+} & \text{et} & \pi_{ij}^{\mathcal{N}} = -\pi_{ij}^{\mathcal{S}^-} \\ H_1 : \pi_{ij}^{\mathcal{P}} \neq \pi_{ij}^{\mathcal{N}} & \text{ou} & \pi_{ij}^{\mathcal{P}} \neq -\pi_{ij}^{\mathcal{S}^+} & \text{ou} & \pi_{ij}^{\mathcal{N}} \neq -\pi_{ij}^{\mathcal{S}^-} \end{cases}.$$

L'application menée dans le chapitre 3 fournit un argument fort en faveur de la décomposition en semi-bêtas réalisés. En effet, pour seulement six des trente-cinq portefeuilles industriels disponibles sur le site de Kenneth French, l'hypothèse H_0 ne peut être rejetée au niveau de 5%.

Bien qu'elle permette une modélisation directe du bêta conditionnel combinée à un modèle de décroissance flexible des coefficients de la (co)variance conditionnelle, la structure du Cholesky-GARCH, et a fortiori la structure du Cholesky-ARCH(∞) rend son utilisation difficile dans les exercices d'évaluation d'actifs avec un large panel. Dans un article récent, Blasques et al. (2022a) proposent une autre approche pour obtenir des bêtas conditionnels dynamiques. En particulier, leur modèle de bêta conditionnel autorégressif (ACB) s'appuie sur la littérature des modèles basés sur les modèles de score autorégressif généralisés (GAS) pour permettre à la fois la modélisation de coefficients de régression dynamiques et une hétéroscédasticité conditionnelle des facteurs. Les modèles GAS ont été introduits par Harvey et Chakravarty (2008) et Creal et al.

(2013) et visent à généraliser les modèles dont les paramètres varient dans le temps. Considérons un processus observable y_t et supposons qu'il suive la densité conditionnelle $p(y_t|\psi_t, \Omega_t, \theta)$ où ψ_t est le paramètre dynamique d'intérêt, θ un paramètre constant, et Ω_t l'ensemble d'information disponible en t . L'équation de mise à jour du paramètre dynamique est donnée par

$$\psi_{t+1} = \bar{\omega} + \xi \underbrace{S(\psi_t) \frac{\partial \log p(y_t|\psi_t, \Omega_t, \theta)}{\partial \psi_t}}_{\text{mise à jour}} + c \psi_t \quad (1.10)$$

où $\bar{\omega}$, ξ et c sont des constantes à estimer. Le terme de mise à jour est décomposé entre le score de la vraisemblance par rapport au paramètre dynamique et un terme d'échelle $S(\psi_t)$. Il est intéressant de noter que de nombreux modèles économétriques peuvent être exprimés comme un modèle *score-driven*. Par exemple, si nous supposons que la densité conditionnelle des observations est gaussienne et que le paramètre d'intérêt est la variance dynamique, avec un choix approprié de la fonction d'échelle, l'équation (1.10) donne le modèle GARCH(1,1) usuel.

Soient $(r_{i,t})$ la série temporelle des rendements de l'actif $i = 1, \dots, N$ dans un panel de N actifs et $\mathbf{f}_t = (f_{1,t}, \dots, f_{m,t})'$ un ensemble de m facteurs observables. On considère une régression conditionnelle de la forme

$$r_{i,t} = \alpha_{i,t} + \beta_{i,t}' \mathbf{f}_t + \varepsilon_{i,t}. \quad (1.11)$$

En outre, supposons que chaque facteur dynamique f_j suive un GARCH(1,1) avec une moyenne conditionnelle constante

$$\begin{aligned} f_{j,t} &= \mu_j + \sigma_{j,t} \eta_{j,t} \\ \sigma_{j,t}^2 &= \omega_j + a_j (f_{j,t-1} - \mu_j)^2 + b_j \sigma_{j,t-1}^2 \end{aligned} \quad (1.12)$$

où $\eta_{j,t}$ est une suite de variables aléatoires iid centrées et de variance unitaire. De même, supposons que les résidus de la régression conditionnelle suivent un modèle GARCH(1,1) avec une variance conditionnelle

$$g_{i,t}^2 = \omega_{\varepsilon_i} + a_{\varepsilon_i} (r_{i,t-1} - \alpha_{i,t-1} - \beta_{i,t-1}' \mathbf{f}_{t-1})^2 + b_{\varepsilon_i} g_{i,t-1}^2. \quad (1.13)$$

En supposant que pour tout i , $\varepsilon_{i,t} = g_{i,t} \zeta_{i,t}$ où $\zeta_{i,t}$ sont gaussiens iid, Blasques et al. (2022a) dérivent les équations de mise à jour suivantes pour les coefficients de la régression conditionnelle

$$\begin{aligned} \alpha_{i,t+1} &= \bar{\omega}_{\alpha_i} + \xi_{\alpha_i} \varepsilon_{i,t} + c_{\alpha_i} \alpha_{i,t} \\ \beta_{i,j,t+1} &= \bar{\omega}_{\beta_{i,j}} + \xi_{\beta_{i,j}} \frac{f_{j,t} \varepsilon_{i,t}}{\mu_j^2 + \sigma_{j,t}^2} + c_{\beta_{i,j}} \beta_{i,j,t} \end{aligned} \quad (1.14)$$

pour $i = 1, \dots, N$ and $j = 1, \dots, m$. Les auteurs établissent la convergence et la normalité asymptotique d'un QMLE à plusieurs étapes. Le chapitre 4 s'appuie sur le modèle ACB pour développer un modèle conditionnel d'évaluation d'actifs et dériver une procédure d'estimation des primes de risque dynamiques associées à l'ensemble des facteurs \mathbf{f}_t . Tout d'abord, les distributions asymptotiques de statistiques de test usuelles sont dérivées pour évaluer la constance d'un paramètre dans la régression conditionnelle. Ces statistiques permettent notamment de

tester la pertinence d'un facteur donné. Dans le cas d'un modèle de régression inconditionnelle, un facteur est dit significatif si le bêta associé à ce facteur est significativement différent de zéro. Nous proposons donc d'évaluer la significativité d'un facteur j en testant que son bêta conditionnel est à la fois constant et égal à zéro. D'après (1.14), la constance de $\beta_{i,j,t}$ est obtenue lorsque $\xi_{0ij} = 0$. Cependant, tester cette contrainte sur le vecteur de paramètres $\boldsymbol{\vartheta}_0^{(i)}$ est non-trivial. En effet, sous $H_{0,ij} : \xi_{0ij} = 0$, le bêta conditionnel $\beta_{i,j,t}$ tend vers $\bar{\omega}_{ij}/(1 - c_{ij})$ et il existe une infinité de couples $(\bar{\omega}_{ij}, c_{ij})$ tels que $\beta_{i,j,t} = \bar{\omega}_{ij}$, ce qui entraîne un modèle non identifiable sous l'hypothèse nulle. Un problème similaire se produit lorsque l'on teste la significativité d'un facteur dans la régression conditionnelle, ce qui équivaut à $\xi_{0ij} = 0$ et $\bar{\omega}_{ij} = 0$. Pour contourner ce problème d'identification, nous proposons de fixer le paramètre c_{0ij} à une valeur arbitraire \bar{c}_{ij} . Cela résout le problème d'identification et permet de tester l'hypothèse de contrainte linéaire $H_{0,ij}(\bar{c}_{ij}) : \mathbf{R}\boldsymbol{\vartheta}_0^{(i)} = 0$ où \mathbf{R} est la matrice des contraintes.

Le théorème 4.1 établit que, sous $H_{0,ij}(\bar{c}_{ij})$, la statistique de Wald, notée W_n , et la statistique de test de Rao, notée R_n , suivent asymptotiquement une distribution χ_r^2 et leurs régions critiques au niveau asymptotique ν sont données par

$$\{W_n > \chi_r^2(1 - \nu)\}, \{R_n > \chi_r^2(1 - \nu)\}.$$

Cependant, la statistique du rapport de quasi-vraisemblance a une distribution asymptotique différente de la distribution usuelle χ^2 des statistiques de Wald et de Rao. La région critique au niveau asymptotique ν pour L_n peut cependant être facilement obtenue en suivant la méthode de Imhof (1961). De plus, une procédure bootstrap basée sur les résidus est proposée pour approximer la distribution asymptotique de la statistique de Wald. Le théorème 4.2 établit la validité du schéma de rééchantillonnage et garantit que la distribution de la statistique bootstrap, étant donné $(r_{i,t}, \mathbf{f}_t)$, imite bien la distribution inconditionnelle de la statistique de Wald.

Le chapitre 4 détaille ensuite la procédure d'estimation permettant de récupérer les primes de risque dynamiques. La procédure s'appuie sur Gagliardini et al. (2016) qui dérivent des restrictions de non-arbitrage pour une économie avec un continuum d'actifs. En particulier, ils montrent que la restriction d'évaluation d'actifs, basée sur une hypothèse d'absence d'opportunités d'arbitrages, est telle que, pour tout t , il existe un vecteur aléatoire unique $\mathbf{v}_t = (v_{1,t}, \dots, v_{m,t})'$ mesurable par rapport à la filtration naturelle \mathcal{F}_{t-1} tel que pour presque tous les i ,

$$\alpha_{i,t} = \boldsymbol{\beta}_{i,t}' \mathbf{v}_t \text{ presque sûrement.} \quad (1.15)$$

Le modèle dynamique d'évaluation d'actifs (1.11) combiné à (1.15) donne alors que pour presque tous les i ,

$$\mathbb{E}[r_{i,t} | \mathcal{F}_{t-1}] = \boldsymbol{\beta}_{i,t}' \boldsymbol{\lambda}_t \text{ presque sûrement,} \quad (1.16)$$

où $\boldsymbol{\lambda}_t$, le vecteur des primes de risque dynamiques, est donné par $\boldsymbol{\lambda}_t = \mathbf{v}_t + \boldsymbol{\mu}$ avec $\boldsymbol{\mu} = \mathbb{E}[\mathbf{f}_t] = (\mu_1, \dots, \mu_m)$. Il convient de noter que l'équation (1.16) encapsule de nombreux modèles d'évaluation d'actifs dont les coefficients sont invariants. Par exemple, dans le CAPM où le seul facteur

est le portefeuille de marché ou dans le modèle à trois facteurs de Fama et French (1993), nous avons $\mathbf{v}_t = 0$.

En s'appuyant sur la restriction d'évaluation d'actifs (1.15), nous proposons la procédure en deux étapes suivante pour estimer les primes de risque dynamiques à partir du modèle de régression conditionnelle

1. Pour chaque actif i du système, $i = 1, \dots, N$, on estime $\hat{\alpha}_{i,t} := \hat{\alpha}_{i,t}(\hat{\varphi}_n^{(i)})$ et $\hat{\beta}_{i,t} := \hat{\beta}_{i,t}(\hat{\varphi}_n^{(i)})$ où $\hat{\varphi}_n^{(i)}$ désigne le QMLE à plusieurs étapes de Blasques et al. (2022a)
2. Pour chaque t , on estime $\hat{\mathbf{v}}_t = \arg \min_{\mathbf{v}} \left\| \hat{\beta}'_t \mathbf{v} - \hat{\alpha}_t \right\|_2$ par la méthode des moindres carrés ordinaire, où $\hat{\beta}_t = (\hat{\beta}_{i,t})_{i=1, \dots, N}$ et $\hat{\alpha}_t = (\hat{\alpha}_{i,t})_{i=1, \dots, N}$.

En combinant $\hat{\mathbf{v}}_t$ avec le paramètre $\hat{\boldsymbol{\mu}}_n$ estimé dans $\hat{\boldsymbol{\theta}}_n$, on obtient l'estimation des primes de risque $\hat{\boldsymbol{\lambda}}_t = \hat{\boldsymbol{\mu}}_n + \hat{\mathbf{v}}_t$.

Le chapitre 4 se termine par une application visant à évaluer l'existence d'une prime de risque dynamique associée à un facteur lié au risque carbone. Nous trouvons une prime dynamique significative et discutons de la relation entre ses variations et les informations liées à la finance durable.

1.2 Summary of the thesis works

The purpose of this manuscript is to study conditional volatility models. More precisely, this thesis will focus on two areas of research in the financial econometrics literature. First, the modeling of time series exhibiting highly persistent volatilities. To that extent, new ARCH(∞) extensions, both in the univariate and multivariate settings, will be introduced. The probabilistic and statistical aspects of the processes will be discussed. Second, the modeling of time-varying regression coefficients in linear factor models. Although the two research topics appears disconnected, we will see throughout this manuscript that they are both related to the modeling of the conditional (co-)variance. The specifications introduced in the following chapters will be illustrated on real financial data applications, providing strong arguments in favor of the proposed extensions.

1.2.1 Context

The study of time-varying phenomena has been at the heart of statistics and probability theory for the last century. In its most prominent framework, a discrete-time stochastic process $(X_t)_{t \in \mathbb{Z}}$ is assumed to follow a dynamic of the form

$$X_t = \varphi(X_{t-1}, X_{t-2}, \dots; \eta_t)$$

where φ describes the temporal dependence of the random variable X_t and $(\eta_t)_{t \in \mathbb{Z}}$ is a sequence of independent and identically distributed (iid) random variables. Time series analysis first focused on linear φ functions, following the seminal work of Yule (1927), almost a century ago. In

such case, X_t is called an autoregressive process (AR). Theoretical properties of AR processes have been the source of a rich literature (see for example the reference textbook of Brockwell and Davis (2009)). Linear time series applications have been widespread, from economic to climate variables. However, their dynamics are of little use when modeling financial time series.

Although being some of the most widely scrutinized data, financial time series are complex to model. This difficulty stems from the existence of similar statistical properties, observed in a majority of assets independently of their asset class (corporate bonds, equities, commodities...). The first of these so-called *stylized facts* is the absence of temporal linear dependence of financial returns. This property clearly rules out the use of AR specifications for modeling price increments or returns and was at the root of the *Random Walk Hypothesis* of price series and the *Efficient Market* definition of Fama (1970, 1991). However, although not autocorrelated, financial returns are not independent as non-linear transformations of the return series exhibit time dependence. This observation led to the rejection of the *Random Walk Hypothesis* (see for example Lo and MacKinlay (1988)) and the emergence of a second stylized fact: volatility clustering. Returns large deviations from their mean appear in clusters leading to alternating periods of high and low volatility. This phenomenon is recurrent but not periodic and can be directly observed on time series path as shown in Figure 1.1 presenting the CAC 40 Index daily returns from January 1995 to May 2020.

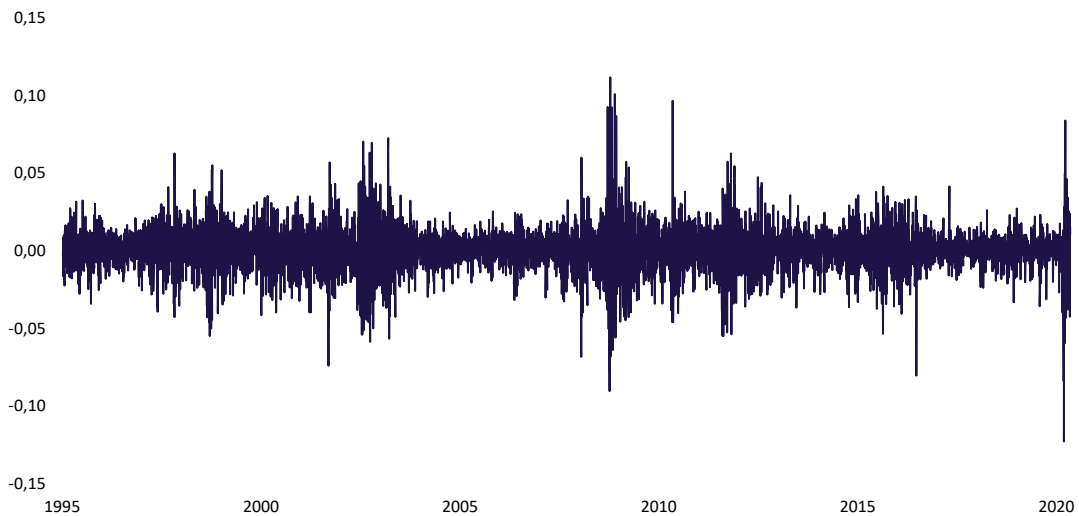


Figure 1.1: *Daily returns of the CAC 40 Index from January 1995 to May 2020*

This empirical fact is incompatible with the assumption of constant conditional variance and has led to the introduction of conditionally heteroskedastic models with the seminal autoregressive conditionally heteroskedastic (ARCH) model of Engle (1982). In this non-linear model, the dynamic of X_t is written in the multiplicative form

$$X_t = \sigma_t \eta_t \tag{1.1}$$

where σ_t is a positive measurable function with respect to the σ -field $\mathcal{F}_{t-1} = \{X_u, u < t\}$. In particular, in the ARCH(q) model of Engle, $0 < q < \infty$, the function σ_t is a linear function of the first q squared past values of X_t . This model was extended by Bollerslev (1986), yielding the famous Generalized ARCH (GARCH), which arguably became one of the most prominent models in the financial econometrics literature. In this specification, the conditional variance of X_t is given by the following dynamic

$$\sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i X_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 \quad (1.2)$$

where ω is a non-negative constant, $(\alpha_i)_{i=1,\dots,q}$ and $(\beta_j)_{j=1,\dots,p}$ are positive constants, and η_t are centered independent and identically distributed random variables with unit variance.

While GARCH models are well-suited to model time series with both absence of autocorrelation and volatility clustering, the particular structure of the conditional variance imposes an exponential decay of the autocovariance of the squared returns which is in contradiction with their empirical autocorrelograms. Indeed, as first noted by Ding and Granger (1996), the volatility process empirically exhibits strong persistence which contradicts the use of standard GARCH models. Figure 1.2 presents the autocorrelogram of the returns, squared returns, and absolute value of the returns of the CAC 40 Index. It is seen that even after a hundred lags, the autocorrelations remain far from 0 for the transformed return series⁴.

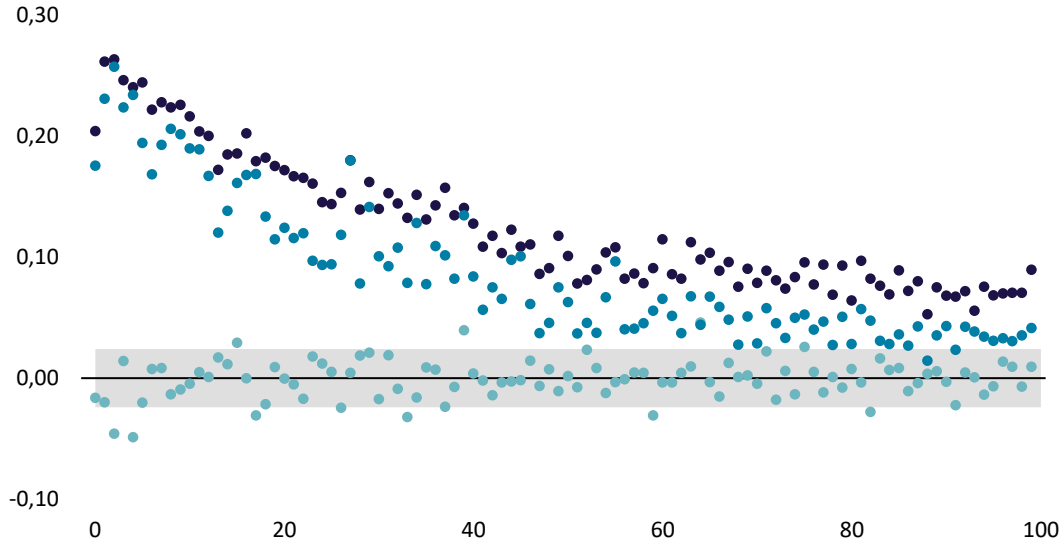


Figure 1.2: *Autocorrelations of the returns (light blue), squared returns (blue), and absolute value of the returns (dark blue) of the CAC 40 Index*

To allow for a more flexible form of decay of the coefficients in the volatility equation, Robinson

⁴In addition to the autocorrelogram of power transformations of returns, Luati et al. (2012) and Luati and Proietti (2015) advocate the use of transformation of the spectral density function when studying the memory pattern of stationary time series

(1991) introduced the ARCH(∞) model as an extension to the *short memory* GARCH model. In these models, the time series dynamic is given by

$$\begin{cases} X_t = \sigma_t \eta_t \\ \sigma_t^2 = \omega + \sum_{i=1}^{\infty} \alpha_i X_{t-i}^2 \end{cases} \quad (1.3)$$

with $\omega > 0$, α_i , $i = 1, \dots$, a sequence of positive constants, and η_t are centered independent and identically distributed random variables with unit variance. By relaxing the exponential decrease of autocorrelations of squared GARCH processes, ARCH(∞) are suitable for series with longer memory. Remarkably, by inverting the lag polynomial in Equation (1.2), one can show that a GARCH(p, q) model admits an ARCH(∞) representation. In particular, the GARCH(1,1) conditional variance equation writes as

$$\sigma_t^2 = \frac{\omega}{1 - \beta_1} + \sum_{i=1}^{\infty} \alpha_1 \beta_1^{i-1} X_{t-i}^2$$

which uncovers the exponential decay of the coefficients imposed by the GARCH structure. To circumvent this exponential decay, a popular ARCH(∞) specification is the fractionally integrated generalized autoregressive conditionally heteroskedastic (FIGARCH) of Baillie et al. (1996). Building upon the fractionary autoregressive integrated moving average (ARFIMA) model introduced by Granger and Joyeux (1980) and Hosking (1981)⁵, the FIGARCH(p, d, q) model is defined as follows

$$\begin{cases} X_t = \sigma_t \eta_t \\ \sigma_t^2 = \omega_0 + \left[1 - (1-L)^d \frac{\zeta(L)}{\psi(L)} \right] X_t^2 \end{cases} \quad (1.4)$$

where ψ and ζ are polynomials of degrees p and q , $d \in (0, 1)$, and $(1-L)^d$ denotes the fractionary difference operator

$$(1-L)^d = \sum_{k=0}^{\infty} \pi_k(d) L^k$$

with L the standard lag-operator such that $L^k X_t = X_{t-k}$ for any $k \in \mathbb{Z}$ and

$$\pi_k(d) = \frac{\Gamma(k-d)}{\Gamma(k+1)\Gamma(-d)}.$$

To illustrate the slower decay of the coefficients in the FIGARCH model, one can consider the simple FIGARCH(0, d , 0) obtained with $\psi(L) = \zeta(L) = 1$. The model then writes as

$$\sigma_t^2 = \omega_0 - \sum_{k=1}^{\infty} \pi_k(d) X_{t-k}^2.$$

⁵See Giraitis et al. (2012) for a review of linear memory model.

Thus, by identification of the α_i coefficients in Equation (1.3), we have

$$\alpha_i = -\pi_i(d) = -\frac{\Gamma(i-d)}{\Gamma(i+1)\Gamma(-d)}$$

and Stirling's formula yields $\alpha_i \sim i^{-d-1}$.

The existence of a strictly stationary and nonanticipative solution to the ARCH(∞) model has been proved by Giraitis et al. (2000), Kazakevičius and Leipus (2002), and Douc et al. (2008) under the condition

$$A_s \mu_{2s} < 1 \tag{1.5}$$

for some $s \in (0, 1]$, where $A_s = \sum_{i=1}^{\infty} \alpha_i^s$ and $\mu_{2s} = E|\eta_t|^{2s}$. This condition is however not compatible with the usual concept of long memory as it entails summability of the autocovariances of X_t^2 (see Giraitis et al. (2000) and Zaffaroni (2004)). Note that this condition also rules out the existence of a stationary Integrated ARCH(∞) model for which $\sum_{i=1}^{\infty} \alpha_i = 1$. The question of the existence of a stationary solution with a finite fourth order moment for the FIGARCH model has been the subject of a long academic discussion which amounted to the recent paper by Giraitis et al. (2018) in which the existence of such solution was established. Nevertheless, ARCH(∞) models remain compatible with a slow decay of the autocorrelation function of the squares and are sometimes referred to as *moderate memory models*.

1.2.2 Motivations

Although moderate memory models are natural candidates to model financial time series, their use amongst practitioners has been regrettably limited. The aim of this dissertation is to buttress the use of ARCH(∞)-type models. To that effect, we will propose several extensions to the ARCH(∞) framework to fill the gap between popular GARCH-type extensions and moderate memory models.

In particular, Chapter 2 considers a generalization of the classical ARCH(∞) model, allowing for conditional asymmetry, namely the higher impact on conditional volatility of negative observations rather than positive observation of the same magnitude. Figure 1.3 illustrates this stylized fact of financial series⁶. In addition to the absence of autocorrelation of returns (in light blue) and significant positive autocorrelation of absolute values of returns (in dark blue) already shown in Figure 1.2, it presents the correlation between X_t^+ and X_{t-h} for $h = 1, \dots, 100$ where X_t is the daily return of the CAC 40 Index and $X_t^+ = \max(X_t, 0)$ (in red). Admitting that the process X_t is second-order stationary and admits a dynamic as Equation (1.1), we have

$$\rho(X_t^+, X_{t-h}) = K \text{cov}(\sigma_t, X_{t-h}) = K[\text{cov}(\sigma_t, X_{t-h}^+) + \text{cov}(\sigma_t, X_{t-h}^-)]$$

where $\rho(Y, Z)$ denotes the correlation between random variables Y and Z . As shown in Figure

⁶This illustration is inspired by Chapter 10 of Francq and Zakoïan (2019).

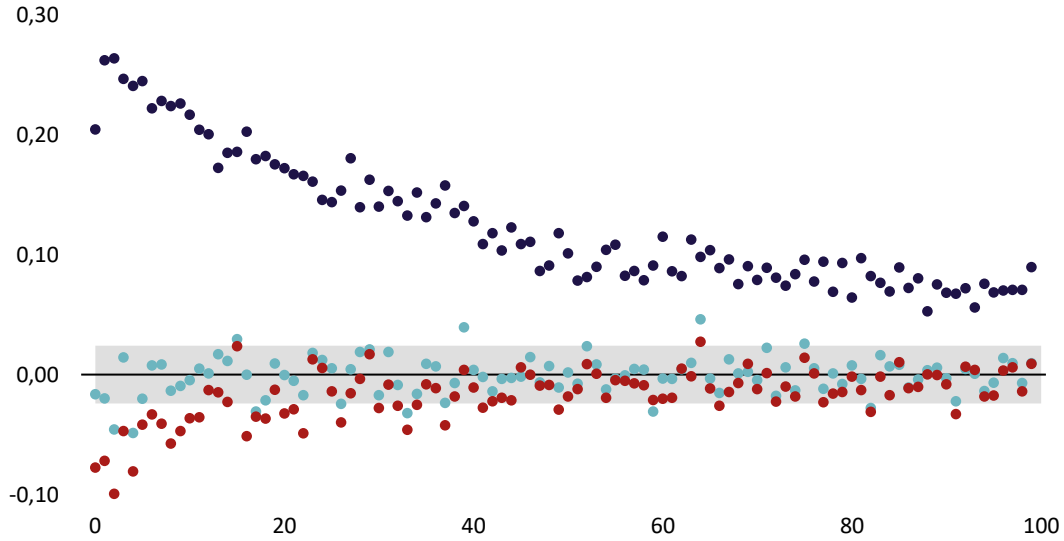


Figure 1.3: *Illustration of the conditional asymmetry on the CAC 40 Index daily returns*

1.3, for several lags, the correlation between X_t^+ and X_{t-h} is clearly negative which means

$$\text{cov}(\sigma_t, X_{t-h}^+) < \text{cov}(\sigma_t, -X_{t-h}^-)$$

underlying the higher impact of past negative values of X_t on the conditional variance compared to positive values. This asymmetry was first observed by Black (1976) and Christie (1982) and is often referred to as *leverage effect* since the authors suggest this phenomenon arises due to the change in the equity-debt ratio of the firm after a large equity price drop leading to a higher risk and thus a higher volatility. It is however unlikely that this sole mechanism accounts for the magnitude of the asymmetry (see for example Schwert (1989)). Another popular explanation is the *volatility feedback* which explains asymmetry as the result of time-varying risk premia and volatility anticipations (e.g. French et al. (1987), Campbell and Hentschel (1992), and Bekaert and Wu (2000)). Although numerous extensions to the GARCH models have been introduced in the literature (popular specifications include the GJR-GARCH model of Glosten et al. (1993) or the Threshold-GARCH model of Zakoïan (1994)), to the best of our knowledge, attempts to capture both the asymmetry and the memory properties of financial time series have been scarce. Chapter 2 proposes a novel ARCH(∞) model to take this *leverage effect* into account.

Although conditional volatilities were initially developed to capture heteroskedasticity of univariate time series, numerous extensions have been proposed to model multivariate systems. These multivariate GARCH models are of great interest in risk management or portfolio construction where the conditional covariance matrix is key to the investment decision process. In these models, a real-valued vector process \mathbf{X}_t of dimension m , $\mathbf{X}_t = (X_{1,t}, \dots, X_{m,t})'$ is defined as

$$\mathbf{X}_t = \mathbf{H}_t^{1/2} \boldsymbol{\eta}_t$$

where $(\boldsymbol{\eta}_t)$ is a sequence of iid centered \mathbb{R}^m -valued process with identity covariance matrix. Various specifications of the conditional covariance matrix \mathbf{H}_t have been proposed, the first one being the diagonal model of Engle et al. (1984). Popular extensions noticeably include the Constant Conditional Correlation (CCC) GARCH of Bollerslev (1990) and the extension of Jeantheau (1998), the BEKK-model of Engle and Kroner (1995) as well as the Dynamic Conditional Correlation (DCC) model of Engle (2002) and Tse and Tsui (2002).⁷ Stationarity conditions for the CCC and DCC model were respectively obtained by Aue et al. (2009) and Fermanian and Malongo (2017) while asymptotic inference results for quasi maximum likelihood (QML) estimation were derived by Jeantheau (1998), Ling and McAleer (2003), Comte and Lieberman (2003) and Francq and Zakoïan (2012).

Although a large literature has been devoted to the study of short memory multivariate models, their ARCH(∞) counterparts have rarely been explored. Stationarity conditions of multivariate ARCH(∞) model were nevertheless discussed by Doukhan et al. (2006) while Bardet and Wintenberger (2009) established the consistency and asymptotic normality of the QML estimator (QMLE) for a very general class of multidimensional causal processes. The latter contribution however does not directly aim at modeling multivariate moderate memory models and the condition of higher moments for the data generating process may be difficult to conciliate with ARCH(∞) sequences. Chapter 3 introduces a novel multivariate volatility models in which covolatilities can exhibit higher persistence than the standard multivariate GARCH equation. Conditions for existence of a unique stationary solution are established as well as the consistency and asymptotic normality of the QMLE. In particular, we build upon the Cholesky GARCH model of Darolles et al. (2018) which allows to directly model the conditional betas between assets in the multivariate system through a Cholesky decomposition of the conditional covariance matrix.

The modeling of conditional linear regression is of great interest in practice. Consider the conditional regression of Y_t on a set of m exogenous variables $\mathbf{X}_t = (X_{1,t}, \dots, X_{m,t})'$

$$Y_t = \boldsymbol{\beta}_t' \mathbf{X}_t + \eta_t = \sum_{k=1}^m \beta_{k,t} X_{k,t} + \eta_t$$

where $\boldsymbol{\beta}_t$ are the conditional slope coefficients and η_t are error terms that are orthogonal to \mathbf{X}_t . In asset pricing models, Y_t usually describes the excess returns of a financial asset or a portfolio, and the exogenous variables \mathbf{X}_t are systematic risk factors impacting the cross-section of assets. For example, the Capital Asset Pricing Model (CAPM) of Sharpe (1964) and Lintner (1965) use the the excess market portfolio returns in their single factor model, while Fama and French (1993) consider two additional factors capturing size and value effects. In these factor models, a recurrent yet problematic assumption is the time invariance of the slope parameters (the so-called betas). Such an hypothesis is difficult to justify empirically and is often rejected on real-data applications (see for example Engle (2016)). To our knowledge, Goldfeld and Quandt (1973) and Hamilton (1989) were the first to consider dynamic regression

⁷For a review of multivariate GARCH models, see Bauwens et al. (2006).

coefficients through regime-switching models. However, in such models, the slope coefficients are considered constant in each state which can be restrictive when considering regimes with high duration. Shanken (1990) and Ferson and Harvey (1991) proposed to use financial or economics instrumental variables to recover time-varying betas while more recently Gagliardini et al. (2016) established theoretical results on the estimation of time-varying risk premia in this setting. This method however requires the selection of exogenous variables to drive the entire dynamics of the betas. In particular, Ghysels (1998) noted that the obtained slope coefficients can be highly sensitive to the choice of instrumental variables. Building upon the work of Bollerslev et al. (1988), Engle (2016) recently proposed a different approach to recovering time-varying betas. Assuming

$$\begin{pmatrix} Y_t \\ \mathbf{X}_t \end{pmatrix} \Big| \mathcal{F}_{t-1} \sim \mathcal{N} \left(\mathbf{0}, \mathbf{H}_t := \begin{bmatrix} H_{YY,t} & \mathbf{H}_{YX,t} \\ \mathbf{H}_{XY,t} & \mathbf{H}_{XX,t} \end{bmatrix} \right)$$

where \mathbf{H}_t is measurable with respect to the σ -field \mathcal{F}_{t-1} , the conditional distribution of Y given X is then

$$Y_t | \mathbf{X}_t, \mathcal{F}_{t-1} \sim \mathcal{N}(\mathbf{H}_{YX,t} \mathbf{H}_{XX,t}^{-1} \mathbf{X}_t, H_{YY,t} - \mathbf{H}_{YX,t} \mathbf{H}_{XX,t}^{-1} \mathbf{H}_{XY,t})$$

and the time-varying betas are directly obtained by

$$\beta_t = \mathbf{H}_{XX,t}^{-1} \mathbf{H}_{XY,t}.$$

It thus suffices to choose a specification of \mathbf{H}_t to recover time-varying slope coefficients. Although easy to implement, the method suffers numerous caveats. The dynamic parameter of interests are not directly modeled but obtained by a transformation of the conditional covariance matrix, rendering tractability and financial interpretation difficult. More problematic, the obtained betas are often found to be highly volatile as noted by Grassi and Violante (2021). To circumvent these issues, Darolles et al. (2018) introduced the Cholesky-GARCH model allowing to directly model the conditional betas in a multivariate GARCH framework. Chapter 3 presents an ARCH(∞) extension to this model. However, it requires estimating the full multivariate system even if only one equation is of interest, which can prove cumbersome in an asset pricing exercise on a large panel. To mitigate this shortcoming of the Cholesky-GARCH model, Grassi and Violante (2021) propose a Block-Cholesky GARCH model, based on Block-Cholesky decomposition between the set of factors and the set of individual assets. Although better suited to an empirical asset pricing exercise, the model imposes the constancy of the risk premia, which can be restrictive. As an alternative, Chapter 4 proposes to leverage the recently introduced Autoregressive Conditional Beta model (ACB) of Blasques et al. (2022a).

1.2.3 Main results

Before exposing the main results of the thesis, we specify the notation that will generally be used in the rest of the manuscript to avoid confusions with the general notations used in the first subsections of the introduction. From now on, we denote by ε_t a financial returns series, centered with conditional volatility process σ_t a measurable function with respect to the σ -field $\mathcal{F}_{t-1} = \{\varepsilon_u, u < t\}$.

Chapter 2 focuses on univariate conditional volatility models. In particular, it aims at extending the classical ARCH(∞) literature by allowing for a different response of the conditional volatility process to negative or positive returns to take into account the aforementioned *leverage effect*. The proposed model builds upon the popular and very general Asymmetric Power GARCH (APARCH(p, q)) of Ding et al. (1993), where the conditional volatility process follows

$$\sigma_t^\delta = \omega + \sum_{i=1}^q \alpha_i^+ |\varepsilon_{t-i}|^\delta \mathbf{1}_{\varepsilon_{t-i} \geq 0} + \alpha_i^- |\varepsilon_{t-i}|^\delta \mathbf{1}_{\varepsilon_{t-i} < 0} + \sum_{j=1}^p \beta_j \sigma_{t-j}^\delta$$

with $\omega > 0$, and where the coefficients are positive constants and δ is a strictly positive constant. In particular, we define a Asymmetric Power ARCH(∞) process as follows

$$\begin{cases} \varepsilon_t = \sigma_t \eta_t \\ \sigma_t^\delta = \omega + \sum_{i=1}^{\infty} \alpha_i^+ |\varepsilon_{t-i}|^\delta \mathbf{1}_{\varepsilon_{t-i} \geq 0} + \alpha_i^- |\varepsilon_{t-i}|^\delta \mathbf{1}_{\varepsilon_{t-i} < 0} \end{cases}$$

where (η_t) is an iid sequence of centered random variables with unit variance, $\omega > 0$, $\delta > 0$, and α_i^+ and α_i^- , $i = 1, \dots$, are sequences of positive constants. Interestingly, APARCH(∞) processes nest numerous short memory models. For example, the symmetric standard GARCH(1,1) admit an APARCH(∞) representation from Equation (1.3). By analogous computations, one can show that the Threshold-GARCH of Zakoian (1994) or the GJR-GARCH of Glosten et al. (1993) admits the following APARCH(∞) representation

$$\begin{cases} \varepsilon_t = \sigma_t \eta_t \\ \sigma_t^\delta = \frac{\omega}{1-\beta} + \sum_{i=1}^{\infty} \beta^{i-1} (\alpha^+ \mathbf{1}_{\varepsilon_{t-i} \geq 0} + \alpha^- \mathbf{1}_{\varepsilon_{t-i} < 0}) |\varepsilon_{t-i}|^\delta \end{cases}$$

where $\delta = 1$ and 2 respectively. Similarly to the GARCH model, the exponential decay of the coefficients sequences in these representations entails short memory of the volatility process.

The first theoretical result of Chapter 2 addresses the existence of a stationary solution of the proposed model. For any $s > 0$, let

$$A_s^+ = \sum_{i=1}^{\infty} (\alpha_i^+)^s, \quad A_s^- = \sum_{i=1}^{\infty} (\alpha_i^-)^s$$

and

$$\mu_{\delta s}^+ = \mathbb{E}[\mathbf{1}_{\eta_t \geq 0} |\eta_t|^{\delta s}], \quad \mu_{\delta s}^- = \mathbb{E}[\mathbf{1}_{\eta_t < 0} |\eta_t|^{\delta s}].$$

Theorem 2.1 establishes that if there exists $s \in (0, 1]$ such that $A_s^+ \mu_{\delta s}^+ + A_s^- \mu_{\delta s}^- < 1$, then there exists a unique, strictly stationary, ergodic, and nonanticipative solution of the APARCH(∞) equation such that $\mathbb{E}|\varepsilon_t|^{\delta s} < \infty$. The proof is based on a Volterra expansion and, in this sense, follows the work of Giraitis et al. (2000), Kazakevičius and Leipus (2002), and Douc et al. (2008).

In particular, the solution takes the form of the following Volterra series

$$\begin{cases} \varepsilon_t = \sigma_t \eta_t \\ \sigma_t^\delta = \omega + \omega \sum_{k=1}^{\infty} \sum_{i_1, \dots, i_k \geq 1} a_{i_1, t-i_1} \dots a_{i_k, t-i_1-\dots-i_k} |\eta_{t-i_1}|^\delta \dots |\eta_{t-i_1-\dots-i_k}|^\delta \end{cases}$$

with $a_{i, t-j} = \alpha_i^+ \mathbb{1}_{\eta_{t-j} \geq 0} + \alpha_i^- \mathbb{1}_{\eta_{t-j} < 0}$. Interestingly, the stationarity condition established in Theorem 2.1 reduces to the stationary condition (1.5) in the ARCH(∞) case where $\delta = 2$ and $A_s^+ = A_s^- = A_s$ since $\mu_{2s}^+ + \mu_{2s}^- = \mu_{2s}$.

On the statistical front, this chapter addresses the QML estimation of the model. Asymptotic results for the QMLE were first established in the context of ARCH(∞) models by Robinson and Zaffaroni (2006). Building upon the aforementioned authors, Chapter 2 introduces the following parametric form

$$\begin{cases} \varepsilon_t = \sigma_t(\boldsymbol{\theta}_0) \eta_t \\ \sigma_t^\delta(\boldsymbol{\theta}_0) = \omega_0 + \sum_{i=1}^{\infty} \alpha_i^+(\boldsymbol{\phi}_0) |\varepsilon_{t-i}|^\delta \mathbb{1}_{\varepsilon_{t-i} \geq 0} + \alpha_i^-(\boldsymbol{\phi}_0) |\varepsilon_{t-i}|^\delta \mathbb{1}_{\varepsilon_{t-i} < 0} \end{cases}$$

where $\alpha_i^+(\cdot), \alpha_i^-(\cdot) : \Phi \rightarrow [0, \infty]$ are known functions and $\boldsymbol{\phi}_0$ is a $r \times 1$ unknown vector of parameters, and ω_0 is an unknown positive constant. We wish to estimate $\boldsymbol{\theta}_0 = (\omega_0, \boldsymbol{\phi}_0)'$ over a parameter space Θ , on the basis of n observations $\varepsilon_1, \dots, \varepsilon_n$.

Theorem 2.2 states the strong consistency of the QMLE, denoted $\tilde{\boldsymbol{\theta}}_n$. To derive the consistency of the estimator, the assumptions presented in Chapter 2 are mostly in line with the ones proposed by Robinson and Zaffaroni (2006). However, the assumptions on η_t are noticeably milder as we do not specify a particular restriction on the distribution as opposed to Robinson and Zaffaroni (2006) in which the probability density function of η_t is assumed to be well behaved near 0. Furthermore, our assumptions on α_i are also milder as we allow our coefficients to be equal to 0 and do not impose $\alpha_i(\boldsymbol{\phi}_0) \leq K \alpha_j(\boldsymbol{\phi}_0)$ for $i \geq j \geq 1$.

The asymptotic normality of the QMLE is established in Theorem 2.3 stating

$$\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, (\kappa_\eta - 1) \mathbf{J}^{-1})$$

where

$$\kappa_\eta = \mathbb{E}(\eta_t^4) \quad \text{and} \quad \mathbf{J} = \frac{4}{\delta^2} \mathbb{E}_{\boldsymbol{\theta}_0} \left[\frac{1}{\sigma_t^{2\delta}(\boldsymbol{\theta}_0)} \frac{\partial \sigma_t^\delta(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \frac{\partial \sigma_t^\delta(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} \right]$$

is a positive definite matrix.

The structure of the proofs of the theorem follows the ones for the earlier results in the classical GARCH case and is similar to the one presented in Francq and Zakoian (2019). However, the demonstrations of the intermediate results differ and, noticeably, the impact of the initial values requires a particular attention. Asymptotic irrelevance of the initial values is proved using a

similar argument as in Robinson and Zaffaroni (2006) and rest upon the rate of decay of the α_i functions detailed in Assumptions **A3(ii)** and **A8**. A particular difficulty in the setting of this chapter is that the existence of moments for the ratio $\sigma_t^\delta(\boldsymbol{\theta}_0)/\sigma_t^\delta(\boldsymbol{\theta})$ is not ensured as the $\alpha_i^{+(-)}$ functions are allowed to be equal to zero for some i .

The asymptotic distribution of $\tilde{\boldsymbol{\theta}}_n$ has been obtained under the assumption that the true parameter values belong to the interior of the parameter space Θ . This assumption is too restrictive when some parameters are on the boundary. Theorem 2.4 generalizes the results in Theorem 2.3 by establishing the asymptotic distribution of $\tilde{\boldsymbol{\theta}}_n$ when $\boldsymbol{\theta}_0$ may be on the boundary. In such case, similarly to Francq and Zakoian (2007), the asymptotic distribution is the orthogonal projection of a normal vector distribution onto a convex cone. It is worth emphasizing that the asymptotic distribution of the QMLE is obtained without any additional assumption on the moments of ε_t .

In addition to existence and inference results, Chapter 2 introduces test procedures. In particular, the asymptotic distribution of a Portmanteau goodness-of-fit test is established in Theorem 2.6. Since their introduction by Box and Pierce (1970), Portmanteau tests, have become widely used in to test the adequacy of econometric specifications and were extended to volatility models by Li and Mak (1994). Asymptotic properties of these tests have been established by Berkes et al. (2003a) for standard GARCH(p, q) models and by Carbon and Francq (2011) in the APARCH(p, q) case. To our best knowledge, these results had not yet been extended to the ARCH(∞) literature. Let us consider the null hypothesis of goodness-of-fit H_0^{GoF} that a process (ε_t) satisfies the APARCH(∞) model. The empirical autocovariance at lag $h \leq n - 1$ of the squared residuals is given by

$$\hat{r}_h = n^{-1} \sum_{t=h+1}^n (\hat{\eta}_t^2 - 1)(\hat{\eta}_{t-h}^2 - 1), \text{ with } \hat{\eta}_t^2 = \varepsilon_t^2 / \tilde{\sigma}_t^2(\tilde{\boldsymbol{\theta}}_n),$$

and let $\hat{\boldsymbol{r}}_m = (\hat{r}_1, \dots, \hat{r}_m)$ for any $1 \leq m \leq n$, and $\hat{\boldsymbol{C}}_m$ the $m \times (r + 1)$ matrix whose elements (h, k) are given by

$$\hat{C}_m(h, k) = -\frac{2}{\delta n} \sum_{t=h+1}^n (\hat{\eta}_{t-h}^2 - 1) \frac{1}{\tilde{\sigma}_t^\delta(\tilde{\boldsymbol{\theta}}_n)} \frac{\partial \tilde{\sigma}_t^\delta(\tilde{\boldsymbol{\theta}}_n)}{\partial \theta_k}.$$

Letting

$$\hat{\boldsymbol{J}}_n = \frac{4}{\delta^2} \frac{1}{n} \sum_{t=1}^n \frac{1}{\tilde{\sigma}_t^{2\delta}(\tilde{\boldsymbol{\theta}}_n)} \frac{\partial \tilde{\sigma}_t^\delta(\tilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}} \frac{\partial \tilde{\sigma}_t^\delta(\tilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}'} \text{ and } \hat{\kappa}_\eta = \frac{1}{n} \sum_{t=1}^n \frac{\varepsilon_t^4}{\tilde{\sigma}_t^4(\tilde{\boldsymbol{\theta}}_n)}$$

be consistent estimators of \boldsymbol{J} and κ_η , the Portmanteau test statistic is then given by

$$n \hat{\boldsymbol{r}}_m' \hat{\boldsymbol{D}}^{-1} \hat{\boldsymbol{r}}_m \text{ with } \hat{\boldsymbol{D}} = (\hat{\kappa}_\eta - 1)^2 \boldsymbol{I}_m - (\hat{\kappa}_\eta - 1) \hat{\boldsymbol{C}}_m \hat{\boldsymbol{J}}_n^{-1} \hat{\boldsymbol{C}}_m'$$

and its χ_m^2 asymptotic distribution is established in Theorem 2.6.

Remarkably, in APARCH(∞) models, the symmetry hypothesis is generally a particular constrained representation. Testing for the significance of asymmetry can thus be achieved by testing an implied restriction on $\boldsymbol{\theta}_0$. Chapter 2 presents two tests for linear constraints on the parameters. Consider a general hypothesis of the form

$$H_0 : \mathbf{R} \boldsymbol{\theta}_0 = \mathbf{k}, \quad H_1 : \mathbf{R} \boldsymbol{\theta}_0 \neq \mathbf{k}$$

where \mathbf{R} is the constraints matrix and \mathbf{k} is a constant vector. Let c be the rank of the matrix \mathbf{R} . Proposition 2.1 establishes the asymptotic distribution of the triptych of the standard Wald, Rao-score, and Quasi Likelihood Ratio (LR) statistics, allowing the parameters of interest to be on the boundary of the parameter space. Similar constrained representations are highly common when testing for asymmetry in parametric models, see for example Nelson (1991).

The second test for linear constraints studied in Chapter 2 is more challenging. For simplicity, consider the following symmetric ARCH(∞) representation

$$\begin{cases} \varepsilon_t = \sigma_t(\boldsymbol{\theta}_0)\eta_t \\ \sigma_t^2(\boldsymbol{\theta}_0) = \frac{\omega_0}{1 - \beta_0} + \sum_{i=1}^{\infty} (\alpha_0\beta_0^{i-1} + \gamma_0 i^{-(d_0+1)})\varepsilon_{t-i}^2 \end{cases}$$

with $\alpha_0 > 0$, $\beta_0 > 0$, $\gamma_0 \geq 0$, and $d_0 > 0$. From Equation (1.4), this dynamic reduces to the GARCH(1,1) if the parameter γ_0 is null. Testing the validity of a GARCH(1,1) representation can then be achieved by testing

$$H_0^{\text{GARCH}} : \gamma_0 = 0, \quad H_1^{\text{ARCH}(\infty)} : \gamma_0 > 0.$$

While this test may seem standard, it poses a major difficulty. Indeed, the parameter d_0 is not identified under the null hypothesis, thus we cannot directly use Proposition 2.1(ii) to obtain the asymptotic distribution of the test statistics when the parameter lies on the boundary. Proposition 2.1 establishes the distribution of the Wald statistic when the parameter d_0 is assumed known and fixed at a value $\bar{d} > 0.5$. Under this assumption, the Wald statistic distribution is a mixture of a χ_1^2 and a Dirac measure at 0. As an alternative, we introduce a recursive residual-based bootstrap procedure to approximate the Wald statistic asymptotic distribution without fixing the value of d_0 . The validity of this bootstrap is confirmed through Monte Carlo experiments.

Chapter 2 concludes with an extensive application of the proposed tests on a set of various international equity indices daily returns over the period of January 1995 to December 2020. Unsurprisingly, the vast majority of indices reject the symmetry assumption, which is a classical result in the financial literature. Less expected is the fact that almost half of the thirty indices reject the hypothesis of a GARCH(1,1) specification, and eight reject the GJR-GARCH(1,1) or the TGARCH(1,1) model at the 5% level. Remarkably, all the indices that reject the hypothesis of short memory are from emerging markets. This suggests that the level of development of a financial market has implications on the persistence patterns exhibited by its assets. A

possible explanation stems from the difficulty to invest in peripheral markets with fewer investors and with less liquid instruments, which leads to a slower integration of shocks and ultimately higher persistence. Additionally, an out-of-sample exercise illustrates the usefulness of the APARCH(∞) model when computing daily one-day ahead forecasts for the 95%-Value-at-Risk.

Taking into account the higher persistence of the conditional volatility process has proved valuable in Chapter 2. However, the setting considered does not allow the modeling of multivariate systems. Probabilistic and statistical aspects of multivariate ARCH(∞) have rarely been studied in the academic literature. To that end, Chapter 3 leverages the Cholesky-GARCH model of Darolles et al. (2018). More precisely, let $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, \dots, \varepsilon_{m,t})'$ a vector of $m \geq 2$ return series verifying a general multivariate volatility model

$$\boldsymbol{\varepsilon}_t = \boldsymbol{\Sigma}_t^{1/2} \boldsymbol{\eta}_t$$

where $(\boldsymbol{\eta}_t)$ is iid $(\mathbf{0}, \mathbf{I}_m)$. Pourahmadi (1999) Cholesky decomposition of the conditional covariance matrix yields

$$\boldsymbol{\Sigma}_t = \mathbf{L}_t \mathbf{G}_t \mathbf{L}_t'$$

where $\mathbf{G}_t = \text{diag}(\mathbf{g}_t)$ is a diagonal matrix, and \mathbf{L}_t is a lower unitriangular matrix. Let $\mathbf{v}_t = (v_{1,t}, \dots, v_{m,t})'$ with $v_{k,t} = \sqrt{g_{k,t}} \eta_{k,t}$ for $k = 1, \dots, m$, be the recursively obtained orthogonal factors from $\boldsymbol{\varepsilon}_t$. By taking $\boldsymbol{\Sigma}_t^{1/2} = \mathbf{L}_t \mathbf{G}_t^{1/2}$, we obtain

$$\boldsymbol{\varepsilon}_t = \boldsymbol{\Sigma}_t^{1/2} \boldsymbol{\eta}_t = \mathbf{L}_t \mathbf{v}_t$$

and thus $\mathbf{B}_t \boldsymbol{\varepsilon}_t = \mathbf{v}_t$ where $\mathbf{B}_t = \mathbf{L}_t^{-1}$. Let $\ell_{ij,t}$ (resp. $-\beta_{ij,t}$) denote the row i and column j element of \mathbf{L}_t (resp. \mathbf{B}_t). Thus, the Cholesky-GARCH model allows a direct joint modeling of the dynamic conditional betas and the variance of the conditional orthogonal factors $v_{k,t}$. In this model, the conditional volatilities \mathbf{g}_t of the orthogonal factors are usually modeled through a multivariate GARCH equation, such as the Extended-CCC GARCH model of Jeantheau (1998) where

$$\mathbf{g}_t = \boldsymbol{\omega} + \sum_{i=1}^q \mathbf{A}_i \mathbf{v}_{t-i}^2 + \sum_{j=1}^p \mathbf{b}_j \mathbf{g}_{t-j} \quad (1.6)$$

with $\mathbf{v}_t^2 = (v_{1,t}^2, \dots, v_{m,t}^2)$, and the vector $\boldsymbol{\omega} \in \mathbb{R}^m$ and the $m \times m$ matrices \mathbf{A}_i and \mathbf{b}_j have nonnegative entries. However, similarly to the univariate GARCH model, under that structure the Cholesky-GARCH cannot account for the strong persistence of some financial returns series. To alleviate the short-memory feature of model (1.6) we consider a multivariate ARCH(∞) representation of the form

$$\mathbf{g}_t = \boldsymbol{\omega} + \sum_{i=1}^{\infty} \mathbf{A}_i \mathbf{v}_{t-i}^2 + \sum_{i=1}^{\infty} \boldsymbol{\Pi}_i \mathbf{X}_{t-i} \quad (1.7)$$

where $\mathbf{v}_t^2 = (v_{1,t}^2, \dots, v_{m,t}^2)$, and the vector $\boldsymbol{\omega} \in \mathbb{R}^m$ and the $m \times m$ matrices \mathbf{A}_i and $m \times P$ matrices $\boldsymbol{\Pi}_i$, $i \in \{1, 2, \dots\}$ have nonnegative entries. The vector \mathbf{X}_t contains P positive exogenous

variables such that $(\boldsymbol{\eta}'_t, \mathbf{X}'_t)'$ is stationary and ergodic, with $\boldsymbol{\eta}_t$ independent of $\{\mathbf{X}_{t-i}, i > 0\}$. The addition of covariates has proven useful in both univariate (see e.g. Francq and Thieu (2019)) and multivariate conditional (co-)variances (see e.g. Bollerslev et al. (2020b)), but to our knowledge it had never been studied in the ARCH(∞) literature. Note that we do not prevent the covariates from being correlated together as well as correlated with the innovations process. The exogenous variables are thus neither weakly nor strongly exogenous in the sense of Engle et al. (1983), but exogenous in the sense that their dynamics are not specified by the Cholesky-ARCH(∞) model.

To complete the definition of the Cholesky-ARCH(∞) model, the dynamic of the conditional betas has to be specified and we propose a process of the general form

$$\boldsymbol{\beta}_t = \mathbf{c}(\mathbf{v}_{t-1}, \dots, \mathbf{v}_{t-q}, \mathbf{g}_{t-1}^{1/2}, \dots, \mathbf{g}_{t-q}^{1/2}, \mathbf{X}_{t-1}, \dots, \mathbf{X}_{t-q}) + \sum_{k=1}^p \mathbf{C}_k \boldsymbol{\beta}_{t-k}, \quad (1.8)$$

for two integers p and q , where \mathbf{c} is any measurable function from $\mathbb{R}^{(2m+P)q}$ to $\mathbb{R}^{m(m-1)/2}$.

The theoretical results of Chapter 3 start with the probabilistic properties of the model. Theorem 3.1 establishes mild conditions for the existence of a unique strictly stationary, ergodic and nonanticipative solution to the Cholesky-ARCH(∞) model, extending the results of Doukhan et al. (2006). In particular, the conditions proposed in the aforementioned contribution depend on the choice of a multiplicative matrix norm, which is not the case in Theorem 3.1. An additional necessary condition for the finiteness of moments of the returns process $(\boldsymbol{\varepsilon}_t)$ is provided in Proposition 3.1. Chapter 3 then presents statistical inference results. Asymptotic properties of the QMLE are derived and presented in Theorem 3.2 showing the consistency and asymptotic normality of the estimator.

An interesting feature of the proposed multivariate ARCH(∞) defined by the couple of Equations (1.7)-(1.8) is its versatility. In particular, building upon Chapter 2, a potential specification for the individual conditional variances is given by

$$g_{k,t} = \omega + \underbrace{\sum_{i=1}^{\infty} (\alpha_k \beta_k^{i-1} + \gamma_k i^{-d-1}) v_{k,t-i}^2}_{\text{specific ARCH}(\infty) \text{ effect}} + \underbrace{\sum_{j \neq k} \delta_{k,j} v_{j,t-1}^2}_{\text{spillover}} + \underbrace{\boldsymbol{\pi}'_k \mathbf{X}_{t-1}}_{\text{exogeneity}}$$

which nests the Cholesky-GARCH model as well as the diagonal-ARCH(∞) model. Popular covariates in the conditional volatility equations include daily traded volumes, implied volatility measures obtained from quoted options on financial exchanges, as well as realized volatility measures computed using high-frequency intraday data (see for example Engle and Patton (2001) and Francq and Thieu (2019)). The selection of exogenous variables in the conditional betas equation has however never been discussed. Building upon the realized volatility literature, we propose to use realized betas measures. Realized betas, first introduced by Barndorff-Nielsen and Shephard (2004) and Andersen et al. (2006), are computed using observed returns over

fixed time intervals. Let $\boldsymbol{\varepsilon}_{w,t} = (\varepsilon_{w,1,t}, \dots, \varepsilon_{w,m,t})'$ the vector of returns over the time interval w . The realized covariance matrix $\mathbf{C}_t^{w,W}$ is then defined as the sum of the W outer products of the return vectors. For example, w could be a day and W a month, or w could be a 5-minute interval and W a day. For ease of notation, we denote the realized covariance matrix

$$\mathbf{C}_t := \mathbf{C}_t^{w,W} = \sum_{w=1}^W \boldsymbol{\varepsilon}_{w,t} \boldsymbol{\varepsilon}_{w,t}'$$

In addition, for $1 < k \leq m$, for any $m \times m$ square matrix \mathbf{M} , let $\mathbf{M}_{[kk]}$ denote the square sub-matrix formed by the first $k-1$ rows and columns and $\mathbf{M}_{[k]}$ the sub-vector formed by the first $k-1$ rows and the k -th column of \mathbf{M} . For $1 < i \leq m$, the realized betas between asset i and assets $j = 1, \dots, i-1$ are then defined by

$$\bar{\boldsymbol{\beta}}_{i,t} = (\bar{\beta}_{ij,t})_{j=1, \dots, i-1} = (\mathbf{C}_{t[ii]})^{-1} \mathbf{C}_{t[i]}.$$

In the spirit of the Heterogeneous Autoregressive model of Realized Volatility (HAR-RV) introduced by Corsi (2009), the daily conditional beta could be depending on realized beta measures computed on lower frequencies

$$\beta_{ij,t} = \varpi_{ij} + \pi_{ij}^{\mathcal{W}} \bar{\beta}_{ij,t-1}^{\mathcal{W}} + \pi_{ij}^{\mathcal{M}} \bar{\beta}_{ij,t-1}^{\mathcal{M}} + c_{ij} \beta_{ij,t-1}$$

where $\bar{\beta}_{ij,t}^{\mathcal{W}}$ denotes the weekly realized beta at time t computed over 5 trading days, and $\bar{\beta}_{ij,t}^{\mathcal{M}}$ the monthly realized beta at time t computed over 21 trading days. Alternatively, potential useful exogenous variables in the conditional betas dynamics could be the recently introduced realized semibetas of Bollerslev et al. (2022). Consider the realized covariance matrix decomposition

$$\mathbf{C}_t = \mathcal{P}_t + \mathcal{N}_t + \mathcal{S}_t^+ + \mathcal{S}_t^-$$

into the semicovariance components introduced by Bollerslev et al. (2020a)

$$\begin{aligned} \mathcal{P}_t &= \sum_{w=1}^W \boldsymbol{\varepsilon}_{w,t}^+ \boldsymbol{\varepsilon}_{w,t}^{+'}, & \mathcal{S}_t^+ &= \sum_{w=1}^W \boldsymbol{\varepsilon}_{w,t}^- \boldsymbol{\varepsilon}_{w,t}^{+'}, \\ \mathcal{N}_t &= \sum_{w=1}^W \boldsymbol{\varepsilon}_{w,t}^- \boldsymbol{\varepsilon}_{w,t}^{-'}, & \mathcal{S}_t^- &= \sum_{w=1}^W \boldsymbol{\varepsilon}_{w,t}^+ \boldsymbol{\varepsilon}_{w,t}^{-'}. \end{aligned}$$

where $\boldsymbol{\varepsilon}_{w,t}^{+(-)}$ denotes the componentwise positive (respectively negative) part of $\boldsymbol{\varepsilon}_{w,t}$. Each realized semicovariance component then yields a realized semibeta

$$\begin{aligned} \bar{\boldsymbol{\beta}}_{i,t}^{\mathcal{P}} &= (\mathbf{C}_{t[ii]})^{-1} \mathcal{P}_{t[i]}, & \bar{\boldsymbol{\beta}}_{i,t}^{\mathcal{S}^+} &= -(\mathbf{C}_{t[ii]})^{-1} \mathcal{S}_{t[i]}^+, \\ \bar{\boldsymbol{\beta}}_{i,t}^{\mathcal{N}} &= (\mathbf{C}_{t[ii]})^{-1} \mathcal{N}_{t[i]}, & \bar{\boldsymbol{\beta}}_{i,t}^{\mathcal{S}^-} &= -(\mathbf{C}_{t[ii]})^{-1} \mathcal{S}_{t[i]}^-. \end{aligned}$$

providing a four-way decomposition of the traditional realized betas

$$\bar{\boldsymbol{\beta}}_{i,t} = \bar{\boldsymbol{\beta}}_{i,t}^{\mathcal{P}} + \bar{\boldsymbol{\beta}}_{i,t}^{\mathcal{N}} - \bar{\boldsymbol{\beta}}_{i,t}^{\mathcal{S}^+} - \bar{\boldsymbol{\beta}}_{i,t}^{\mathcal{S}^-}.$$

The conditional beta dynamics could then be given by

$$\beta_{ij,t} = \varpi_{ij} + \pi_{ij}^{\mathcal{P}} \bar{\beta}_{ij,t-1}^{\mathcal{P}} + \pi_{ij}^{\mathcal{N}} \bar{\beta}_{ij,t-1}^{\mathcal{N}} + \pi_{ij}^{\mathcal{S}^+} \bar{\beta}_{ij,t-1}^{\mathcal{S}^+} + \pi_{ij}^{\mathcal{S}^-} \bar{\beta}_{ij,t-1}^{\mathcal{S}^-} + c_{ij} \beta_{ij,t-1} \quad (1.9)$$

allowing for a more accurate contribution of downside and upside risk.

Remarkably, Equation (1.9) reduces to

$$\beta_{ij,t} = \varpi_{ij} + \pi_{ij}^{\mathcal{P}} \bar{\beta}_{ij,t} + c_{ij} \beta_{ij,t}$$

if $\pi_{ij}^{\mathcal{P}} = \pi_{ij}^{\mathcal{N}} = -\pi_{ij}^{\mathcal{S}^+} = -\pi_{ij}^{\mathcal{S}^-}$. Testing the informational benefit of including realized semibetas instead of standard realized betas can thus be achieved by testing

$$\begin{cases} H_0 : \pi_{ij}^{\mathcal{P}} = \pi_{ij}^{\mathcal{N}} & \text{and} & \pi_{ij}^{\mathcal{P}} = -\pi_{ij}^{\mathcal{S}^+} & \text{and} & \pi_{ij}^{\mathcal{N}} = -\pi_{ij}^{\mathcal{S}^-} \\ H_1 : \pi_{ij}^{\mathcal{P}} \neq \pi_{ij}^{\mathcal{N}} & \text{or} & \pi_{ij}^{\mathcal{P}} \neq -\pi_{ij}^{\mathcal{S}^+} & \text{or} & \pi_{ij}^{\mathcal{N}} \neq -\pi_{ij}^{\mathcal{S}^-} \end{cases} .$$

The application conducted in Chapter 3 provides a strong argument in favor the realized semi-betas decomposition. Indeed, for only six out of the thirty-five industry portfolios available on Kenneth French website the hypothesis H_0 cannot be rejected at the 5% level.

Although it allows for a direct modeling of the conditional beta combined with a flexible decay pattern of the conditional covariance coefficients, the structure of the Cholesky-GARCH, and *a fortiori* the structure of the Cholesky-ARCH(∞) renders its use cumbersome in large panel asset pricing exercises. In a recent paper, Blasques et al. (2022a) propose another approach to obtain dynamic conditional betas. In particular, their Autoregressive Conditional Beta (ACB) model builds upon the Score-Driven models literature introduced by Harvey and Chakravarty (2008) and Creal et al. (2013) to allow for both time-varying regression coefficients and conditional heteroskedasticity of the factors. Score-Driven models, also known as Generalized Autoregressive Score models (GAS) were introduced by Harvey and Chakravarty (2008) and Creal et al. (2013) and aim at generalizing models with time-varying parameters. Consider an observable process y_t and assume it follows the conditional density $p(y_t|\psi_t, \Omega_t, \theta)$ where ψ_t is the time varying parameter of interest, θ a constant parameter, and Ω_t the information set available at time t . The updating equation of the dynamic parameter is given by

$$\psi_{t+1} = \bar{\omega} + \xi \underbrace{S(\psi_t) \frac{\partial \log p(y_t|\psi_t, \Omega_t, \theta)}{\partial \psi_t}}_{\text{updating term}} + c \psi_t \quad (1.10)$$

where $\bar{\omega}$, ξ and c are unknown parameters to be estimated. The updating term is decomposed into the score of the likelihood with respect of the time-varying parameter and a scaling term $S(\psi_t)$. The dynamics of the parameter of interest is thus driven by the scaled score. Interestingly, numerous econometric models can be expressed as a Score-Driven model. For example, if we assume that the observation conditional density is Gaussian and the parameter of interest is the time-varying variance, under an appropriate choice of the scaling function, Equation (1.10)

yields the standard GARCH(1,1) model.

Let us denote $(r_{i,t})$ the excess returns time series of asset $i = 1, \dots, N$ in a system of N assets and $\mathbf{f}_t = (f_{1,t}, \dots, f_{m,t})'$ a set of m observable factors. We consider a conditional regression model of the form

$$r_{i,t} = \alpha_{i,t} + \boldsymbol{\beta}_{i,t}' \mathbf{f}_t + \varepsilon_{i,t}. \quad (1.11)$$

Additionally, assume that each factor f_j dynamic is given by a GARCH(1,1) with constant conditional mean

$$\begin{aligned} f_{j,t} &= \mu_j + \sigma_{j,t} \eta_{j,t} \\ \sigma_{j,t}^2 &= \omega_j + a_j (f_{j,t-1} - \mu_j)^2 + b_j \sigma_{j,t-1}^2 \end{aligned} \quad (1.12)$$

where $\eta_{j,t}$ is an iid centered random variable with unit variance. Similarly, assume that the residuals of the conditional regression follows a GARCH(1,1) with conditional variance

$$g_{i,t}^2 = \omega_{\varepsilon_i} + a_{\varepsilon_i} (r_{i,t-1} - \alpha_{i,t-1} - \boldsymbol{\beta}_{i,t-1}' \mathbf{f}_{t-1})^2 + b_{\varepsilon_i} g_{i,t-1}^2. \quad (1.13)$$

Assuming for all i , $\varepsilon_{i,t} = g_{i,t} \zeta_{i,t}$ where $\zeta_{i,t}$ are iid Gaussian, Blasques et al. (2022a) derive the updating equations for the conditional regression coefficients

$$\begin{aligned} \alpha_{i,t+1} &= \bar{\omega}_{\alpha_i} + \xi_{\alpha_i} \varepsilon_{i,t} + c_{\alpha_i} \alpha_{i,t} \\ \beta_{i,j,t+1} &= \bar{\omega}_{i,j} + \xi_{i,j} \frac{f_{j,t} \varepsilon_{i,t}}{\mu_j^2 + \sigma_{j,t}^2} + c_{i,j} \beta_{i,j,t} \end{aligned} \quad (1.14)$$

for $i = 1, \dots, N$ and $j = 1, \dots, m$. The authors establish the consistence and asymptotic normality of a multi-step QMLE. Chapter 4 leverages the ACB model to develop a conditional asset pricing model and to derive an estimation procedure for the dynamic risk premia associated to the set of factors \mathbf{f}_t . First, the asymptotic distributions of standard test statistics are derived to assess the constancy of a parameter in the conditional regression. These statistics noticeably allow to test the relevance of a given factor. In the case of an unconditional regression model, a factor is said to be significant if the beta associated to this factor is significantly different from zero. We thus propose to assess a factor j significance by testing that its conditional beta is both constant and equal to zero. From (1.14), constancy of $\beta_{i,j,t}$ is obtained when $\xi_{0ij} = 0$. However, testing for this constraint on the true parameter vector $\boldsymbol{\theta}_0^{(i)}$ is difficult. Indeed, under $H_{0,ij} : \xi_{0ij} = 0$, the conditional beta $\beta_{i,j,t}$ tends to $\bar{\omega}_{ij}/(1 - c_{ij})$ and there exists an infinity of pairs $(\bar{\omega}_{ij}, c_{ij})$ such that $\beta_{i,j,t} = \bar{\omega}_{0,ij}$, resulting in a non identifiable model under the null hypothesis. A similar problem occurs when testing for the significance of a factor in the conditional regression, which is equivalent to $\xi_{0ij} = 0$ and $\bar{\omega}_{i,j} = 0$. To circumvent this identification problem, we set the parameter c_{0ij} at an arbitrary value \bar{c}_{ij} . This solves the identification issue and we can now test $H_{0,ij}(\bar{c}_{ij}) : \mathbf{R} \boldsymbol{\theta}_0^{(i)} = 0$ using standard techniques where \mathbf{R} is the constraint matrix.

Theorem 4.1 establishes that, under $H_{0,ij}(\bar{c}_{ij})$, the Wald statistic W_n and Rao-score test statistic R_n asymptotically follow a χ_r^2 distribution and their critical regions at the asymptotic level ν

are given by

$$\{W_n > \chi_r^2(1 - \nu)\}, \{R_n > \chi_r^2(1 - \nu)\}.$$

However, the Quasi Likelihood Ratio statistic has a different asymptotic distribution than the standard χ^2 distribution of the Wald and Rao-score statistics. The critical region at the asymptotic level ν for L_n can however be easily obtained following the method of Imhof (1961). In addition, a residual-based bootstrap procedure is proposed to approximate the statistics asymptotic distribution. Theorem 4.2 establishes the validity of the bootstrap scheme and ensures that the distribution of the bootstrap Wald test statistic given $(r_{i,t}, \mathbf{f}_t)$ well mimics the unconditional distribution of the Wald statistic.

Chapter 4 then details the estimation procedure to recover the time-varying risk premia. The procedure builds upon Gagliardini et al. (2016) who derive no-arbitrage restrictions for a multi-period economy with a continuum of assets. In particular, they show that the asset pricing restriction is such that, for any t , there exists a unique random vector $\mathbf{v}_t = (v_{1,t}, \dots, v_{m,t})'$ measurable with respect to the natural filtration \mathcal{F}_{t-1} such that for almost all i ,

$$\alpha_{i,t} = \beta_{i,t}' \mathbf{v}_t \text{ almost surely.} \quad (1.15)$$

The dynamic asset pricing model (1.11) combined with (1.15) yields that for almost all i ,

$$\mathbb{E}[r_{i,t} | \mathcal{F}_{t-1}] = \beta_{i,t}' \boldsymbol{\lambda}_t \text{ almost surely,} \quad (1.16)$$

where $\boldsymbol{\lambda}_t$, the vector of time-varying risk premia, is given by $\boldsymbol{\lambda}_t = \mathbf{v}_t + \boldsymbol{\mu}$ with $\boldsymbol{\mu} = \mathbb{E}[\mathbf{f}_t] = (\mu_1, \dots, \mu_m)$. Note that Equation (1.16) nests favored asset pricing models with time-invariant coefficients. For example, in the CAPM where the only factor is the Market portfolio excess returns or in the three factor model of Fama and French (1993), we have $\mathbf{v}_t = 0$.

Building upon the asset pricing restriction (1.15), we propose the following two step procedure to recover time-varying risk premia from the Score-Driven conditional regression model.

1. For each asset i in the system, $i = 1, \dots, N$, we estimate $\hat{\alpha}_{i,t} := \hat{\alpha}_{i,t}(\hat{\boldsymbol{\varphi}}_n^{(i)})$ and $\hat{\boldsymbol{\beta}}_{i,t} := \hat{\boldsymbol{\beta}}_{i,t}(\hat{\boldsymbol{\varphi}}_n^{(i)})$ where $\hat{\boldsymbol{\varphi}}_n^{(i)}$ denotes the multi-step QMLE of Blasques et al. (2022a)
2. For each t , we estimate $\hat{\mathbf{v}}_t = \arg \min_{\mathbf{v}} \left\| \hat{\boldsymbol{\beta}}_t' \mathbf{v} - \hat{\boldsymbol{\alpha}}_t \right\|_2$ by OLS, where $\hat{\boldsymbol{\beta}}_t = (\hat{\boldsymbol{\beta}}_{i,t})_{i=1, \dots, N}$ and $\hat{\boldsymbol{\alpha}}_t = (\hat{\alpha}_{i,t})_{i=1, \dots, N}$.

Combining $\hat{\mathbf{v}}_t$ with parameter $\hat{\boldsymbol{\mu}}_n$ estimated in $\hat{\boldsymbol{\theta}}_n$ yields the estimation of the time-varying risk premia $\hat{\boldsymbol{\lambda}}_t = \hat{\boldsymbol{\mu}}_n + \hat{\mathbf{v}}_t$.

Chapter 4 concludes by an application aiming at assessing the existence of a time-varying risk premium associated to a factor related to Carbon risk. We find a significant dynamic premium and discuss the relationship between its variations and news related to sustainable finance.

Chapter 2

Conditional asymmetry in Power ARCH(∞) models¹

Abstract: We consider an extension of ARCH(∞) models to account for conditional asymmetry in the presence of high persistence. After stating existence and stationarity conditions, this paper develops the statistical inference of such models and proves the consistency and asymptotic distribution of a Quasi Maximum Likelihood estimator. Some particular specifications are studied and we introduce a Portmanteau goodness-of-fit test. Additionally, test procedures for asymmetry and GARCH validity are derived. Finally, we present an application on a set of equity indices to reexamine the preeminence of GARCH(1,1) specifications. We find strong evidence that the short memory feature of such models is not suitable for peripheral assets.

Keywords: Quasi Maximum Likelihood Estimation, Moderate memory, Testing parameters on the boundary, Recursive design bootstrap

¹A shorter version of this chapter is forthcoming in the Journal of Econometrics and can be found via the following link <https://doi.org/10.1016/j.jeconom.2021.10.013>.

2.1 Introduction

Empirical autocorrelations of squared financial returns are often significantly positive and decay slowly. The ARCH(∞) model, introduced by Robinson (1991) as an extension of the standard GARCH(p, q) model, allows for a flexible form of decay of the coefficients in the volatility equation. In particular, by relaxing the exponential decrease of autocorrelations of squared GARCH processes, ARCH(∞) are suitable for series with longer memory. In these models, financial returns (ε_t) and volatilities write as

$$\begin{cases} \varepsilon_t = \sigma_t \eta_t, & (\eta_t) \text{ iid } (0, 1) \\ \sigma_t^2 = \omega + \sum_{i=1}^{\infty} \alpha_i \varepsilon_{t-i}^2 \end{cases} \quad (2.1)$$

with $\omega > 0$, and $\alpha_i, i = 1, \dots$, a sequence of nonnegative constants. The existence of a strictly stationary and nonanticipative solution has been proved by Giraitis et al. (2000), Kazakevičius and Leipus (2002), and Douc et al. (2008) under the condition

$$A_s \mu_{2s} < 1 \quad (2.2)$$

for some $s \in (0, 1]$, where $A_s = \sum_{i=1}^{\infty} \alpha_i^s$ and $\mu_{2s} = E|\eta_t|^{2s}$. Condition (2.2) entails summability of the autocovariances of ε_t^2 and thus is not compatible with the usual concept of long memory (see Giraitis et al. (2000) and Zaffaroni (2004)). However, this condition is compatible with a slow decay of the autocorrelation function of the squares and is sometimes referred to as moderate memory models².

Another well-documented empirical fact concerning stock returns is the leverage effect, namely the higher impact on current volatility of past price decreases rather than increases of the same magnitude. To the best of our knowledge, attempts to capture both the asymmetry and the memory properties of financial time series have been scarce. A noticeable exception is the fractionally integrated EGARCH model introduced by Bollerslev and Mikkelsen (1996). The estimation of such models has been particularly studied by Zaffaroni (2009).

Although long or moderate memory models are natural candidates to model financial time series, their use amongst practitioners has been regrettably limited. The aim of this paper is to buttress the use of ARCH(∞) models. More precisely, we introduce a new model aiming at capturing both the asymmetry and memory effect in financial returns. This new specification can be seen as the natural extension of the APARCH model of Ding et al. (1993) to the ARCH(∞) framework. In particular, it nests a large number of short memory volatility models, including the standard GARCH(p, q), or favored asymmetric extensions such as the GJR-GARCH of Glosten et al. (1993). We give a condition for the existence of a unique stationary solution and derive the asymptotic distribution of the Quasi Maximum Likelihood estimator (QMLE), allow-

²The question of existence of a stationary solution with a finite fourth order moment for the FIGARCH model has given rise to a long academic discussion, until the paper by Giraitis et al. (2018) in which the existence of such solution was established.

ing the parameter to belong to the boundary of the parameter space. In addition to existence and inference results, we introduce test procedures. In particular, we establish the asymptotic distribution of a Portmanteau goodness-of-fit test that had only been considered for short memory models. We also introduce a test procedure to validate GARCH(1,1)-type specifications by allowing the infinite sequence of coefficients to have a slower than exponential decay. This test having a non-standard asymptotic distribution, we propose a residual bootstrap procedure. Carrying this test on a set of equity indices, we find strong evidence that short memory models are not suitable for peripheral assets.

Our work is organized as follows. In Section 1, we introduce our new specification and establish a condition for the existence of a stationary solution. In Section 2, we focus on statistical inference. In Section 3, we present procedures to test for goodness-of-fit, asymmetry, and the adequacy of GARCH(1,1)-type specifications. Monte Carlo experiments are conducted in Section 4. Section 5 presents an application on a wide set of equity indices to reexamine the preeminence of GARCH(1,1)-type models. Finally, Section 6 concludes. Proofs and technical results are relegated to an appendix. Additional simulations and applications are available in a supplementary file.

2.2 Asymmetric Power ARCH(∞) model

Modeling asymmetry has led to the introduction of numerous specifications of the conditional volatility process. Among them, a popular and very general class of models is the Asymmetric Power GARCH (APARCH(p, q)) of Ding et al. (1993), defined by

$$\begin{cases} \varepsilon_t = \sigma_t \eta_t \\ \sigma_t^\delta = \omega + \sum_{i=1}^q (\alpha_i^+ |\varepsilon_{t-i}|^\delta \mathbf{1}_{\varepsilon_{t-i} \geq 0} + \alpha_i^- |\varepsilon_{t-i}|^\delta \mathbf{1}_{\varepsilon_{t-i} < 0}) + \sum_{j=1}^p \beta_j \sigma_{t-j}^\delta \end{cases}$$

where $\omega > 0$, the coefficients are nonnegative constants, and δ is a positive constant. We propose an ARCH(∞) extension of this model defined as follows.

Definition 2.1. *Let (η_t) be an iid sequence of random variables such that $\mathbb{E}\eta_0 = 0$ and $\mathbb{E}\eta_0^2 = 1$. Then, (ε_t) is called an APARCH(∞) process if it satisfies an equation of the form*

$$\begin{cases} \varepsilon_t = \sigma_t \eta_t \\ \sigma_t^\delta = \omega + \sum_{i=1}^{\infty} (\alpha_i^+ |\varepsilon_{t-i}|^\delta \mathbf{1}_{\varepsilon_{t-i} \geq 0} + \alpha_i^- |\varepsilon_{t-i}|^\delta \mathbf{1}_{\varepsilon_{t-i} < 0}) \end{cases} \quad (2.3)$$

with $\omega > 0$, $\delta > 0$, and where α_i^+ and α_i^- , $i = 1, \dots$, are sequences of nonnegative constants.

Note that this specification is very general and includes standard ARCH(∞) as well as the Threshold-ARCH(∞) model which corresponds to $\delta = 2$. TARARCH(∞) models were first considered by Bardet and Wintenberger (2009) as a particular example of a more general causal process.

The following theorem gives a condition for the existence of a strictly stationary and nonanticipative solution to an APARCH(∞) model defined by (2.3). For any $s > 0$, let

$$A_s^+ = \sum_{i=1}^{\infty} (\alpha_i^+)^s, \quad A_s^- = \sum_{i=1}^{\infty} (\alpha_i^-)^s \quad \text{and} \quad \mu_{\delta s}^+ = \mathbb{E}|\mathbb{1}_{\eta_t \geq 0} \eta_t|^{\delta s}, \quad \mu_{\delta s}^- = \mathbb{E}|\mathbb{1}_{\eta_t < 0} \eta_t|^{\delta s}.$$

Theorem 2.1. *If there exists $s \in (0, 1]$ such that*

$$A_s^+ \mu_{\delta s}^+ + A_s^- \mu_{\delta s}^- < 1, \tag{2.4}$$

there exists a unique, strictly stationary, ergodic, and nonanticipative solution of (2.3) such that $\mathbb{E}|\varepsilon_t|^{\delta s} < \infty$. This solution is given by

$$\begin{cases} \varepsilon_t = \sigma_t \eta_t \\ \sigma_t^\delta = \omega + \omega \sum_{k=1}^{\infty} \sum_{i_1, \dots, i_k \geq 1} a_{i_1, t-i_1} \dots a_{i_k, t-i_1-\dots-i_k} |\eta_{t-i_1}|^\delta \dots |\eta_{t-i_1-\dots-i_k}|^\delta \end{cases} \tag{2.5}$$

with $a_{i, t-j} = \alpha_i^+ \mathbb{1}_{\eta_{t-j} \geq 0} + \alpha_i^- \mathbb{1}_{\eta_{t-j} < 0}$.

Remark 2.1.

- *In the ARCH(∞) case, where $\delta = 2$ and $A_s^+ = A_s^- = A_s$, Assumption (2.4) reduces to (2.2) since $\mu_{2s}^+ + \mu_{2s}^- = \mu_{2s}$. For the TARARCH(∞), Bardet and Wintenberger (2009) establish the sufficient second order stationarity condition $\sum_{i=1}^{\infty} \max(\alpha_i^+, \alpha_i^-) < 1$ which is stronger than (2.4) since $A_1^+ \mu_2^+ + A_1^- \mu_2^- \leq \sum_{i=1}^{\infty} \max(\alpha_i^+, \alpha_i^-) \mu_2$.*

It is worth noticing that the process introduced in (2.3) nests some widely used models in the financial industry. For example, the ARCH(∞) representation of the classical GARCH(1,1) process

$$\varepsilon_t = \sigma_t \eta_t, \quad \sigma_t^2 = \frac{\omega}{1-\beta} + \sum_{i=1}^{\infty} \alpha \beta^{i-1} \varepsilon_{t-i}^2, \tag{2.6}$$

where α and β , are positive constants, $\beta < 1$, and $\omega > 0$ is obviously a particular (symmetrical) specification of (2.3) with $\delta = 2$. Of course, this specification has short memory as the ARCH(∞) coefficients decay exponentially to zero. A more persistent specification of (2.3) based on the GARCH(1,1) model (2.6) is

$$\varepsilon_t = \sigma_t \eta_t, \quad \sigma_t^2 = \frac{\omega}{1-\beta} + \sum_{i=1}^{\infty} (\alpha \beta^{i-1} + \gamma i^{-d-1}) \varepsilon_{t-i}^2 \tag{2.7}$$

with $\gamma > 0$ and $d > 0$, where the coefficients have a hyperbolic decay. Figure 2.1 presents the effect of a shock on the conditional variance of a GARCH(1,1) and on an ARCH(∞) process specified as (2.7) for the same simulation of the iid process. It is seen that the shock at $t = 500$ is less persistent for a GARCH(1,1) process than for the ARCH(∞) one. Even if the β used in this illustration is fairly high (0.85), the effect of the shock has almost entirely disappeared after a hundred lags in the GARCH(1,1) case, while it remains clearly observable on the ARCH(∞) process.

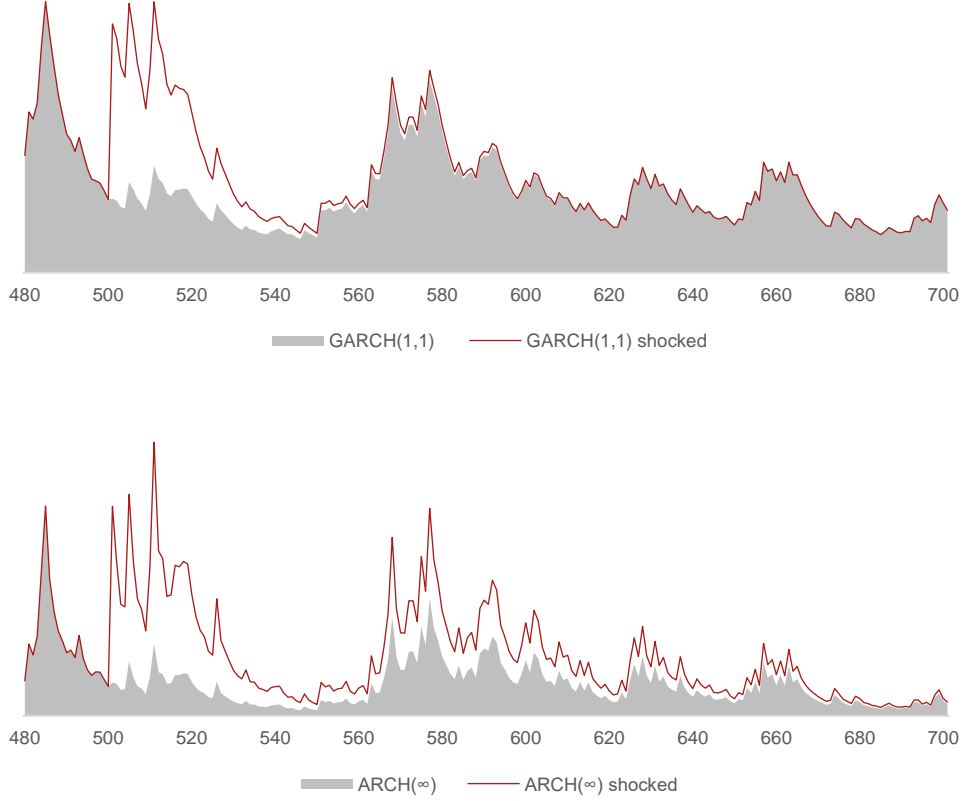


Figure 2.1: *Effect of a shock on η_t at $t = 500$ on the conditional variance of a $GARCH(1,1)$ process and an $ARCH(\infty)$ process with $\alpha_i = \alpha\beta^{i-1} + \gamma i^{-(d+1)}$, where $\omega = 0.01$, $\alpha = 0.1$, $\beta = 0.85$, $\gamma = 0.15$, and $d = 1$, and with $\eta_t \sim \mathcal{N}(0, 1)$*

Some well known asymmetric extensions to the $GARCH(1,1)$ are also particular specifications of model (2.3). Consider the following $APARCH(\infty)$ specification

$$\varepsilon_t = \sigma_t \eta_t, \quad \sigma_t^\delta = \frac{\omega}{1 - \beta} + \sum_{i=1}^{\infty} \beta^{i-1} (\alpha^+ \mathbf{1}_{\varepsilon_{t-i} \geq 0} + \alpha^- \mathbf{1}_{\varepsilon_{t-i} < 0}) |\varepsilon_{t-i}|^\delta$$

which is the rewriting of an $APARCH(1,1)$ as an $APARCH(\infty)$. The GJR- $GARCH(1,1)$ model introduced by Glosten et al. (1993) is obtained when $\delta = 2$, and the Threshold $GARCH$ (TGARCH) model of Zakoïan (1994) is obtained when $\delta = 1$. In the spirit of (2.7), an extension to the $APARCH(1,1)$ model to allow for higher persistence is then

$$\begin{cases} \varepsilon_t = \sigma_t \eta_t \\ \sigma_t^\delta = \frac{\omega}{1 - \beta} + \sum_{i=1}^{\infty} \beta^{i-1} (\alpha^+ \mathbf{1}_{\varepsilon_{t-i} \geq 0} + \alpha^- \mathbf{1}_{\varepsilon_{t-i} < 0}) |\varepsilon_{t-i}|^\delta + \gamma i^{-d-1} |\varepsilon_{t-i}|^\delta. \end{cases} \quad (2.8)$$

The models introduced in (2.7) and (2.8) are particularly interesting as they allow to nest $GARCH$ -type specifications in highly persistent volatility models. They will be used throughout

the paper to illustrate the assumptions required to establish asymptotic results.

2.3 Statistical inference of an APARCH(∞) process

Direct estimation of the models defined in (2.1) and (2.3) is not feasible without constraining the infinite sequence of coefficients and requires considering a parametrization. Building upon Robinson and Zaffaroni (2006), we introduce the parametric form of Model (2.3)

$$\begin{cases} \varepsilon_t = \sigma_t(\boldsymbol{\theta}_0)\eta_t \\ \sigma_t^\delta(\boldsymbol{\theta}_0) = \omega_0 + \sum_{i=1}^{\infty} \alpha_i^+(\boldsymbol{\phi}_0)|\varepsilon_{t-i}|^\delta \mathbb{1}_{\varepsilon_{t-i} \geq 0} + \alpha_i^-(\boldsymbol{\phi}_0)|\varepsilon_{t-i}|^\delta \mathbb{1}_{\varepsilon_{t-i} < 0} \end{cases} \quad (2.9)$$

where $\alpha_i^+(\cdot), \alpha_i^-(\cdot) : \Phi \rightarrow [0, \infty]$ are known functions, $\boldsymbol{\phi}_0$ is a $r \times 1$ unknown vector of parameters, ω_0 is an unknown positive constant, and $\delta > 0$ is a known parameter. We wish to estimate $\boldsymbol{\theta}_0 = (\omega_0, \boldsymbol{\phi}_0)'$ over a parameter space Θ , on the basis of n observations $\varepsilon_1, \dots, \varepsilon_n$. For example, the parametric form of Model (2.8) would then be $\alpha_i^{+(-)}(\boldsymbol{\phi}_0) = \alpha_0^{+(-)}\beta_0 + \gamma_0 i^{-d_0-1}$ with $\boldsymbol{\phi}_0 = (\alpha_0^+, \alpha_0^-, \beta_0, \gamma_0, d_0)$.

Following the works of Berkes et al. (2003b) and Francq and Zakoïan (2004) for the GARCH(p, q) process, asymptotic properties of the QMLE for APARCH(p, q) models have been established by Hamadeh and Zakoïan (2011), and extended by Francq and Thieu (2019). In Hamadeh and Zakoïan (2011), the authors show that the empirical estimation of the power parameter δ , although theoretically possible, is difficult to achieve. Following Francq and Thieu (2019), we therefore consider that the parameter δ is fixed and known. In general, this parameter is fixed to 1 (TGARCH) or 2 (GJR-GARCH) by practitioners. A comment on how to choose this parameter is however provided at the end of this section.

Estimation of the parameters of ARCH(∞) models has first been studied by Giraitis and Robinson (2001) who proposed a Whittle estimation of $\boldsymbol{\theta}_0$. However, this method presents some drawbacks as discussed by the authors³. Linton and Mammen (2005) studied semi-parametric estimation of an ARCH(∞) model without parametric specification of the effect of past returns on the conditional variance, but their method requires the existence of a fourth moment for ε_t . Robinson and Zaffaroni (2006) proposed to estimate the parameter $\boldsymbol{\theta}_0$ by QML under milder assumptions on the observed process. For different assumptions, see also Hafner and Preminger (2017). Finally, Bardet and Wintenberger (2009) studied the QMLE for ARCH(∞) and TARCH(∞) processes under mild assumptions but at the cost of imposing higher moments on ε_t . In the spirit of Robinson and Zaffaroni (2006), we study the QMLE in the case of an APARCH(∞) process.

³For example, the existence of a fourth-moment of ε_t is required for consistency and an eighth-moment for asymptotic normality.

Let us rewrite the volatility in (2.9) as

$$\sigma_t^\delta(\boldsymbol{\theta}_0) = \omega_0 + \sum_{i=1}^{\infty} a_{i,t-i}(\boldsymbol{\phi}_0) |\varepsilon_{t-i}|^\delta \quad (2.10)$$

where $a_{i,t-i}(\boldsymbol{\phi}) = \alpha_i^+(\boldsymbol{\phi}) \mathbb{1}_{\varepsilon_{t-i} \geq 0} + \alpha_i^-(\boldsymbol{\phi}) \mathbb{1}_{\varepsilon_{t-i} < 0}$ and note that for all i , any t , and any $\boldsymbol{\phi}$ in Φ , $a_{i,t-i}(\boldsymbol{\phi}) \leq \max(\alpha_i^+(\boldsymbol{\phi}), \alpha_i^-(\boldsymbol{\phi}))$. We define the QMLE as

$$\tilde{\boldsymbol{\theta}}_n = \underset{\boldsymbol{\theta} \in \Theta}{\text{Argmin}} \tilde{Q}_n(\boldsymbol{\theta}), \quad \tilde{Q}_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{t=1}^n \tilde{l}_t(\boldsymbol{\theta}), \quad \tilde{l}_t(\boldsymbol{\theta}) = \log \tilde{\sigma}_t^2(\boldsymbol{\theta}) + \frac{\varepsilon_t^2}{\tilde{\sigma}_t^2(\boldsymbol{\theta})}$$

where, for any admissible value $\boldsymbol{\theta}$ of $\boldsymbol{\theta}_0$, $\tilde{\sigma}_t^\delta$ is defined as ω for $t = 1$ and for $t > 1$

$$\tilde{\sigma}_t^\delta(\boldsymbol{\theta}) = \omega + \sum_{i=1}^{t-1} \alpha_i^+(\boldsymbol{\phi}) |\varepsilon_{t-i}|^\delta \mathbb{1}_{\varepsilon_{t-i} \geq 0} + \alpha_i^-(\boldsymbol{\phi}) |\varepsilon_{t-i}|^\delta \mathbb{1}_{\varepsilon_{t-i} < 0} = \omega + \sum_{i=1}^{t-1} a_{i,t-i}(\boldsymbol{\phi}) |\varepsilon_{t-i}|^\delta. \quad (2.11)$$

To show strong consistency, the following assumptions are used, and we denote from now on by K a generic positive constant.

[A1] The parameter space is of the form $\Theta = [\omega_L, \omega_U] \times \Phi$ where $0 < \omega_L < \omega_U < \infty$, and $\Phi \subset \mathbb{R}^r$ is a compact space.

[A2] The η_t are iid with $\mathbb{E}\eta_0 = 0$, $\mathbb{E}\eta_0^2 = 1$ and the distribution of the positive (resp. negative) part of (η_t) is non-degenerate.

[A3] (i) For any $\boldsymbol{\phi}$ and $\boldsymbol{\phi}^* \in \Phi$ such that $\boldsymbol{\phi} \neq \boldsymbol{\phi}^*$, there exists $k \geq 1$ such that $\alpha_k^+(\boldsymbol{\phi}) \neq \alpha_k^+(\boldsymbol{\phi}^*)$ and $\alpha_k^-(\boldsymbol{\phi}) \neq \alpha_k^-(\boldsymbol{\phi}^*)$.

(ii) For all $i \geq 1$, $\sup_{\boldsymbol{\phi} \in \Phi} \max(\alpha_i^+(\boldsymbol{\phi}), \alpha_i^-(\boldsymbol{\phi})) \leq K i^{-d-1}$ for some $d > 0$.

[A4] There exists a solution (ε_t) of equation (2.9) such that $\mathbb{E}|\varepsilon_t|^{(2 \wedge \delta)\rho} < \infty$ for $\rho > \frac{1}{d+1}$.

[A5] $\mathbb{E}_{\boldsymbol{\theta}_0} \sup_{\boldsymbol{\theta} \in \Theta} \frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2(\boldsymbol{\theta})} < \infty$.

Remark 2.2.

– The compactness assumption **A1** is standard for QML estimation. Assumptions **A2** and **A3(i)** are needed for identifiability. The former is slightly stronger than needed in the ARCH(∞) case where only the distribution of (η_t^2) needs to be non-degenerate. Assumption **A3(ii)** along with Assumption **A4** entail the existence of $\sigma_t^\delta(\boldsymbol{\theta})$ for any $\boldsymbol{\theta}$. Note that the d in assumptions **A3(ii)** and **A4** is the same and that these assumptions may be stronger than (2.4). Nevertheless, Assumption **A4** is quite mild as, for a large value of d , it would only imply the existence of a small moment. For example, it is the case for the GARCH(1,1) model where the α_i are exponentially decaying. Note that a sufficient condition for Assumption **A5** is of course $\mathbb{E}\varepsilon_t^2 < \infty$. Proposition 2.2 in the appendix gives a different sufficient condition for **A5** to hold without additional moment condition for ε_t .

– In the classical ARCH(∞) case where $\delta = 2$ and α_i^+ and α_i^- are equal, though our assumptions are mostly in line with the ones proposed by Robinson and Zaffaroni (2006), they are noticeably milder concerning the distribution of η_t . Indeed, we do not specify that the density of η_t is well-behaved near 0. Furthermore, our assumptions on α_i are also milder as we allow our coefficients to be equal to 0 and do not impose $\alpha_i(\boldsymbol{\phi}_0) \leq K\alpha_j(\boldsymbol{\phi}_0)$ for $i \geq j \geq 1$. Note that Robinson and Zaffaroni considered a slightly more general model $y_t = \mu + \varepsilon_t$ allowing for a drift. Our model could similarly be extended to take into account this parameter but for the sake of clarity we assume that this drift parameter is known and equal to 0.

Notice that Model (2.8) where $\alpha_i^{+(-)}(\boldsymbol{\phi}_0) = \alpha_0^{+(-)}\beta_0^{i-1} + \gamma_0 i^{-(d_0+1)}$ and $\boldsymbol{\Phi} \subset (0, \infty)^5$ satisfies the proposed assumptions. In particular, Assumption **A3(i)** is satisfied if $\alpha_0^{+(-)}$, β_0 , and γ_0 are positive, which ensures Assumption **A5** using Proposition 2.2.

The following result states the strong consistency of $\tilde{\boldsymbol{\theta}}_n$.

Theorem 2.2. *Under assumptions **A1-A5**, almost surely*

$$\tilde{\boldsymbol{\theta}}_n \rightarrow \boldsymbol{\theta}_0, \quad \text{as } n \rightarrow \infty.$$

To show the asymptotic normality, the following additional assumptions are considered.

[A6] $\boldsymbol{\theta}_0$ belongs to the interior of Θ .

[A7] $1 < \kappa_\eta = \mathbb{E}\eta_0^4 < \infty$.

[A8] For all $i \geq 1$, $\max(\alpha_i^+(\boldsymbol{\phi}_0), \alpha_i^-(\boldsymbol{\phi}_0)) \leq K i^{-d^*-1}$ for some $d^* > \frac{1}{2}$.

[A9] $\mathbb{E}|\varepsilon_t|^{(2\wedge\delta)\rho} < \infty$ for some $\rho > \frac{4}{2d^*+3}$.

[A10] (i) For all j , α_j has continuous k th derivative on $\boldsymbol{\Phi}$, $k \leq 3$, such that, denoting ϕ_i the i th element of $\boldsymbol{\phi}$,

$$\left| \frac{\partial^k \alpha_j^+(\boldsymbol{\phi})}{\partial \phi_{i_1} \dots \partial \phi_{i_k}} \right| \leq K(\alpha_j^+)^{1-\xi}(\boldsymbol{\phi}) \quad \text{and} \quad \left| \frac{\partial^k \alpha_j^-(\boldsymbol{\phi})}{\partial \phi_{i_1} \dots \partial \phi_{i_k}} \right| \leq K(\alpha_j^-)^{1-\xi}(\boldsymbol{\phi})$$

for all $\xi > 0$ and all $i_h = 1, \dots, r$, $h = 1, \dots, k$.

(ii) There exists $i_h^+ = i_h^+(\boldsymbol{\phi}_0)$ and $i_h^- = i_h^-(\boldsymbol{\phi}_0)$, $h = 1, \dots, r$, such that $1 \leq i_1^{+(-)} < \dots < i_r^{+(-)} < \infty$ and

$$\text{rank} \left[\frac{\partial \alpha_{i_1^+}^+(\boldsymbol{\phi}_0)}{\partial \boldsymbol{\phi}} \dots \frac{\partial \alpha_{i_r^+}^+(\boldsymbol{\phi}_0)}{\partial \boldsymbol{\phi}} \right] = \text{rank} \left[\frac{\partial \alpha_{i_1^-}^-(\boldsymbol{\phi}_0)}{\partial \boldsymbol{\phi}} \dots \frac{\partial \alpha_{i_r^-}^-(\boldsymbol{\phi}_0)}{\partial \boldsymbol{\phi}} \right] = r.$$

[A11] For all $k > 0$, there exists a neighborhood $V(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ such that,

$$\mathbb{E}_{\boldsymbol{\theta}_0} \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left[\frac{\sigma_t^\delta(\boldsymbol{\theta}_0)}{\sigma_t^\delta(\boldsymbol{\theta})} \right]^k < \infty.$$

Remark 2.3.

- Assumption **A6** is required for asymptotic normality. Assumption **A7** is necessary for the existence of the variance of the score vector $\partial l_t(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}$. Assumptions **A8** and **A9** are stronger than Assumptions **A3(ii)** and **A4** and impose a higher rate of convergence for $\alpha_i^{+(-)}$. Assumption **A10(i)** is similar to Assumption **A3(ii)** and allows the summability of the derivatives of the $\alpha_i^{+(-)}$ functions, while Assumption **A10(ii)** ensures non singularity of the matrix **J**. The particular rates of convergence of the $\alpha_i^{+(-)}$ functions and their derivatives imposed in Assumptions **A3(ii)**, **A8** and **A10(i)** are crucial to show the asymptotic irrelevance of the initial values and the integrability of the derivatives in a neighborhood of $\boldsymbol{\theta}_0$. Proposition 2.3 in the appendix gives an example of a sufficient condition for **A11**.

Note again that Model (2.8) satisfies the set of additional assumptions if $d_0 > 0.5$. In particular, Assumption **A10** is satisfied and Assumption **A11** holds from Proposition 2.3.

Theorem 2.3. Under assumptions A1-A11,

$$\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, (\kappa_\eta - 1)\mathbf{J}^{-1}) \quad (2.12)$$

where

$$\mathbf{J} = \frac{4}{\delta^2} \mathbb{E}_{\boldsymbol{\theta}_0} \left[\frac{1}{\sigma_t^{2\delta}(\boldsymbol{\theta}_0)} \frac{\partial \sigma_t^\delta(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \frac{\partial \sigma_t^\delta(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} \right]$$

is a positive definite matrix.

Next, we consider an extension allowing the true parameter value to lie on the boundary of Θ , relaxing Assumption **A6**. Let $\Phi = [\phi_1, \bar{\phi}_1] \times \cdots \times [\phi_r, \bar{\phi}_r]$, define $\partial\Phi = \{\boldsymbol{\phi}_0 \in \Phi : \phi_{0,i} = \underline{\phi}_i \text{ for some } i > 0\}$ and let $\boldsymbol{\phi}_0(\varepsilon)$ the vector obtained by replacing $\phi_{0,i}$ by $\underline{\phi}_i + \varepsilon$ for all i such that $\phi_{0,i} = \underline{\phi}_i$. Similarly to Francq and Zakoïan (2007) for the case of GARCH(p, q) models, we make the following assumption to prevent $\boldsymbol{\phi}_0$ from reaching the upper bound of Φ .

[A6’] There exists $\varepsilon > 0$ such that $\boldsymbol{\theta}_0(\varepsilon) = [\omega_0, \boldsymbol{\phi}_0(\varepsilon)]'$ belongs to the interior of Θ .

The following theorem establishes the asymptotic distribution of $\tilde{\boldsymbol{\theta}}_n$ when $\boldsymbol{\theta}_0$ may be on the boundary.

Theorem 2.4. Under the assumptions of Theorem 2.3 where A6 is replaced by A6’,

$$\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{\mathcal{L}} \boldsymbol{\lambda}^\Lambda := \arg \inf_{\boldsymbol{\lambda} \in \Lambda} [\boldsymbol{\lambda} - \mathbf{Z}]' \mathbf{J} [\boldsymbol{\lambda} - \mathbf{Z}] \quad (2.13)$$

with $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, (\kappa_\eta - 1)\mathbf{J}^{-1})$, $\Lambda = \Lambda(\boldsymbol{\theta}_0) = \Lambda_1 \times \cdots \times \Lambda_{r+1}$, where $\Lambda_1 = \mathbb{R}$ and for $i = 2, \dots, r+1$, $\Lambda_i = \mathbb{R}$ if $\phi_{0,i} \neq \underline{\phi}_i$ and $\Lambda_i = [0, \infty)$ otherwise.

Remark 2.4.

- We emphasize that the asymptotic distribution of the QMLE is obtained without any additional assumption on the moments of ε_t . Similarly, Francq and Zakoïan (2007) establish the asymptotic distribution of the QMLE without assuming additional moment assumption by ensuring that assumptions **A5** and **A11** are satisfied.

– The asymptotic distribution in (2.13) is the orthogonal projection of a normal vector distribution onto a convex cone, see Francq and Zakoïan (2007) for a practical derivation of this limiting distribution.

Note that in Model (2.8), Assumption **A3(i)** does not allow for the parameters β_0 or γ_0 to be equal to zero for identification reasons. A particular discussion on how to test $\gamma_0 = 0$ is proposed in Section 3.3.

Asymptotic results for $\tilde{\boldsymbol{\theta}}_n$ have been obtained under the assumption that δ was known. Although the choice of the power parameter may have little influence on the fitted volatility (see Francq and Thieu (2019) and references therein), a practitioner might be unsure of which model to select. As the number of unknown parameters in $\tilde{\boldsymbol{\theta}}_n$ is the same for different choices of δ , it seems natural to select the model with the highest quasi likelihood. Theorem 2.5, which we introduce below, justifies this approach.

Let us denote by δ_0 the true value of the power parameter and replace $\sigma_t^\delta(\boldsymbol{\theta}_0)$ in (2.9) and $\tilde{\sigma}_t^\delta(\boldsymbol{\theta}_0)$ in (2.11) by $\sigma_{\delta_0,t}^{\delta_0}(\boldsymbol{\theta}_0)$ and $\tilde{\sigma}_{\delta_0,t}^{\delta_0}(\boldsymbol{\theta}_0)$ to underline that they actually depend on the value of δ_0 . Consider a set \mathcal{D} of k candidates for δ_0 (e.g. $\delta_0 = 1$ for the APARCH(∞) extension of the TGARCH, or $\delta_0 = 2$ for a TARCH(∞)) such as

$$\delta_0 \in \mathcal{D} = \{\delta_1, \dots, \delta_k\}, \quad \delta_i > 0, \quad i = 1, \dots, k \quad (2.14)$$

and let

$$(\tilde{\delta}_n, \tilde{\boldsymbol{\theta}}_n) = \underset{\delta \in \mathcal{D}, \boldsymbol{\theta} \in \Theta}{\text{Argmin}} \tilde{Q}_{\delta,n}(\boldsymbol{\theta}), \quad \tilde{Q}_{\delta,n}(\boldsymbol{\theta}) = \frac{1}{n} \sum_{t=1}^n \tilde{l}_{\delta,t}(\boldsymbol{\theta}), \quad \tilde{l}_{\delta,t}(\boldsymbol{\theta}) = \log \tilde{\sigma}_{\delta,t}^2(\boldsymbol{\theta}) + \frac{\varepsilon_t^2}{\tilde{\sigma}_{\delta,t}^2(\boldsymbol{\theta})}. \quad (2.15)$$

We need to slightly strengthen the assumption on the distribution of η_t .

[A2'] The η_t are iid with $\mathbb{E}\eta_0 = 0$, $\mathbb{E}\eta_0^2 = 1$ and the positive (resp. negative) part of (η_t) takes more than 3 values.

Theorem 2.5. *Under the assumptions of Theorem 2.2, Assumption A2', and (2.14), almost surely $\tilde{\boldsymbol{\theta}}_n \rightarrow \boldsymbol{\theta}_0$ as $n \rightarrow \infty$ and $\tilde{\delta}_n = \delta_0$ for n large enough. Moreover, under the corresponding additional assumptions, the asymptotic distribution of $\tilde{\boldsymbol{\theta}}_n$ is still given by Theorem 2.3 or Theorem 2.4.*

2.4 Specification tests

The presence of asymmetry and memory in financial time series has been well documented. However, in order to select the most parsimonious model, it is critical to test their statistical significance and the adequacy of the chosen model. This section introduces simple test procedures for goodness-of-fit, asymmetry and strong (non-exponentially decaying) memory.

2.4.1 Portmanteau goodness-of-fit test for APARCH(∞) models

Since their introduction by Box and Pierce (1970), tests based on residuals autocorrelations, the so-called Portmanteau tests, have become widely used in econometrics. To test the adequacy of conditional volatility models, Li and Mak (1994) proposed to use Portmanteau tests based on squared residuals autocorrelations. Asymptotic properties of these tests have been established by Berkes et al. (2003a) for standard GARCH(p, q) models and by Carbon and Francq (2011) in the APARCH(p, q) case. To our best knowledge, these results have not yet been extended to the ARCH(∞) literature. This section therefore aims at filling that gap. One should note that other kinds of goodness-of-fit tests exist. In particular, Hidalgo and Zaffaroni (2007) propose a goodness-of-fit test based on the estimated spectral distribution function. However, contrary to Portmanteau tests, their statistic has a nonstandard asymptotic distribution and requires bootstrap procedures to compute critical values.

Let us consider the null hypothesis H_0^{GoF} that the process (ε_t) satisfies model (2.3). We define the autocovariances of the squared residuals by

$$\hat{r}_h = n^{-1} \sum_{t=h+1}^n (\hat{\eta}_t^2 - 1)(\hat{\eta}_{t-h}^2 - 1), \text{ with } \hat{\eta}_t^2 = \varepsilon_t^2 / \hat{\sigma}_t^2(\tilde{\boldsymbol{\theta}}_n),$$

and let $\hat{\mathbf{r}}_m = (\hat{r}_1, \dots, \hat{r}_m)$ for any $1 \leq m \leq n$, and $\hat{\mathbf{C}}_m$ the $m \times (r+1)$ matrix whose elements (h, k) are given by

$$\hat{\mathbf{C}}_m(h, k) = -\frac{2}{\delta n} \sum_{t=h+1}^n (\hat{\eta}_{t-h}^2 - 1) \frac{1}{\hat{\sigma}_t^2(\tilde{\boldsymbol{\theta}}_n)} \frac{\partial \hat{\sigma}_t^2(\tilde{\boldsymbol{\theta}}_n)}{\partial \theta_k}.$$

In addition, let

$$\hat{\mathbf{J}}_n = \frac{4}{\delta^2} \frac{1}{n} \sum_{t=1}^n \frac{1}{\hat{\sigma}_t^{2\delta}(\tilde{\boldsymbol{\theta}}_n)} \frac{\partial \hat{\sigma}_t^2(\tilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}} \frac{\partial \hat{\sigma}_t^2(\tilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}'} \text{ and } \hat{\kappa}_\eta = \frac{1}{n} \sum_{t=1}^n \frac{\varepsilon_t^4}{\hat{\sigma}_t^4(\tilde{\boldsymbol{\theta}}_n)}$$

be consistent estimators of \mathbf{J} and κ_η respectively (from arguments in the proofs of Theorem 2.3, Lemma 2.1, and the ergodic theorem).

The following theorem establishes the asymptotic distribution of the Portmanteau test statistic.

Theorem 2.6. *Under H_0^{GoF} , under the assumptions of Theorem 2.3 and assumption **A2'**,*

$$n\hat{\mathbf{r}}_m' \hat{\mathbf{D}}^{-1} \hat{\mathbf{r}}_m \xrightarrow{L} \chi_m^2,$$

with $\hat{\mathbf{D}} = (\hat{\kappa}_\eta - 1)^2 \mathbf{I}_m - (\hat{\kappa}_\eta - 1) \hat{\mathbf{C}}_m \hat{\mathbf{J}}_n^{-1} \hat{\mathbf{C}}_m'$.

The adequacy of the APARCH(∞) model (2.3) is then rejected at the asymptotic level ν when $n\hat{\mathbf{r}}_m' \hat{\mathbf{D}}^{-1} \hat{\mathbf{r}}_m > \chi_m^2(1 - \nu)$ where $\chi_m^2(1 - \nu)$ is the $(1 - \nu)$ -quantile of the χ^2 distribution with m degrees of freedom.

2.4.2 Testing for linear constraints on the parameters

We are now interested in testing for a general hypothesis of the form

$$H_0 : \mathbf{R} \boldsymbol{\theta}_0 = \mathbf{k}, \quad H_1 : \mathbf{R} \boldsymbol{\theta}_0 \neq \mathbf{k} \quad (2.16)$$

where \mathbf{R} is the constraints matrix and \mathbf{k} is a constant vector. Let c be the rank of the matrix \mathbf{R} . The triptych of the Wald, Rao-score, and Quasi Likelihood Ratio (LR) statistics to test (2.16) is given by

$$\begin{aligned} W_n &= (\mathbf{R} \tilde{\boldsymbol{\theta}}_n - \mathbf{k})' \left(\mathbf{R} \left(\frac{(\hat{\kappa}_\eta - 1)}{n} \hat{\mathbf{J}}_n^{-1} \right) \mathbf{R}' \right)^{-1} (\mathbf{R} \tilde{\boldsymbol{\theta}}_n - \mathbf{k}) \\ R_n &= \frac{n}{\hat{\kappa}_{\eta|H_0} - 1} \frac{\partial \tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{n|H_0})}{\partial \boldsymbol{\theta}'} \hat{\mathbf{J}}_{n|H_0}^{-1} \frac{\partial \tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{n|H_0})}{\partial \boldsymbol{\theta}} \\ L_n &= \frac{2n}{\hat{\kappa}_{\eta|H_0} - 1} \left[\tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{n|H_0}) - \tilde{Q}_n(\tilde{\boldsymbol{\theta}}_n) \right] \end{aligned} \quad (2.17)$$

where $\tilde{\boldsymbol{\theta}}_{n|H_0}$ is the QMLE restricted by H_0 and

$$\hat{\mathbf{J}}_{n|H_0} = \frac{4}{\delta^2} \frac{1}{n} \sum_{t=1}^n \frac{1}{\tilde{\sigma}_t^{2\delta}(\tilde{\boldsymbol{\theta}}_{n|H_0})} \frac{\partial \tilde{\sigma}_t^\delta(\tilde{\boldsymbol{\theta}}_{n|H_0})}{\partial \boldsymbol{\theta}} \frac{\partial \tilde{\sigma}_t^\delta(\tilde{\boldsymbol{\theta}}_{n|H_0})}{\partial \boldsymbol{\theta}'}, \quad \text{and} \quad \hat{\kappa}_{\eta|H_0} = \frac{1}{n} \sum_{t=1}^n \frac{\varepsilon_t^4}{\tilde{\sigma}_t^4(\tilde{\boldsymbol{\theta}}_{n|H_0})}$$

are consistent estimators of \mathbf{J} and κ_η respectively, under H_0 .

Proposition 2.1. *Under $H_0 : \mathbf{R} \boldsymbol{\theta}_0 = \mathbf{k}$,*

(i) *under the assumptions of Theorem 2.3,*

$$W_n \xrightarrow{\mathcal{L}} \chi_c^2, \quad R_n \xrightarrow{\mathcal{L}} \chi_c^2, \quad \text{and} \quad L_n \xrightarrow{\mathcal{L}} \chi_c^2,$$

(ii) *under the assumptions of Theorem 2.4,*

$$\begin{aligned} W_n &\xrightarrow{\mathcal{L}} \boldsymbol{\lambda}' \mathbf{R}' [(\kappa_\eta - 1) \mathbf{R} \mathbf{J}^{-1} \mathbf{R}']^{-1} \boldsymbol{\lambda}' \mathbf{R}, \quad R_n \xrightarrow{\mathcal{L}} \chi_c^2, \quad \text{and} \\ L_n &\xrightarrow{\mathcal{L}} -\frac{1}{2} (\boldsymbol{\lambda}^\Lambda - \mathbf{Z})' \mathbf{J} (\boldsymbol{\lambda}^\Lambda - \mathbf{Z}) + \frac{1}{2} \mathbf{Z}' \mathbf{R}' [\mathbf{R} \mathbf{J}^{-1} \mathbf{R}']^{-1} \mathbf{R} \mathbf{Z}. \end{aligned}$$

Note that in Model (2.9), the symmetry hypothesis is generally a particular constrained representation. Testing for the significance of asymmetry can thus be achieved by testing an implied restriction on $\boldsymbol{\theta}_0$. For example, if we consider the parametric version of specification (2.8) obtained by setting $\boldsymbol{\phi}_0 = (\alpha_0^+, \alpha_0^-, \beta_0, \gamma_0, d_0)$ and $\alpha_i^{+(-)}(\boldsymbol{\phi}_0) = \beta_0^{i-1} \alpha_0^{+(-)} + \gamma_0 i^{-(d_0+1)}$, the symmetry hypothesis is given by

$$H_0^{\text{sym}} : \alpha_0^+ = \alpha_0^-, \quad H_1^{\text{sym}} : \alpha_0^+ \neq \alpha_0^-$$

which is a particular form of (2.16). Testing for a constrained representation is highly common when testing for asymmetry in parametric models, see for example Nelson (1991).

2.4.3 Testing for GARCH(1,1) specifications

Despite the development of multiple extensions, the GARCH(1,1) model remains preeminent in the financial industry and literature. Although this model admits an ARCH(∞) representation, it imposes an exponential decay on its coefficients. We propose to study the validity of a GARCH(1,1) representation by allowing these coefficients to decay in a slower manner. In order to do so, consider the following ARCH(∞) parametrization

$$\begin{cases} \varepsilon_t = \sigma_t(\boldsymbol{\theta}_0)\eta_t \\ \sigma_t^2(\boldsymbol{\theta}_0) = \frac{\omega_0}{1-\beta_0} + \sum_{i=1}^{\infty} (\alpha_0\beta_0^{i-1} + \gamma_0 i^{-(d_0+1)})\varepsilon_{t-i}^2 \end{cases} \quad (2.18)$$

with $\alpha_0 > 0$, $\beta_0 > 0$, $\gamma_0 \geq 0$, and $d_0 > 0$. Testing the validity of a GARCH(1,1) representation can then be achieved by testing

$$H_0^{\text{GARCH}} : \gamma_0 = 0, \quad H_1^{\text{ARCH}(\infty)} : \gamma_0 > 0, \quad (2.19)$$

which can be rewritten as $H_0^{\text{GARCH}} : \mathbf{R} \boldsymbol{\theta}_0 = 0$, and $H_1^{\text{ARCH}(\infty)} : \mathbf{R} \boldsymbol{\theta}_0 > 0$ with $\mathbf{R} = (0, 0, 0, 1, 0)$. While this test may seem standard, it poses a major difficulty. Indeed, the parameter d_0 is not identified under the null hypothesis, thus we cannot directly use Proposition 2.1(ii) to obtain the asymptotic distribution of the test statistics when the parameter is on the boundary. A simple solution could be to assume that the parameter d_0 is known and fixed at a value $\bar{d} > 0.5$. Under this assumption, the Wald statistic distribution is a mixture of a χ_1^2 and a Dirac measure at 0, both with weight 1/2 (see Proposition 4 in the supplementary file). In addition, Francq and Zakoïan (2009) show that, when testing the nullity of only one coefficient, the Wald test is locally asymptotically more powerful than the standard score test. Although setting the unidentified under the null parameter at an arbitrary value facilitates the derivation of the asymptotic distribution, choosing a value \bar{d} that is far from d_0 may lead to spurious results⁴. Asymptotic results, when the presence of a coefficient on the boundary of the parameter space involves the non identification of a second parameter, have been established by Andrews (2001). However, the limiting distributions in such case are highly non-standard. Instead, we propose to use a residual-based bootstrap procedure to approximate the statistic asymptotic distribution. Using the terminology of Beutner et al. (2018), we propose the following recursive design bootstrap procedure for testing (2.19) on a sample of n observations $\varepsilon_1, \dots, \varepsilon_n$. Let us denote by $\tilde{\boldsymbol{\theta}}_n^c = (\tilde{\omega}^c, \tilde{\alpha}^c, \tilde{\beta}^c)'$ the QMLE of a GARCH(1,1) model and let $\tilde{\sigma}_t^c(\tilde{\boldsymbol{\theta}}_n^c)$ the estimated volatility process.

1. On the observations, compute $\tilde{\boldsymbol{\theta}}_n^c$ and the standardized residuals $\tilde{\eta}_t^c = \hat{\eta}_t^c / s_n^c$ where $\hat{\eta}_t^c = \varepsilon_t / \tilde{\sigma}_t^c(\tilde{\boldsymbol{\theta}}_n^c)$ and $(s_n^c)^2 = n^{-1} \sum_{t=1}^n (\hat{\eta}_t^c)^2$. Denote by F_n^* the empirical distribution of $\tilde{\eta}_t^c$.
2. Simulate a trajectory of length n of a GARCH(1,1) with parameter $\tilde{\boldsymbol{\theta}}_n^c$ and where the innovations $\eta_t^* \stackrel{\text{iid}}{\sim} F_n^*$. On this simulation, compute the unconstrained estimator $\boldsymbol{\theta}_n^*$ of an ARCH(∞) and compute the statistic $W_n^{\text{GARCH}^*}$.

⁴See the supplementary file for a study of the empirical power of the test when \bar{d} is misspecified.

3. On the observations, compute the unconstrained estimator $\tilde{\boldsymbol{\theta}}_n$ of an ARCH(∞) and compute the statistic W_n .
4. Repeat B times Step 2 and denote by $W_n^{\text{GARCH}*1} \dots W_n^{\text{GARCH}*B}$ the obtained bootstrap test statistics. Approximate the p -value of the test H_0^{GARCH} against $H_1^{\text{ARCH}(\infty)}$ by $B^{-1} \sum_{b=1}^B \mathbb{1}_{W_n^{\text{GARCH}*b} \geq W_n^{\text{GARCH}}}$.

This test can easily be extended to an asymmetric volatility model with a different δ_0 . Consider, the APARCH(∞) specification presented in (2.8). Testing for the adequacy of the GJR-GARCH model ($\delta_0 = 2$) or the TGARCH ($\delta_0 = 1$) can then be achieved by testing for $H_0 : \gamma_0 = 0$ against $H_1^{\text{APARCH}(\infty)} : \gamma_0 > 0$. By changing the constrained model, in steps 1 and 2, and the unconstrained model in step 3, we obtain the corresponding test procedures. In the case of ARCH(∞) models, the validity of this approach has been established by Hidalgo and Zaffaroni (2007) and we assume it holds in our setting. Simulations conducted in Section 4 confirm that the bootstrap scheme works well. Note that the proposed bootstrap might be invalid when other parameters than γ_0 are on the boundary. However, the procedure can be modified to account for such problems using the recent technique introduced by Cavaliere et al. (2022).

2.5 Simulations

In order to assess the finite sample properties of the QMLE in the different settings studied in this paper and to study the empirical behavior of the test statistics defined in Section 3, we carry out some Monte Carlo experiments. In the following simulations, we use Gaussian innovations ($\eta_t \sim \mathcal{N}(0, 1)$).

We focus on specifications (2.7) and (2.8) that nest several favored volatility models. We want to estimate $\boldsymbol{\theta}_0 = (\omega_0, \alpha_0^+, \alpha_0^-, \beta_0, \gamma_0, d_0)$. We start by simulating a thousand samples of size $n = 5000$ of different specifications including symmetric models (i.e. with $\alpha_0^+ = \alpha_0^-$) either for $\delta_0 = 1$ or $\delta_0 = 2$. The empirical mean and RMSE of the obtained QMLE are reported in Table 2.1 as well as the empirical mean of $\tilde{\delta}_n$ obtained from Proposition 1 when the candidates for δ_0 range from 0.5 to 3 with a 0.25 step. In order to assess the finite sample properties of the asymptotic variance estimator, given by (2.12), we can compare $V_n^{1/2} = \text{diag}[(\hat{\kappa}_\eta - 1)\hat{\mathbf{J}}_n^{-1}]^{1/2}/\sqrt{n}$ with the RMSE. On that matter, the results in Table 2.1 are quite satisfactory. Note that $d_0 = 1$ allows to easily derive sufficient stationary conditions for model (2.8) as the Riemann sum $\sum_{i=1}^{\infty} i^{-2} = \pi^2/6$ and thus, for $\delta = 2$, Theorem 2.1 entails the existence of a second order stationary solution if $\max(\alpha_0^+, \alpha_0^-)/(1 - \beta_0) + \gamma_0\pi^2/6 < 1$, which is verified for the $\boldsymbol{\theta}_0$ reported in Table 2.1.

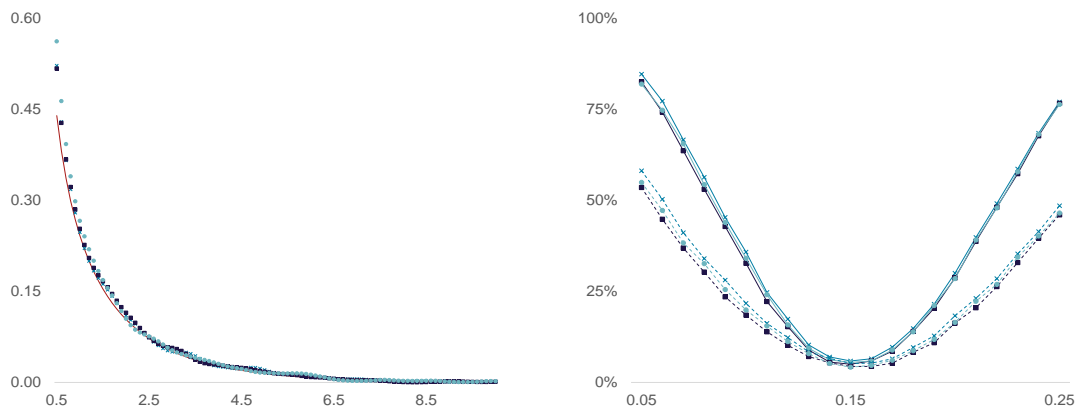
We then turn to the asymptotic properties of the tests statistics introduced in Section 3. Note that in model (2.8), the null hypothesis $H_0^{\text{sym}} : \alpha_0^+ = \alpha_0^-$, is a linear constraint on $\boldsymbol{\theta}_0$ with $\mathbf{R} = (0, 1, -1, 0, 0, 0)$. We denote by W_n^{sym} , R_n^{sym} and L_n^{sym} the Wald, Rao, and Quasi-Likelihood Ratio test statistics derived from (2.17). Figure 2.2a presents kernel density estimators of the three test statistics for $n = 5000$ under H_0^{sym} obtained with 5000 replications for

	θ_0	$\delta_0 = 2$				$\delta_0 = 1$			
		$\tilde{\theta}_n$	Bias	RMSE	$V_n^{\frac{1}{2}}$	$\tilde{\theta}_n$	Bias	RMSE	$V_n^{\frac{1}{2}}$
ω	0.20	0.209	0.009	0.053	0.029	0.206	0.006	0.046	0.024
α^+	0.05	0.057	0.007	0.038	0.021	0.055	0.005	0.038	0.017
α^-	0.15	0.158	0.008	0.045	0.026	0.155	0.005	0.043	0.019
β	0.70	0.688	-0.012	0.067	0.035	0.691	-0.009	0.060	0.029
γ	0.15	0.140	-0.010	0.049	0.029	0.142	-0.008	0.048	0.022
d	1.0	0.983	-0.017	0.383	0.157	0.980	-0.020	0.388	0.115
				$\bar{\delta}_n = 2.001$				$\bar{\delta}_n = 0.988$	
ω	0.20	0.211	0.011	0.057	0.029	0.213	0.013	0.064	0.019
α^+	0.10	0.108	0.008	0.038	0.021	0.108	0.008	0.040	0.013
α^-	0.10	0.108	0.008	0.037	0.021	0.109	0.009	0.039	0.013
β	0.75	0.739	-0.011	0.053	0.027	0.736	-0.014	0.064	0.017
γ	0.20	0.189	-0.011	0.049	0.029	0.189	-0.011	0.050	0.019
d	1.0	0.989	-0.011	0.372	0.129	0.973	-0.027	0.398	0.090
				$\bar{\delta}_n = 2.005$				$\bar{\delta}_n = 0.997$	

Table 2.1: *Estimation results for 1000 simulations of size 5000 of an APARCH(∞) process defined as (2.8) with different specifications and for $\delta_0 = 1$ and 2*

$\theta_0 = (0.2, 0.1, 0.1, 0.75, 0.2, 1)$. All kernel estimators are close to the asymptotic distribution χ_1^2 . Additionally, the relative rejection frequency of the Wald, Rao-score, and LR test statistics, at the asymptotic level 5%, are respectively 5.38%, 5.70% and 5.78%. When using 5000 independent replications, the empirical level should belong to the confidence interval [4.40%, 5.60%], hence the Wald statistic seems to better control the error of first kind. To study the empirical behavior of these statistics under H_1^{asym} , we also performed the tests on each realization of a TARARCH(∞) simulations sample when $\theta_0 = (0.2, \alpha_0^+, 0.15, 0.5, 0.25, 1)$ and α_0^+ ranges from 0.05 to 0.25. Figure 2.2b compares the observed powers of the three tests, that is, the relative frequency of rejection of the null hypothesis of symmetry on the 1000 independent realizations of length $n = 2500$ and $n = 5000$, as a function of α_0^+ . On these simulations, we see that the three test statistics seem powerful but may require a large number of observations to capture a weak asymmetry.

Figure 2.3a presents kernel estimators for the bootstrap-Wald test statistics defined in Section 3.3 when testing for a GARCH(1,1), a GJR-GARCH(1,1) and a TGARCH(1,1) against an APARCH(∞) model of form (2.8) with $\delta_0 = 2$ and 1 respectively under H_0 . The statistics have been obtained by adapting the "Warp-Speed" bootstrap techniques introduced by Giacomini et al. (2013) to reduce the computational burden of the bootstrap procedure. The parameters used for the simulations are $\theta_0 = (0.2, 0.15, 0.75)$ for the GARCH model and $\theta_0 = (0.2, 0.05, 0.2, 0.75)$ for both the GJR-GARCH and the TGARCH. All kernels estimators are obtained from 1000 replications of the experiment. We clearly see that the estimated distributions are different from the theoretical asymptotic distributions when assuming d is fixed at some arbitrary value \bar{d} . The relative rejection frequency of the test statistics, at the asymptotic levels 5%, are respectively 3.70%, 4.60% and 5.50%. We then repeat the experience under H_1 with θ_0 similar to the top part of Table 2.1. The obtained empirical power of the three test statistics, at the asymptotic levels 5%, are respectively 95.5%, 71.4% and 69.3%. It thus appears that on these realizations, the GARCH(1-1) test has a better power but seems to have



(a) Comparison between kernel density estimators and the χ_1^2 density on $[0.5, \infty)$ (red solid line) on 5000 simulations of a symmetric $ARCH(\infty)$ process for sample size $n = 5000$

(b) Observed powers as a function of α_0^+ when $\alpha_0^- = 0.15$, on 1000 simulations with $n = 2500$ (dashed line) and $n = 5000$ (solid line)

Figure 2.2: Empirical behavior of the Wald (dark blue square), the Rao-score (light blue dot), and the LR (blue cross) test statistics

a lower control of the error of first kind.

Finally, Figure 2.3b presents the empirical kernels of the Portmanteau statistic for the goodness-of-fit test presented in Section 3.1. The kernels are obtained from 1000 simulations of an $APARCH(\infty)$ with $\delta = 1$ and $\theta_0 = (0.2, 0.05, 0.15, 0.7, 0.15, 1.0)$, for $m = 5, 10$ and 20 lags. All are close to the theoretical asymptotic distributions. The relative rejection frequency of the test statistics, at the asymptotic levels 5%, are 5.0%, 5.8%, 3.2% and 5.9% for 5, 10, 20 and 50 lags respectively.

Additional simulation results are presented in the supplementary file.

2.6 Application: Are GARCH(1,1)-type models suitable for peripheral markets?

Despite the development of numerous extensions, short memory models, and in particular GARCH(1,1) specifications, remain the preferred choice for most academics and practitioners when studying volatility. However, the weak persistence they impose might be too restrictive to accurately model some financial time series. We propose to test the GARCH(1,1), TGARCH(1,1), and GJR-GARCH(1,1) specifications on a broad set of equity indices to verify whether their preeminence is justified.

Our dataset contains daily returns from January 1995 to December 2020⁵ of 30 indices in their local currency, from five regions with the following breakdown: 4 in North America (S&P500,

⁵Data for the FTSE MIB and the MOEX start respectively in September 1997 and January 1998.

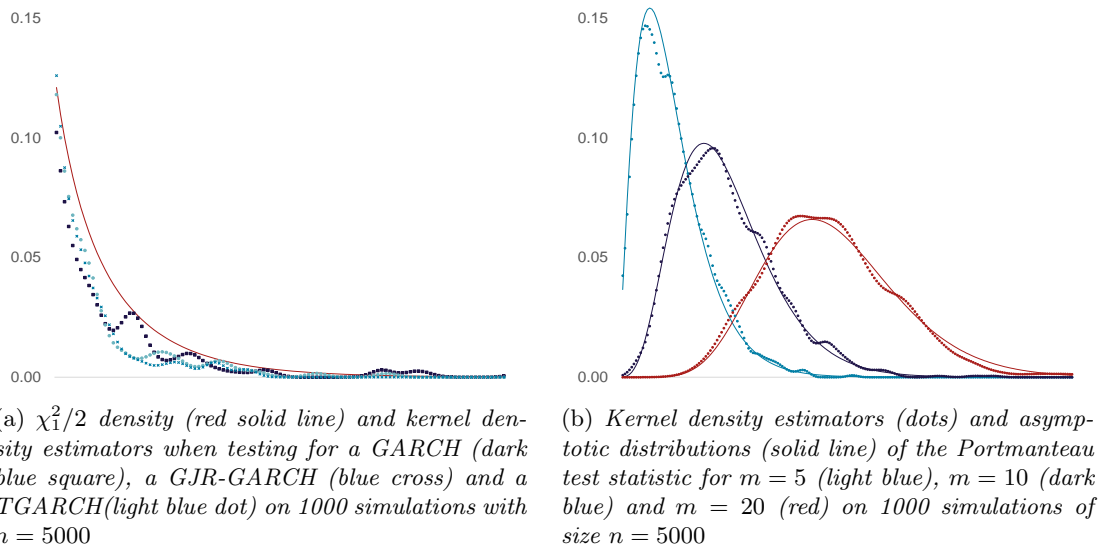


Figure 2.3: Kernel density estimators for the GARCH-type test statistics and for the goodness-of-fit test under their respective null hypothesis

Nasdaq, TSX, Mexico IPC), 11 in Europe (FTSE, DAX, CAC, SMI, AEX, FTSE MIB, IBEX, MOEX, WIG, BUX, TA-125), 10 in Asia (Nikkei, KOSPI, Hang Seng, TAIEX, MSCI Singapore, BSET, PSEi, IDX, KLCI, NIFTY), 2 in Oceania (ASX AO, MSCI New Zealand), and 3 in South America (Merval, Bovespa, IGBVL).

Table 2.2 presents the p-values of the statistics for the symmetry test, and the GARCH-type tests presented in Section 3. The vast majority of indices reject the symmetry assumption, which is a classical result in the financial literature. However, almost half of the thirty indices reject the hypothesis of a GARCH(1,1) specification, and eight reject the GJR-GARCH(1,1) or the TGARCH(1,1) model at the 5% level. Interestingly, all the indices that reject the hypothesis of short memory are from emerging markets. This suggests that the level of development of a financial market has implications on the persistence patterns exhibited by its assets. A possible explanation stems from the difficulty to invest in peripheral markets with fewer investors and with less liquid instruments, which leads to a slower integration of shocks and ultimately higher persistence.

In addition, we propose to study the ability of our model to forecast tail risk measures. We study six competing models, corresponding to the last columns of Table 2.2. The first column is the standard GARCH(1,1) process, the second is an ARCH(∞) model that nests the GARCH(1,1) similarly to equation (2.7), the third column corresponds to the GJR-GARCH(1,1), the fourth to the TARARCH(∞) model (2.8) with $\delta = 2$, the fifth column corresponds to the TGARCH(1,1) and finally the last column corresponds to the APARCH(∞) model (2.8) with $\delta = 1$. For each specification, we fit the model on the sample from 1995 to 2017, and compute daily one-day

ahead forecasts for the 95%-Value-at-Risk using the residuals obtained from each models

$$\text{VaR}_{95\%} = \hat{\sigma}_{t+1}^\delta F_{\hat{\eta}_t}^{-1}(0.05) \text{ with } \hat{\eta}_t = \varepsilon_t / \hat{\sigma}_t$$

where $F_{\hat{\eta}_t}$ is a non parametric estimator of the distribution of the residuals. We thus obtain approximately 750 forecasts for each index ranging from January 2018 to December 2020. The last six columns of Table 2.2 give the frequency of violation of the VaR forecasts in each model for every index. Aside from the Bovespa index, all the competing models have a frequency of violations that is not statistically different from 5% using Kupiec's test (Kupiec, 1995) at the 95% confidence level. From this perspective, it thus seems difficult to choose between the studied specifications. We therefore propose to use the Model Confidence Set (MCS) procedure of Hansen et al. (2011) to select the best predictive models amongst our competitors. The idea of the MCS procedure is to sequentially eliminate competitors until the set of remaining models does not reject the hypothesis of equal predictive ability. To perform this series of tests, we used the R package developed by Bernardi and Catania (2018) and used the asymmetric VaR loss function of González-Rivera et al. (2004) to compute the losses associated with the VaR forecasts. The results of the MCS procedures are also presented in Table 2.2. For each index, the models included in the Superior Set of Models (SSM) at the 80%-confidence level are marked with a star. A notable result is that for more than a third of the studied indices, the GARCH(1,1) and ARCH(∞) models are excluded from the SSM. This is a clear argument for the use of asymmetric models, even if the impact of asymmetry is less obvious on tail measures than on volatility. Surprisingly, however, asymmetric short memory models are often included in the SSM even if the hypothesis of nullity of γ_0 is rejected for some indices. Finally, the APARCH(∞) model with $\delta = 1$ is always included in the SSM, which seems to validate the pertinence of our model.

2.7 Concluding remarks

Although the leverage effect and the long memory of volatility are well established empirical facts of financial time series, models aiming at capturing both properties have been scarce. In this paper, we proposed an extension of the ARCH(∞) model of Robinson (1991) to account for high persistence in power-transformed returns and conditional asymmetry. We proved the existence of a stationary solution and derived statistical inference results. In particular, we proved the consistency and asymptotic normality of QMLE. We showed that the APARCH(∞) representation nests some of the most used models in the financial industry and introduced a Portmanteau type goodness-of-fit test to verify the adequacy of such models. We derived test procedures for conditional asymmetry and to verify whether GARCH(1,1)-type memory patterns are sufficient to model financial returns. In this regard, the results of the application on real data provide a remarkable argument for the use of moderate memory models when studying peripheral assets. We showed that in our database, most of the emerging markets equity indices exhibit a stronger persistence than the standard APARCH(1,1) allows for. The study of conditional Value-at-Risk measures seems to validate the pertinence of the proposed

Index	Country	Symmetry tests			GARCH-type tests			Frequency of rejections of the VaR _{95%} (in %)					
		W_n^{sym}	R_n^{sym}	L_n^{sym}	W_n^{GARCH}	W_n^{GJR}	W_n^{TGARCH}	GARCH	ARCH(∞)	GJR	TARCH(∞)	TGARCH	APARCH(∞)
S&P 500	USA	0.000	0.000	0.000	0.958	0.991	0.996	5.04	5.04	4.24*	4.38	5.17*	5.17*
Nasdaq 100	USA	0.000	0.000	0.000	0.858	0.996	0.993	5.17*	5.17*	4.77*	4.77*	5.44*	5.44*
S&P/TSX	Canada	0.000	0.000	0.000	0.392	0.614	0.955	5.59*	5.59*	4.79*	4.79*	4.92*	4.92*
Mexico IPC	Mexico	0.000	0.000	0.000	0.000	0.156	0.066	5.60*	4.93*	5.20*	5.20*	4.93*	4.93*
FTSE 100	UK	0.000	0.000	0.000	0.204	0.875	0.890	5.15*	5.15*	5.15*	5.15*	4.88*	4.88*
DAX 30	Germany	0.000	0.000	0.000	0.966	0.995	0.999	5.56	5.56	4.90*	4.90*	5.30*	5.30*
CAC 40	France	0.000	0.000	0.000	0.943	1.000	0.997	5.75	5.88	4.71*	4.71*	5.36*	5.36*
SMI	Switzerland	0.000	0.000	0.000	0.930	0.993	0.991	4.67	4.67	5.07*	5.07*	5.34*	5.34*
AEX	Netherlands	0.000	0.000	0.000	0.960	0.995	0.998	5.75	5.75	5.10*	5.10*	5.36*	5.36*
FTSE MIB	Italy	0.000	0.000	0.000	0.947	0.993	0.996	5.80	5.80	5.01*	5.01*	5.54*	5.54*
IBEX 35	Spain	0.000	0.000	0.000	0.953	0.992	0.998	5.49	5.49	4.71*	4.71*	4.84*	4.84*
MOEX	Russia	0.003	0.000	0.000	0.068	0.066	0.146	3.72*	3.72*	3.72*	3.59	3.86*	3.86*
WIG	Poland	0.000	0.000	0.000	0.034	0.160	0.158	5.50	5.63*	5.36*	5.23*	5.50*	5.50*
BUX	Hungary	0.000	0.000	0.000	0.000	0.000	0.000	3.92*	4.59*	4.19*	4.05*	4.05*	4.59*
TAI25	Israel	0.000	0.000	0.000	0.031	0.942	0.972	5.31	5.18	5.31	5.31	5.18*	5.18*
Nikkei 225	Japan	0.000	0.000	0.000	0.625	0.978	0.987	4.81	4.81	4.54	4.54	4.95*	4.95*
KOSPI	South Korea	0.000	0.000	0.000	0.402	0.937	0.937	5.03*	5.03*	5.30	5.30	5.16*	5.16*
Hang Seng	Hong Kong	0.000	0.000	0.000	0.967	0.995	0.998	6.23*	6.23*	5.28*	5.28*	5.42*	5.42*
TAIEX	Taiwan	0.000	0.000	0.000	0.530	0.986	0.997	4.52	4.52	4.52	4.52	4.79*	4.79*
MSCI Singapore	Singapore	0.000	0.000	0.000	0.035	0.199	0.259	4.66*	5.06*	4.93*	5.06*	4.93*	4.93*
BSET	Thailand	0.000	0.000	0.000	0.003	0.049	0.110	6.02	6.29*	5.61*	5.75*	5.20*	5.20*
PSEi	Philippines	0.001	0.000	0.000	0.112	0.004	0.021	5.36*	5.36*	5.50*	5.36*	4.81*	4.95*
IDX	Indonesia	0.001	0.000	0.000	0.000	0.000	0.000	5.81*	5.67*	4.84*	4.98*	4.98*	4.98*
FTSE KLCI	Malaysia	0.000	0.000	0.000	0.000	0.002	0.000	6.55*	6.28*	6.14*	6.28*	5.87*	6.14*
NIFTY 50	India	0.000	0.000	0.000	0.020	0.440	0.656	5.02*	5.02*	4.75*	4.75*	4.61*	4.75*
ASX AO	Australia	0.000	0.000	0.000	0.019	0.643	0.778	5.93*	5.53*	5.93*	6.06*	6.32*	6.32*
MSCI NZ	New Zealand	0.032	0.054	0.047	0.000	0.000	0.000	5.05*	5.45*	4.92*	5.32*	5.19*	5.85*
Merval	Argentina	0.005	0.000	0.000	0.000	0.000	0.002	6.74	6.88	6.74*	6.74*	7.02*	6.74*
Bovespa	Brazil	0.000	0.000	0.000	0.187	0.994	0.989	3.64*	3.64*	3.64*	3.64*	3.78*	3.78*
S&P/BVL	Peru	0.002	0.000	0.000	0.000	0.000	0.000	4.51*	5.17*	4.77*	4.77*	4.91*	5.31*

Table 2.2: Symmetry tests, GARCH-memory tests, and validation of VaR forecasts on different equity indices.
* indicate models included in the Superior Set of Models for VaR forecasts

extensions. Although it would be of interest to derive asymptotic results for such quantities under stronger persistence, we leave this problem for future research.

2.8 Appendix A: Proofs and technical results

2.8.1 Existence of a stationary APARCH(∞) solution

We develop in this section the proof of Theorem 2.1. The proof is based on a Volterra expansion and, in this sense, follows the work of Giraitis et al. (2000), Kazakevičius and Leipus (2002), and Douc et al. (2008).

Proof of Theorem 2.1. First, let us remark that $\sigma_t > 0$ which implies for any t , $\mathbb{1}_{\varepsilon_t \geq 0} = \mathbb{1}_{\eta_t \geq 0}$, and consider the random variable

$$S_t = \omega + \omega \sum_{k=1}^{\infty} \sum_{i_1, \dots, i_k \geq 1} a_{i_1, t-i_1} \dots a_{i_k, t-i_1-\dots-i_k} |\eta_{t-i_1}|^{\delta} \dots |\eta_{t-i_1-\dots-i_k}|^{\delta}$$

defined in $[0, +\infty]$. From the independence of (η_t) and since $s \in (0, 1]$, we have

$$\begin{aligned} \mathbb{E} S_t^s &\leq \omega^s + \omega^s \sum_{k=1}^{\infty} \sum_{i_1, \dots, i_k \geq 1} \mathbb{E} \left([(\alpha_{i_1}^+)^s \mathbb{1}_{\eta_{t-i_1} \geq 0} + (\alpha_{i_1}^-)^s \mathbb{1}_{\eta_{t-i_1} < 0}] |\eta_{t-i_1}|^{\delta s} \right) \dots \\ &\quad \mathbb{E} \left([(\alpha_{i_k}^+)^s \mathbb{1}_{\eta_{t-i_1-\dots-i_k} \geq 0} + (\alpha_{i_k}^-)^s \mathbb{1}_{\eta_{t-i_1-\dots-i_k} < 0}] |\eta_{t-i_1-\dots-i_k}|^{\delta s} \right), \end{aligned}$$

and thus

$$\mathbb{E} S_t^s \leq \omega^s \left[1 + \sum_{k=1}^{\infty} (A_s^+ \mu_{\delta s}^+ + A_s^- \mu_{\delta s}^-)^k \right] \leq \frac{\omega^s}{1 - (A_s^+ \mu_{\delta s}^+ + A_s^- \mu_{\delta s}^-)} < \infty,$$

whence S_t is finite almost surely. In addition, we have

$$\sum_{i=1}^{\infty} a_{i, t-i} S_{t-i} |\eta_{t-i}|^{\delta} = \omega \sum_{k=0}^{\infty} \sum_{i_0, \dots, i_k \geq 1} a_{i_0, t-i_0} \dots a_{i_k, t-i_0-\dots-i_k} |\eta_{t-i_0}|^{\delta} \dots |\eta_{t-i_0-\dots-i_k}|^{\delta}$$

and thus we obtain the recursive equation $S_t = \omega + \sum_{i=1}^{\infty} a_{i, t-i} S_{t-i} |\eta_{t-i}|^{\delta}$. By setting $\varepsilon_t = S_t^{1/\delta} \eta_t$, we obtain a strictly stationary and nonanticipative solution of (2.3) and $\mathbb{E} |\varepsilon_t|^{\delta s} \leq \mu_{\delta s} \omega^s / (1 - (A_s^+ \mu_{\delta s}^+ + A_s^- \mu_{\delta s}^-))$. In addition, Theorem 36.4 in Billingsley (1995) entails the ergodicity of the stationary solution.

Now denote by (ε_t^*) any strictly stationary and nonanticipative solution of (2.3) such that $\mathbb{E} |\varepsilon_t^*|^{\delta s} < \infty$. For all $q \geq 1$, by q recursive substitutions of the $\varepsilon_{t-i}^{*\delta}$, we obtain

$$\begin{aligned} \sigma_t^{\delta} &= \omega + \sum_{i=1}^{\infty} a_{i, t-i} |\varepsilon_{t-i}^*|^{\delta} \\ &= \left\{ \omega + \omega \sum_{k=1}^q \sum_{i_1, \dots, i_k \geq 1} a_{i_1, t-i_1} \dots a_{i_k, t-i_1-\dots-i_k} |\eta_{t-i_1}|^{\delta} \dots |\eta_{t-i_1-\dots-i_k}|^{\delta} \right\} \\ &\quad + \sum_{i_1, \dots, i_{q+1} \geq 1} a_{i_1, t-i_1} \dots a_{i_{q+1}, t-i_1-\dots-i_{q+1}} |\eta_{t-i_1}|^{\delta} \dots |\eta_{t-i_1-\dots-i_{q+1}}|^{\delta} |\varepsilon_{t-i_1-\dots-i_{q+1}}^*|^{\delta} \\ &:= \{S_{t,q}\} + R_{t,q}. \end{aligned}$$

Since (ε_t^*) is nonanticipative, it is independent of $\eta_{t'}$ for any $t' > t$. Hence, since $s \in (0, 1]$,

$$\mathbb{E}R_{t,q}^s \leq (A_s^+ \mu_{\delta s}^+ + A_s^- \mu_{\delta s}^-)^q (A_s^+ \mathbb{E}|\mathbb{1}_{\eta_t \geq 0} \varepsilon_t^*|^{\delta s} + A_s^- \mathbb{E}|\mathbb{1}_{\eta_t < 0} \varepsilon_t^*|^{\delta s})$$

Since $A_s^+ \mu_{\delta s}^+ + A_s^- \mu_{\delta s}^- < 1$, we have $\sum_{q \geq 1} \mathbb{E}R_{t,q}^s < \infty$, whence $R_{t,q}$ tends to 0 almost surely as $q \rightarrow \infty$. Furthermore, $S_{t,q}$ tends to S_t almost surely as $q \rightarrow \infty$, which implies $\sigma_t^\delta = S_t$ almost surely and yields $\varepsilon_t^* = \varepsilon_t$ almost surely, hence concluding the proof. \square

2.8.2 Statistical inference of an APARCH(∞) process

We develop in this section the proofs of the main results of Section 2 on consistency and asymptotic normality of the QMLE in our model. Note that in the following proofs, it will not be restrictive to assume $\rho < 1$.

Let us define the theoretical criterion function and its corresponding minimizer

$$Q_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{t=1}^n l_t(\boldsymbol{\theta}), \quad l_t(\boldsymbol{\theta}) = \log \sigma_t^2(\boldsymbol{\theta}) + \frac{\varepsilon_t^2}{\sigma_t^2(\boldsymbol{\theta})}, \quad \hat{\boldsymbol{\theta}}_n = \underset{\boldsymbol{\theta} \in \Theta}{\text{Argmin}} Q_n(\boldsymbol{\theta}).$$

The theoretical QMLE $\hat{\boldsymbol{\theta}}_n$ is infeasible, and we will thus study the feasible estimator $\tilde{\boldsymbol{\theta}}_n$, which is conditional to initial values. We will show that the choice of the initial values is unimportant for the asymptotic properties of the QMLE.

In the following, we denote $\mathcal{I}^+(\boldsymbol{\phi})$ (respectively $\mathcal{I}^-(\boldsymbol{\phi})$) the sets $\{i \text{ such that } \alpha_i^{+(-)}(\boldsymbol{\phi}) \neq 0\}$, and we define \mathcal{I}_t^+ (respectively \mathcal{I}_t^-) as $\mathcal{I}_t^{+(-)} = \{i \text{ such that } \varepsilon_{t-i} \geq 0 \text{ (resp. } < 0)\}$, yielding the following rewriting of (2.9) as

$$\sigma_t^\delta(\boldsymbol{\theta}_0) = \omega_0 + \sum_{i \in \mathcal{I}_t^+} \alpha_i^+(\boldsymbol{\phi}_0) |\varepsilon_{t-i}|^\delta + \sum_{j \in \mathcal{I}_t^-} \alpha_j^-(\boldsymbol{\phi}_0) |\varepsilon_{t-j}|^\delta. \quad (2.20)$$

We first state and prove the property mentioned in the remark about assumption **A5**.

Proposition 2.2. *Under assumptions A1-A4, if there exists $0 < \tau < \rho - (d+1)^{-1}$ such that*

$$\sup_{i \in \mathcal{I}^+(\boldsymbol{\phi}_0)} \sup_{\boldsymbol{\phi} \in \Phi} \frac{\alpha_i^+(\boldsymbol{\phi}_0)}{(\alpha_i^+)^{1-\tau}(\boldsymbol{\phi})} \leq K \text{ and } \sup_{i \in \mathcal{I}^-(\boldsymbol{\phi}_0)} \sup_{\boldsymbol{\phi} \in \Phi} \frac{\alpha_i^-(\boldsymbol{\phi}_0)}{(\alpha_i^-)^{1-\tau}(\boldsymbol{\phi})} \leq K, \quad (2.21)$$

then

$$\mathbb{E}_{\boldsymbol{\theta}_0} \sup_{\boldsymbol{\theta} \in \Theta} \frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2(\boldsymbol{\theta})} < \infty.$$

Condition (2.21) limits the set of eligible functions $\alpha_i^{+(-)}$ that allows the integrability of $\varepsilon_t^2/\sigma_t^2(\boldsymbol{\theta})$ without additional moment assumptions on ε_t . In particular, functions such that $\inf_{\boldsymbol{\phi} \in \Phi} \alpha_i^{+(-)}(\boldsymbol{\phi}) = 0$ when $\alpha_i^{+(-)}(\boldsymbol{\phi}_0) \neq 0$ are excluded. Note however that the latter case does not exclude every $\alpha_i^{+(-)}$ that can be equal to 0. For example, $\alpha_i(\boldsymbol{\phi}) = [(1 + \sin(i\pi/4))/2] \gamma i^{-(d+1)}$ where $\boldsymbol{\phi} = (\gamma, d)$ and $\Phi \subset (0, \infty)^2$ is periodically equal to 0 and still verifies (2.21). The requirement for higher

moment for ε_t is also required for the estimation of ARCH(q) models or GARCH models when coefficients are equal to 0 (see for example Francq and Zakoian (2007) or Cavaliere et al. (2022)).

Proof of Proposition 2.2. Let us first note that if $c > 0$ and for all i in a set I , $a_i \geq 0$ and $b_i \geq 0$ then $\frac{\sum_{i \in I} a_i}{c + \sum_{j \in I} b_j} \leq \sum_{i \in I} \frac{a_i}{c + b_i}$. Since $\omega_L > 0$ and for all $\theta \in \Theta$ we have $\alpha_i^{+(-)}(\theta) \geq 0$, using the previous elementary inequality and the fact that for any $s > 0$, we have $x/(1+x) \leq x^s$, equation (2.20) gives

$$\begin{aligned} & \sup_{\theta \in \Theta} \frac{\sigma_t^\delta(\theta_0)}{\sigma_t^\delta(\theta)} \\ & \leq \frac{\omega_0}{\omega} + \sup_{\theta \in \Theta} \sum_{i \in \mathcal{I}_t^+ \cap \mathcal{I}^+(\phi_0)} \frac{\alpha_i^+(\phi_0) |\varepsilon_{t-i}|^\delta}{\omega + \alpha_i^+(\phi) |\varepsilon_{t-i}|^\delta} + \sup_{\theta \in \Theta} \sum_{i \in \mathcal{I}_t^- \cap \mathcal{I}^-(\phi_0)} \frac{\alpha_i^-(\phi_0) |\varepsilon_{t-i}|^\delta}{\omega + \alpha_i^-(\phi) |\varepsilon_{t-i}|^\delta} \\ & \leq K + K \sup_{\theta \in \Theta} \sum_{i \in \mathcal{I}_t^+ \cap \mathcal{I}^+(\phi_0)} (\alpha_i^+)^{s-\tau}(\phi) |\varepsilon_{t-i}|^{\delta s} + K \sup_{\theta \in \Theta} \sum_{i \in \mathcal{I}_t^- \cap \mathcal{I}^-(\phi_0)} (\alpha_i^-)^{s-\tau}(\phi) |\varepsilon_{t-i}|^{\delta s} \\ & \leq K + K \sum_{i=1}^{\infty} i^{-(d+1)(s-\tau)} |\varepsilon_{t-i}|^{\delta s} \end{aligned}$$

using assumptions **A3(ii)** and (2.21). This yields

$$\mathbb{E}_{\theta_0} \sup_{\theta \in \Theta} \frac{\sigma_t^\delta(\theta_0)}{\sigma_t^\delta(\theta)} \leq K + \omega^{-s} K' \sum_{i=1}^{\infty} i^{-(d+1)(s-\tau)} \mathbb{E}_{\theta_0} |\varepsilon_{t-i}|^{\delta s}.$$

By taking $s = \rho$ we have that $(d+1)(s-\tau) > 1$ by assumption **A4**, we thus obtain

$$\mathbb{E}_{\theta_0} \sup_{\theta \in \Theta} \frac{\sigma_t^\delta(\theta_0)}{\sigma_t^\delta(\theta)} < \infty.$$

If $\delta < 2$, using Minkowski inequality and assumption **A4**, we obtain

$$\left[\mathbb{E}_{\theta_0} \sup_{\theta \in \Theta} \left(\frac{\sigma_t^\delta(\theta_0)}{\sigma_t^\delta(\theta)} \right)^{2/\delta} \right]^{\delta/2} \leq K + K \sum_{i=1}^{\infty} i^{-(d+1)(\rho-\tau)} [\mathbb{E}_{\theta_0} |\varepsilon_{t-i}|^{2\rho}]^{\delta/2} < \infty$$

from assumption **A3(ii)**, and if $\delta \geq 2$, Jensen inequality yields $\mathbb{E}_{\theta_0} \sup_{\theta \in \Theta} \frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)} < \infty$, which concludes the proof. \square

The following lemma shows the asymptotic irrelevance of the initial values on Q_n .

Lemma 2.1. *Under assumptions A1-A5, $\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta} |Q_n(\theta) - \tilde{Q}_n(\theta)| = 0$.*

Proof of Lemma 2.1. Consider

$$Q_n(\theta) - \tilde{Q}_n(\theta) = \frac{1}{n} \sum_{t=1}^n \log \frac{\sigma_t^2(\theta)}{\tilde{\sigma}_t^2(\theta)} + \frac{1}{n} \sum_{t=1}^n \varepsilon_t^2 \left(\frac{1}{\sigma_t^2(\theta)} - \frac{1}{\tilde{\sigma}_t^2(\theta)} \right) := A_n(\theta) + B_n(\theta).$$

and remark that $\sigma_t^2(\theta) \geq \tilde{\sigma}_t^2(\theta)$, since we have $\sigma_t^\delta(\theta) = \tilde{\sigma}_t^\delta(\theta) + \sum_{i=t}^{\infty} a_{i,t-i}(\phi) |\varepsilon_{t-i}|^\delta$. We denote

$\chi_t = \sup_{\boldsymbol{\theta} \in \Theta} |\sigma_t^\delta(\boldsymbol{\theta}) - \tilde{\sigma}_t^\delta(\boldsymbol{\theta})|$, and we have from assumption **A3(ii)**

$$\chi_t = \sup_{\boldsymbol{\theta} \in \Theta} \sum_{i=t}^{\infty} a_{i,t-i}(\boldsymbol{\phi}) |\varepsilon_{t-i}|^\delta \leq K \sum_{i=0}^{\infty} (i+t)^{-(d+1)} |\varepsilon_{-i}|^\delta,$$

whence $\mathbb{E}\chi_t^\rho \leq K \sum_{i=0}^{\infty} (i+t)^{-(d+1)\rho} \mathbb{E}|\varepsilon_{-i}|^{\delta\rho}$. Since from assumption **A4**, $\mathbb{E}|\varepsilon_t|^{\delta\rho} < \infty$, with $\rho(d+1) > 1$, and since for any $k > 1$ we have $\int_t^\infty x^{-k} dx = (k-1)^{-1} t^{-k+1}$, we obtain $\mathbb{E}\chi_t^\rho \leq K t^{-(d+1)\rho+1}$. This shows that χ_t has a finite moment of order ρ and thus is finite almost surely. Furthermore, since $\rho(d+1) > 1$, the dominated convergence theorem entails $\lim_{t \rightarrow \infty} \chi_t = 0$ almost surely. Then, we have

$$|A_n(\boldsymbol{\theta})| = \frac{2}{\delta n} \sum_{t=1}^n \log \left[1 + \frac{\sigma_t^\delta(\boldsymbol{\theta}) - \tilde{\sigma}_t^\delta(\boldsymbol{\theta})}{\tilde{\sigma}_t^2(\boldsymbol{\theta})} \right] \leq \frac{K}{n} \sum_{t=1}^n \sigma_t^\delta(\boldsymbol{\theta}) - \tilde{\sigma}_t^\delta(\boldsymbol{\theta})$$

since $\log(1+x) \leq x$ for $x \geq 0$ and, for all t , $\tilde{\sigma}_t^2(\boldsymbol{\theta}) \geq \omega$. Therefore, we obtain $\sup_{\boldsymbol{\theta} \in \Theta} |A_n(\boldsymbol{\theta})| \leq K n^{-1} \sum_{t=1}^n \chi_t$ and from Cesaro mean convergence theorem, we obtain $\limsup_{n \rightarrow \infty} \sup_{\boldsymbol{\theta} \in \Theta} |A_n(\boldsymbol{\theta})| = 0$ almost surely.

Consider now

$$|B_n(\boldsymbol{\theta})| \leq \frac{K}{n} \sum_{t=1}^n \eta_t^2 \frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2(\boldsymbol{\theta})} \frac{\max[\sigma_t^{2-\delta}(\boldsymbol{\theta}), \tilde{\sigma}_t^{2-\delta}(\boldsymbol{\theta})]}{\tilde{\sigma}_t^2(\boldsymbol{\theta})} [\sigma_t^\delta(\boldsymbol{\theta}) - \tilde{\sigma}_t^\delta(\boldsymbol{\theta})]$$

whence

$$\sup_{\boldsymbol{\theta} \in \Theta} |B_n(\boldsymbol{\theta})| \leq \frac{K}{n} \sum_{t=1}^n \eta_t^2 \sup_{\boldsymbol{\theta} \in \Theta} \frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2(\boldsymbol{\theta})} \frac{\max[\sigma_t^{2-\delta}(\boldsymbol{\theta}), \tilde{\sigma}_t^{2-\delta}(\boldsymbol{\theta})]}{\tilde{\sigma}_t^2(\boldsymbol{\theta})} \chi_t.$$

If $\delta \geq 2$, $\sigma_t^{2-\delta}(\boldsymbol{\theta}) \leq \tilde{\sigma}_t^{2-\delta}(\boldsymbol{\theta})$ and since $\tilde{\sigma}_t^{-\delta}(\boldsymbol{\theta}) \leq \omega^{-\delta} < \infty$ from **A1**, we have

$$\sup_{\boldsymbol{\theta} \in \Theta} |B_n(\boldsymbol{\theta})| \leq \frac{K}{n} \sum_{t=1}^n \eta_t^2 \sup_{\boldsymbol{\theta} \in \Theta} \frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2(\boldsymbol{\theta})} [\tilde{\sigma}_t^2(\boldsymbol{\theta})]^{-\delta} \chi_t \leq \frac{K}{n} \sum_{t=1}^n \eta_t^2 \sup_{\boldsymbol{\theta} \in \Theta} \frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2(\boldsymbol{\theta})} \chi_t.$$

If $\delta < 2$, $\sigma_t^{2-\delta}(\boldsymbol{\theta}) \geq \tilde{\sigma}_t^{2-\delta}(\boldsymbol{\theta})$ and we have

$$\sup_{\boldsymbol{\theta} \in \Theta} |B_n(\boldsymbol{\theta})| \leq \frac{K}{n} \sum_{t=1}^n \eta_t^2 \sup_{\boldsymbol{\theta} \in \Theta} \frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2(\boldsymbol{\theta})} \frac{\sigma_t^2(\boldsymbol{\theta})}{\tilde{\sigma}_t^2(\boldsymbol{\theta})} \chi_t.$$

From assumptions **A3(ii)** and **A4**, we have

$$\begin{aligned} \sup_{\boldsymbol{\theta} \in \Theta} \eta_t^2 \frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2(\boldsymbol{\theta})} \frac{\sigma_t^2(\boldsymbol{\theta})}{\tilde{\sigma}_t^2(\boldsymbol{\theta})} &= \sup_{\boldsymbol{\theta} \in \Theta} \eta_t^2 \frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2(\boldsymbol{\theta})} \left[\frac{\sigma_t^\delta(\boldsymbol{\theta})}{\tilde{\sigma}_t^\delta(\boldsymbol{\theta})} \right]^{2/\delta} \\ &\leq K \sup_{\boldsymbol{\theta} \in \Theta} \eta_t^2 \frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2(\boldsymbol{\theta})} \left[1 + \sum_{i=0}^{\infty} i^{-d-1} |\varepsilon_{-i}|^\delta \right]^{2/\delta} \\ &\leq K \sup_{\boldsymbol{\theta} \in \Theta} \eta_t^2 \frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2(\boldsymbol{\theta})} \end{aligned} \quad (2.22)$$

where K is finite almost surely and does not depend on t since $\sum_{i=0}^{\infty} i^{-(d+1)} |\varepsilon_{-i}|^\delta$ admits a moment of order ρ and thus is finite almost surely.

Thus, we have

$$\sup_{\boldsymbol{\theta} \in \Theta} |B_n(\boldsymbol{\theta})| \leq \frac{K}{n} \sum_{t=1}^n \eta_t^2 \sup_{\boldsymbol{\theta} \in \Theta} \frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2(\boldsymbol{\theta})} \chi_t.$$

By ergodicity and independance of η_t^2 with σ_t^2 , we have that $\frac{1}{n} \sum_{t=1}^n \eta_t^2 \sup_{\boldsymbol{\theta} \in \Theta} \frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2(\boldsymbol{\theta})}$ tends to

$\mathbb{E} \eta_t^2 \mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta} \frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2(\boldsymbol{\theta})}$ almost surely as n tends to infinity. Since $\chi_t \rightarrow 0$ almost surely and

$\mathbb{E} \eta_t^2 \mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta} \frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2(\boldsymbol{\theta})} < \infty$ by assumptions **A2** and **A5**, from Toeplitz lemma we obtain $\lim_{n \rightarrow \infty} \sup_{\boldsymbol{\theta} \in \Theta} |B_n(\boldsymbol{\theta})| = 0$ almost surely, which concludes the proof. \square

Proof of Theorem 2.2. The proof of the strong consistency of the QMLE is achieved by proving the four following intermediate results and using a compactness argument:

- (a) $\lim_{n \rightarrow \infty} \sup_{\boldsymbol{\theta} \in \Theta} |Q_n(\boldsymbol{\theta}) - \tilde{Q}_n(\boldsymbol{\theta})| = 0$
- (b) $\exists t \in \mathbb{Z}$ such that $\sigma_t^\delta(\boldsymbol{\theta}) = \sigma_t^\delta(\boldsymbol{\theta}_0)$ a.s. $\Rightarrow \boldsymbol{\theta} = \boldsymbol{\theta}_0$
- (c) $\mathbb{E}_{\boldsymbol{\theta}_0} |l_t(\boldsymbol{\theta}_0)| < \infty$, and if $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, $\mathbb{E}_{\boldsymbol{\theta}_0} l_t(\boldsymbol{\theta}) > \mathbb{E}_{\boldsymbol{\theta}_0} l_t(\boldsymbol{\theta}_0)$
- (d) For any $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, there exists a neighborhood $V(\boldsymbol{\theta})$ such that $\liminf_{n \rightarrow \infty} \inf_{\boldsymbol{\theta}^* \in V(\boldsymbol{\theta})} \tilde{Q}_n(\boldsymbol{\theta}^*) > \mathbb{E}_{\boldsymbol{\theta}_0} l_1(\boldsymbol{\theta}_0)$ a.s.

(a) is directly obtained from Lemma 2.1.

Now let $\boldsymbol{\theta} \in \Theta$, such that, for some $t \in \mathbb{Z}$, we have $\sigma_t^\delta(\boldsymbol{\theta}) = \sigma_t^\delta(\boldsymbol{\theta}_0)$ almost surely. Assume $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, and suppose that

$$\alpha_1^+(\boldsymbol{\phi}_0) \mathbb{1}_{\varepsilon_{t-1} \geq 0} + \alpha_1^-(\boldsymbol{\phi}_0) \mathbb{1}_{\varepsilon_{t-1} < 0} \neq \alpha_1^+(\boldsymbol{\phi}) \mathbb{1}_{\varepsilon_{t-1} \geq 0} + \alpha_1^-(\boldsymbol{\phi}) \mathbb{1}_{\varepsilon_{t-1} < 0}. \quad (2.23)$$

Then $([\alpha_1^+(\boldsymbol{\phi}_0) - \alpha_1^+(\boldsymbol{\phi})] \mathbb{1}_{\eta_{t-1} \geq 0} + [\alpha_1^-(\boldsymbol{\phi}_0) - \alpha_1^-(\boldsymbol{\phi})] \mathbb{1}_{\eta_{t-1} < 0}) |\eta_{t-1}|^\delta$ belongs to the σ -field \mathcal{F}_{t-2} generated by $\{\eta_s : s \leq t-2\}$ and thus, by independence, is almost surely constant. Since, from **A2**, η_1 takes at least two positive (respectively negative) values, this implies almost surely $\alpha_1^+(\boldsymbol{\phi}_0) = \alpha_1^+(\boldsymbol{\phi})$ and $\alpha_1^-(\boldsymbol{\phi}_0) = \alpha_1^-(\boldsymbol{\phi})$, which contradicts (2.23). Recursively, we obtain that $\sigma_t^2(\boldsymbol{\theta}) = \sigma_t^2(\boldsymbol{\theta}_0)$ implies that, for all i , $\alpha_i^+(\boldsymbol{\phi}_0) = \alpha_i^+(\boldsymbol{\phi})$ and $\alpha_i^-(\boldsymbol{\phi}_0) = \alpha_i^-(\boldsymbol{\phi})$ and thus, from assumption **A3(i)**, $\boldsymbol{\phi}^+ = \boldsymbol{\phi}_0^+$ and $\boldsymbol{\phi}^- = \boldsymbol{\phi}_0^-$ a.s., whence $\omega = \omega_0$ a.s., and thus $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ almost surely, which proves (b).

We now turn to (c). First, notice that, even if the limit criterion may not be integrable at some point of Θ , it is well defined in $\mathbb{R} \cup \{+\infty\}$. Indeed

$$\mathbb{E}_{\boldsymbol{\theta}_0} [l_t^-(\boldsymbol{\theta})] = \mathbb{E}_{\boldsymbol{\theta}_0} \max[0; -l_t(\boldsymbol{\theta})] \leq \mathbb{E}_{\boldsymbol{\theta}_0} \max[0; -\log \sigma_t^2(\boldsymbol{\theta})] < \infty.$$

Furthermore, we can show that it is integrable at θ_0 . Using Jensen inequality and assumption **A3(ii)**, we obtain

$$\mathbb{E}_{\theta_0} [l_t(\theta_0)] = 1 + \mathbb{E}_{\theta_0} \log \sigma_t^2(\theta_0) \leq 1 + \frac{2}{\delta\rho} \log \left(\omega^\rho + K \sum_{i=1}^{\infty} i^{-(d+1)\rho} \mathbb{E}_{\theta_0} |\varepsilon_{t-i}|^{\delta\rho} \right) < \infty$$

since, from assumption **A4**, $\mathbb{E} |\varepsilon_t|^{\delta\rho} < \infty$ and $\rho(d+1) > 1$. Thus, $\mathbb{E}_{\theta_0} |l_t(\theta_0)|$ is well defined in \mathbb{R} . In addition, we have

$$\begin{aligned} \mathbb{E}_{\theta_0} [l_t(\theta)] - \mathbb{E}_{\theta_0} [l_t(\theta_0)] &= \mathbb{E}_{\theta_0} \left[\log \frac{\sigma_t^2(\theta)}{\sigma_t^2(\theta_0)} \right] + \mathbb{E}_{\theta_0} \left[\frac{\sigma_t^2(\theta_0)\eta_t^2}{\sigma_t^2(\theta)} - \eta_t^2 \right] \\ &\geq -\log \left[\mathbb{E}_{\theta_0} \frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)} \right] + \mathbb{E}_{\theta_0} \left[\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)} \right] - 1 \geq 0 \end{aligned}$$

since, for any $x > 0$, $\log x \leq x - 1$. We can conclude by noticing that $\mathbb{E}_{\theta_0} [l_t(\theta)] = \mathbb{E}_{\theta_0} [l_t(\theta_0)]$ if and only if $\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)} = 1$ almost surely, and thus, by identifiability of the parameter, if and only if $\theta = \theta_0$.

Finally, we turn to the proof of **(d)**. For all $\theta \in \Theta$, and any positive integer k , let $V_k(\theta)$ be the open ball of center θ and radius $1/k$. Because of the asymptotic irrelevance of the initial values, we have

$$\begin{aligned} \liminf_{n \rightarrow \infty} \inf_{\theta^* \in V_k(\theta) \cap \Theta} \tilde{Q}_n(\theta^*) &\geq \liminf_{n \rightarrow \infty} \inf_{\theta^* \in V_k(\theta) \cap \Theta} Q_n(\theta^*) \\ &\quad - \limsup_{n \rightarrow \infty} \sup_{\theta \in \Theta} |Q_n(\theta) - \tilde{Q}_n(\theta)| \\ &\geq \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \inf_{\theta^* \in V_k(\theta) \cap \Theta} l_t(\theta^*). \end{aligned}$$

The sequences $(l_t(\theta^*))$ and $\left(\inf_{\theta^* \in V_k(\theta) \cap \Theta} l_t(\theta^*) \right)$ being measurable functions of ε_t and its past values, they are ergodic and strictly stationary, and admit an expectation in $\mathbb{R} \cup \{\infty\}$. Using the ergodic theorem for non-integrable processes ⁶, we have

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \inf_{\theta^* \in V_k(\theta) \cap \Theta} l_t(\theta^*) = \mathbb{E}_{\theta_0} \inf_{\theta^* \in V_k(\theta) \cap \Theta} l_1(\theta^*).$$

By Beppo-Levi theorem, $\mathbb{E}_{\theta_0} \inf_{\theta^* \in V_k(\theta) \cap \Theta} l_1(\theta^*)$ increases to $\mathbb{E}_{\theta_0} l_1(\theta)$ as $k \rightarrow \infty$. The limit criterion being minimized at the true value θ_0 , we obtain

$$\liminf_{n \rightarrow \infty} \inf_{\theta^* \in V_k(\theta) \cap \Theta} \tilde{Q}_n(\theta^*) > \mathbb{E}_{\theta_0} l_1(\theta_0).$$

The conclusion of the proof uses a compactness argument. First note that for any neighborhood

⁶If (X_t) is an ergodic and strictly stationary process and if $\mathbb{E}X_1$ exists in $\mathbb{R} \cup \{+\infty\}$ then

$$\frac{1}{n} \sum_{t=1}^n X_t \xrightarrow[n \rightarrow \infty]{} \mathbb{E}X_1 \text{ a.s.}$$

$V(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$,

$$\begin{aligned} \limsup_{n \rightarrow \infty} \inf_{\boldsymbol{\theta}^* \in V(\boldsymbol{\theta}_0)} \tilde{Q}_n(\boldsymbol{\theta}^*) &\leq \lim_{n \rightarrow \infty} \tilde{Q}_n(\boldsymbol{\theta}_0) \\ &\leq \lim_{n \rightarrow \infty} Q_n(\boldsymbol{\theta}_0) \\ &\leq \mathbb{E}_{\boldsymbol{\theta}_0} l_1(\boldsymbol{\theta}_0). \end{aligned}$$

The compact set Θ is covered by the union of an arbitrary neighborhood $V(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ and the set of the neighborhoods $V(\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta/V(\boldsymbol{\theta}_0)$, satisfying $\liminf_{n \rightarrow \infty} \inf_{\boldsymbol{\theta}^* \in V(\boldsymbol{\theta})} \tilde{Q}_n(\boldsymbol{\theta}^*) \geq \mathbb{E}_{\boldsymbol{\theta}_0} l_1(\boldsymbol{\theta}_0)$. Thus, there exists a finite subcover of Θ of the form $V(\boldsymbol{\theta}_0), V(\boldsymbol{\theta}_1), \dots, V(\boldsymbol{\theta}_k)$, whence

$$\inf_{\boldsymbol{\theta} \in \Theta} \tilde{Q}_n(\boldsymbol{\theta}) = \min_{i=0,1,\dots,k} \inf_{\boldsymbol{\theta}^* \in V(\boldsymbol{\theta}_i) \cap \Theta} \tilde{Q}_n(\boldsymbol{\theta}^*).$$

We obtain that, for n large enough, $\tilde{\boldsymbol{\theta}}_n$ belongs to $V(\boldsymbol{\theta}_0)$ almost surely. Since this is true for any neighborhood $V(\boldsymbol{\theta}_0)$, we have shown that, almost surely,

$$\tilde{\boldsymbol{\theta}}_n \xrightarrow[n \rightarrow \infty]{} \boldsymbol{\theta}_0.$$

□

We now state and prove the property mentioned in the remark about assumption **A11**.

Proposition 2.3. *Under assumptions A1-A4, if for all $\tau > 0$, there exists a neighborhood $V(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ such that*

$$\sup_{i \in \mathcal{I}^+(\boldsymbol{\phi}_0)} \sup_{\boldsymbol{\phi} \in V(\boldsymbol{\phi}_0)} \frac{\alpha_i^+(\boldsymbol{\phi}_0)}{(\alpha_i^+)^{1-\tau}(\boldsymbol{\phi})} \leq K \text{ and } \sup_{i \in \mathcal{I}^-(\boldsymbol{\phi}_0)} \sup_{\boldsymbol{\phi} \in V(\boldsymbol{\phi}_0)} \frac{\alpha_i^-(\boldsymbol{\phi}_0)}{(\alpha_i^-)^{1-\tau}(\boldsymbol{\phi})} \leq K. \quad (2.24)$$

then, for all $k > 0$, there exists some neighborhood $V(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ such that

$$\mathbb{E}_{\boldsymbol{\theta}_0} \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left[\frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2(\boldsymbol{\theta})} \right]^k < \infty.$$

Proof of Proposition 2.3. For all $s \in (0, 1]$ and $k > s$, (2.10) and Hölder inequality yield

$$\begin{aligned} \sigma_t^\delta(\boldsymbol{\theta}_0) &= \omega_0 \omega^{\frac{s}{k}-1} \omega^{1-\frac{s}{k}} + \sum_{i=1}^{\infty} a_{i,t-i}(\boldsymbol{\phi}_0) a_{i,t-i}^{\frac{s}{k}-1}(\boldsymbol{\phi}) a_{i,t-i}^{1-\frac{s}{k}}(\boldsymbol{\phi}) |\varepsilon_{t-i}|^{\delta \frac{s}{k}} |\varepsilon_{t-i}|^{\delta - \delta \frac{s}{k}} \\ &\leq \left(\sum_{i=0}^{\infty} x_i^{\frac{k}{s}} \right)^{\frac{s}{k}} \left(\sum_{i=0}^{\infty} y_i^{\frac{1}{1-s/k}} \right)^{1-\frac{s}{k}} \\ &\leq K \left[\omega_0^{\frac{k}{s}} \omega^{1-\frac{k}{s}} + \sum_{i=1}^{\infty} a_{i,t-i}^{\frac{k}{s}}(\boldsymbol{\phi}_0) a_{i,t-i}^{1-\frac{k}{s}}(\boldsymbol{\phi}) |\varepsilon_{t-i}|^\delta \right]^{\frac{s}{k}} [\sigma_t^\delta(\boldsymbol{\theta})]^{1-\frac{s}{k}}. \end{aligned}$$

with $x_0 = \omega_0 \omega^{\frac{s}{k}-1}$, $x_i = a_{i,t-i}(\boldsymbol{\phi}_0) a_{i,t-i}^{\frac{s}{k}-1}(\boldsymbol{\phi}) |\varepsilon_{t-i}|^{\delta \frac{s}{k}}$, $y_0 = \omega^{1-\frac{s}{k}}$, $y_i = a_{i,t-i}^{1-\frac{s}{k}}(\boldsymbol{\phi}) |\varepsilon_{t-i}|^{\delta - \delta \frac{s}{k}}$. Since $[\sigma_t^\delta(\boldsymbol{\theta})]^{-\frac{s}{k}} \leq K$, we obtain $[\sigma_t^\delta(\boldsymbol{\theta}_0)/\sigma_t^\delta(\boldsymbol{\theta})]^k \leq$

$$K \left[1 + \sum_{\mathcal{I}_t^+ \cap \mathcal{I}^+(\boldsymbol{\phi}_0)} \frac{(\alpha_i^+)^k(\boldsymbol{\phi}_0)}{(\alpha_i^+)^k(\boldsymbol{\phi})} (\alpha_i^+)^s(\boldsymbol{\phi}) |\varepsilon_{t-i}|^{\delta s} + \sum_{\mathcal{I}_t^- \cap \mathcal{I}^-(\boldsymbol{\phi}_0)} \frac{(\alpha_i^-)^k(\boldsymbol{\phi}_0)}{(\alpha_i^-)^k(\boldsymbol{\phi})} (\alpha_i^-)^s(\boldsymbol{\phi}) |\varepsilon_{t-i}|^{\delta s} \right],$$

whence, from (2.24) and assumptions **A3(ii)** and **A4**, by taking $s = \rho$, there exists a neighborhood such that

$$\mathbb{E}_{\theta_0} \sup_{\theta \in V(\theta_0)} \left[\frac{\sigma_t^\delta(\theta_0)}{\sigma_t^\delta(\theta)} \right]^k \leq K \left[1 + \sum_{i=1}^{\infty} i^{-(d+1)(\rho-k\tau)} \mathbb{E}_{\theta_0} |\varepsilon_{t-i}|^{\delta\rho} \right] < \infty.$$

Indeed, from the arbitrariness of τ , we can find a τ such that $(d+1)(\rho-k\tau) > 1$. \square

Before developing the proofs of Theorems 2.3 and 2.4, it is useful to state the following lemmas. Note that the function $l_t(\theta)$ may be non-defined in a neighborhood of θ_0 when $\theta_0 \in \partial\Theta$ since the volatility process $\sigma_t^\delta(\theta)$ can take negative values. For ease of notation, we denote by $\partial\sigma_t^\delta(\theta_0)/\partial\theta$ the vector of partial derivatives $(\partial\sigma_t^\delta(\theta_0)/\partial\theta_i)_{i=1,\dots,r+1}$ where the j -th derivatives is replaced by the right derivative when $\phi_{0,j} = \underline{\phi}_j$. The same convention is applied to the derivatives of l_t , Q_t , $\tilde{\sigma}_t^\delta$, \tilde{l}_t , and \tilde{Q}_t .

Lemma 2.2. *Under assumptions A1-A10, for all $i_h = 1, \dots, r+1$, $h = 1, \dots, k$, $k \leq 3$, and for all $p > 0$, we have*

$$\mathbb{E}_{\theta_0} \sup_{\theta \in \Theta} \left| \frac{1}{\sigma_t^\delta(\theta)} \frac{\partial^k \sigma_t^\delta(\theta)}{\partial\theta_{i_1} \dots \partial\theta_{i_k}} \right|^p < \infty.$$

Proof of Lemma 2.2. From (2.20) and assumption **A10(i)**, we have, for all $j_1 \in \{1, \dots, r\}$,

$$\frac{\partial\sigma_t^\delta}{\partial\theta_1} = \frac{\partial\sigma_t^\delta}{\partial\omega} = 1 \quad \text{and} \quad \frac{\partial\sigma_t^\delta}{\partial\theta_{1+j_1}} = \frac{\partial\sigma_t^\delta}{\partial\phi_{j_1}} = \sum_{i \in \mathcal{I}_t^+} \frac{\partial\alpha_i^+}{\partial\phi_{j_1}} |\varepsilon_{t-i}|^\delta + \sum_{i \in \mathcal{I}_t^-} \frac{\partial\alpha_i^-}{\partial\phi_{j_1}} |\varepsilon_{t-i}|^\delta. \quad (2.25)$$

It is thus sufficient to show that for all $j_h \in \{1, \dots, r\}$, $h = 1, \dots, k$, $k \leq 3$, we have

$$\mathbb{E}_{\theta_0} \sup_{\theta \in \Theta} \left| \frac{1}{\sigma_t^\delta(\theta)} \frac{\partial^k \sigma_t^\delta(\theta)}{\partial\phi_{j_1} \dots \partial\phi_{j_k}} \right|^p < \infty$$

From (2.25), and assumptions **A3(ii)** and **A10(i)** we have

$$\begin{aligned} \sup_{\theta \in \Theta} \left| \frac{\partial^k \sigma_t^\delta(\theta)}{\partial\phi_{j_1} \dots \partial\phi_{j_k}} \right| &\leq \sup_{\theta \in \Theta} \sum_{i \in \mathcal{I}_t^+} \left| \frac{\partial^k \alpha_i^+(\phi)}{\partial\phi_{j_1} \dots \partial\phi_{j_k}} \right| |\varepsilon_{t-i}|^\delta + \sup_{\theta \in \Theta} \sum_{i \in \mathcal{I}_t^-} \left| \frac{\partial^k \alpha_i^-(\phi)}{\partial\phi_{j_1} \dots \partial\phi_{j_k}} \right| |\varepsilon_{t-i}|^\delta \\ &\leq K \sum_{i \in \mathcal{I}_t^+} \sup_{\theta \in \Theta} (\alpha_i^+)^{(1-\xi)}(\theta) |\varepsilon_{t-i}|^\delta + K \sum_{i \in \mathcal{I}_t^-} \sup_{\theta \in \Theta} (\alpha_i^-)^{(1-\xi)}(\theta) |\varepsilon_{t-i}|^\delta \\ &\leq K \sum_{i=1}^{\infty} i^{-(d+1)(1-\xi)} |\varepsilon_{t-i}|^\delta \end{aligned}$$

and from the Hölder inequality we obtain, for all $p > \rho$

$$\begin{aligned} &\sup_{\theta \in \Theta} \left| \frac{\partial^k \sigma_t^\delta(\theta)}{\partial\phi_{j_1} \dots \partial\phi_{j_k}} \right| \\ &\leq K \left[\sum_{i=1}^{\infty} [i^{-(d+1)(1-\xi)}]^\frac{p}{\rho} a_{i,t-i}^{1-\frac{p}{\rho}}(\phi) |\varepsilon_{t-i}|^\delta \right]^\frac{p}{\rho} \left[\sum_{i=1}^{\infty} a_{i,t-i}(\phi) |\varepsilon_{t-i}|^\delta \right]^{1-\frac{p}{\rho}} \\ &\leq K \left[\sum_{i=1}^{\infty} [i^{-(d+1)(1-\xi)}]^\frac{p}{\rho} a_{i,t-i}^{1-\frac{p}{\rho}}(\phi) |\varepsilon_{t-i}|^\delta \right]^\frac{p}{\rho} [\sigma_t^\delta(\theta)]^{1-\frac{p}{\rho}}, \end{aligned}$$

whence, from assumptions **A3(ii)** and **A10(i)**,

$$\mathbb{E}_{\theta_0} \sup_{\theta \in \Theta} \left| \frac{1}{\sigma_t^2(\theta)} \frac{\partial^k \sigma_t^2(\theta)}{\partial \phi_{j_1,1}^+ \dots \partial \phi_{j_1,k}^+} \right|^p \leq K \sum_{i=1}^{\infty} i^{-(d+1)(\rho-p\xi)} \mathbb{E}_{\theta_0} |\varepsilon_{t-i}|^{\delta\rho}$$

for all $\xi > 0$. Since $\rho > \frac{1}{d+1}$, we may choose ξ such that $(d+1)(\rho-p\xi) > 1$ and thus we have

$$\mathbb{E}_{\theta_0} \sup_{\theta \in \Theta} \left| \frac{1}{\sigma_t^\delta(\theta)} \frac{\partial^k \sigma_t^\delta(\theta)}{\partial \theta_{i_1} \dots \partial \theta_{i_k}} \right|^p < \infty.$$

□

The following lemma shows the integrability of the criterion derivatives at θ_0 .

Lemma 2.3. *Under the assumptions of Theorem 2.3 or Theorem 2.4,*

$$\mathbb{E}_{\theta_0} \left\| \frac{\partial l_t(\theta_0)}{\partial \theta} \frac{\partial l_t(\theta_0)}{\partial \theta'} \right\| < \infty \quad \text{and} \quad \mathbb{E}_{\theta_0} \left\| \frac{\partial^2 l_t(\theta_0)}{\partial \theta \partial \theta'} \right\| < \infty$$

Proof. We have $l_t(\theta) = \log \sigma_t^2(\theta) + \frac{\varepsilon_t^2}{\sigma_t^2(\theta)}$, thus we obtain

$$\begin{aligned} \frac{\partial l_t(\theta)}{\partial \theta} &= \frac{2}{\delta} \left[1 - \frac{\varepsilon_t^2}{\sigma_t^2} \right] \left[\frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \theta} \right] \\ \frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta'} &= \frac{2}{\delta} \left[1 - \frac{\varepsilon_t^2}{\sigma_t^2} \right] \left[\frac{1}{\sigma_t^\delta} \frac{\partial^2 \sigma_t^\delta}{\partial \theta \partial \theta'} \right] + \frac{2}{\delta} \left[\frac{\delta + 2}{\delta} \frac{\varepsilon_t^2}{\sigma_t^2} - 1 \right] \left[\frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \theta} \right] \left[\frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \theta'} \right]. \end{aligned} \quad (2.26)$$

Note that at θ_0 , $\frac{\varepsilon_t^2}{\sigma_t^2(\theta_0)} = \eta_t^2$ is independent of σ_t^2 and its derivatives. It thus suffices to show

$$\mathbb{E}_{\theta_0} \left\| \frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \theta}(\theta_0) \right\| < \infty, \quad \mathbb{E}_{\theta_0} \left\| \frac{1}{\sigma_t^\delta} \frac{\partial^2 \sigma_t^\delta}{\partial \theta \partial \theta'}(\theta_0) \right\| < \infty \quad \text{and} \quad \mathbb{E}_{\theta_0} \left\| \frac{1}{\sigma_t^{2\delta}} \frac{\partial \sigma_t^\delta}{\partial \theta} \frac{\partial \sigma_t^\delta}{\partial \theta'}(\theta_0) \right\| < \infty.$$

The first two inequalities directly follow from Lemma 2.2. Since for all j , $\sigma_t^{-\delta} \partial \sigma_t^\delta / \partial \theta_j$ is bounded at θ_0 , we obtain the last inequality, which concludes the proof. □

The following lemma shows the non-singularity of J and how it connects with the variance of the criterion derivatives.

Lemma 2.4. *Under the assumptions of Theorem 2.3 or Theorem 2.4,*

$$J \text{ is invertible and } \mathbb{V}_{\theta_0} \left[\frac{\partial l_t(\theta_0)}{\partial \theta} \right] = (\kappa_\eta - 1) J$$

Proof. Since at θ_0 , $\varepsilon_t^2 / \sigma_t^2(\theta_0) = \eta_t^2$ is independent of σ_t^2 and its derivatives, we have

$$\mathbb{E}_{\theta_0} \left[\frac{\partial l_t}{\partial \theta}(\theta_0) \right] = \frac{2}{\delta} \mathbb{E}_{\theta_0} [1 - \eta_t^2] \mathbb{E}_{\theta_0} \left[\frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \theta}(\theta_0) \right] = 0$$

from **A2**. Moreover, in view of integrability of the derivatives of the criterion at θ_0 , $J =$

$\mathbb{E}_{\theta_0} \left[\frac{\partial^2 l_t(\theta_0)}{\partial \theta \partial \theta'} \right]$ exists, and from assumption **A7** we can write

$$\mathbb{V}_{\theta_0} \left[\frac{\partial l_t}{\partial \theta}(\theta_0) \right] = \frac{4}{\delta^2} \mathbb{E}_{\theta_0} [(1 - \eta_t^2)^2] \mathbb{E}_{\theta_0} \left[\frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \theta} \frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \theta'}(\theta_0) \right] = (\kappa_\eta - 1) \mathbf{J}.$$

Assume now that \mathbf{J} is singular, then there exists a non-zero vector $\boldsymbol{\Lambda} = [\lambda_0, \boldsymbol{\lambda}']'$, with $\boldsymbol{\lambda} \in \mathbb{R}^r$, such that almost surely $\boldsymbol{\Lambda}' \mathbf{J} \boldsymbol{\Lambda} = 0$, which is equivalent to

$$\lambda_0 + \sum_{i=1}^{\infty} \left[\sum_{j=1}^r \lambda_j \frac{\partial \alpha_i^+(\phi_0)}{\partial \phi_j} \mathbb{1}_{\varepsilon_{t-i} \geq 0} + \sum_{k=1}^r \lambda_k \frac{\partial \alpha_i^-(\phi_0)}{\partial \phi_k} \mathbb{1}_{\varepsilon_{t-i} < 0} \right] |\varepsilon_{t-i}|^\delta = 0.$$

Now, assume $\sum_{j=1}^r \lambda_j \frac{\partial \alpha_1^+(\phi_0)}{\partial \phi_j} \mathbb{1}_{\varepsilon_{t-1} \geq 0} + \sum_{k=1}^r \lambda_k \frac{\partial \alpha_1^-(\phi_0)}{\partial \phi_k} \mathbb{1}_{\varepsilon_{t-1} < 0} \neq 0$, then it follows

$$\begin{aligned} & \left[\sum_{j=1}^r \lambda_j \frac{\partial \alpha_1^+(\phi_0)}{\partial \phi_j} \mathbb{1}_{\eta_{t-1} \geq 0} + \sum_{k=1}^r \lambda_k \frac{\partial \alpha_1^-(\phi_0)}{\partial \phi_k} \mathbb{1}_{\eta_{t-1} < 0} \right] |\eta_{t-1}|^\delta \sigma_{t-1}^\delta(\theta_0) \\ &= -\lambda_0 - \sum_{i=2}^{\infty} \left[\sum_{j=1}^r \lambda_j \frac{\partial \alpha_i^+(\phi_0)}{\partial \phi_j} \mathbb{1}_{\eta_{t-i} \geq 0} + \sum_{k=1}^r \lambda_k \frac{\partial \alpha_i^-(\phi_0)}{\partial \phi_k} \mathbb{1}_{\eta_{t-i} < 0} \right] |\eta_{t-i}|^\delta \sigma_{t-i}^\delta(\theta_0) \end{aligned}$$

whence $\eta_{t-1}^\delta \in \mathcal{F}(\eta_{t-2}^\delta, \dots)$ and thus, by independence, $\sum_{j=1}^r \lambda_j \frac{\partial \alpha_1^+(\phi_0)}{\partial \phi_j} \mathbb{1}_{\eta_{t-1} \geq 0} |\eta_{t-1}|^\delta$ is constant almost surely and thus $\boldsymbol{\lambda}' \frac{\partial \alpha_1^+(\phi_0)}{\partial \phi} = 0$ almost surely since, from assumption **A2**, η_1 takes at least two positive values. Iterating this argument for $\alpha_i^{+(-)}$, we obtain that for all $i_h^{+(-)} = i_h^{+(-)}(\phi_0)$, $i_h^{+(-)} = 1, \dots, r$, we have $\boldsymbol{\lambda}' \frac{\partial \alpha_{i_h}^{+(-)}(\phi_0)}{\partial \phi} = 0$ and thus from assumption **A10(ii)** we must have $\boldsymbol{\lambda} = \mathbf{0}$. This implies $\lambda_0 = 0$ and contradicts the singularity of \mathbf{J} . \square

The following lemma shows the uniform integrability of the second and third order of the criterion derivatives.

Lemma 2.5. *Under the assumptions of Theorem 2.3 or Theorem 2.4, for any $\varepsilon > 0$, there exists a neighborhood $\mathcal{V}(\theta_0)$ of θ_0 such that for all $k_1, k_2, k_3 \in \{1, \dots, r+1\}$,*

$$\mathbb{E}_{\theta_0} \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 l_t(\theta_0)}{\partial \theta_{k_1} \partial \theta_{k_2}} \right| < \infty \text{ and } \mathbb{E}_{\theta_0} \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^3 l_t(\theta_0)}{\partial \theta_{k_1} \partial \theta_{k_2} \partial \theta_{k_3}} \right| < \infty \text{ a.s.}$$

Proof. We have

$$\begin{aligned}
\frac{\partial^3 l_t(\boldsymbol{\theta})}{\partial \theta_{i_1} \partial \theta_{i_2} \partial \theta_{i_3}} &= \frac{2}{\delta} \left\{ \left[1 - \frac{\varepsilon_t^2}{\sigma_t^2} \right] \left[\frac{1}{\sigma_t^\delta} \frac{\partial^3 \sigma_t^\delta}{\partial \theta_{i_1} \partial \theta_{i_2} \partial \theta_{i_3}} \right] \right. \\
&\quad + \left[\frac{\delta + 2}{2} \frac{\varepsilon_t^2}{\sigma_t^2} - 1 \right] \left[\frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \theta_{i_1}} \right] \left[\frac{1}{\sigma_t^\delta} \frac{\partial^2 \sigma_t^\delta}{\partial \theta_{i_2} \partial \theta_{i_3}} \right] \\
&\quad + \left[\frac{\delta + 2}{2} \frac{\varepsilon_t^2}{\sigma_t^2} - 1 \right] \left[\frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \theta_{i_2}} \right] \left[\frac{1}{\sigma_t^\delta} \frac{\partial^2 \sigma_t^\delta}{\partial \theta_{i_1} \partial \theta_{i_3}} \right] \\
&\quad + \left[\frac{\delta + 2}{2} \frac{\varepsilon_t^2}{\sigma_t^2} - 1 \right] \left[\frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \theta_{i_3}} \right] \left[\frac{1}{\sigma_t^\delta} \frac{\partial^2 \sigma_t^\delta}{\partial \theta_{i_1} \partial \theta_{i_2}} \right] \\
&\quad \left. + 2 \left[1 - \frac{\delta^2 + 3\delta + 2}{\delta^2} \frac{\varepsilon_t^2}{\sigma_t^2} \right] \left[\frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \theta_{i_1}} \right] \left[\frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \theta_{i_2}} \right] \left[\frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \theta_{i_3}} \right] \right\} (\boldsymbol{\theta}).
\end{aligned}$$

From assumptions **A7** and **A11**, and the triangular inequality, there exists a neighborhood $V(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ such that,

$$\left\| \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \frac{\varepsilon_t^2}{\sigma_t^2(\boldsymbol{\theta})} \right\|_2 = \sqrt{\kappa_\eta} \left\| \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2(\boldsymbol{\theta})} \right\|_2 < \infty.$$

Using Lemma 2.2, the Cauchy-Schwarz inequality and the Hölder inequality, we have for all $i_1, i_2, i_3 \in \{1, \dots, r+1\}$

$$\mathbb{E}_{\boldsymbol{\theta}_0} \sup_{\boldsymbol{\theta} \in V_\tau(\boldsymbol{\theta}_0)} \left\| \left[1 - \frac{\varepsilon_t^2}{\sigma_t^2(\boldsymbol{\theta})} \right] \left[\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})} \frac{\partial^3 \sigma_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_1} \partial \theta_{i_2} \partial \theta_{i_3}} \right] \right\| < \infty,$$

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}_0} \sup_{\boldsymbol{\theta} \in V_\tau(\boldsymbol{\theta}_0)} \left\| \left[\frac{\delta + 2}{2} \frac{\varepsilon_t^2}{\sigma_t^2(\boldsymbol{\theta})} - 1 \right] \left[\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})} \frac{\partial \sigma_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_1}} \right] \left[\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})} \frac{\partial^2 \sigma_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_2} \partial \theta_{i_3}} \right] \right\| \\
&\leq \left\| \sup_{\boldsymbol{\theta} \in V_\tau(\boldsymbol{\theta}_0)} \left| \frac{\delta + 2}{2} \frac{\varepsilon_t^2}{\sigma_t^2(\boldsymbol{\theta})} - 1 \right| \right\|_2 \left\| \sup_{\boldsymbol{\theta} \in \Theta} \left| \frac{1}{\sigma_t^\delta(\boldsymbol{\theta})} \frac{\partial \sigma_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_1}} \right| \right\|_4 \left\| \sup_{\boldsymbol{\theta} \in \Theta} \left| \frac{1}{\sigma_t^\delta(\boldsymbol{\theta})} \frac{\partial^2 \sigma_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_2} \partial \theta_{i_3}} \right| \right\|_4 \\
&< \infty,
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}_0} \sup_{\boldsymbol{\theta} \in V_\tau(\boldsymbol{\theta}_0)} \left\| \left[1 - \frac{\delta^2 + 3\delta + 2}{\delta^2} \frac{\varepsilon_t^2}{\sigma_t^2(\boldsymbol{\theta})} \right] \left[\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})} \frac{\partial \sigma_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_1}} \right] \left[\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})} \frac{\partial \sigma_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_2}} \right] \left[\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})} \frac{\partial \sigma_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_3}} \right] \right\| \\
&\leq \left\| \sup_{\boldsymbol{\theta} \in V_\tau(\boldsymbol{\theta}_0)} \left| 1 - \frac{\delta^2 + 3\delta + 2}{\delta^2} \frac{\varepsilon_t^2}{\sigma_t^2(\boldsymbol{\theta})} \right| \right\|_2 \max_h \left\| \sup_{\boldsymbol{\theta} \in \Theta} \left| \frac{1}{\sigma_t^\delta(\boldsymbol{\theta})} \frac{\partial \sigma_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_h}} \right| \right\|_6^3 \\
&< \infty,
\end{aligned}$$

which concludes the proof. \square

The following lemma shows the asymptotic irrelevance of the initial values on the derivatives of the criterion.

Lemma 2.6. *Under the assumptions of Theorem 2.3 or Theorem 2.4,*

$$\left\| \frac{1}{\sqrt{n}} \sum_{t=1}^n \left(\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} - \frac{\partial \tilde{l}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right) \right\| \xrightarrow{P} 0 \quad \text{and} \quad \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\| \frac{1}{n} \sum_{t=1}^n \left(\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} - \frac{\partial^2 \tilde{l}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right) \right\| \xrightarrow{P} 0$$

Proof. First, remark that, from assumption **A3(ii)** and **A11**, on a neighborhood $V(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$,

we have similarly to (2.22)

$$\sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \frac{\varepsilon_t^2}{\tilde{\sigma}_t^2(\boldsymbol{\theta})} = \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \eta_t^2 \frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2(\boldsymbol{\theta})} \left[\frac{\sigma_t^\delta(\boldsymbol{\theta})}{\tilde{\sigma}_t^\delta(\boldsymbol{\theta})} \right]^{2/\delta} \leq K \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \eta_t^2 \frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2(\boldsymbol{\theta})} \quad (2.27)$$

where K is finite almost surely and does not depend on t since $\sum_{i=0}^{\infty} i^{-(d+1)} \varepsilon_{-i}^2$ admits a moment of order ρ and thus is finite almost surely.

We have

$$\frac{\partial \tilde{l}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{2}{\delta} \left[1 - \frac{\varepsilon_t^2}{\tilde{\sigma}_t^2} \right] \left[\frac{1}{\tilde{\sigma}_t^\delta} \frac{\partial \tilde{\sigma}_t^\delta}{\partial \boldsymbol{\theta}} \right] (\boldsymbol{\theta}) = \frac{2}{\delta} \left[1 - \eta_t^2 \frac{\sigma_t^2}{\tilde{\sigma}_t^2} \right] \left[\frac{1}{\tilde{\sigma}_t^\delta} \frac{\partial \tilde{\sigma}_t^\delta}{\partial \boldsymbol{\theta}} \right] (\boldsymbol{\theta}),$$

therefore we can write

$$\begin{aligned} \left| \frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \theta_k} - \frac{\partial \tilde{l}_t(\boldsymbol{\theta}_0)}{\partial \theta_k} \right| &= \frac{2}{\delta} \left[\left[\frac{\varepsilon_t^2}{\tilde{\sigma}_t^2} - \frac{\varepsilon_t^2}{\sigma_t^2} \right] \left[\frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \theta_k} \right] + \left[1 - \frac{\varepsilon_t^2}{\tilde{\sigma}_t^2} \right] \left[\frac{1}{\tilde{\sigma}_t^\delta} - \frac{1}{\sigma_t^\delta} \right] \left[\frac{\partial \sigma_t^\delta}{\partial \theta_k} \right] \right. \\ &\quad \left. + \left[1 - \frac{\varepsilon_t^2}{\tilde{\sigma}_t^2} \right] \left[\frac{1}{\tilde{\sigma}_t^\delta} \right] \left[\frac{\partial \sigma_t^\delta}{\partial \theta_k} - \frac{\partial \tilde{\sigma}_t^\delta}{\partial \theta_k} \right] \right| (\boldsymbol{\theta}_0) \\ &= \frac{2}{\delta} |A_t + B_t + C_t| (\boldsymbol{\theta}_0) \end{aligned}$$

From the Markov inequality we have

$$\begin{aligned} \mathbb{P} \left[\left| \frac{1}{\sqrt{n}} \sum_{t=1}^n \left[\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \theta_k} - \frac{\partial \tilde{l}_t(\boldsymbol{\theta}_0)}{\partial \theta_k} \right] \right| > \varepsilon \right] \\ \leq \frac{1}{\varepsilon} \mathbb{E}_{\boldsymbol{\theta}_0} \left[\left| \frac{1}{\sqrt{n}} \sum_{t=1}^n \left[\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \theta_k} - \frac{\partial \tilde{l}_t(\boldsymbol{\theta}_0)}{\partial \theta_k} \right] \right| \right] \\ \leq \frac{1}{\varepsilon} \frac{2}{\delta} \left[\mathbb{E}_{\boldsymbol{\theta}_0} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n A_t(\boldsymbol{\theta}_0) \right| + \mathbb{E}_{\boldsymbol{\theta}_0} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n B_t(\boldsymbol{\theta}_0) \right| + \mathbb{E}_{\boldsymbol{\theta}_0} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n C_t(\boldsymbol{\theta}_0) \right| \right] \end{aligned} \quad (2.28)$$

From (2.27), we have

$$|A_t(\boldsymbol{\theta}_0)| = \left| \left[\frac{\varepsilon_t^2}{\tilde{\sigma}_t^2} - \frac{\varepsilon_t^2}{\sigma_t^2} \right] \left[\frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \theta_k} \right] \right| (\boldsymbol{\theta}_0) \leq K \eta_t^2 \left[\sum_{i=t}^{\infty} a_{i,t-i}(\boldsymbol{\phi}_0) |\varepsilon_{t-i}|^\delta \right] \left| \frac{1}{\sigma_t^\delta(\boldsymbol{\theta}_0)} \frac{\partial \sigma_t^\delta}{\partial \theta_k}(\boldsymbol{\theta}_0) \right|.$$

Using the independence of η_t^2 with σ_t^δ and its derivatives at $\boldsymbol{\theta}_0$, **A2**, **A8** and **A9**, Hölder inequality, and Lemma 2.2,

$$\begin{aligned} \mathbb{E}_{\boldsymbol{\theta}_0} |A_t(\boldsymbol{\theta}_0)|^\rho &\leq K \left(\mathbb{E}_{\boldsymbol{\theta}_0} \left[\sum_{i=t}^{\infty} i^{-(d^*+1)} |\varepsilon_{t-i}|^\delta \right]^{\rho(1+\xi)} \right)^{\frac{1}{1+\xi}} \left(\mathbb{E}_{\boldsymbol{\theta}_0} \left| \frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \theta_k}(\boldsymbol{\theta}_0) \right|^{\rho \frac{\xi+1}{\xi}} \right)^{\frac{\xi}{1+\xi}} \\ &\leq K \sum_{i=0}^{\infty} (t+i)^{-(d^*+1)\rho} \left(\mathbb{E}_{\boldsymbol{\theta}_0} |\varepsilon_{-i}|^{\delta \rho(1+\xi)} \right)^{\frac{1}{1+\xi}} \\ &\leq K t^{-(d^*+1)\rho+1}, \end{aligned} \quad (2.29)$$

for some $\xi > 0$ such that $\rho(1 + \xi) \leq 1$ and thus

$$\mathbb{E}_{\boldsymbol{\theta}_0} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n A_t(\boldsymbol{\theta}_0) \right|^\rho \leq K n^{-(d^* + \frac{3}{2})\rho + 2} \xrightarrow{n \rightarrow \infty} 0$$

since from assumption **A9** we have $(d^* + \frac{3}{2})\rho - 2 > 0$. Using Markov inequality, we can conclude $\frac{1}{\sqrt{n}} \sum_{t=1}^n |A_t(\boldsymbol{\theta}_0)|$ tends to 0 in probability. Similar arguments yield $\frac{1}{\sqrt{n}} \sum_{t=1}^n |B_t(\boldsymbol{\theta}_0)|$ tends to 0 in probability. In addition, from (2.25), and from assumptions **A10(i)** and **A3(ii)**, we have for all $\xi > 0$,

$$|C_t(\boldsymbol{\theta}_0)| = \left| \left[1 - \frac{\varepsilon_t^2}{\bar{\sigma}_t^2} \right] \left[\frac{1}{\bar{\sigma}_t^\delta} \right] \left[\frac{\partial \sigma_t^\delta}{\partial \theta_k} - \frac{\partial \bar{\sigma}_t^\delta}{\partial \theta_k} \right] \right|(\boldsymbol{\theta}_0) \leq K \eta_t^2 \sum_{i=0}^{\infty} (t+i)^{-(d+1)(1-\xi)} |\varepsilon_{-i}|^\delta,$$

and thus

$$\mathbb{E}_{\boldsymbol{\theta}_0} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n C_t(\boldsymbol{\theta}_0) \right|^\rho \leq n^{-\frac{1}{2}\rho} \sum_{t=1}^n \mathbb{E}_{\boldsymbol{\theta}_0} |C_t(\boldsymbol{\theta}_0)|^\rho \leq K n^{-(d^*+1)\rho(1-\xi)+2} \xrightarrow{n \rightarrow \infty} 0$$

since, from assumption **A8** and **A9**, there exists a ξ such that $(d^* + 1)\rho(1 - \xi) - 2 > 0$. Using Markov inequality, we can conclude $\frac{1}{\sqrt{n}} \sum_{t=1}^n |C_t(\boldsymbol{\theta}_0)|$ tends to 0 in probability. Hence (2.28) yields $\mathbb{P} \left[\left| \frac{1}{\sqrt{n}} \sum_{t=1}^n \left[\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \theta_k} - \frac{\partial \tilde{l}_t(\boldsymbol{\theta}_0)}{\partial \theta_k} \right] \right| > \varepsilon \right] \rightarrow 0$ for all $\varepsilon > 0$ which concludes the proof of the first inequality.

Now consider the asymptotic impact of the initial values on the second-order derivatives of the criterion in a neighborhood of $\boldsymbol{\theta}_0$. Let us denote

$$\chi_t := \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} |\sigma_t^\delta(\boldsymbol{\theta}) - \bar{\sigma}_t^\delta(\boldsymbol{\theta})| = \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \sum_{i=t}^{\infty} a_{i,t-i}(\boldsymbol{\phi}) |\varepsilon_{t-i}|^\delta \leq K \sum_{i=0}^{\infty} (i+t)^{-(d+1)} |\varepsilon_{-i}|^\delta \quad (2.30)$$

from assumption **A3(ii)**, whence $\mathbb{E}\chi_t^\rho \leq K t^{-(d+1)\rho+1}$ from **A4**. This shows that χ_t has a finite moment of order ρ and thus is finite almost surely. Furthermore, since $\rho(d+1) > 1$, the dominated convergence theorem entails $\lim_{t \rightarrow \infty} \chi_t = 0$ almost surely. Let us now denote

$$\chi_t^{(i_1)} = \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left| \frac{\partial \sigma_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_1}} - \frac{\partial \bar{\sigma}_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_1}} \right| \quad \text{and} \quad \chi_t^{(i_1, i_2)} = \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left| \frac{\partial^2 \sigma_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_1} \partial \theta_{i_2}} - \frac{\partial^2 \bar{\sigma}_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_1} \partial \theta_{i_2}} \right| \quad (2.31)$$

where $V(\boldsymbol{\theta}_0)$ is a neighborhood of $\boldsymbol{\theta}_0$ and $i_1, i_2 \in \{1, \dots, r\}$. From assumptions **A10(i)** and **A3(ii)**, we have for all $\xi > 0$,

$$\chi_t^{(1+i_1)} \leq \sum_{i=t}^{\infty} \sup_{\boldsymbol{\phi} \in V(\boldsymbol{\phi}_0)} \max \left(\left| \frac{\partial \alpha_i^+(\boldsymbol{\phi})}{\partial \phi_{i_1}} \right|, \left| \frac{\partial \alpha_i^-(\boldsymbol{\phi})}{\partial \phi_{i_1}} \right| \right) |\varepsilon_{t-i}|^\delta \leq K \sum_{i=0}^{\infty} (i+t)^{-(d+1)(1-\xi)} |\varepsilon_{-i}|^\delta,$$

whence $\mathbb{E} \left(\chi_t^{(i_1)} \right)^\rho \leq K t^{-(d+1)\rho(1-\xi)+1}$ from **A4**. This shows that for any i_1 , $\chi_t^{(i_1)}$ has a finite

moment of order ρ and thus is finite almost surely. Furthermore, since $\rho(d+1) > 1$, we can find a $\xi > 0$ such that $\rho(d+1)(1-\xi) > 1$, and thus the dominated convergence theorem entails $\lim_{t \rightarrow \infty} \chi_t^{(i_1)} = 0$ almost surely. The same arguments yield $\lim_{t \rightarrow \infty} \chi_t^{(i_1, i_2)} = 0$ almost surely for any i_1, i_2 .

Consider now

$$\begin{aligned}
& \sup_{\boldsymbol{\theta} \in \mathcal{V}(\boldsymbol{\theta}_0)} \left| \frac{1}{n} \sum_{t=1}^n \left[\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \theta_{i_1} \partial \theta_{i_2}} - \frac{\partial^2 \tilde{l}_t(\boldsymbol{\theta})}{\partial \theta_{i_1} \partial \theta_{i_2}} \right] \right| \\
\leq & \frac{1}{n} \sum_{t=1}^n \frac{2}{\delta} \sup_{\boldsymbol{\theta} \in \mathcal{V}(\boldsymbol{\theta}_0)} \left[\frac{\varepsilon_t^2}{\tilde{\sigma}_t^2} - \frac{\varepsilon_t^2}{\sigma_t^2} \right] \left[\frac{1}{\sigma_t^\delta} \frac{\partial^2 \sigma_t^\delta}{\partial \theta_{i_1} \partial \theta_{i_2}} \right] \\
& + \left[1 - \frac{\varepsilon_t^2}{\tilde{\sigma}_t^2} \right] \left[\left(\frac{1}{\sigma_t^\delta} - \frac{1}{\tilde{\sigma}_t^\delta} \right) \frac{\partial^2 \sigma_t^\delta}{\partial \theta_{i_1} \partial \theta_{i_2}} + \frac{1}{\tilde{\sigma}_t^\delta} \left(\frac{\partial^2 \sigma_t^\delta}{\partial \theta_{i_1} \partial \theta_{i_2}} - \frac{\partial^2 \tilde{\sigma}_t^\delta}{\partial \theta_{i_1} \partial \theta_{i_2}} \right) \right] \\
& + \left[\frac{2 + \delta}{\delta} \frac{\varepsilon_t^2}{\sigma_t^2} - \frac{2 + \delta}{\delta} \frac{\varepsilon_t^2}{\tilde{\sigma}_t^2} \right] \left[\frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \theta_{i_1}} \right] \left[\frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \theta_{i_2}} \right] \\
& + \left[\frac{2 + \delta}{\delta} \frac{\varepsilon_t^2}{\tilde{\sigma}_t^2} - 1 \right] \left[\left(\frac{1}{\sigma_t^\delta} - \frac{1}{\tilde{\sigma}_t^\delta} \right) \frac{\partial \sigma_t^\delta}{\partial \theta_{i_1}} + \frac{1}{\tilde{\sigma}_t^\delta} \left(\frac{\partial \sigma_t^\delta}{\partial \theta_{i_1}} - \frac{\partial \tilde{\sigma}_t^\delta}{\partial \theta_{i_1}} \right) \right] \left[\frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \theta_{i_2}} \right] \\
& + \left[\frac{2 + \delta}{\delta} \frac{\varepsilon_t^2}{\tilde{\sigma}_t^2} - 1 \right] \left[\left(\frac{1}{\sigma_t^\delta} - \frac{1}{\tilde{\sigma}_t^\delta} \right) \frac{\partial \sigma_t^\delta}{\partial \theta_{i_2}} + \frac{1}{\tilde{\sigma}_t^\delta} \left(\frac{\partial \sigma_t^\delta}{\partial \theta_{i_2}} - \frac{\partial \tilde{\sigma}_t^\delta}{\partial \theta_{i_2}} \right) \right] \left[\frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \theta_{i_1}} \right] \quad (\boldsymbol{\theta}),
\end{aligned}$$

which yields

$$\begin{aligned}
& \sup_{\boldsymbol{\theta} \in \mathcal{V}(\boldsymbol{\theta}_0)} \left| \frac{1}{n} \sum_{t=1}^n \left[\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \theta_{i_1} \partial \theta_{i_2}} - \frac{\partial^2 \tilde{l}_t(\boldsymbol{\theta})}{\partial \theta_{i_1} \partial \theta_{i_2}} \right] \right| \\
\leq & \frac{K}{n} \sum_{t=1}^n \eta_t^2 \sup_{\boldsymbol{\theta} \in \mathcal{V}(\boldsymbol{\theta}_0)} \left| \frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2(\boldsymbol{\theta})} \frac{1}{\sigma_t^\delta(\boldsymbol{\theta})} \frac{\partial^2 \sigma_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_1} \partial \theta_{i_2}} \right| \chi_t \\
& + \frac{K}{n} \sum_{t=1}^n \eta_t^2 \sup_{\boldsymbol{\theta} \in \mathcal{V}(\boldsymbol{\theta}_0)} \left| \frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2(\boldsymbol{\theta})} \frac{1}{\sigma_t^\delta(\boldsymbol{\theta})} \frac{\partial \sigma_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_1}} \right| \left[\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})} \frac{\partial \sigma_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_2}} \right] \chi_t \\
& + \frac{K}{n} \sum_{t=1}^n \eta_t^2 \sup_{\boldsymbol{\theta} \in \mathcal{V}(\boldsymbol{\theta}_0)} \left| \frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2(\boldsymbol{\theta})} \frac{1}{\tilde{\sigma}_t^\delta(\boldsymbol{\theta})} \frac{\partial \tilde{\sigma}_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_1}} \right| \left[\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})} \frac{\partial \sigma_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_2}} \right] \chi_t \\
& + \frac{K}{n} \sum_{t=1}^n \eta_t^2 \sup_{\boldsymbol{\theta} \in \mathcal{V}(\boldsymbol{\theta}_0)} \left| \frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2(\boldsymbol{\theta})} \frac{1}{\tilde{\sigma}_t^\delta(\boldsymbol{\theta})} \frac{\partial \tilde{\sigma}_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_1}} \right| \chi_t^{(i_2)} \\
& + \frac{K}{n} \sum_{t=1}^n \eta_t^2 \sup_{\boldsymbol{\theta} \in \mathcal{V}(\boldsymbol{\theta}_0)} \left| \frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2(\boldsymbol{\theta})} \frac{1}{\sigma_t^\delta(\boldsymbol{\theta})} \frac{\partial \sigma_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_1}} \right| \chi_t^{(i_1)} \\
& + \frac{K}{n} \sum_{t=1}^n \eta_t^2 \sup_{\boldsymbol{\theta} \in \mathcal{V}(\boldsymbol{\theta}_0)} \left| \frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2(\boldsymbol{\theta})} \right| \chi_t^{(i_1, i_2)}.
\end{aligned}$$

We can first notice that, from the same arguments used to show Lemma 2.2, for all $p > 0$, $i_1, i_2 = 1, \dots, r+1$,

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}_0} \sup_{\boldsymbol{\theta} \in \Theta} \left| \frac{1}{\tilde{\sigma}_t^\delta(\boldsymbol{\theta})} \frac{\partial \tilde{\sigma}_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_1}} \right|^p & < \infty \\
\mathbb{E}_{\boldsymbol{\theta}_0} \sup_{\boldsymbol{\theta} \in \Theta} \left| \frac{1}{\tilde{\sigma}_t^\delta(\boldsymbol{\theta})} \frac{\partial^2 \tilde{\sigma}_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_1} \partial \theta_{i_2}} \right|^p & < \infty.
\end{aligned} \tag{2.32}$$

Then, from independence of η_t^2 with σ_t^δ and its derivatives, assumption **A11**, Lemma 2.2, (2.27),

and (2.32) we have, using Hölder inequality, for all i_1, i_2 ,

$$\begin{aligned}
& \mathbb{E} \left[\eta_t^2 \sup_{\boldsymbol{\theta} \in \mathcal{V}(\boldsymbol{\theta}_0)} \left| \left[\frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2(\boldsymbol{\theta})} \frac{1}{\sigma_t^\delta(\boldsymbol{\theta})} \frac{\partial^2 \sigma_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_1} \partial \theta_{i_2}} \right] \right| \right] < \infty \\
& \mathbb{E} \left[\eta_t^2 \sup_{\boldsymbol{\theta} \in \mathcal{V}(\boldsymbol{\theta}_0)} \left| \frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2(\boldsymbol{\theta})} \left[\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})} \frac{\partial \sigma_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_1}} \right] \left[\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})} \frac{\partial \sigma_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_2}} \right] \right| \right] < \infty \\
& \mathbb{E} \left[\eta_t^2 \sup_{\boldsymbol{\theta} \in \mathcal{V}(\boldsymbol{\theta}_0)} \left| \frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2(\boldsymbol{\theta})} \left[\frac{1}{\tilde{\sigma}_t^\delta(\boldsymbol{\theta})} \frac{\partial \tilde{\sigma}_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_1}} \right] \left[\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})} \frac{\partial \sigma_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_2}} \right] \right| \right] < \infty \\
& \mathbb{E} \left[\eta_t^2 \sup_{\boldsymbol{\theta} \in \mathcal{V}(\boldsymbol{\theta}_0)} \left| \frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2(\boldsymbol{\theta})} \left[\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})} \frac{\partial \sigma_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_1}} \right] \right| \right] < \infty \\
& \mathbb{E} \left[\eta_t^2 \sup_{\boldsymbol{\theta} \in \mathcal{V}(\boldsymbol{\theta}_0)} \left| \frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2(\boldsymbol{\theta})} \left[\frac{1}{\tilde{\sigma}_t^\delta(\boldsymbol{\theta})} \frac{\partial \tilde{\sigma}_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_1}} \right] \right| \right] < \infty.
\end{aligned} \tag{2.33}$$

Since χ_t , $\chi_t^{(i_1)}$, and $\chi_t^{(i_1, i_2)}$ tend to 0 almost surely as t tends to infinity, and (2.33), Toeplitz lemma combined with Markov inequality entail

$$\sup_{\boldsymbol{\theta} \in \mathcal{V}(\boldsymbol{\theta}_0)} \left| \frac{1}{n} \sum_{t=1}^n \left[\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \theta_{i_1} \partial \theta_{i_2}} - \frac{\partial^2 \tilde{l}_t(\boldsymbol{\theta})}{\partial \theta_{i_1} \partial \theta_{i_2}} \right] \right| \xrightarrow[n \rightarrow \infty]{} 0$$

in probability, which concludes the proof. \square

Finally, the following lemma shows the asymptotic normality of the normalized score.

Lemma 2.7. *Under the assumptions of Theorem 2.3 or Theorem 2.4,*

$$\mathbf{Z}_n = -\mathbf{J}_n^{-1} \sqrt{n} \frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \xrightarrow{\mathcal{L}} \mathbf{Z}, \text{ with } \mathbf{Z} \sim \mathcal{N}(\mathbf{0}, (\kappa_\eta - 1)\mathbf{J})$$

where $\mathbf{J}_n^{-1} = \frac{\partial^2 Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$ is an almost surely positive definite matrix for n sufficiently large.

Proof. Using the fact that $\sigma_t^\delta(\boldsymbol{\theta}_0)$ and its derivatives belong to the σ -field generated by $\{\varepsilon_{t-i}, i \geq 0\}$, and the fact that $\mathbb{E}_{\boldsymbol{\theta}_0} [\varepsilon_t^2 | \varepsilon_u, u < t] = \sigma_t^2(\boldsymbol{\theta}_0)$, we have

$$\mathbb{E}_{\boldsymbol{\theta}_0} \left[\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \mid \varepsilon_u, u < t \right] = \frac{1}{\sigma_t^\delta(\boldsymbol{\theta}_0)} \left[\frac{\partial \sigma_t^\delta(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right] \mathbb{E}_{\boldsymbol{\theta}_0} [\sigma_t^2(\boldsymbol{\theta}_0) - \varepsilon_t^2 | \varepsilon_u, u < t] = 0$$

and we have from Lemma 2.4 that $\mathbb{V}_{\boldsymbol{\theta}_0} \left[\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right]$ is finite. In view of the invertibility of \mathbf{J} and the assumptions on the distribution of η_t (which entails $0 < \kappa_\eta - 1 < \infty$), this covariance matrix is non-degenerate. It follows that, for all $\boldsymbol{\lambda} \in \mathbb{R}^{r+1}$, the sequence $\left\{ \boldsymbol{\lambda}' \frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}, \varepsilon_t \right\}_t$ is a square integrable ergodic stationary martingale difference. The Cramer-Wold theorem and the central limit theorem for square-integrable martingale difference of Billingsley (1961) entail $\frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, (\kappa_\eta - 1)\mathbf{J})$. The ergodic theorem entails $\mathbf{J}_n \rightarrow \mathbf{J}$ as $n \rightarrow \infty$ almost surely and thus the conclusion follows from Slutsky lemma. \square

We can now develop the proof of Theorem 2.3.

Proof of Theorem 2.3. From Theorem 2.2, we have that $\tilde{\boldsymbol{\theta}}_n$ converges to $\boldsymbol{\theta}_0$ which, from assumption **A6**, belongs to the interior of Θ , whence the derivative of the criterion is equal to zero at $\tilde{\boldsymbol{\theta}}_n$. It follows that, by a standard Taylor expansion at $\boldsymbol{\theta}_0$, we have

$$0 = \frac{\partial \tilde{Q}_n}{\partial \boldsymbol{\theta}}(\tilde{\boldsymbol{\theta}}_n) = \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\partial \tilde{l}_t}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_0) + \left[\frac{1}{n} \sum_{t=1}^n \frac{\partial^2 \tilde{l}_t}{\partial \theta_i \partial \theta_j}(\boldsymbol{\theta}_{ij}^*) \right] \sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$$

where the $\boldsymbol{\theta}_{ij}^*$ are between $\tilde{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}_0$. We will show the result by proving that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\partial \tilde{l}_t}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_0) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, (\kappa_\eta - 1)\mathbf{J}) \text{ and } \frac{1}{n} \sum_{t=1}^n \frac{\partial^2 \tilde{l}_t}{\partial \theta_i \partial \theta_j}(\boldsymbol{\theta}_{ij}^*) \rightarrow J(i, j) \text{ in probability.} \quad (2.34)$$

Using lemmas 2.3, 2.4, 2.6, and 2.7 along with Slutsky lemma directly yields the first part of (2.34).

Consider now a second Taylor expansion of the criterion at $\boldsymbol{\theta}_0$. We have for all i and j ,

$$\frac{1}{n} \sum_{t=1}^n \frac{\partial^2 l_t}{\partial \theta_i \partial \theta_j}(\boldsymbol{\theta}_{ij}^*) = \frac{1}{n} \sum_{t=1}^n \frac{\partial^2 l_t}{\partial \theta_i \partial \theta_j}(\boldsymbol{\theta}_0) + \frac{1}{n} \sum_{t=1}^n \frac{\partial l_t}{\partial \boldsymbol{\theta}'} \left[\frac{\partial^2 l_t}{\partial \theta_i \partial \theta_j}(\tilde{\boldsymbol{\theta}}_{ij}) \right] (\boldsymbol{\theta}_{ij}^* - \boldsymbol{\theta}_0)$$

where $\tilde{\boldsymbol{\theta}}_{ij}$ is between $\boldsymbol{\theta}_{ij}^*$ and $\boldsymbol{\theta}_0$. The almost sure convergence of $\tilde{\boldsymbol{\theta}}_{ij}$ to $\boldsymbol{\theta}_0$, the ergodic theorem and the uniform integrability of the third-order derivatives of the criterion (from Lemma 2.5) imply that almost surely

$$\begin{aligned} \limsup_{n \rightarrow \infty} \left\| \frac{1}{n} \sum_{t=1}^n \frac{\partial l_t}{\partial \boldsymbol{\theta}'} \left[\frac{\partial^2 l_t}{\partial \theta_i \partial \theta_j}(\tilde{\boldsymbol{\theta}}_{ij}) \right] \right\| &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \sup_{\boldsymbol{\theta} \in \mathcal{V}(\boldsymbol{\theta}_0)} \left\| \frac{\partial l_t}{\partial \boldsymbol{\theta}'} \left[\frac{\partial^2 l_t}{\partial \theta_i \partial \theta_j}(\boldsymbol{\theta}) \right] \right\| \\ &\leq \mathbb{E}_{\boldsymbol{\theta}_0} \sup_{\boldsymbol{\theta} \in \mathcal{V}(\boldsymbol{\theta}_0)} \left\| \frac{\partial l_t}{\partial \boldsymbol{\theta}'} \left[\frac{\partial^2 l_t}{\partial \theta_i \partial \theta_j}(\boldsymbol{\theta}) \right] \right\| \\ &< \infty \end{aligned}$$

Since $\|\boldsymbol{\theta}_{ij}^* - \boldsymbol{\theta}_0\| \rightarrow 0$ almost surely, we have for all $\varepsilon > 0$,

$$\mathbb{P} \left[\left| \frac{1}{n} \sum_{t=1}^n \frac{\partial l_t}{\partial \boldsymbol{\theta}'} \left[\frac{\partial^2 l_t}{\partial \theta_i \partial \theta_j}(\tilde{\boldsymbol{\theta}}_{ij}) \right] (\boldsymbol{\theta}_{ij}^* - \boldsymbol{\theta}_0) \right| \leq \varepsilon \right] = 1$$

and by the ergodic theorem,

$$\frac{1}{n} \sum_{t=1}^n \frac{\partial^2 l_t}{\partial \theta_i \partial \theta_j}(\boldsymbol{\theta}_0) \xrightarrow{\mathbb{P}} J(i, j).$$

Using Slutsky lemma along with the previous lemmas allows us to obtain the last part of (2.34) which ends the proof. \square

The proof of Theorem 2.4 builds upon the one developed by Francq and Zakoian (2007) using the previously established lemmas.

Proof of Theorem 2.4. From the proof of Theorem 2.3, we have

$$\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \stackrel{o_P(1)}{=} \mathbf{Z}_n = -\mathbf{J}_n^{-1} \sqrt{n} \frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}. \quad (2.35)$$

when $\boldsymbol{\theta}_0$ belongs to the interior of Θ . This relation does not hold when $\boldsymbol{\theta}_0 \in \partial\Theta$ since then at least one element of the vector $(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ is a positive random variable. We will show that, in the general case, for all $\boldsymbol{\theta}_0 \in \Theta$, we have

$$\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \stackrel{o_P(1)}{=} \boldsymbol{\lambda}_n^\Lambda = \arg \inf_{\boldsymbol{\lambda} \in \Lambda} [\boldsymbol{\lambda} - \mathbf{Z}_n]' \mathbf{J}_n [\boldsymbol{\lambda} - \mathbf{Z}_n]. \quad (2.36)$$

Note that when $\boldsymbol{\theta}_0$ belongs to the interior of Θ , we have $\boldsymbol{\lambda}_n^\Lambda = \mathbf{Z}_n$ and (2.36) reduces to (2.35). $\boldsymbol{\lambda}_n^\Lambda$ can be seen as the orthogonal projection of \mathbf{Z}_n on Λ for the inner product $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{J}_n} = \mathbf{x}' \mathbf{J}_n \mathbf{y}$. This projection can be approximated by

$$\sqrt{n}(\boldsymbol{\theta}_{\mathbf{J}_n}(\mathbf{Z}_n) - \boldsymbol{\theta}_0) \quad \text{with} \quad \boldsymbol{\theta}_{\mathbf{J}_n}(\mathbf{Z}_n) = \arg \inf_{\boldsymbol{\theta} \in \Theta} \|\mathbf{Z}_n - \sqrt{n}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\|_{\mathbf{J}_n}$$

which is the the projection of \mathbf{Z}_n on the space $\sqrt{n}(\Theta - \boldsymbol{\theta}_0)$ which increases to Λ . Using a Taylor expansion for a function with right partial derivatives, we have for all $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0 \in \Theta$,

$$\begin{aligned} \tilde{Q}_n(\boldsymbol{\theta}) &= \tilde{Q}_n(\boldsymbol{\theta}_0) + \frac{\partial \tilde{Q}_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} (\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)' \frac{\partial^2 \tilde{Q}_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} (\boldsymbol{\theta} - \boldsymbol{\theta}_0) + R_n(\boldsymbol{\theta}) \\ &= \tilde{Q}_n(\boldsymbol{\theta}_0) + \frac{1}{2n} \mathbf{Z}_n' \mathbf{J}_n \sqrt{n} (\boldsymbol{\theta} - \boldsymbol{\theta}_0) - \frac{1}{2n} \sqrt{n} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)' \mathbf{J}_n \mathbf{Z}_n \\ &\quad + \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)' \mathbf{J}_n (\boldsymbol{\theta} - \boldsymbol{\theta}_0) + R_n(\boldsymbol{\theta}) + R_n^*(\boldsymbol{\theta}) \\ &= \tilde{Q}_n(\boldsymbol{\theta}_0) + \frac{1}{2n} \|\mathbf{Z}_n - \sqrt{n}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\|_{\mathbf{J}_n}^2 - \frac{1}{2n} \mathbf{Z}_n' \mathbf{J}_n \mathbf{Z}_n + R_n(\boldsymbol{\theta}) + R_n^*(\boldsymbol{\theta}) \end{aligned} \quad (2.37)$$

where $R_n(\boldsymbol{\theta})$ and $R_n^*(\boldsymbol{\theta})$ are remainder terms. To conclude the proof, we will prove the following intermediate results. For all $\boldsymbol{\theta}_0 \in \Theta$

- (a) $\sqrt{n}(\boldsymbol{\theta}_{\mathbf{J}_n}(\mathbf{Z}_n) - \boldsymbol{\theta}_0) = O_P(1)$
- (b) $\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = O_P(1)$
- (c) For any sequence $(\boldsymbol{\theta}_n)$ such that $\sqrt{n}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0) = O_P(1)$, $R_n(\boldsymbol{\theta}_n) = o_P(n^{-1})$ and $R_n^*(\boldsymbol{\theta}_n) = o_P(n^{-1})$
- (d) $\|\mathbf{Z}_n - \sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\|_{\mathbf{J}_n}^2 \stackrel{o_P(1)}{=} \|\mathbf{Z}_n - \boldsymbol{\lambda}_n^\Lambda\|_{\mathbf{J}_n}^2$
- (e) $\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \stackrel{o_P(1)}{=} \boldsymbol{\lambda}_n^\Lambda$
- (f) $\boldsymbol{\lambda}_n^\Lambda \xrightarrow{\mathcal{L}} \boldsymbol{\lambda}^\Lambda$

Lemma 2.4 ensures that for n sufficiently large, $\|\cdot\|_{\mathbf{J}_n}$ almost surely defines a norm. Using the triangular inequality and the fact that $\boldsymbol{\theta}_{\mathbf{J}_n}(\mathbf{Z}_n)$ minimizes $\|\mathbf{Z}_n - \sqrt{n}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\|_{\mathbf{J}_n}$ over Θ , we have

$$\|\sqrt{n}(\boldsymbol{\theta}_{\mathbf{J}_n}(\mathbf{Z}_n) - \boldsymbol{\theta}_0)\|_{\mathbf{J}_n} \leq \|\mathbf{Z}_n - \sqrt{n}(\boldsymbol{\theta}_{\mathbf{J}_n}(\mathbf{Z}_n) - \boldsymbol{\theta}_0)\|_{\mathbf{J}_n} + \|\mathbf{Z}_n\|_{\mathbf{J}_n} \leq \|\mathbf{Z}_n\|_{\mathbf{J}_n} + \|\mathbf{Z}_n\|_{\mathbf{J}_n}$$

and from Lemma 2.7, we have $\|\mathbf{Z}_n\|_{\mathbf{J}_n} + \|\mathbf{Z}_n\|_{\mathbf{J}_n} = O_P(1)$ which concludes the proof of (a).

By the Taylor expansion

$$\tilde{Q}_n(\boldsymbol{\theta}) = \tilde{Q}_n(\boldsymbol{\theta}_0) + \frac{\partial \tilde{Q}_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} (\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)' \frac{\partial^2 \tilde{Q}_n(\boldsymbol{\theta}_{ij}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)$$

with $\boldsymbol{\theta}_{ij}^*$ between $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}$, we have

$$R_n(\boldsymbol{\theta}) = \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)' \left[\frac{\partial^2 \tilde{Q}_n(\boldsymbol{\theta}_{ij}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} - \frac{\partial^2 \tilde{Q}_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right] (\boldsymbol{\theta} - \boldsymbol{\theta}_0). \quad (2.38)$$

Theorem 2.2, and Lemmas 2.5 and 2.6 ensure $\left[\frac{\partial^2 \tilde{Q}_n(\boldsymbol{\theta}_{ij}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} - \frac{\partial^2 \tilde{Q}_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right] \rightarrow 0$ as n tends to infinity when $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}_n$, and thus $R_n(\tilde{\boldsymbol{\theta}}_n) = o_P(\|\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|_{\mathbf{J}_n}^2)$.

In addition,

$$R_n^*(\boldsymbol{\theta}) = \left[\frac{\partial \tilde{Q}_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} - \frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right] (\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)' \left[\frac{\partial^2 \tilde{Q}_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} - \mathbf{J}_n \right] (\boldsymbol{\theta} - \boldsymbol{\theta}_0). \quad (2.39)$$

and from Lemma 2.6, we have $R_n^*(\tilde{\boldsymbol{\theta}}_n) = o_P(n^{-1/2} \|\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|_{\mathbf{J}_n}) + o_P(\|\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|_{\mathbf{J}_n}^2)$.

Since $\tilde{\boldsymbol{\theta}}_n$ minimizes \tilde{Q}_n over Θ , equation (2.37) yields

$$\begin{aligned} \tilde{Q}_n(\tilde{\boldsymbol{\theta}}_n) - \tilde{Q}_n(\boldsymbol{\theta}_0) &= \frac{1}{2n} \left[\|\mathbf{Z}_n - \sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\|_{\mathbf{J}_n}^2 - \|\mathbf{Z}_n\|_{\mathbf{J}_n}^2 \right. \\ &\quad \left. + o_P(\|n^{1/2}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\|_{\mathbf{J}_n}) + o_P(\|n^{1/2}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\|_{\mathbf{J}_n}^2) \right] \leq 0 \end{aligned}$$

and thus

$$\|\mathbf{Z}_n - \sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\|_{\mathbf{J}_n}^2 \leq \left[\|\mathbf{Z}_n\|_{\mathbf{J}_n} + o_P(\|n^{1/2}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\|_{\mathbf{J}_n}) \right]^2.$$

By the triangular inequality, we obtain

$$\begin{aligned} \|\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\|_{\mathbf{J}_n} &\leq \|\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) - \mathbf{Z}_n\|_{\mathbf{J}_n} + \|\mathbf{Z}_n\|_{\mathbf{J}_n} \\ &\leq 2\|\mathbf{Z}_n\|_{\mathbf{J}_n} + o_P(\|n^{1/2}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\|_{\mathbf{J}_n}) \end{aligned}$$

whence $\|\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\|_{\mathbf{J}_n} [1 + o_P(1)] \leq 2\|\mathbf{Z}_n\|_{\mathbf{J}_n} = O_P(1)$, which proves (b).

From lemmas 2.5 and 2.6, equation (2.38) entail $R_n(\boldsymbol{\theta}_n) = o_P(\|\boldsymbol{\theta}_n - \boldsymbol{\theta}_0\|^2) = o_P(n^{-1})$, which proves the first part of (c), while equation (2.39) similarly yields $R_n^*(\boldsymbol{\theta}_n) = o_P(n^{-1/2} \|\boldsymbol{\theta}_n - \boldsymbol{\theta}_0\|_{\mathbf{J}_n}) + o_P(\|\boldsymbol{\theta}_n - \boldsymbol{\theta}_0\|_{\mathbf{J}_n}^2) = o_P(n^{-1})$ which concludes the proof of (c).

By (a)-(c) and (2.37), we have

$$\begin{aligned}
0 &\leq \|\mathbf{Z}_n - \sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\|_{\mathbf{J}_n}^2 - \|\mathbf{Z}_n - \sqrt{n}(\boldsymbol{\theta}_{\mathbf{J}_n}(\mathbf{Z}_n) - \boldsymbol{\theta}_0)\|_{\mathbf{J}_n}^2 \\
&= 2n[\tilde{Q}_n(\tilde{\boldsymbol{\theta}}_n) - \tilde{Q}_n(\boldsymbol{\theta}_{\mathbf{J}_n}(\mathbf{Z}_n))] - 2n[(R_n + R_n^*)(\tilde{\boldsymbol{\theta}}_n) - (R_n + R_n^*)(\boldsymbol{\theta}_{\mathbf{J}_n})] \\
&\leq -2n[(R_n + R_n^*)(\tilde{\boldsymbol{\theta}}_n) - (R_n + R_n^*)(\boldsymbol{\theta}_{\mathbf{J}_n})] = o_P(1)
\end{aligned}$$

since $\tilde{\boldsymbol{\theta}}_n$ minimizes \tilde{Q}_n and $\boldsymbol{\theta}_{\mathbf{J}_n}$ minimizes $\|\mathbf{Z}_n - \sqrt{n}(\boldsymbol{\theta}_{\mathbf{J}_n}(\mathbf{Z}_n) - \boldsymbol{\theta}_0)\|_{\mathbf{J}_n}$. Since for n sufficiently large, we have $\sqrt{n}(\boldsymbol{\theta}_{\mathbf{J}_n} - \boldsymbol{\theta}_0) = \boldsymbol{\lambda}_n^\Lambda$, (d) holds.

The vector $\boldsymbol{\lambda}_n^\Lambda$ being the projection of \mathbf{Z}_n on the convex set Λ for the scalar product $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{J}_n}$, it is characterized by $\boldsymbol{\lambda}_n^\Lambda \in \Lambda$, $\langle \mathbf{Z}_n \boldsymbol{\lambda}_n^\Lambda, \boldsymbol{\lambda}_n^\Lambda - \boldsymbol{\lambda} \rangle$ for all $\boldsymbol{\lambda} \in \Lambda$ (see for example Lemma 1.1 in Zarantonello (1971)). Thus we obtain

$$\begin{aligned}
\|\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) - \mathbf{Z}_n\|_{\mathbf{J}_n}^2 &= \|\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) - \boldsymbol{\lambda}_n^\Lambda\|_{\mathbf{J}_n}^2 + \|\boldsymbol{\lambda}_n^\Lambda - \mathbf{Z}_n\|_{\mathbf{J}_n}^2 \\
&\quad + 2\langle \sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) - \boldsymbol{\lambda}_n^\Lambda, \boldsymbol{\lambda}_n^\Lambda - \mathbf{Z}_n \rangle_{\mathbf{J}_n} \\
&\geq \|\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) - \boldsymbol{\lambda}_n^\Lambda\|_{\mathbf{J}_n}^2 + \|\boldsymbol{\lambda}_n^\Lambda - \mathbf{Z}_n\|_{\mathbf{J}_n}^2
\end{aligned}$$

whence, by (d),

$$\|\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) - \boldsymbol{\lambda}_n^\Lambda\|_{\mathbf{J}_n}^2 \leq \|\mathbf{Z}_n - \sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\|_{\mathbf{J}_n}^2 - \|\mathbf{Z}_n - \boldsymbol{\lambda}_n^\Lambda\|_{\mathbf{J}_n}^2 = o_P(1)$$

which proves (e).

Finally, Lemma 2.7 entails $(\mathbf{J}_n, \mathbf{Z}_n) \rightarrow (\mathbf{J}, \mathbf{Z})$. In addition, $\boldsymbol{\lambda}_n^\Lambda = f(\mathbf{J}_n, \mathbf{Z}_n)$ and $\boldsymbol{\lambda}^\Lambda = f(\mathbf{J}, \mathbf{Z})$ where f is a continuous function, except on the set of the points $(\mathbf{J}_n, \mathbf{Z}_n)$ such that \mathbf{J}_n is singular, which is a set of $P_{(\mathbf{J}, \mathbf{Z})}$ -probability zero. Thus, the continuous mapping theorem entails (f).

The proof of Theorem 2.4 directly follows from (e) and (f). □

Proof of Theorem 2.5. It suffices to show that $\tilde{\delta}_n = \delta_0$ for n large enough, the other results being easily obtained from the proofs of Theorems 2.2, 2.3 and 2.4. We first show that

$$\frac{\sigma_{\delta, t}(\boldsymbol{\theta})}{\sigma_{\delta_0, t}(\boldsymbol{\theta}_0)} = 1 \text{ almost surely} \Rightarrow \delta = \delta_0. \quad (2.40)$$

We have, denoting $\eta_t^{+(-)} = \eta_t \mathbf{1}_{\eta_t \geq (<) 0}$,

$$\begin{aligned}
\sigma_{\delta, t}^\delta(\boldsymbol{\theta}) &= \omega + \sum_{i=1}^{\infty} \alpha_i^+(\phi) \sigma_{\delta_0, t-i}^\delta(\boldsymbol{\theta}_0) |\eta_{t-i}^+|^\delta + \alpha_i^-(\phi) \sigma_{\delta_0, t-i}^\delta(\boldsymbol{\theta}_0) |\eta_{t-i}^-|^\delta \\
&= \omega_{\delta, t-2}(\boldsymbol{\theta}) + \alpha_1^+(\phi) \sigma_{\delta_0, t-1}^\delta(\boldsymbol{\theta}_0) |\eta_{t-1}^+|^\delta + \alpha_1^-(\phi) \sigma_{\delta_0, t-1}^\delta(\boldsymbol{\theta}_0) |\eta_{t-1}^-|^\delta
\end{aligned}$$

where $\omega_{\delta, t-2}(\boldsymbol{\theta}) = \omega + \sum_{i=2}^{\infty} \alpha_i^+(\phi) \sigma_{\delta_0, t-i}^\delta(\boldsymbol{\theta}_0) |\eta_{t-i}^+|^\delta + \alpha_i^-(\phi) \sigma_{\delta_0, t-i}^\delta(\boldsymbol{\theta}_0) |\eta_{t-i}^-|^\delta$ is measurable with respect to \mathcal{F}_{t-2} . Let $\Psi = (a, b, r, c, d) \in (0, \infty)^3 \times [0, \infty)^2$ and let the function $g_\Psi : [0, \infty) \rightarrow$

$(0; \infty)$ defined by $g_\Psi(x) = (a + bx)^{-1}(c + dx^r)^{1/r}$. We have $g'_\Psi(x) = 0$ if and only if $adx^{r-1} = bc$, whence $g_\Psi(x) = 1$ cannot have more than two solutions, except if i) $r = 1, a = c, b = d$, or ii) $b = d = 0$ and $c = a^r$. Conditionally on \mathcal{F}_{t-1} we have

$$\left[\frac{\sigma_{\delta,t}(\boldsymbol{\theta})}{\sigma_{\delta_0,t}(\boldsymbol{\theta}_0)} \right]^{\delta_0} = g_{\Psi^+}(|\eta_{t-1}|^{\delta_0})\mathbb{1}_{\eta_{t-1} \geq 0} + g_{\Psi^-}(|\eta_{t-1}|^{\delta_0})\mathbb{1}_{\eta_{t-1} < 0} \quad (2.41)$$

where $\Psi^{+(-)} = (\omega_{\delta_0,t-2}(\boldsymbol{\theta}_0), \omega_{\delta,t-2}(\boldsymbol{\theta}), \delta/\delta_0, \alpha_1^{+(-)}(\boldsymbol{\phi}_0)\sigma_{\delta_0,t-1}^{\delta_0}(\boldsymbol{\theta}_0), \alpha_1^{+(-)}(\boldsymbol{\phi})\sigma_{\delta_0,t-1}^\delta(\boldsymbol{\theta}_0))$. Thus $\sigma_{\delta,t}(\boldsymbol{\theta}) = \sigma_{\delta_0,t}(\boldsymbol{\theta}_0)$ implies i) $\delta = \delta_0$ or ii) $\alpha_1^+(\boldsymbol{\phi}) = \alpha_1^+(\boldsymbol{\phi}_0) = 0$ and $\alpha_1^-(\boldsymbol{\phi}) = \alpha_1^-(\boldsymbol{\phi}_0) = 0$. In the latter case, (2.41) holds by replacing η_{t-1} by η_{t-2} . Iterating the arguments, under A2', the first equality in (2.40) entails either i) $\delta = \delta_0$ or ii) $\alpha_i^+(\boldsymbol{\phi}) = \alpha_i^+(\boldsymbol{\phi}_0) = 0$ and $\alpha_i^-(\boldsymbol{\phi}) = \alpha_i^-(\boldsymbol{\phi}_0) = 0$ for all $i \geq 1$. The latter is precluded by Assumption A3(i), thus we have shown (2.40), which concludes the proof using Theorem 2.2 and arguments of its proof. \square

2.8.3 Specification tests

We develop in this section the proofs of the results of Section 3.

Proof of Theorem 2.6. Let us define for $0 < h < n$

$$r_h = n^{-1} \sum_{t=h+1}^n s_t s_{t-h}, \text{ with } s_t = \eta_t^2 - 1,$$

and let $\mathbf{r}_m = (r_1, \dots, r_m)'$ for any $1 \leq m \leq n$. Let $s_t(\boldsymbol{\theta})$ (respectively $\tilde{s}_t(\boldsymbol{\theta})$) be the random variable obtained by replacing η_t by $\eta_t(\boldsymbol{\theta}) = \varepsilon_t/\sigma_t(\boldsymbol{\theta})$ (respectively $\tilde{\eta}_t(\boldsymbol{\theta}) = \varepsilon_t/\tilde{\sigma}_t(\boldsymbol{\theta})$). Let $r_h(\boldsymbol{\theta})$ and $\tilde{r}_h(\boldsymbol{\theta})$ be defined with the same convention.

We first prove the asymptotic irrelevance of the initial values on \mathbf{r}_m

$$\sqrt{n}\|\mathbf{r}_m(\boldsymbol{\theta}_0) - \tilde{\mathbf{r}}_m(\boldsymbol{\theta}_0)\| = o_P(1) \text{ and } \sup_{\boldsymbol{\theta} \in \mathcal{V}(\boldsymbol{\theta}_0)} \left\| \frac{\partial \mathbf{r}_m(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \frac{\partial \tilde{\mathbf{r}}_m(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\| = o_P(1). \quad (2.42)$$

We have

$$\begin{aligned} s_t(\boldsymbol{\theta})s_{t-h}(\boldsymbol{\theta}) - \tilde{s}_t(\boldsymbol{\theta})\tilde{s}_{t-h}(\boldsymbol{\theta}) &= (s_t(\boldsymbol{\theta}) - \tilde{s}_t(\boldsymbol{\theta}))s_{t-h}(\boldsymbol{\theta}) + (s_{t-h}(\boldsymbol{\theta}) - \tilde{s}_{t-h}(\boldsymbol{\theta}))\tilde{s}_t(\boldsymbol{\theta}) \\ &:= A_t(\boldsymbol{\theta}) + B_t(\boldsymbol{\theta}) \end{aligned}$$

Similarly to (2.29), we have

$$\mathbb{E}_{\boldsymbol{\theta}_0} |A_t(\boldsymbol{\theta}_0)|^\rho \leq K \mathbb{E}_{\boldsymbol{\theta}_0} \left| \frac{\sigma_t^2(\boldsymbol{\theta}_0) - \tilde{\sigma}_t^2(\boldsymbol{\theta}_0)}{\tilde{\sigma}_t^2(\boldsymbol{\theta}_0)} \right|^\rho \leq K t^{-(d^*+1)\rho+1}$$

and thus

$$\mathbb{E}_{\boldsymbol{\theta}_0} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n A_t(\boldsymbol{\theta}_0) \right|^\rho \leq K n^{-(d^*+\frac{3}{2})\rho+2} \xrightarrow{n \rightarrow \infty} 0$$

since from assumption **A9** we have $(d^* + \frac{3}{2})\rho - 2 > 0$. Using Markov inequality, we can conclude

$\frac{1}{\sqrt{n}} \sum_{t=1}^n |A_t(\boldsymbol{\theta}_0)|$ tends to 0 in probability. Similar arguments yield that $\frac{1}{\sqrt{n}} \sum_{t=1}^n |B_t(\boldsymbol{\theta}_0)|$ tends to 0 in probability, which proves the first part of (2.42). In addition, we have

$$\frac{\partial s_t}{\partial \boldsymbol{\theta}} - \frac{\partial \tilde{s}_t}{\partial \boldsymbol{\theta}} = \frac{-2}{\delta} \left[\left[\frac{\varepsilon_t^2}{\sigma_t^2} - \frac{\varepsilon_t^2}{\tilde{\sigma}_t^2} \right] \left[\frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \boldsymbol{\theta}} \right] + \frac{\varepsilon_t^2}{\tilde{\sigma}_t^2} \left[\frac{1}{\sigma_t^\delta} - \frac{1}{\tilde{\sigma}_t^\delta} \right] \frac{\partial \sigma_t^\delta}{\partial \boldsymbol{\theta}} + \frac{\varepsilon_t^2}{\tilde{\sigma}_t^2} \frac{1}{\sigma_t^\delta} \left[\frac{\partial \sigma_t^\delta}{\partial \boldsymbol{\theta}} - \frac{\partial \tilde{\sigma}_t^\delta}{\partial \boldsymbol{\theta}} \right] \right]$$

whence, for all $k \in \{1, \dots, r+1\}$, using similar notations as in (2.30) and (2.31),

$$\begin{aligned} \sup_{\boldsymbol{\theta} \in \mathcal{V}(\boldsymbol{\theta}_0)} \left| \frac{\partial s_t}{\partial \theta_k} - \frac{\partial \tilde{s}_t}{\partial \theta_k} \right| s_{t-h} &\leq K \eta_{t-h}^2 \eta_t^2 \sup_{\boldsymbol{\theta} \in \mathcal{V}(\boldsymbol{\theta}_0)} \frac{\sigma_{t-h}^2(\boldsymbol{\theta}_0) \sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_{t-h}^2(\boldsymbol{\theta}) \sigma_t^2(\boldsymbol{\theta})} \left| \frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \theta_k} \right| \chi_t \\ &\quad + K \eta_{t-h}^2 \eta_t^2 \sup_{\boldsymbol{\theta} \in \mathcal{V}(\boldsymbol{\theta}_0)} \frac{\sigma_{t-h}^2(\boldsymbol{\theta}_0) \sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_{t-h}^2(\boldsymbol{\theta}) \tilde{\sigma}_t^2(\boldsymbol{\theta})} \left| \frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \theta_k} \right| \chi_t \\ &\quad + K \eta_{t-h}^2 \eta_t^2 \sup_{\boldsymbol{\theta} \in \mathcal{V}(\boldsymbol{\theta}_0)} \frac{\sigma_{t-h}^2(\boldsymbol{\theta}_0) \sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_{t-h}^2(\boldsymbol{\theta}) \tilde{\sigma}_t^2(\boldsymbol{\theta})} \chi_t^{(k)}. \end{aligned}$$

Then similarly to (2.33), from independence of η_t^2 with σ_t^δ and its derivatives, assumption **A11**, Lemma 2.2, (2.27), and (2.32) we have, using Hölder inequality, and Toeplitz lemma,

$$\sup_{\boldsymbol{\theta} \in \mathcal{V}(\boldsymbol{\theta}_0)} \left| \frac{1}{n} \sum_{t=1}^n \left(\frac{\partial s_t}{\partial \theta_k} - \frac{\partial \tilde{s}_t}{\partial \theta_k} \right) s_{t-h} \right| \xrightarrow{n \rightarrow \infty} 0$$

since χ_t and $\chi_t^{(k)}$ tend to 0 almost surely as t tends to infinity. In a similar manner, we obtain that

$$\mathbb{E} \sup_{\boldsymbol{\theta} \in \mathcal{V}(\boldsymbol{\theta}_0)} \left| \frac{1}{n} \sum_{t=1}^n \frac{\partial s_{t-h}}{\partial \theta_k} (s_t - \tilde{s}_t) \right| \leq \mathbb{E} \left[\frac{K}{n} \sum_{t=1}^n \eta_{t-h}^2 \eta_t^2 \sup_{\boldsymbol{\theta} \in \mathcal{V}(\boldsymbol{\theta}_0)} \frac{\sigma_{t-h}^2(\boldsymbol{\theta}_0)}{\sigma_{t-h}^2(\boldsymbol{\theta})} \left| \frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \theta_k} \right| \chi_t \right] \xrightarrow{n \rightarrow \infty} 0.$$

Using Markov inequality, we thus obtain that $n^{-1} \sum_{t=1}^n \sup_{\boldsymbol{\theta} \in \mathcal{V}(\boldsymbol{\theta}_0)} |\partial A_t(\boldsymbol{\theta}) / \partial \theta_k| \rightarrow 0$ in probability as n tends to infinity. Similar arguments yield the convergence of the term $n^{-1} \sum_{t=1}^n \sup_{\boldsymbol{\theta} \in \mathcal{V}(\boldsymbol{\theta}_0)} |\partial B_t(\boldsymbol{\theta}) / \partial \theta_k|$ and thus we have shown the second part of (2.42).

Using a Taylor expansion of \tilde{r}_h at $\tilde{\boldsymbol{\theta}}_n$ for $h = 1, \dots, m$ along with (2.42) yields

$$\sqrt{n} \tilde{r}_h(\tilde{\boldsymbol{\theta}}_n) = \sqrt{n} \tilde{r}_h(\boldsymbol{\theta}_0) + \frac{\partial \tilde{r}_h(\boldsymbol{\theta}_n^*)}{\partial \boldsymbol{\theta}} \sqrt{n} (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \stackrel{o_P(1)}{=} \sqrt{n} r_h(\boldsymbol{\theta}_0) + \frac{\partial r_h(\boldsymbol{\theta}_n^*)}{\partial \boldsymbol{\theta}} \sqrt{n} (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$$

for some $\boldsymbol{\theta}_n^*$ between $\boldsymbol{\theta}_0$ and $\tilde{\boldsymbol{\theta}}_n$. In addition, assumption **A11** and Lemma 2.2 entail that there exists a neighborhood $\mathcal{V}(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ such that for all $i, j \in \{1, \dots, r+1\}$

$$\sup_{\boldsymbol{\theta} \in \mathcal{V}(\boldsymbol{\theta}_0)} \mathbb{E}_{\boldsymbol{\theta}_0} \left| \frac{\partial^2 s_t(\boldsymbol{\theta}) s_{t-h}(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right| < \infty.$$

Using a second Taylor expansion, the ergodic theorem, and Theorem 2.2, we thus obtain for all $0 < h < n$

$$\frac{\partial r_h(\boldsymbol{\theta}_n^*)}{\partial \boldsymbol{\theta}} \rightarrow \mathbb{E}_{\boldsymbol{\theta}_0} \left[\frac{\partial s_t s_{t-h}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right] = -\frac{2}{\delta} \mathbb{E}_{\boldsymbol{\theta}_0} \left[s_{t-h}(\boldsymbol{\theta}_0) \frac{1}{\sigma_t^\delta(\boldsymbol{\theta}_0)} \frac{\partial \sigma_t^\delta(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right]$$

since $\mathbb{E}_{\theta_0}[s_t \partial s_{t-h}(\theta_0) / \partial \theta] = 0$ and thus we have

$$\sqrt{n} \tilde{\mathbf{r}}_m(\tilde{\theta}_n) \stackrel{o_P(1)}{=} \sqrt{n} \mathbf{r}_m(\theta_0) + \mathbf{C}_m \sqrt{n}(\tilde{\theta}_n - \theta_0). \quad (2.43)$$

We now derive the asymptotic distribution of $\sqrt{n}(\mathbf{r}_m(\theta_0), \tilde{\theta}_n - \theta_0)$. Let us denote $\mathbf{s}_{t-1:t-m} = (s_{t-1}, \dots, s_{t-m})'$ and remark that $\mathbf{r}_m(\theta_0) \stackrel{o_P(1)}{=} n^{-1} \sum_{t=1}^n s_t \mathbf{s}_{t-1:t-m}$. From the proof of Theorem 2.3, we have

$$\sqrt{n}(\tilde{\theta}_n - \theta_0) \stackrel{o_P(1)}{=} \mathbf{J}^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^n (\eta_t^2 - 1) \frac{1}{\sigma_t^2(\theta_0)} \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta},$$

thus the central limit theorem applied to the martingale difference

$$\left\{ \left(s_t \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta'}, s_t \mathbf{s}'_{t-1:t-m} \right)'; \mathcal{F}(\eta_u, u \leq t) \right\}$$

shows that

$$\begin{aligned} \sqrt{n} \begin{pmatrix} \tilde{\theta}_n - \theta_0 \\ \mathbf{r}_m(\theta_0) \end{pmatrix} &\stackrel{o_P(1)}{=} \frac{1}{\sqrt{n}} \sum_{t=1}^n s_t \begin{pmatrix} \mathbf{J}^{-1} \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta} \\ \mathbf{s}_{t-1:t-m} \end{pmatrix} \\ &\stackrel{\mathcal{L}}{\rightarrow} \mathcal{N} \left(\mathbf{0}, \begin{bmatrix} (\kappa_\eta - 1) \mathbf{J}^{-1} & (\kappa_\eta - 1) \mathbf{J}^{-1} \mathbf{C}'_m \\ (\kappa_\eta - 1) \mathbf{C}_m \mathbf{J}^{-1} & (\kappa_\eta - 1)^2 \mathbf{I}_m \end{bmatrix} \right). \end{aligned} \quad (2.44)$$

From (2.43) and (2.44), we obtain

$$\sqrt{n} \tilde{\mathbf{r}}_m(\tilde{\theta}_n) \stackrel{\mathcal{L}}{\rightarrow} \mathcal{N}(\mathbf{0}, \mathbf{D}), \text{ with } \mathbf{D} = (\kappa_\eta - 1)^2 \mathbf{I}_m - (\kappa_\eta - 1) \mathbf{C}_m \mathbf{J}^{-1} \mathbf{C}'_m$$

and we can show that $\hat{\mathbf{D}} \rightarrow \mathbf{D}$ almost surely as $n \rightarrow \infty$. Finally, we show that \mathbf{D} is invertible. From assumption **A2**, the law of η_t^2 is non-degenerated hence $\kappa_\eta > 1$ and it suffices to show the non singularity of

$$(\kappa_\eta - 1) \mathbf{I}_m - \mathbf{C}_m \mathbf{J}^{-1} \mathbf{C}'_m = \mathbb{E}_{\theta_0} \mathbf{V} \mathbf{V}', \text{ with } \mathbf{V} = \mathbf{s}_{-1:-m} + \mathbf{C}_m \mathbf{J}^{-1} \frac{2}{\delta} \frac{1}{\sigma_0^2} \frac{\partial \sigma_0^2(\theta_0)}{\partial \theta}.$$

If this matrix is singular, then there exists $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m)'$ such that $\boldsymbol{\lambda} \neq \mathbf{0}$ and

$$\boldsymbol{\lambda}' \mathbf{s}_{-1:-m} + \boldsymbol{\mu}' \frac{1}{\sigma_0^2} \frac{\partial \sigma_0^2(\theta_0)}{\partial \theta} \text{ a.s., where } \boldsymbol{\mu} = \frac{2}{\delta} \boldsymbol{\lambda}' \mathbf{C}_m \mathbf{J}^{-1}. \quad (2.45)$$

If $\boldsymbol{\mu} = (\mu_1, \dots, \mu_{r+1}) = \mathbf{0}$, then $\boldsymbol{\lambda}' \mathbf{s}_{-1:-m} = 0$ almost surely, and thus there exists $j \in \{1, \dots, m\}$ such that $s_{-j} \in \mathcal{F}(s_t, t \neq -j)$, which is impossible since s_t are independent and non-degenerated, and thus we have $\boldsymbol{\mu} \neq \mathbf{0}$. Denoting by R_t any random variable measurable with respect to

$\mathcal{F}(\eta_u, u \leq t)$, we have

$$\begin{aligned} \boldsymbol{\mu}' \frac{\partial \sigma_0^2(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} &= \mu_1 + \sum_{i=1}^{\infty} \left[\sum_{j=2}^{r+1} \mu_j \frac{\partial \alpha_i^+(\boldsymbol{\phi}_0)}{\partial \phi_j} \mathbb{1}_{\eta_{-i} \geq 0} + \sum_{k=2}^{r+1} \mu_k \frac{\partial \alpha_i^-(\boldsymbol{\phi}_0)}{\partial \phi_k} \mathbb{1}_{\eta_{-i} < 0} \right] \sigma_{-i}^\delta(\boldsymbol{\theta}_0) |\eta_{-i}|^\delta \\ &= \mu_1 + \left[\sum_{j=2}^{r+1} \mu_j \frac{\partial \alpha_1^+(\boldsymbol{\phi}_0)}{\partial \phi_j} |\eta_{-1}^+|^\delta + \sum_{k=2}^{r+1} \mu_k \frac{\partial \alpha_1^-(\boldsymbol{\phi}_0)}{\partial \phi_k} |\eta_{-1}^-|^\delta \right] \sigma_{-1}^\delta(\boldsymbol{\theta}_0) + R_{-2} \end{aligned}$$

where $\eta_t^{+(-)} = \eta_t \mathbb{1}_{\eta_t \geq (<) 0}$. In addition, we have

$$\begin{aligned} \sigma_0^\delta(\boldsymbol{\theta}_0) \boldsymbol{\lambda}' \mathbf{s}_{-1:-m} &= (\omega_0 + a_{1,t-1}(\boldsymbol{\phi}_0) |\eta_{-1}|^\delta \sigma_{-1}^\delta(\boldsymbol{\theta}_0) + R_{-2}) (\lambda_1 \eta_{-1}^2 + R_{-2}) \\ &= \lambda_1 \sigma_{-1}^\delta(\boldsymbol{\theta}_0) a_{1,t-1}(\boldsymbol{\phi}_0) |\eta_{-1}|^{\delta+2} + (\omega_0 \lambda_1 + R_{-2}) \eta_{-1}^2 + R_{-2}. \end{aligned}$$

Thus (2.45) entails almost surely

$$\begin{aligned} 0 &= \lambda_1 \sigma_{-1}^\delta(\boldsymbol{\theta}_0) \alpha_1^{+(-)}(\boldsymbol{\phi}_0) |\eta_{-1}^{+(-)}|^{\delta+2} + \left[\sum_{j=2}^{r+1} \mu_j \frac{\partial \alpha_1^{+(-)}(\boldsymbol{\phi}_0)}{\partial \phi_j} \right] \sigma_{-1}^\delta(\boldsymbol{\theta}_0) |\eta_{-1}^{+(-)}|^\delta \\ &\quad + (\omega_0 \lambda_1 + R_{-2}) |\eta_{-1}^{+(-)}|^2 + R_{-2}. \end{aligned}$$

Since an equation of the form $a|x|^{\delta+2} + b|x|^\delta + c|x|^2 + d = 0$ cannot have more than three positive or more than three negative roots, except if all the coefficients are equal to 0, assumption **A2'** implies $\sum_{j=2}^{r+1} \mu_j \frac{\partial \alpha_1^+(\boldsymbol{\phi}_0)}{\partial \phi_j} = 0$ and $\sum_{j=2}^{r+1} \mu_j \frac{\partial \alpha_1^-(\boldsymbol{\phi}_0)}{\partial \phi_j} = 0$ almost surely. Iterating this argument, we obtain for all $i_h^+(\boldsymbol{\phi}_0)$ and $i_h^-(\boldsymbol{\phi}_0)$ a similar result, and thus from assumption **A10(ii)**, we must have $\boldsymbol{\mu} = 0$ which is impossible and thus contradicts the singularity of \mathbf{D} , concluding the proof. \square

Proof of Proposition 2.1. The first part of the proposition is a standard result for testing linear constraints. See for example Chapter 17 of Gouriéroux and Monfort (1995) for proofs of the asymptotic distributions. We now turn to the second part of the proposition, when the parameter is allowed to be on the boundary of the parameter space. For the ease of notation, we assume that $\phi_i = 0$ for all $i = 1, \dots, r$. This is often the case in conditional volatility models as boundary conditions are necessary to ensure positivity of the conditional variance. In addition, without loss of generality, we consider testing that the last d_2 coefficients of $\boldsymbol{\theta}_0$ are on the boundary. We thus split $\boldsymbol{\theta}_0$ into two components $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}_0^{(1)}, \boldsymbol{\theta}_0^{(2)})'$, where $\boldsymbol{\theta}_0^{(i)} \in \mathbb{R}^{d_i}$, $d_1 + d_2 = 1 + r$. The null hypothesis is thus $H_0 : \boldsymbol{\theta}_0^{(2)} = \mathbf{R}\boldsymbol{\theta}_0 = \mathbf{0}_{d_2 \times 1}$ with $\mathbf{R} = (\mathbf{0}_{d_2 \times d_1}, \mathbf{I}_{d_2})$. Let $H : \boldsymbol{\theta}_0^{(1)} = \bar{\mathbf{R}}\boldsymbol{\theta}_0 > \mathbf{0}_{d_1 \times 1}$ with $\bar{\mathbf{R}} = (\mathbf{I}_{d_1}, \mathbf{0}_{d_1 \times d_2})$ denote the maintained assumption.

From (2.13) and a direct application of the continuous mapping theorem, we have that, under H_0 ,

$$\sqrt{n}(\mathbf{R}\tilde{\boldsymbol{\theta}}_n - \mathbf{k} - \mathbf{R}\boldsymbol{\theta}_0 + \mathbf{k}) = \sqrt{n}\mathbf{R}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{\mathcal{L}} \mathbf{R}\boldsymbol{\lambda}^\Lambda$$

which yields

$$W_n \xrightarrow{\mathcal{L}} \boldsymbol{\lambda}^\Lambda' \mathbf{R}' [(\kappa_\eta - 1) \mathbf{R} \mathbf{J}^{-1} \mathbf{R}']^{-1} \boldsymbol{\lambda}^\Lambda' \mathbf{R}.$$

We now turn to the Rao-score statistic. Since $\tilde{\boldsymbol{\theta}}_{n|H_0}^{(1)}$ is a consistent estimator of $\boldsymbol{\theta}_0^{(1)}$, we have

$\tilde{\boldsymbol{\theta}}_{n|H_0}^{(1)} > 0$ for n large enough. Therefore $\partial\tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{n|H_0})/\partial\theta_i = 0$ for $i = 1, \dots, d_1$, or equivalently

$$\frac{\partial\tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{n|H_0})}{\partial\boldsymbol{\theta}} = \mathbf{R}' \frac{\partial\tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{n|H_0})}{\partial\boldsymbol{\theta}^{(2)}}. \quad (2.46)$$

A Taylor expansion yields

$$\sqrt{n} \frac{\partial\tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{n|H_0})}{\partial\boldsymbol{\theta}} \stackrel{o_P(1)}{=} \sqrt{n} \frac{\partial\tilde{Q}_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} + \mathbf{J}(\tilde{\boldsymbol{\theta}}_{n|H_0} - \boldsymbol{\theta}_0). \quad (2.47)$$

The last d_2 components of this vector relation give

$$\sqrt{n} \frac{\partial\tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{n|H_0})}{\partial\boldsymbol{\theta}^{(2)}} \stackrel{o_P(1)}{=} \sqrt{n} \frac{\partial\tilde{Q}_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}^{(2)}} + \mathbf{R}\mathbf{J}(\tilde{\boldsymbol{\theta}}_{n|H_0} - \boldsymbol{\theta}_0) \quad (2.48)$$

while the first d_1 components give

$$\mathbf{0} \stackrel{o_P(1)}{=} \sqrt{n} \frac{\partial\tilde{Q}_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}^{(1)}} + \sqrt{n} \overline{\mathbf{R}}\mathbf{J}\overline{\mathbf{R}}'(\tilde{\boldsymbol{\theta}}_{n|H_0}^{(1)} - \boldsymbol{\theta}_0^{(1)}) \quad (2.49)$$

using

$$\overline{\mathbf{R}}'(\tilde{\boldsymbol{\theta}}_{n|H_0}^{(1)} - \boldsymbol{\theta}_0^{(1)}) = \tilde{\boldsymbol{\theta}}_{n|H_0} - \boldsymbol{\theta}_0. \quad (2.50)$$

In view of (2.49), we have

$$\sqrt{n}(\tilde{\boldsymbol{\theta}}_{n|H_0}^{(1)} - \boldsymbol{\theta}_0^{(1)}) \stackrel{o_P(1)}{=} (\overline{\mathbf{R}}\tilde{\mathbf{J}}_{n|H_0}\overline{\mathbf{R}}')^{-1} \sqrt{n} \frac{\partial\tilde{Q}_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}^{(1)}}. \quad (2.51)$$

From (2.46), (2.48), (2.50) and (2.51), we obtain

$$\begin{aligned} R_n &= \frac{n}{\tilde{\kappa}_{n|H_0} - 1} \frac{\partial\tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{n|H_0})}{\partial\boldsymbol{\theta}^{(2)'}} \mathbf{R}\tilde{\mathbf{J}}_{n|H_0}^{-1} \mathbf{R}' \frac{\partial\tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{n|H_0})}{\partial\boldsymbol{\theta}^{(2)}} \\ &\stackrel{o_P(1)}{=} \frac{n}{\kappa_\eta - 1} \left\| \frac{\partial\tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{n|H_0})}{\partial\boldsymbol{\theta}^{(2)}} \right\|_{\mathbf{R}\mathbf{J}^{-1}\mathbf{R}'}^2 \\ &\stackrel{o_P(1)}{=} \frac{n}{\kappa_\eta - 1} \left\| \frac{\partial\tilde{Q}_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}^{(2)}} + \mathbf{R}\mathbf{J}\overline{\mathbf{R}}'(\tilde{\boldsymbol{\theta}}_{n|H_0}^{(1)} - \boldsymbol{\theta}_0^{(1)}) \right\|_{\mathbf{R}\mathbf{J}^{-1}\mathbf{R}'}^2 \\ &\stackrel{o_P(1)}{=} \frac{n}{\kappa_\eta - 1} \left\| \frac{\partial\tilde{Q}_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}^{(2)}} + \mathbf{R}\mathbf{J}\overline{\mathbf{R}}'(\overline{\mathbf{R}}\tilde{\mathbf{J}}_{n|H_0}\overline{\mathbf{R}}')^{-1} \frac{\partial\tilde{Q}_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}^{(1)}} \right\|_{\mathbf{R}\mathbf{J}^{-1}\mathbf{R}'}^2. \end{aligned}$$

Now recall that under H_0 ,

$$\begin{pmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \end{pmatrix} := \sqrt{\frac{n}{\kappa_\eta - 1}} \begin{pmatrix} \frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}^{(1)}} \\ \frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}^{(2)}} \end{pmatrix} \stackrel{\mathcal{L}}{\rightarrow} \mathcal{N} \left(\mathbf{0}, \mathbf{J} = \begin{pmatrix} \mathbf{J}_{11} & \mathbf{J}_{12} \\ \mathbf{J}_{21} & \mathbf{J}_{22} \end{pmatrix} \right). \quad (2.52)$$

As $\mathbf{R}\mathbf{J}^{-1}\mathbf{R}' = (\mathbf{J}_{22} - \mathbf{J}_{21}\mathbf{J}_{11}^{-1}\mathbf{J}_{12})^{-1}$, the asymptotic distribution of R_n is that of $(\mathbf{W}_2 - \mathbf{J}_{21}\mathbf{J}_{11}^{-1}\mathbf{W}_1)'(\mathbf{J}_{22} - \mathbf{J}_{21}\mathbf{J}_{11}^{-1}\mathbf{J}_{12})^{-1}(\mathbf{W}_2 - \mathbf{J}_{21}\mathbf{J}_{11}^{-1}\mathbf{W}_1)$ under H_0 , which follows a $\chi_{d_2}^2$ since $\mathbf{W}_2 - \mathbf{J}_{21}\mathbf{J}_{11}^{-1}\mathbf{W}_1 \sim \mathcal{N}(0, \mathbf{J}_{22} - \mathbf{J}_{21}\mathbf{J}_{11}^{-1}\mathbf{J}_{12})$.

Finally, we turn to L_n . Using Taylor expansions, we get

$$\tilde{Q}_n(\tilde{\boldsymbol{\theta}}_n) \stackrel{o_P(1)}{=} \tilde{Q}_n(\boldsymbol{\theta}_0) + \frac{\partial \tilde{Q}_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + \frac{1}{2} (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' J (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$$

and

$$\tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{n|H_0}) \stackrel{o_P(1)}{=} \tilde{Q}_n(\boldsymbol{\theta}_0) + \frac{\partial \tilde{Q}_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} (\tilde{\boldsymbol{\theta}}_{n|H_0} - \boldsymbol{\theta}_0) + \frac{1}{2} (\tilde{\boldsymbol{\theta}}_{n|H_0} - \boldsymbol{\theta}_0)' J (\tilde{\boldsymbol{\theta}}_{n|H_0} - \boldsymbol{\theta}_0),$$

hence, by subtraction,

$$\begin{aligned} (\hat{\kappa}_{\eta|H_0} - 1) L_n &\stackrel{o_P(1)}{=} 2n \frac{\partial \tilde{Q}_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} (\tilde{\boldsymbol{\theta}}_{n|H_0} - \tilde{\boldsymbol{\theta}}_n) + n (\tilde{\boldsymbol{\theta}}_{n|H_0} - \boldsymbol{\theta}_0)' J (\tilde{\boldsymbol{\theta}}_{n|H_0} - \boldsymbol{\theta}_0) \\ &\quad - n (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' J (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0). \end{aligned}$$

We thus have

$$\begin{aligned} \frac{\hat{\kappa}_{\eta|H_0} - 1}{2} L_n &\stackrel{o_P(1)}{=} -n \left[\frac{1}{2} \frac{\partial \tilde{Q}_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^{(1)'}} (\bar{\mathbf{R}} \mathbf{J} \bar{\mathbf{R}}^{-1}) \frac{\partial \tilde{Q}_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^{(1)}} \right. \\ &\quad \left. + \frac{\partial \tilde{Q}_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + \frac{1}{2} (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' J (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \right]. \end{aligned}$$

Under H_0 , by showing $\sqrt{n} \begin{pmatrix} \frac{\partial \tilde{Q}_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \\ \tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} -\mathbf{J}\mathbf{Z} \\ \boldsymbol{\lambda}^\Lambda \end{pmatrix}$, it can be seen that the asymptotic distribution of L_n is the law of

$$L = \frac{1}{2} \mathbf{Z}' \mathbf{J}' \bar{\mathbf{R}}' \mathbf{J}_{11}^{-1} \bar{\mathbf{R}} \mathbf{J} \mathbf{Z} + \mathbf{Z}' \mathbf{J}' \boldsymbol{\lambda}^{\Lambda'} - \frac{1}{2} \boldsymbol{\lambda}^{\Lambda'} \mathbf{J} \boldsymbol{\lambda}^\Lambda.$$

Since,

$$\mathbf{J}' \bar{\mathbf{R}}' \mathbf{J}_{11}^{-1} \bar{\mathbf{R}} \mathbf{J} = \mathbf{J} - (\kappa_\eta - 1) \boldsymbol{\Omega} \quad \text{with } (\kappa_\eta - 1) \boldsymbol{\Omega} = \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{J}_{22} - \mathbf{J}_{21} \mathbf{J}_{11}^{-1} \mathbf{J}_{12} \end{pmatrix},$$

we have

$$\begin{aligned} L &= -\frac{1}{2} \mathbf{Z}' \mathbf{J} \mathbf{Z} + \frac{1}{2} \mathbf{Z}' (\kappa_\eta - 1) \boldsymbol{\Omega} \mathbf{Z} + \mathbf{Z}' \mathbf{J}' \boldsymbol{\lambda}^\Lambda - \frac{1}{2} \boldsymbol{\lambda}^{\Lambda'} \mathbf{J} \boldsymbol{\lambda}^\Lambda \\ &= -\frac{1}{2} (\boldsymbol{\lambda}^\Lambda - \mathbf{Z})' \mathbf{J} (\boldsymbol{\lambda}^\Lambda - \mathbf{Z}) + \frac{\kappa_\eta - 1}{2} \mathbf{Z}' \boldsymbol{\Omega} \mathbf{Z} \end{aligned}$$

which concludes the proof. \square

2.9 Appendix B: GARCH(1,1)-type test when the parameter d is considered fixed

2.9.1 Asymptotics for the Wald statistic

Non-identification of the parameter d under H_0^{GARCH} makes the derivation of the Wald statistic asymptotic distribution infeasible. In Section 3.3 of the paper, we proposed to estimate this

distribution by bootstrap. Another solution is to set the unidentified parameter to an arbitrary value $\bar{d} > \frac{1}{2}$. Under this assumption, the identification problem disappears and we can use Theorem 2.4 to obtain the asymptotic distribution of the Wald statistic.

Proposition 2.4. *Under the assumptions of Theorem 2.4, under H_0^{GARCH} ,*

$$W_n^{GARCH}(\bar{d}) \xrightarrow{\mathcal{L}} \frac{1}{2}\Delta_0 + \frac{1}{2}\chi_1^2,$$

where Δ_0 is the Dirac measure at 0. Thus, the critical region of asymptotic level ν is given by $\{W_n^{GARCH} > \chi_1^2(1 - 2\nu)\}$.

This test can easily be extended to an asymmetric volatility model with a different δ . Consider, the APARCH(∞) specification presented in (2.8). Testing for the adequacy of the GJR-GARCH model ($\delta = 2$) or the TGARCH ($\delta = 1$) can then be achieved by testing for $H_0 : \gamma_0 = 0$ against $H_1^{\text{APARCH}(\infty)} : \gamma_0 > 0$. We can thus define the Wald statistic in a similar manner and derive its asymptotic distribution under the null.

Proof of Proposition 2.4. First, note that, under H_0^{GARCH} , the continuous mapping theorem and Theorem 2.4 entail

$$W_n^{\text{GARCH}} \xrightarrow{\mathcal{L}} \boldsymbol{\lambda}^{\Lambda'} \boldsymbol{\Omega} \boldsymbol{\lambda}^{\Lambda}$$

with $\boldsymbol{\Omega} = \mathbf{R}'[(\kappa_\eta - 1)\mathbf{R}\mathbf{J}^{-1}\mathbf{R}']\mathbf{R}$. In addition, since we are in the case where only one coefficient is at the boundary of the parameter space, $\boldsymbol{\lambda}^{\Lambda}$ has a trivial form. We have $\boldsymbol{\Lambda} = \mathbb{R}^{r-1} \times [0, \infty)$ and $\mathbf{R} = (0, \dots, 0, 1)$. We thus get

$$\boldsymbol{\lambda}^{\Lambda'} = \mathbf{Z}\mathbb{1}_{Z_r \geq 0} + \mathbf{P}\mathbf{Z}\mathbb{1}_{Z_r < 0}, \text{ with } \mathbf{P} = \mathbf{I}_r - \mathbf{J}^{-1}\mathbf{R}'[\mathbf{R}\mathbf{J}^{-1}\mathbf{R}']^{-1}\mathbf{R}$$

where Z_d denotes the last component of vector \mathbf{Z} , and thus it follows that

$$\boldsymbol{\lambda}^{\Lambda'} = \mathbf{Z} - [Z_r\mathbb{1}_{Z_r < 0}]\mathbf{c}$$

where vector $\mathbf{c} = \mathbb{E}[Z_r\mathbf{Z}]/\mathbb{V}[Z_d]$ is the last column of \mathbf{J}^{-1} divided by its (r, r) -component. We thus obtain

$$W_n^{\text{GARCH}} \xrightarrow{\mathcal{L}} \frac{(\mathbf{R}\boldsymbol{\lambda}^{\Lambda})^2}{\mathbf{R}[(\kappa_\eta - 1)\mathbf{J}^{-1}]\mathbf{R}'} = \frac{(\mathbf{R}\boldsymbol{\lambda}^{\Lambda})^2}{\mathbb{V}[Z_r]} = U^2\mathbb{1}_{U \geq 0} \sim \frac{1}{2}\Delta_0 + \frac{1}{2}\chi_1^2$$

where $U \sim \mathcal{N}(0, 1)$. □

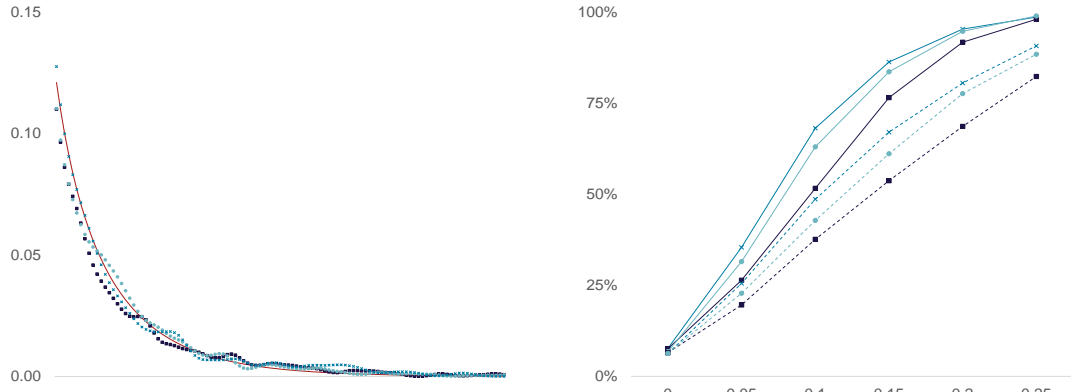
2.9.2 Finite sample properties of the statistic

We propose to study the empirical behavior of the test statistics defined in Section B.1. In the following simulations, we use Gaussian innovations ($\eta_t \sim \mathcal{N}(0, 1)$).

We first study the empirical level of the statistic. Figure 2.4a presents kernel estimators for $W_n(\bar{d})$ when testing for a GARCH(1,1), a GJR-GARCH(1,1) and a TGARCH(1,1) against an APARCH(∞) model of form (2.8) with $\delta = 2$ and 1 respectively under H_0 . The parameters used

for the simulations are $\theta_0 = (0.2, 0.15, 0.75)$ for the GARCH model and $\theta_0 = (0.2, 0.05, 0.2, 0.75)$ for both the GJR-GARCH and the TGARCH. All kernel estimators are obtained from 1000 replications and are close to their theoretical asymptotic distributions. The relative rejection frequency of the test statistics, at the asymptotic levels 5%, are respectively 5.70%, 5.10% and 5.70% which is not significantly different from 5%.

We now turn to the empirical properties of the statistics under $H_1 : \gamma_0 > 0$. We first study the empirical power of the statistics under the assumption that parameters \bar{d} and δ are well specified. We consider the parametric form of Model (8) with $\theta_0 = (\omega_0, \alpha_0^+, \alpha_0^-, \beta_0, \gamma_0) = (0.25, 0.05, 0.15, 0.7, \gamma_0)$ where γ_0 ranges from 0 to 0.25, $\bar{d} = 1$, and either $\delta = 2$ or $\delta = 1$ which corresponds to testing for a GJR-GARCH(1,1) or a TGARCH(1,1). In addition we consider an ARCH(∞) specified as Model (7) where $\theta_0 = (\omega_0, \alpha_0, \beta_0, \gamma_0) = (0.25, 0.1, 0.7, \gamma_0)$ where γ_0 ranges from 0 to 0.25 and $\bar{d} = 1$ which allows us to test for a GARCH(1,1). Figure 2.4b compares the observed powers of the three tests, that is, the relative frequency of rejection of the null hypothesis $H_0 : \gamma_0 = 0$ on the 1000 independent realizations of length $n = 2500$ and $n = 5000$, as a function of γ_0 . On these simulations, we see that the three test statistics seem powerful even for low values of γ_0 .



(a) $\chi_1^2/2$ density (red solid line) and kernel density estimators when testing for a GARCH (dark blue square), a GJR-GARCH (blue cross) and a TGARCH (light blue dot) on 1000 simulations with $n = 5000$

(b) Observed powers of the GARCH test (dark blue square), GJR-GARCH test (light blue dot), and TGARCH test (blue cross) as a function of γ_0 , on 1000 simulations with $n = 2500$ (dashed line) and $n = 5000$ (solid line)

Figure 2.4: Empirical behavior of the Wald statistics when d is fixed

We then consider the empirical power of the test statistics when either \bar{d} or δ is assumed known but is misspecified. We simulate 1000 replications of size 5000 of the parametric form of Model (8) with $\theta_0 = (\omega_0, \alpha_0^+, \alpha_0^-, \beta_0, \gamma_0, d_0) = (0.25, 0.05, 0.1, 0.7, 0.25, 1.0)$ and $\delta = 2$. We then conduct the GJR-GARCH(1,1) test assuming $\gamma = 2$ is well specified but \bar{d} is misspecified. Table 2.3a presents the observed power of the test at different asymptotic level ν and for different values of \bar{d} . Even when \bar{d} is far from its true value, the empirical power remains high. In addition we conduct the test assuming, this time, that $\bar{d} = 1.0$ is well specified but the power δ is

misspecified. This is equivalent to testing that a short memory APARCH(1,1) with power δ is suited to model a persistent TARARCH(∞). Table 2.3b presents the observed power of the test at different asymptotic level ν and for different values of δ . Again, even when δ is misspecified, the empirical power remains high, meaning our test statistic appears robust to misspecifications in \bar{d} and δ .

	ν				ν				
	1%	5%	10%		1%	5%	10%		
\bar{d}	0.55	70.9%	89.3%	94.6%	δ	0.5	82.1%	91.1%	93.7%
	1.0	91.3%	96.6%	98.3%		1.0	86.9%	94.1%	96.4%
	1.5	92.7%	96.8%	98.4%		1.5	89.4%	95.7%	97.9%
	2.0	92.3%	96.6%	97.9%		2.0	91.3%	96.6%	98.3%
	2.5	91.2%	96.3%	97.6%		2.5	90.5%	97.0%	98.4%
	5.0	89.6%	95.7%	97.3%		3.0	85.8%	95.5%	97.9%
	7.5	89.5%	95.5%	97.3%					
10.0	89.5%	95.5%	97.3%						

(a) Observed power of the GJR-GARCH(1,1) test when $d = 1.0$ and \bar{d} is misspecified

(b) Observed power of the APARCH(1,1) test when δ is misspecified

Table 2.3: Observed power of the GARCH(1,1)-type test under $H_1 : \gamma_0 > 0$ for an asymptotic level ν of 1%, 5%, and 10%, when the model is misspecified

2.10 Appendix C: Additional Monte Carlo experiments

We assess the finite sample properties of the QML estimator in the case where $\delta = 2$ and $\alpha_i^{+(-)}(\boldsymbol{\theta}) = \gamma^{+(-)}i^{-(d^{+(-)}+1)}$ with $\boldsymbol{\theta} = (\omega, \gamma^+, d^+, \gamma^-, d^-)$, which corresponds to an hyperbolic decay. We have simulated a thousand samples of different sizes n for $\boldsymbol{\theta}_0 = (1.0, 0.40, 0.85, 0.40, 0.85)$, which generates a standard ARCH(∞) process, as well as $\boldsymbol{\theta}_0 = (1.0, 0.40, 1.2, 0.20, 0.75)$, which generates a TARARCH(∞) process. On each realisation, we fitted a TARARCH(∞) by QML, which gave us a thousand estimators $\tilde{\boldsymbol{\theta}}_n$. Table 2.4 presents the empirical mean and root mean squared error (RMSE) of these estimators (in brackets). We can note that the estimations results are satisfactory, although the parameters $d^{+(-)}$ may require a large sample size to be precise.

	ARCH(∞)					TARARCH(∞)				
	ω	γ^+	d^+	γ^-	d^-	ω	γ^+	d^+	γ^-	d^-
$\boldsymbol{\theta}_0$	1.00	0.40	0.85	0.40	0.85	1.00	0.40	1.20	0.20	0.75
$\tilde{\boldsymbol{\theta}}_{1000}$	1.08 (0.19)	0.40 (0.07)	0.85 (0.53)	0.40 (0.07)	0.87 (0.50)	1.04 (0.15)	0.40 (0.08)	1.20 (0.59)	0.20 (0.06)	0.89 (0.73)
$\tilde{\boldsymbol{\theta}}_{2000}$	1.06 (0.14)	0.40 (0.05)	0.85 (0.47)	0.40 (0.05)	0.86 (0.43)	1.03 (0.11)	0.40 (0.05)	1.19 (0.48)	0.20 (0.04)	0.82 (0.70)
$\tilde{\boldsymbol{\theta}}_{5000}$	1.04 (0.08)	0.40 (0.03)	0.85 (0.38)	0.40 (0.03)	0.85 (0.38)	1.01 (0.07)	0.40 (0.04)	1.19 (0.39)	0.20 (0.03)	0.75 (0.70)

Table 2.4: Estimation results for 1000 simulations of size n of a symmetric ARCH(∞) process and a TARARCH(∞) process with $\alpha_i^{+(-)}(\boldsymbol{\phi}) = \gamma^{+(-)}i^{-d^{+(-)}-1}$

2.11 Appendix D: Additional application – Are GARCH(1,1)-type specifications suitable to model peripheral stocks?

The results in Table 2.2 provide a compelling argument in favor of applying models with strong persistence to less developed markets. Additionally, if this persistence stems from illiquidity, it should also imply heterogeneity at the stock level within developed markets. Assets that are less traded should therefore exhibit stronger persistence than highly traded ones. We propose to test this hypothesis by studying the Fama and French (1993) Size equity portfolios. Every year, the authors sort in ascending order of market equity all the NYSE, AMEX, and NASDAQ stocks and construct 10 decile portfolios labelled "Dec1", "Dec2", etc. Our dataset contains their daily returns from January 1975 to March 2020 and was obtained from Kenneth French's website⁷. For each portfolio, we compute the Wald statistic $W_n^{GJR}(0.55)$ from Proposition 4 to test the hypothesis that the GJR-GARCH(1,1) is well suited to model returns series. The results are presented in Figure 2.5. It is clear that on our whole sample, the GJR-GARCH(1,1) specification is strongly rejected for smaller Size portfolios. Moreover, we find that the portfolios composed of large stocks do not exhibit strong persistence, which confirms our finding for the US indices in Table 2.2. However, when focusing on the 1995-onward period, the results are quite different. Indeed, we do not reject the GJR-GARCH(1,1) for the low Size portfolios, which would imply that Size stocks have become less peripheral. A possible explanation could be that the search for higher returns has sparked investors' interest in small stocks, resulting in more efficient market conditions. Actually, the Size premium's existence has been disputed since the early nineties (see for example Van Dijk (2011)), which would support our tests results on the above mentioned time frame.

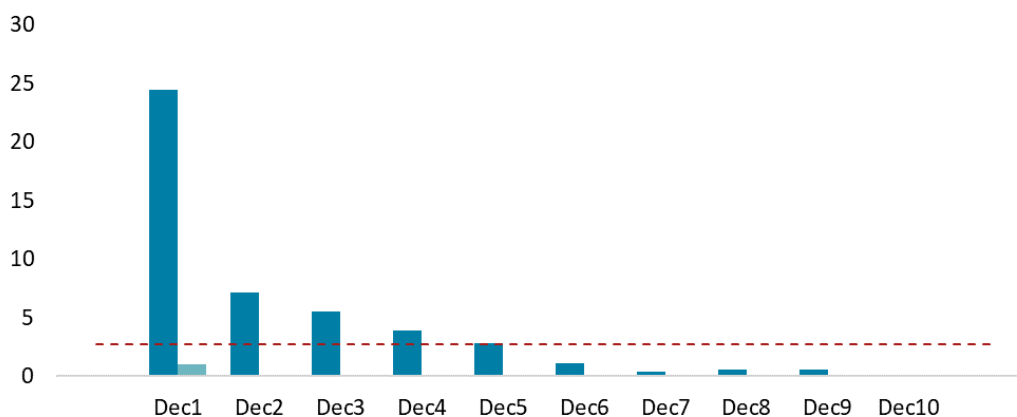


Figure 2.5: Wald statistics $W_n^{GJR}(0.55)$ computed on the portfolios formed on Size from 1975 (in blue) and from 1995 (in light blue). The rejection threshold of H_0^{GJR} at the 5% asymptotic level is represented by the red dashed line

⁷https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Chapter 3

A multivariate ARCH(∞) model with exogenous variables and dynamic conditional betas

Christian Francq

Julien Royer

Jean-Michel Zakoïan

Abstract: Factor models are highly common in the financial literature. Recent advances allow to relax the constancy of slope coefficients (the so-called betas) by considering conditional regressions. The theory on the estimation of these dynamic conditional betas however usually relies on short memory volatility models, which can be restrictive in empirical applications. Moreover, exogenous variables have proven useful in recent studies on volatility modeling. In this paper, we introduce a multivariate framework allowing for time-varying betas in which covolatilities can exhibit higher persistence than the standard exponential decay. Covariates are included in the dynamics of both conditional variances and betas. We establish stationarity conditions for the proposed model and prove the consistency and asymptotic normality of the QML estimator. Monte Carlo experiments are conducted to assess the performance of the estimation procedure in finite sample. Finally, we discuss the choice of potential relevant exogenous variables and illustrate the pertinence of the model on real data applications.

Keywords: Dynamic factors model, Long memory, Conditional betas

3.1 Introduction

An important area of the financial literature rests upon factor models. In static factor models, the regression coefficients are assumed to be constant over time. However, the justification for such parameters constancy are scarce and this assumption is often challenged in empirical studies. As wrongfully assuming betas constancy may lead to false financial conclusions, it is crucial to develop models accounting for their dynamic behavior. Time-varying parameter regressions were first considered through regime-switching models by Goldfeld and Quandt (1973) and Hamilton (1989) and more recently extended to the case of endogenous switching by Kim et al. (2008). However, in such models, the slope coefficients is considered constant in each state which can be restrictive when considering regimes with high duration. As an alternative, Shanken (1990) and Ferson and Harvey (1991) considered using financial or economics instrumental variables to recover time-varying betas. The obtained factor loadings are however highly sensitive to the choice of instrumental variables as noted by Ghysels (1998). In addition, Ang and Kristensen (2012) consider a family of deterministic paths for the regression coefficients while Hansen et al. (2014) considered a multivariate GARCH specification where asset returns are modeled through their conditional betas on the market that is assumed to follow a GARCH equation. Although the latter takes advantage of intraday information by using realized measures of volatilities and correlations, the beta is not directly modeled.

More recently, Engle (2016) proposed to recover the time-varying conditional betas indirectly from the conditional covariance matrix modeled using a dynamic conditional correlation specification. Building upon the seminal idea of Bollerslev et al. (1988) that conditional beta can be computed as the ratio of the conditional covariance to the conditional variance, the author introduces a general multivariate GARCH framework with dynamic betas. Although easy to implement, the model presents some drawbacks. First, the multivariate GARCH structure imposes an exponential decay of volatility shocks that might be problematic for empirical applications. Indeed, numerous studies have documented the high persistence of volatility in stock returns, see e.g. Ding and Granger (1996) or Conrad et al. (2011). More recently, Royer (2022) shows that peripheral assets, such as equity indices from emerging financial markets, exhibit higher persistence than allowed in standard GARCH models. In addition, the model proposed by Engle (2016) lacks tractability as the dynamics of slope coefficients are not directly specified, rendering statistical testing and financial interpretability difficult.

In this paper, we propose a new multivariate ARCH(∞) model that allows for a direct specification of the dynamic conditional betas. Theoretical analysis of multivariate long-memory volatility models have been scarce and to our knowledge, none considers the case of time-varying regression models. Our model builds on the Cholesky-GARCH (CHAR) model of Darolles et al. (2018), relaxing its inherent short memory feature. Moreover, we allow the inclusion of exogenous variables in the dynamics of both conditional variances and betas. The inclusion of exogenous variables, and more precisely the inclusion of volatility measures computed intraday data, have proven useful in modeling individual GARCH (see e.g. Francq and Thieu (2019))

and multivariate GARCH models (Bollerslev et al. (2020b)). Our work is organized as follows. In Section 2, we present the Cholesky-ARCH-X(∞) model. We establish conditions for the existence of a stationary solution. In Section 3, we focus on statistical inference. We prove the strong consistency and asymptotic normality of the quasi maximum likelihood estimator (QMLE). Finally, Section 4 presents a preliminary financial application to illustrate our model.

3.2 ARCH-X(∞) extension of the CHAR model

Directly modeling betas dynamics provides numerous advantages as it allows to gain economic and financial interpretability on the time-varying dependency between assets or factors. To that extent, Darolles et al. (2018) proposed a method based on a Cholesky-GARCH (CHAR) model to obtain a direct specification of the conditional betas. More precisely, let $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, \dots, \varepsilon_{m,t})$ a vector of $m \geq 2$ return series verifying a general multivariate volatility model

$$\boldsymbol{\varepsilon}_t = \boldsymbol{\Sigma}_t^{1/2} \boldsymbol{\eta}_t \quad (3.1)$$

where $(\boldsymbol{\eta}_t)$ is iid $(\mathbf{0}, \mathbf{I}_m)$. Standard Cholesky decomposition¹ of the covariance matrix yields

$$\boldsymbol{\Sigma}_t = \mathbf{L}_t \mathbf{G}_t \mathbf{L}_t' \quad (3.2)$$

where $\mathbf{G}_t = \text{diag}(\mathbf{g}_t)$ is a positive definite diagonal matrix, and \mathbf{L}_t is a lower unitriangular matrix. Following Pourahmadi (1999), let $\mathbf{v}_t = (v_{1,t}, \dots, v_{m,t})$, with $v_{k,t} = \sqrt{g_{k,t}} \eta_{k,t}$ for $k = 1, \dots, m$, the recursively obtained orthogonal factors from $\boldsymbol{\varepsilon}_t$. By taking $\boldsymbol{\Sigma}_t^{1/2} = \mathbf{L}_t \mathbf{G}_t^{1/2}$, we obtain

$$\boldsymbol{\varepsilon}_t = \boldsymbol{\Sigma}_t^{1/2} \boldsymbol{\eta}_t = \mathbf{L}_t \mathbf{v}_t$$

whence $\mathbf{B}_t \boldsymbol{\varepsilon}_t = \mathbf{v}_t$ where $\mathbf{B}_t = \mathbf{L}_t^{-1}$. Let $\ell_{ij,t}$ (resp. $-\beta_{ij,t}$) denote the row i and column j element of \mathbf{L}_t (resp. \mathbf{B}_t). Thus, the CHAR model writes as

$$\mathbf{B}_t \boldsymbol{\varepsilon}_t = \mathbf{G}_t^{1/2} \boldsymbol{\eta}_t \quad (3.3)$$

and allows a direct joint modeling of the dynamic conditional betas and the variance of the conditional orthogonal factors $v_{k,t}$.

Theoretical results for the CHAR model have been established under the assumption that the conditional volatilities \mathbf{g}_t are given by a multivariate GARCH equation, such as the Extended Constant Conditional Correlation GARCH model of Jeantheau (1998) where

$$\mathbf{g}_t = \boldsymbol{\omega} + \sum_{i=1}^q \mathbf{A}_i \mathbf{v}_{t-i}^2 + \sum_{j=1}^p \mathbf{B}_j \mathbf{g}_{t-j} \quad (3.4)$$

with $\mathbf{v}_t^2 = (v_{1,t}^2, \dots, v_{m,t}^2)$, and the vector $\boldsymbol{\omega} \in \mathbb{R}^m$ and the $m \times m$ matrices \mathbf{A}_i and \mathbf{B}_j have

¹see for example Meyer (2000)

nonnegative entries².

The CHAR model, as it is based on a multivariate GARCH equation, cannot account for the strong persistence of some financial returns series. In addition, although the betas dynamics are explicit, they do not exploit additional information than the one captured in the assets time series. Our proposed model alleviates the short-memory feature of model (3.4) and allows for exogenous variables. We consider a multivariate ARCH(∞) representation of the form

$$\mathbf{g}_t = \boldsymbol{\omega} + \sum_{i=1}^{\infty} \mathbf{A}_i \mathbf{v}_{t-i}^2 + \sum_{i=1}^{\infty} \boldsymbol{\Pi}_i \mathbf{X}_{t-i} \quad (3.5)$$

where $\mathbf{v}_t^2 = (v_{1,t}^2, \dots, v_{m,t}^2)$, and the vector $\boldsymbol{\omega} \in \mathbb{R}^m$ and the $m \times m$ matrices \mathbf{A}_i and $m \times P$ matrices $\boldsymbol{\Pi}_i$, $i \in \{1, 2, \dots\}$ have nonnegative entries. The vector \mathbf{X}_t contains P positive exogenous variables. We assume that $(\boldsymbol{\eta}'_t, \mathbf{X}'_t)'$ is stationary and ergodic, with $\boldsymbol{\eta}_t$ independent of $\{\mathbf{X}_{t-i}, i > 0\}$. Note that Model (3.4) actually admits a representation of the form (3.5) as shown by Conrad and Karanasos (2010), however each term in matrices \mathbf{A}_i is exponentially decaying as the ARCH(∞) kernel of a GARCH model.

In addition, the conditional betas dynamic is of the form

$$\boldsymbol{\beta}_t = -\text{vech}^0 \mathbf{B}_t = \mathbf{c}(\mathbf{v}_{t-1}, \dots, \mathbf{v}_{t-q}, \mathbf{g}_{t-1}^{1/2}, \dots, \mathbf{g}_{t-q}^{1/2}, \mathbf{X}_{t-1}, \dots, \mathbf{X}_{t-q}) + \sum_{k=1}^p \mathbf{C}_k \boldsymbol{\beta}_{t-k}, \quad (3.6)$$

for two integers p and q , where \mathbf{c} is any measurable function from $\mathbb{R}^{(2m+P)q}$ to $\mathbb{R}^{m(m-1)/2}$ and where vech^0 denotes the operator stacking the sub-columns of a square matrix. Alternatively, one can assume a similar dynamic model for the coefficients of the matrix \mathbf{L}_t , as

$$\boldsymbol{\ell}_t = \text{vech}^0 \mathbf{L}_t = \mathbf{c}(\mathbf{v}_{t-1}, \dots, \mathbf{v}_{t-q}, \mathbf{g}_{t-1}^{1/2}, \dots, \mathbf{g}_{t-q}^{1/2}, \mathbf{X}_{t-1}, \dots, \mathbf{X}_{t-q}) + \sum_{k=1}^p \mathbf{C}_k \boldsymbol{\ell}_{t-k}. \quad (3.7)$$

3.2.1 Stationarity conditions

We first note that (3.5) can be equivalently written as

$$\mathbf{g}_t = \boldsymbol{\omega}_t + \sum_{i=1}^{\infty} \mathbf{A}_i \boldsymbol{\Upsilon}_{t-i} \mathbf{g}_{t-i}, \quad (3.8)$$

where

$$\boldsymbol{\omega}_t = \boldsymbol{\omega} + \sum_{i=1}^{\infty} \boldsymbol{\Pi}_i \mathbf{X}_{t-i}, \quad \mathbf{v}_t^2 = \boldsymbol{\Upsilon}_t \mathbf{g}_t, \quad \boldsymbol{\Upsilon}_t = \begin{pmatrix} \eta_{1t}^2 & 0 & \dots & 0 \\ 0 & \ddots & & \\ \vdots & & \ddots & \\ 0 & \dots & & \eta_{mt}^2 \end{pmatrix}.$$

²Conrad and Karanasos (2010) relax the hypothesis of nonnegativity for the components of the matrices \mathbf{B}_j under the condition that the terms in the obtained ARCH(∞) kernel remain positive.

Let \odot denote the Hadamard product and for any matrix $\mathbf{A} = (a_{ij})$ with nonnegative entries, let $\mathbf{A}^{\odot r} = (a_{ij}^r)$ for $r \in (0, 1]$. For any square matrix \mathbf{B} , let $\rho(\mathbf{B})$ the spectral radius of \mathbf{B} . Let the matrices

$$\mathbf{A}^{\odot r} = \sum_{i=1}^{\infty} \mathbf{A}_i^{\odot r}, \quad \mathcal{P}^{\odot r} = \sum_{i=1}^{\infty} \mathbf{\Pi}_i^{\odot r},$$

with entries in $[0, \infty]$.

Theorem 3.1. *Suppose that for some $r \in (0, 1]$,*

(i) $\rho(\mathbf{A}^{\odot r} \mathbf{E} \mathbf{\Upsilon}_t^{\odot r}) < 1$, $\mathcal{P}^{\odot r} < \infty$ and $\mathbf{E} \mathbf{X}_t^{\odot r} < \infty$, componentwise, and

(ii) $\det \mathbf{C}(z) \neq 0$ for all $|z| < 1$, where $\mathbf{C}(z) = \left[\mathbf{I}_m - \sum_{k=1}^s \mathbf{C}_k z^k \right]$.

Then there exists a unique strictly stationary, ergodic and nonanticipative solution to the Cholesky-ARCH(∞) model (3.1)-(3.6) (or (3.1)-(3.5),(3.7)) satisfying $\mathbf{E}(\mathbf{g}_t^{\odot r}) < \infty$.

Remark 3.1. *Previous results on the existence of a stationary solution to multivariate ARCH(∞) model can be found in Doukhan et al. (2006). Their condition however depends on the choice of a multiplicative matrix norm, which is not the case here.*

Finiteness of moments of the returns process $(\boldsymbol{\varepsilon}_t)$ requires an additional assumption on the function \mathbf{c} in (3.6). In the next result, $\|\cdot\|$ denotes the Euclidean norm.

Proposition 3.1. *Suppose that there exists $K > 0$ such that $\|\mathbf{c}(\mathbf{x})\| \leq K(\|\mathbf{x}\| + 1)$ for any $\mathbf{x} \in \mathbb{R}^{(2m+P)q}$. Then, under the assumptions of Theorem 3.1, the unique strictly stationary, ergodic and nonanticipative solution to the Cholesky-ARCH(∞) model (3.1)-(3.5) and (3.7) satisfies $\|\boldsymbol{\varepsilon}_t\|_r < \infty$ with $r \leq 1/2$.*

3.2.2 Particular specifications

In a recent article, Royer (2022) proposes an extension to the GARCH model to account for a higher persistence. A particular specification is given by the ARCH(∞) model

$$\varepsilon_t = \sigma_t \eta_t, \quad \sigma_t^2 = \omega + \sum_{i=1}^{\infty} (\alpha \beta^{i-1} + \gamma i^{-d-1}) \varepsilon_{t-i}^2 \quad (3.9)$$

which nests the classical GARCH(1,1). Building upon this volatility specification, and taking into account possible exogenous covariates, a particular form of model (3.5) could be for any asset $k = 1, \dots, m$

$$g_{k,t} = \omega + \underbrace{\sum_{i=1}^{\infty} (\alpha_k \beta_k^{i-1} + \gamma_k i^{-d-1}) v_{k,t-i}^2}_{\text{specific ARCH}(\infty) \text{ effect}} + \underbrace{\sum_{j \neq k} \delta_{k,j} v_{j,t-1}^2}_{\text{spillover}} + \underbrace{\boldsymbol{\pi}'_k \mathbf{X}_{t-1}}_{\text{exogeneity}}$$

with $j = 1, \dots, m$. Remark that in the case $\delta_{k,j} > 0$ for all k, j , each asset will have an effect on all the conditional volatilities of the system. Moreover, if one of the asset has long (*moderate*) memory, all of the conditional volatility processes will have a persistent component. This could

be problematic when modeling liquid assets as they seem to present rather short memory, as noted by Royer (2022). To avoid this problem, it could be interesting to define a triangular system similar to the one proposed in Darolles et al. (2018). In the latter, the dynamics of the asset on the k -th row depends only on the previous rows, leading to

$$g_{k,t}^2 = \omega_k + \underbrace{\sum_{i=1}^{\infty} [\alpha_k \beta_k^{i-1} + \gamma_k i^{-d-1}] v_{k,t-i}^2}_{\text{ARCH}(\infty) \text{ effect}} + \underbrace{\sum_{j=1}^{k-1} \delta_{k,j} v_{j,t-1}^2}_{\text{spillover}} + \underbrace{\boldsymbol{\pi}'_k \mathbf{X}_{t-1}}_{\text{exogeneity}}. \quad (3.10)$$

Ordering the assets by their order of persistence would thus allow to have short memory for the assets on the top rows while having higher persistence for the assets on the bottom rows. This approach relates to the ordering based on liquidity proposed by Boudt et al. (2017) for the CholCov model.

Royer (2022) proposed an asymmetric extension of (3.9) to take into account the presence of both conditional asymmetry and high persistence of return volatility. In Model (3.10), we prefer to introduce asymmetry effects via the exogenous variables. For instance we could take

$$\boldsymbol{\pi}'_k \mathbf{X}_{t-1} = \pi_{k,+} X_{t-1} \mathbf{1}_{X_{t-1} \leq X_{t-2}} + \pi_{k,-} X_{t-1} \mathbf{1}_{X_{t-1} > X_{t-2}}$$

where X_t is an implied volatility index of the corresponding market (for example the VIX), or a realized measure of volatility as proposed by Barndorff-Nielsen and Shephard (2002). Another popular approach is to use two exogenous variables, respectively capturing a downside or an upside risk. Possible candidates could be realized semivariances as advocated by Barndorff-Nielsen et al. (2010). The reason for introducing asymmetry via exogenous variables is both empirical and statistical. Indeed, when studying multivariate GARCH models, Bollerslev et al. (2020b) argue that modeling asymmetry through realized semicovariance matrices improve model fit compared to classical threshold models. Moreover, a volatility equation with terms of the form $\mathbf{1}_{v_{k,t-j} > 0}$ would lead to major technical difficulties due to the non-differentiability with respect to the parameters when the factor $v_{k,t}$ is not observed (which is the case for $k > 1$).

An advantage of introducing a dynamic model on the Cholesky decomposition of the conditional variance is that the positive definiteness of $\boldsymbol{\Sigma}_t$ is very easily guaranteed. Only the positivity of the elements of \mathbf{g}_t is needed. There is no sign constraint on the elements of $\boldsymbol{\beta}_t$. We can thus consider conditional betas of the form

$$\beta_{ij,t} = \varpi_{ij} + \boldsymbol{\pi}'_{ij} \mathbf{X}_{ij,t-1} + \alpha_{ij} v_{i,t-1}^2 + \gamma_{ij} v_{j,t-1}^2 + c_{ij} \beta_{ij,t-1} \quad (3.11)$$

where $\mathbf{X}_{ij,t-1}$ is a vector of potential beta predictors.

An interesting specification of (3.11) could be to include realized betas lower frequencies in $\mathbf{X}_{ij,t-1}$. Building upon the realized volatility literature, realized betas were first introduced by Barndorff-Nielsen and Shephard (2004) and Andersen et al. (2006) and are computed using

higher frequency returns over fixed time intervals. Let $\boldsymbol{\varepsilon}_{w,t} = (\varepsilon_{w,1,t}, \dots, \varepsilon_{w,m,t})'$ the vector of returns over the time interval w . The realized covariance matrix $\mathbf{C}_t^{w,W}$ is then defined as the sum of the W outer products of the high-frequency return vectors. For example, w could be a day and W a month, or w could be a 5-minute interval and W a day.

$$\mathbf{C}_t = \mathbf{C}_t^{w,W} = \sum_{w=1}^W \boldsymbol{\varepsilon}_{w,t} \boldsymbol{\varepsilon}_{w,t}'$$

In addition, for $1 < k \leq m$, for any $m \times m$ square matrix \mathbf{M} , let $\mathbf{M}_{[kk]}$ denote the square sub-matrix formed by the first $k-1$ rows and columns and $\mathbf{M}_{[k]}$ the sub-vector formed by the first $k-1$ rows and the k -th column of \mathbf{M} . For $1 < i \leq m$, the realized betas between asset i and assets $j = 1, \dots, i-1$ are then defined by

$$\bar{\boldsymbol{\beta}}_{i,t} = (\bar{\beta}_{ij,t})_{j=1,\dots,i-1} = (\mathbf{C}_{t[ii]})^{-1} \mathbf{C}_{t[i]}$$

We denote $\bar{\beta}_{ij,t}^{\mathcal{W}}$ the weekly realized beta at time t computed over 5 trading days, and $\bar{\beta}_{ij,t}^{\mathcal{M}}$ the monthly realized beta at time t computed over 21 trading days. Ignoring the squared terms in (3.11) and letting $\mathbf{X}_{ij,t-1} = (\bar{\beta}_{ij,t-1}^{\mathcal{W}}, \bar{\beta}_{ij,t-1}^{\mathcal{M}})$ yields

$$\beta_{ij,t} = \varpi_{ij} + \pi_{ij}^{\mathcal{W}} \bar{\beta}_{ij,t-1}^{\mathcal{W}} + \pi_{ij}^{\mathcal{M}} \bar{\beta}_{ij,t-1}^{\mathcal{M}} + c_{ij} \beta_{ij,t-1} \quad (3.12)$$

allowing to model the conditional beta as an exponential smoothing of a combination of realized betas at lower frequencies. In that sense, Equation (3.12) is related to the fast growing literature on component volatility models and in particular to the HAR model of Corsi (2009)³.

Alternatively, potential useful exogenous variables in the conditional betas dynamics could be the recently introduced realized semibetas of Bollerslev et al. (2022). Consider the realized covariance matrix decomposition

$$\mathbf{C}_t = \mathbf{P}_t + \mathbf{N}_t + \mathbf{S}_t^+ + \mathbf{S}_t^-$$

into the semicovariance components introduced by Bollerslev et al. (2020a)

$$\begin{aligned} \mathbf{P}_t &= \sum_{w=1}^W \boldsymbol{\varepsilon}_{w,t}^+ \boldsymbol{\varepsilon}_{w,t}^{+'}, & \mathbf{S}_t^+ &= \sum_{w=1}^W \boldsymbol{\varepsilon}_{w,t}^- \boldsymbol{\varepsilon}_{w,t}^{+'}, \\ \mathbf{N}_t &= \sum_{w=1}^W \boldsymbol{\varepsilon}_{w,t}^- \boldsymbol{\varepsilon}_{w,t}^{-'}, & \mathbf{S}_t^- &= \sum_{w=1}^W \boldsymbol{\varepsilon}_{w,t}^+ \boldsymbol{\varepsilon}_{w,t}^{-'}. \end{aligned}$$

where $\boldsymbol{\varepsilon}_{w,t}^{+(-)}$ denotes the componentwise positive (respectively negative) part of $\boldsymbol{\varepsilon}_{w,t}$. Each realized semicovariance component then yields realized semibetas

$$\begin{aligned} \bar{\boldsymbol{\beta}}_{i,t}^{\mathcal{P}} &= (\mathbf{C}_{t[ii]})^{-1} \mathbf{P}_{t[i]}, & \bar{\boldsymbol{\beta}}_{i,t}^{\mathcal{S}^+} &= -(\mathbf{C}_{t[ii]})^{-1} \mathbf{S}_{t[i]}^+, \\ \bar{\boldsymbol{\beta}}_{i,t}^{\mathcal{N}} &= (\mathbf{C}_{t[ii]})^{-1} \mathbf{N}_{t[i]}, & \bar{\boldsymbol{\beta}}_{i,t}^{\mathcal{S}^-} &= -(\mathbf{C}_{t[ii]})^{-1} \mathbf{S}_{t[i]}^-. \end{aligned}$$

³Equation (3.12) can also be related to the Midas model of Ghysels et al. (2006) although the step function of the HAR model cannot be reproduced using the Beta function lag polynomial of Midas regressions.

providing a four-way decomposition of the traditional realized betas

$$\bar{\beta}_{i,t} = \bar{\beta}_{i,t}^{\mathcal{P}} + \bar{\beta}_{i,t}^{\mathcal{N}} - \bar{\beta}_{i,t}^{\mathcal{S}^+} - \bar{\beta}_{i,t}^{\mathcal{S}^-}. \quad (3.13)$$

Letting $\mathbf{X}_{ij,t-1} = (\bar{\beta}_{ij,t-1}^{\mathcal{P}}, \bar{\beta}_{ij,t-1}^{\mathcal{N}}, \bar{\beta}_{ij,t-1}^{\mathcal{S}^+}, \bar{\beta}_{ij,t-1}^{\mathcal{S}^-})$, the conditional beta dynamics could then be given by

$$\beta_{ij,t} = \varpi_{ij} + \pi_{ij}^{\mathcal{P}} \bar{\beta}_{ij,t-1}^{\mathcal{P}} + \pi_{ij}^{\mathcal{N}} \bar{\beta}_{ij,t-1}^{\mathcal{N}} + \pi_{ij}^{\mathcal{S}^+} \bar{\beta}_{ij,t-1}^{\mathcal{S}^+} + \pi_{ij}^{\mathcal{S}^-} \bar{\beta}_{ij,t-1}^{\mathcal{S}^-} + c_{ij} \beta_{ij,t-1} \quad (3.14)$$

allowing for a more accurate contribution of downside and upside risk.

3.3 Statistical inference

We now assume that the conditional variance and its Cholesky decomposition

$$\Sigma_t = \Sigma_t(\boldsymbol{\theta}_0) = \mathbf{L}_t \mathbf{G}_t \mathbf{L}_t' = \mathbf{L}_t(\boldsymbol{\theta}_0) \mathbf{G}_t(\boldsymbol{\theta}_0) \mathbf{L}_t'(\boldsymbol{\theta}_0) \quad (3.15)$$

are parameterized by a d -dimensional parameter $\boldsymbol{\theta}_0$ belonging to a compact parameter space Θ , and that, for all $\boldsymbol{\theta} \in \Theta$, $\mathbf{L}_t(\boldsymbol{\theta})$, $\mathbf{G}_t(\boldsymbol{\theta})$, and thus $\Sigma_t(\boldsymbol{\theta}) = \mathbf{L}_t(\boldsymbol{\theta}) \mathbf{G}_t(\boldsymbol{\theta}) \mathbf{L}_t'(\boldsymbol{\theta})$, are \mathcal{F}_{t-1} measurable, where \mathcal{F}_t is the sigma-field generated by $\{\boldsymbol{\eta}_u, \mathbf{X}_u, u \leq t\}$.

We also introduce the vector $\mathbf{g}_t(\boldsymbol{\theta})$ of generic element $g_{it}(\boldsymbol{\theta})$ such that

$$\mathbf{g}_t = \mathbf{g}_t(\boldsymbol{\theta}_0) = \mathbf{g}(\boldsymbol{\varepsilon}_{t-1}, \boldsymbol{\varepsilon}_{t-2}, \dots, \mathbf{X}_{t-1}, \mathbf{X}_{t-2}, \dots; \boldsymbol{\theta}_0),$$

and the vectors $\boldsymbol{\beta}_t(\boldsymbol{\theta})$ and $\boldsymbol{\ell}_t(\boldsymbol{\theta})$ of \mathbb{R}^{m_0} , with $m_0 = (m-1)m/2$, such that

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_t(\boldsymbol{\theta}_0) = \boldsymbol{\beta}(\boldsymbol{\varepsilon}_{t-1}, \boldsymbol{\varepsilon}_{t-2}, \dots, \mathbf{X}_{t-1}, \mathbf{X}_{t-2}, \dots; \boldsymbol{\theta}_0)$$

and $\boldsymbol{\ell}_t = \boldsymbol{\ell}_t(\boldsymbol{\theta}_0) = \boldsymbol{\ell}(\boldsymbol{\varepsilon}_{t-1}, \boldsymbol{\varepsilon}_{t-2}, \dots, \mathbf{X}_{t-1}, \mathbf{X}_{t-2}, \dots; \boldsymbol{\theta}_0)$. Given the observations $\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_n$, $\mathbf{X}_1, \dots, \mathbf{X}_n$, and arbitrary fixed initial values $\tilde{\boldsymbol{\varepsilon}}_i$ and $\tilde{\mathbf{X}}_i$ for $i \leq 0$, let

$$\tilde{\Sigma}_t(\boldsymbol{\theta}) = \Sigma(\boldsymbol{\varepsilon}_{t-1}, \dots, \boldsymbol{\varepsilon}_1, \tilde{\boldsymbol{\varepsilon}}_0, \tilde{\boldsymbol{\varepsilon}}_{-1}, \dots, \mathbf{X}_{t-1}, \mathbf{X}_{t-2}, \dots; \boldsymbol{\theta})$$

and similarly define $\tilde{\mathbf{L}}_t(\boldsymbol{\theta})$, $\tilde{\mathbf{G}}_t(\boldsymbol{\theta}) = \text{diag}\{\tilde{\mathbf{g}}_t(\boldsymbol{\theta})\}$, $\tilde{\mathbf{B}}_t(\boldsymbol{\theta})$, $\tilde{\boldsymbol{\ell}}_t(\boldsymbol{\theta}) = \text{vec}^0 \tilde{\mathbf{L}}_t(\boldsymbol{\theta})$ and $\tilde{\boldsymbol{\beta}}_t(\boldsymbol{\theta}) = -\text{vec}^0 \tilde{\mathbf{B}}_t(\boldsymbol{\theta})$. A QMLE of $\boldsymbol{\theta}_0$ is defined as any measurable solution $\hat{\boldsymbol{\theta}}_n$ of

$$\hat{\boldsymbol{\theta}}_n = \arg \min_{\boldsymbol{\theta} \in \Theta} \tilde{O}_n(\boldsymbol{\theta}), \quad \tilde{O}_n(\boldsymbol{\theta}) = n^{-1} \sum_{t=1}^n \tilde{q}_t(\boldsymbol{\theta}), \quad (3.16)$$

where

$$\tilde{q}_t(\boldsymbol{\theta}) = \boldsymbol{\varepsilon}_t' \tilde{\Sigma}_t^{-1}(\boldsymbol{\theta}) \boldsymbol{\varepsilon}_t + \log |\tilde{\Sigma}_t(\boldsymbol{\theta})| = \boldsymbol{\varepsilon}_t' \tilde{\mathbf{B}}_t'(\boldsymbol{\theta}) \tilde{\mathbf{G}}_t^{-1}(\boldsymbol{\theta}) \tilde{\mathbf{B}}_t(\boldsymbol{\theta}) \boldsymbol{\varepsilon}_t + \sum_{i=1}^m \log \tilde{g}_{it}(\boldsymbol{\theta}).$$

Note that the Cholesky decomposition facilitates the computation of the QMLE, the diagonal matrix $\tilde{\mathbf{G}}_t(\boldsymbol{\theta})$ being much easier to invert than $\tilde{\boldsymbol{\Sigma}}_t(\boldsymbol{\theta})$.

Darolles et al. (2018) (DFL hereafter) gave a set of regularity conditions which entails the strong consistency and asymptotic normality (CAN) of the QML estimator $\hat{\boldsymbol{\theta}}_n$. However, their framework does not allow for exogenous variables. Moreover their assumptions (in particular their assumption **A2**) preclude the ARCH(∞) models that we consider in the present paper.

We will thus give a set of alternative conditions for CAN of the QMLE. In the sequel ρ denotes a generic constant belonging to $[0, 1)$, and K denotes a positive constant or a positive random variable which is \mathcal{F}_0 -measurable. We take the spectral norm as matrix norm and the Euclidean norm as vector norm. Assume

$$\mathbf{A1:} \sup_{\boldsymbol{\theta} \in \Theta} \left\| \tilde{\mathbf{G}}_t^{-1}(\boldsymbol{\theta}) \right\| \leq K, \quad \sup_{\boldsymbol{\theta} \in \Theta} \left\| \mathbf{G}_t^{-1}(\boldsymbol{\theta}) \right\| \leq K, \quad \text{a.s.}$$

$$\mathbf{A2:} \sup_{\boldsymbol{\theta} \in \Theta} \left\| \boldsymbol{\beta}_t(\boldsymbol{\theta}) - \tilde{\boldsymbol{\beta}}_t(\boldsymbol{\theta}) \right\| \leq K\rho_t \quad \text{where the random variable } \rho_t \text{ satisfies} \\ \sum_{t=1}^{\infty} \{E\rho_t^{s_1}\}^{1/p} < \infty \text{ for all } s_1 \in (0, s_0] \text{ and some } s_0 > 0 \text{ and some } p > 1.$$

$$\mathbf{A3:} E \left\{ \|\boldsymbol{\ell}_t(\boldsymbol{\theta}_0)\|^{s_0} + \|\mathbf{g}_t(\boldsymbol{\theta}_0)\|^{s_0} + \sup_{\boldsymbol{\theta} \in \Theta} \|\boldsymbol{\beta}_t(\boldsymbol{\theta})\|^{s_0} \right\} < \infty \text{ for some } s_0 > 0.$$

$$\mathbf{A4:} \text{For } \boldsymbol{\theta} \in \Theta, \{\mathbf{g}_t(\boldsymbol{\theta}), \boldsymbol{\beta}_t(\boldsymbol{\theta})\} = \{\mathbf{g}_t(\boldsymbol{\theta}_0), \boldsymbol{\beta}_t(\boldsymbol{\theta}_0)\} \text{ a.s. implies } \boldsymbol{\theta} = \boldsymbol{\theta}_0.$$

A5: For any sequence $\mathbf{y}_1, \mathbf{y}_2, \dots$ of vectors of \mathbb{R}^m and any sequence $\mathbf{x}_1, \mathbf{x}_2, \dots$ of vectors of \mathbb{R}^r , the functions $\boldsymbol{\theta} \mapsto \mathbf{g}(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{x}_1, \mathbf{x}_2, \dots; \boldsymbol{\theta})$ from Θ to $(0, +\infty)^m$ and $\boldsymbol{\theta} \mapsto \boldsymbol{\beta}(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{x}_1, \mathbf{x}_2, \dots; \boldsymbol{\theta})$ from Θ to \mathbb{R}^{m_0} are continuous on Θ .

A6: $\boldsymbol{\theta}_0$ belongs to the interior $\mathring{\Theta}$ of Θ .

A7: For any sequence $\mathbf{y}_1, \mathbf{y}_2, \dots$ of vectors of \mathbb{R}^m and any sequence $\mathbf{x}_1, \mathbf{x}_2, \dots$ of vectors of \mathbb{R}^r , the functions $\boldsymbol{\theta} \mapsto \mathbf{g}(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{x}_1, \mathbf{x}_2, \dots; \boldsymbol{\theta})$ and $\boldsymbol{\theta} \mapsto \boldsymbol{\beta}(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{x}_1, \mathbf{x}_2, \dots; \boldsymbol{\theta})$ admit continuous second-order derivatives.

A8: For some neighborhood $V(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$, ρ_t as in **A2** and $s_0 > 0$

$$\sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\| \frac{\partial \boldsymbol{\beta}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \frac{\partial \tilde{\boldsymbol{\beta}}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\| \leq K\rho_t, \quad E \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\| \frac{\partial \boldsymbol{\beta}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\|^{s_0} + \left\| \frac{\partial \mathbf{g}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\|^{s_0} < \infty.$$

A9: For some neighborhood $V(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$, for all $i, j \in \{1, \dots, m\}$ and $p > 0$, $q > 0$ and $r > 0$ such that $2q^{-1} + 2r^{-1} = 1$ and $p^{-1} + 2r^{-1} = 1$, we have

$$E \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\| \boldsymbol{\Sigma}_t^{-1/2'}(\boldsymbol{\theta}) \frac{\partial^2 \boldsymbol{\Sigma}_t(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}) \right\|^p < \infty, \\ E \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\| \boldsymbol{\Sigma}_t^{-1/2'}(\boldsymbol{\theta}) \frac{\partial \boldsymbol{\Sigma}_t(\boldsymbol{\theta})}{\partial \theta_i} \boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}) \right\|^q < \infty, \\ E \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\| \boldsymbol{\Sigma}_t^{1/2'}(\boldsymbol{\theta}_0) \boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}) \right\|^r < \infty,$$

where θ_i denotes the i -th element of the vector $\boldsymbol{\theta}$.

A10: $E\|\boldsymbol{\eta}_t\|^4 < \infty$.

A11: The matrices $\{\partial\Sigma_t(\boldsymbol{\theta}_0)/\partial\theta_i, i = 1, \dots, d\}$ are linearly independent with nonzero probability.

The previous assumptions are essentially the same as those of Darolles et al. (2018). Some of them have been extended to incorporate exogenous variables. To deal with these covariates we assume the following.

A1*: (\mathbf{X}_t) is a sequence of positive random vectors such that $(\boldsymbol{\eta}'_t, \mathbf{X}'_t)'$ is stationary and ergodic, with $\boldsymbol{\eta}_t$ independent of $\{\mathbf{X}_{t-i}, i > 0\}$.

DFL made the assumptions **A2** and **A8** by replacing $\boldsymbol{\beta}_t(\boldsymbol{\theta})$ and $\tilde{\boldsymbol{\beta}}_t(\boldsymbol{\theta})$ by $\mathbf{g}_t(\boldsymbol{\theta})$ and $\tilde{\mathbf{g}}_t(\boldsymbol{\theta})$. Such assumptions are not satisfied for volatilities of the form (3.10). We thus replace these assumptions by the following.

A2*: As $t \rightarrow \infty$, $\sup_{\boldsymbol{\theta} \in \Theta} \|\mathbf{g}_t(\boldsymbol{\theta}) - \tilde{\mathbf{g}}_t(\boldsymbol{\theta})\| \rightarrow 0$ a.s.

To simplify the notation we omit " $\boldsymbol{\theta}$ ", and write for instance \mathbf{B}_t and \mathbf{L}_{0t} instead of $\mathbf{B}_t(\boldsymbol{\theta})$ and $\mathbf{L}_t(\boldsymbol{\theta}_0)$.

A3*: There exist three conjugate numbers $p > 0$, $q > 0$ and $r > 0$ (such that $p^{-1} + q^{-1} + r^{-1} = 1$) and a neighborhood $V(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ such that

$$E \left\{ \|\mathbf{L}_{0t}\|^{2p} + \|\mathbf{G}_{0t}\|^{2q} + \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \|\mathbf{B}_t\|^{2r} \right\} < \infty$$

A4*: For $i = 1, \dots, m$, there exist $\xi > 0$, $\rho \in (0, 1]$, $a > 1 - \rho/2$, three conjugate numbers $p > 0$, $q > 0$, $r > 0$, and a neighborhood $V(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ such that

$$E \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\{ \|\mathbf{B}_t \mathbf{L}_{0t}\|^{2p\rho(\frac{1+\xi}{\xi})} + \left\| \mathbf{G}_t^{-1} \frac{\partial \mathbf{G}_t}{\partial \theta_i} \right\|^{q\rho(\frac{1+\xi}{\xi})} + \|\mathbf{G}_{0t}\|^{2r\rho(\frac{1+\xi}{\xi})} \right\} < \infty \quad (3.17)$$

and

$$\sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \|\mathbf{g}_t - \tilde{\mathbf{g}}_t\|_{1+\xi}^\rho \leq \frac{K}{t^a}.$$

A5*: **A4*** holds when (3.17) is replaced by

$$E \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\{ \|\mathbf{B}_t \mathbf{L}_{0t}\|^{p\rho(\frac{1+\xi}{\xi})} + \left\| \frac{\partial \mathbf{B}_t}{\partial \theta_i} \mathbf{L}_{0t} \right\|^{q\rho(\frac{1+\xi}{\xi})} + \|\mathbf{G}_{0t}\|^{2r\rho(\frac{1+\xi}{\xi})} \right\} < \infty$$

Theorem 3.2 (CAN of the QMLE in the general case). *Let $(\boldsymbol{\varepsilon}_t)$ be a volatility model (3.1) satisfying the \mathcal{F}_{t-1} -measurable Cholesky decomposition (3.15).⁴ Let $(\hat{\boldsymbol{\theta}}_n)$ be a sequence of QML estimators satisfying (3.16). Under **A1-A5** and **A1*-A3*** we have*

$$\hat{\boldsymbol{\theta}}_n \rightarrow \boldsymbol{\theta}_0, \quad \text{almost surely as } n \rightarrow \infty.$$

⁴In particular, under **A1***, $\boldsymbol{\varepsilon}_t$ is stationary and ergodic as fixed measurable function of $\{\boldsymbol{\eta}_u, \mathbf{X}_u, u < t\}$

Under the additional assumptions **A6-A10** and **A4*-A5***, we have the existence of the $d \times d$ matrix

$$\mathbf{J} = E\mathbf{D}'_t \{ \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0) \otimes \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0) \} \mathbf{D}_t, \quad \mathbf{D}_t = \frac{\partial \text{vec} \boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'},$$

and of the $d \times d$ matrix \mathbf{I} of generic term

$$\mathbf{I}(i, j) = \text{Tr} \{ \mathbf{K} E \mathbf{C}_{j,t} \mathbf{C}'_{i,t} \},$$

with $\mathbf{K} = E \text{vec}(\mathbf{I}_m - \boldsymbol{\eta}_t \boldsymbol{\eta}'_t) \text{vec}'(\mathbf{I}_m - \boldsymbol{\eta}_t \boldsymbol{\eta}'_t)$, \mathbf{I}_m the identity matrix of size m and

$$\mathbf{C}_{i,t} = \left\{ \boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0) \otimes \boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0) \right\} \text{vec} \frac{\partial \boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)}{\partial \theta_i}.$$

Moreover, under the additional assumption **A11**, \mathbf{J} is invertible and

$$\sqrt{n} \left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \right) \xrightarrow{L} \mathcal{N} \{ 0, \mathbf{J}^{-1} \mathbf{I} \mathbf{J}^{-1} \} \text{ as } n \rightarrow \infty. \quad (3.18)$$

We also have the Bahadur representation

$$\sqrt{n} \left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \right) = \mathbf{J}^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^n \nabla_t \text{vec}(\boldsymbol{\eta}_t \boldsymbol{\eta}'_t - \mathbf{I}_m) + o_P(1), \quad (3.19)$$

where $\nabla_t = \mathbf{D}'_t \left\{ \boldsymbol{\Sigma}_t^{-1/2'}(\boldsymbol{\theta}_0) \otimes \boldsymbol{\Sigma}_t^{-1/2'}(\boldsymbol{\theta}_0) \right\}$.

3.4 Simulations

In order to assess the finite sample properties of the QMLE, we carry out some Monte Carlo experiments. In the following simulations, we use Gaussian innovations ($\boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$).

To conduct our experiments, we first focus on specification (3.10)-(3.11) where we do not include exogenous variables in the volatility and where we consider the monthly realized beta between the Agricultural industry and the Market portfolio as an exogenous variable in the conditional betas dynamics. We simulate a thousand samples of size 5000 of the Cholesky-ARCH(∞)-X and on each sample we compute the QMLE using (i) the realized beta as exogenous variables (ii) the realized semibetas as exogenous variables. The empirical mean and Bias of the obtained QMLE are reported in Table 3.1. We see that the QMLE performs well. In particular, since the realized betas can be decomposed into semibetas, we see that the four coefficients estimated on the exogenous variables are well estimated and equal in absolute value to the coefficient π_{21} used in the simulation.

We then focus on specification (3.10)-(3.14) where we use realized semibetas between the Agricultural industry and the Market portfolio as an exogenous variable in the conditional betas dynamics. Results are reported in Table 3.2 and show similar good properties of the estimator.

	θ_0	$\hat{\theta}_n$	Bias	$\hat{\theta}_n$	Bias
ω_1	0.10	0.1014	0.0014	0.1014	0.0014
α_1	0.10	0.0940	-0.0060	0.0940	-0.0060
β_1	0.85	0.8541	0.0041	0.8541	0.0041
γ_1	0.00	0.0108	0.0108	0.0108	0.0108
ω_2	0.20	0.2007	0.0007	0.2005	0.0005
α_2	0.10	0.1073	0.0073	0.1074	0.0074
β_2	0.70	0.6969	-0.0031	0.6967	-0.0033
γ_2	0.10	0.0917	-0.0083	0.0917	-0.0083
δ_{21}	0.10	0.0999	-0.0001	0.0998	-0.0002
$\bar{\omega}_{21}$	0.70	0.7342	0.0342	0.7398	0.0398
c_{21}	0.70	0.6852	-0.0148	0.6823	-0.0177
π_{21}	0.20	0.2098	0.0098	-	-
$\pi_{21}^{\mathcal{P}}$	-	-	-	0.2121	0.0121
$\pi_{21}^{\mathcal{N}}$	-	-	-	0.2123	0.0123
$\pi_{21}^{S^+}$	-	-	-	-0.2074	-0.0074
$\pi_{21}^{S^-}$	-	-	-	-0.2081	-0.0081

Table 3.1: *Estimation results for 1000 simulations of size 5000 of a Cholesky ARCH(∞) with realized betas*

	θ_0	$\hat{\theta}_n$	Bias
ω_1	0.10	0.1026	0.0026
α_1	0.10	0.0994	-0.0006
β_1	0.85	0.8491	-0.0009
ω_2	0.20	0.2015	0.0015
α_2	0.10	0.0991	-0.0009
β_2	0.70	0.6996	-0.0004
δ_{21}	0.10	0.1004	0.0004
$\bar{\omega}_{21}$	0.70	0.7059	0.0059
c_{21}	0.70	0.6970	-0.0030
$\pi_{21}^{\mathcal{P}}$	0.20	0.2022	0.0022
$\pi_{21}^{\mathcal{N}}$	0.50	0.5059	0.0059
$\pi_{21}^{S^+}$	0.00	0.0018	0.0018
$\pi_{21}^{S^-}$	0.00	0.0025	0.0025

Table 3.2: *Estimation results for 1000 simulations of size 5000 of a Cholesky GARCH with realized semibetas*

3.5 Application: Do semibetas matter when modeling conditional betas?

In this section we consider the modeling of the conditional betas between 35 industries portfolios and the Market portfolio. Our dataset contains daily returns from January 2000 to December 2020 and are obtained from Kenneth French's website.

For any couple of time series $(\varepsilon_{M,t}, \varepsilon_{i,t})$ where M denotes the Market portfolio and i an industry portfolio, we fit the bivariate Cholesky ARCH(∞) model defined by the volatility processes

$$\begin{aligned}
g_{M,t}^2 &= \omega_M + \sum_{k=1}^{\infty} [\alpha_M \beta_M^{k-1} + \gamma_M k^{-2}] \varepsilon_{M,t-k}^2 \\
g_{i,t}^2 &= \omega_i + \sum_{k=1}^{\infty} [\alpha_i \beta_i^{k-1} + \gamma_i k^{-2}] v_{i,t-k}^2 + \delta_i \varepsilon_{M,t-1}^2
\end{aligned}$$

where

$$v_{i,t} = \varepsilon_{i,t} - \beta_{i,M,t} \varepsilon_{M,t}.$$

In particular, we consider the following two alternative conditional beta dynamics

$$\beta_{i,M,t} = \varpi_{i,M} + \pi_{i,M}^{\mathcal{RB}} \bar{\beta}_{i,M,t} + c_{i,M} \beta_{i,M,t} \quad (3.20)$$

and

$$\beta_{i,M,t} = \varpi_{i,M} + \pi_{i,M}^{\mathcal{P}} \bar{\beta}_{i,M,t}^{\mathcal{P}} + \pi_{i,M}^{\mathcal{N}} \bar{\beta}_{i,M,t}^{\mathcal{N}} + \pi_{i,M}^{\mathcal{S}^+} \bar{\beta}_{i,M,t}^{\mathcal{S}^+} + \pi_{i,M}^{\mathcal{S}^-} \bar{\beta}_{i,M,t}^{\mathcal{S}^-} + c_{i,M} \beta_{i,M,t}. \quad (3.21)$$

Figures 3.1a and 3.1b show the obtained conditional betas for the Agriculture industry portfolio and the Paper industry portfolio. We see clearly that all conditional betas appear very different from the constant conditional betas. However, the betas obtained from dynamic (3.21) are relatively close to the ones obtained from (3.20) using only the realized betas as exogenous variables.

We propose to test that the decomposition of the realized beta into its semibetas components improve the modeling of the conditional betas. From (3.13), equation (3.21) is equivalent to (3.20) if $\pi_{ij}^{\mathcal{P}} = \pi_{ij}^{\mathcal{N}} = -\pi_{ij}^{\mathcal{S}^+} = -\pi_{ij}^{\mathcal{S}^-}$. Testing the informational benefit of including realized semibetas instead of standard realized betas can thus be achieved by testing

$$\begin{cases} H_0 : \pi_{ij}^{\mathcal{P}} = \pi_{ij}^{\mathcal{N}} & \text{and} & \pi_{ij}^{\mathcal{P}} = -\pi_{ij}^{\mathcal{S}^+} & \text{and} & \pi_{ij}^{\mathcal{N}} = -\pi_{ij}^{\mathcal{S}^-} \\ H_1 : \pi_{ij}^{\mathcal{P}} \neq \pi_{ij}^{\mathcal{N}} & \text{or} & \pi_{ij}^{\mathcal{P}} \neq -\pi_{ij}^{\mathcal{S}^+} & \text{or} & \pi_{ij}^{\mathcal{N}} \neq -\pi_{ij}^{\mathcal{S}^-} \end{cases}$$

which can be rewritten as the linear constraint on the parameter vector $\boldsymbol{\theta}_0$

$$\begin{cases} H_0 : \mathbf{R}\boldsymbol{\theta}_0 = (0, 0, 0)' \\ H_1 : \mathbf{R}\boldsymbol{\theta}_0 \neq (0, 0, 0)'. \end{cases}$$

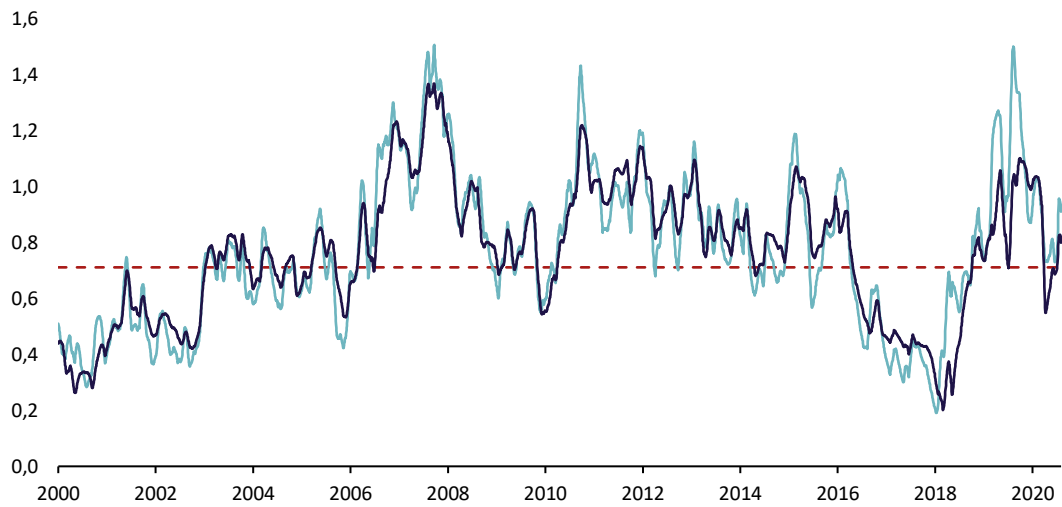
We thus can use the standard Wald statistic

$$W_n = n(\mathbf{R} \hat{\boldsymbol{\theta}}_n)' (\mathbf{R} \mathbf{J}_n^{-1} \mathbf{I}_n \mathbf{J}_n^{-1} \mathbf{R}')^{-1} (\mathbf{R} \hat{\boldsymbol{\theta}}_n)$$

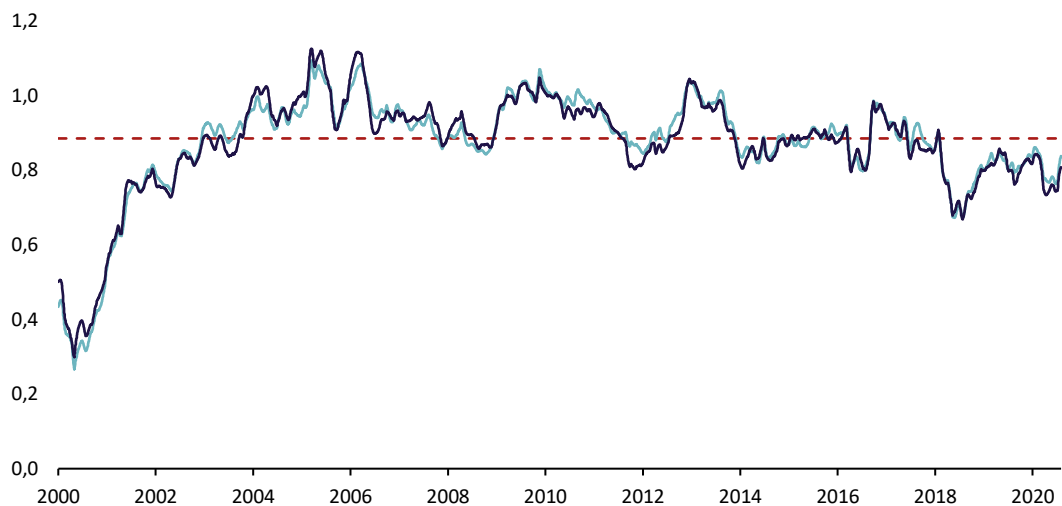
which asymptotically follows a χ^2 distribution with 3 degrees of freedom from standard results (see for example Gouriéroux and Monfort (1995)).

Table 3.3 shows p -values for each industry portfolio and provides a strong argument in favor of using semibetas rather than realized betas as covariates when modeling conditional betas.

In addition, Table 3.4 gives estimated values of γ_{Mkt} and γ_{indus} parameters for each industry portfolio. The γ parameter in Equation (3.10) is of particular interest as it allows to differentiate between a GARCH(1,1) and a more persistent volatility process. Interestingly, we see that the Market portfolio seems to exhibit short memory as γ_{Mkt} is equal to 0. This result is similar to the finding of Royer (2022) when considering capitalization-weighted equity indices from developed markets. However the picture is very different for industry portfolios once dynamically hedged from the Market effect. Indeed, most of the estimated γ_{indus} are far from 0, which seems to



(a) *Conditional betas between the Market portfolio and the Agriculture industry portfolio*



(b) *Conditional betas between the Market portfolio and the Paper industry portfolio*

Figure 3.1: *Conditional betas using the realized betas (in dark blue) or the realized semibetas (in light blue) as exogenous variables. Constant conditional betas are represented by the red dashed lines*

validate the use of an ARCH(∞) equation.

3.6 Conclusion

Models accounting for the time variation of slope coefficients in linear regression have recently attracted substantial attention in the econometric literature. However, most of this literature relies on short memory models for the conditional volatilities in the system. Such hypothesis may prove inadequate on empirical applications. In this paper, we propose a novel multivari-

Industry	W_n	Industry	W_n	Industry	W_n
Agric	0.000	Print	0.005	Manuf	0.000
Mines	0.000	Chems	0.082	Trans	0.000
Oil	0.000	Ptrlm	0.000	Phone	0.000
Stone	0.009	Rubbr	0.000	TV	0.091
Cnstr	0.000	Lethr	0.000	Utils	0.000
Food	0.000	Glass	0.000	Garbg	0.000
Smoke	0.000	Metal	0.000	Whsl	0.002
Txtls	0.000	MtlPr	0.030	Rtail	0.000
Apprl	0.000	Machn	0.000	Money	0.000
Wood	0.000	Elctr	0.000	Srvc	0.000
Chair	0.000	Cars	0.979	Other	0.054
Paper	0.958	Instr	0.977		

Table 3.3: p -values of the Wald test on the importance of the semibetas for each industry portfolios

	γ_{Mkt}	γ_{indus}		γ_{Mkt}	γ_{indus}		γ_{Mkt}	γ_{indus}
Agric	0.000	0.122	Print	0.000	0.078	Manuf	0.000	0.001
Mines	0.000	0.050	Chems	0.000	0.143	Trans	0.000	0.091
Oil	0.000	0.065	Ptrlm	0.000	0.038	Phone	0.000	0.102
Stone	0.000	0.108	Rubbr	0.000	0.042	TV	0.000	0.075
Cnstr	0.000	0.053	Lethr	0.000	0.070	Utils	0.000	0.113
Food	0.000	0.105	Glass	0.000	0.086	Garbg	0.000	0.218
Smoke	0.000	0.166	Metal	0.000	0.034	Whsl	0.000	0.115
Txtls	0.000	0.198	MtlPr	0.000	0.154	Rtail	0.000	0.029
Apprl	0.000	0.080	Machn	0.000	0.080	Money	0.000	0.164
Wood	0.000	0.101	Elctr	0.000	0.103	Srvc	0.000	0.121
Chair	0.000	0.119	Cars	0.000	0.121	Other	0.000	0.067
Paper	0.000	0.123	Instr	0.000	0.078			

Table 3.4: γ_{Mkt} and γ_{indus} parameters for each industry portfolios

ate ARCH(∞) model with dynamic betas where exogenous variables can be introduced in the dynamics of both the conditional covariances and the dynamic slope coefficients. In that sense, we provide two major extensions to the Cholesky GARCH model of Darolles et al. (2018). We prove the existence of a stationary solution which contributes to the scarce literature on multivariate ARCH(∞) models. Additionally, we establish statistical inference results and prove the consistency and asymptotic normality of Quasi Maximum Likelihood estimator. Finite sample properties of the QMLE are assessed through Monte Carlo experiments, while an application on various industry portfolios investigates the role of the recently introduced semibetas of Bollerslev et al. (2022) in the dynamic of conditional betas. This application provides arguments in favor of these new realized measures while underlying the need to consider more persistent volatility processes than standard multivariate GARCH models allow. Future work could be devoted to capture the effects of a global equity index and a country-specific equity index on individual stocks. This *global*-CAPM model has been the subject of a vast financial literature on emerging markets (see for example, Bekaert and Harvey (1995), Bekaert et al. (2005) or Chaieb et al. (2021)). Since emerging equity indices present higher persistence than standard GARCH models, as illustrated in Chapter 2, the use of a multivariate ARCH(∞) model may prove more adequate to this particular asset pricing exercise.

3.7 Appendix A: Proofs and technical results

Proof of Theorem 3.1. In view of (3.8), consider the random vector with (possibly infinite) nonnegative components,

$$\mathbf{S}_t = \boldsymbol{\omega}_t + \sum_{k=1}^{\infty} \sum_{i_1, \dots, i_k \geq 1} \mathbf{A}_{i_1} \boldsymbol{\Upsilon}_{t-i_1} \dots \mathbf{A}_{i_k} \boldsymbol{\Upsilon}_{t-i_1-\dots-i_k} \boldsymbol{\omega}_{t-i_1-\dots-i_k}. \quad (3.22)$$

For any matrices \mathbf{A} and \mathbf{B} with positive entries and conformable dimensions, the c_r -inequality entails $(\mathbf{A}\mathbf{B})^{\odot r} \leq \mathbf{A}^{\odot r} \mathbf{B}^{\odot r}$ and $(\mathbf{A} + \mathbf{B})^{\odot r} \leq \mathbf{A}^{\odot r} + \mathbf{B}^{\odot r}$ for $r \in (0, 1]$. We thus have,

$$E\boldsymbol{\omega}_t^{\odot r} = \boldsymbol{\omega}^{\odot r} + \sum_{i=1}^{\infty} \Pi_i^{\odot r} E\mathbf{X}_t^{\odot r} < \infty,$$

componentwise, and for \mathbf{S}_t satisfying (3.22),

$$\begin{aligned} E\mathbf{S}_t^{\odot r} &\leq E\boldsymbol{\omega}_t^{\odot r} + \sum_{k=1}^{\infty} \sum_{i_1, \dots, i_k \geq 1} \mathbf{A}_{i_1}^{\odot r} E\boldsymbol{\Upsilon}_{t-i_1}^{\odot r} \dots \mathbf{A}_{i_k}^{\odot r} E\boldsymbol{\Upsilon}_{t-i_1-\dots-i_k}^{\odot r} E\boldsymbol{\omega}_t^{\odot r} \\ &= \left[\mathbf{I}_m + \sum_{k=1}^{\infty} (\mathcal{A}^{\odot r} E\boldsymbol{\Upsilon}_t^{\odot r})^k \right] E\boldsymbol{\omega}_t^{\odot r} < \infty, \end{aligned}$$

componentwise. It follows that \mathbf{S}_t is finite a.s. The process (\mathbf{S}_t) is thus strictly stationary and nonanticipative as a function of the past values of $\boldsymbol{\eta}_t$. In addition, it is clear that (\mathbf{S}_t) satisfies the stochastic recurrence equation

$$\mathbf{S}_t = \boldsymbol{\omega}_t + \sum_{i=1}^{\infty} \mathbf{A}_i \boldsymbol{\Upsilon}_{t-i} \mathbf{S}_{t-i}. \quad (3.23)$$

Now we turn to uniqueness. Denote by $(\boldsymbol{\varepsilon}_t^*)$ any strictly stationary and nonanticipative solutions of the model. For all $N \geq 1$, we obtain

$$\begin{aligned} \mathbf{g}_t^* - \mathbf{S}_t &= \left\{ \boldsymbol{\omega}_t + \sum_{k=1}^N \sum_{i_1, \dots, i_k \geq 1} \mathbf{A}_{i_1} \boldsymbol{\Upsilon}_{t-i_1} \dots \mathbf{A}_{i_k} \boldsymbol{\Upsilon}_{t-i_1-\dots-i_k} \boldsymbol{\omega}_{t-i_1-\dots-i_k} - \mathbf{S}_t \right\} \\ &\quad + \sum_{i_1, \dots, i_{N+1} \geq 1} \mathbf{A}_{i_1} \boldsymbol{\Upsilon}_{t-i_1} \dots \mathbf{A}_{i_{N+1}} \boldsymbol{\Upsilon}_{t-i_1-\dots-i_{N+1}} \mathbf{g}_{t-i_1-\dots-i_{N+1}}^* \\ &:= \{\mathbf{g}_{t,N}^* - \mathbf{S}_t\} + \mathbf{R}_{t,N}. \end{aligned}$$

We have

$$\mathbf{S}_t - \mathbf{g}_{t,N}^* = \sum_{k=N+1}^{\infty} \sum_{i_1, \dots, i_k \geq 1} \mathbf{A}_{i_1} \boldsymbol{\Upsilon}_{t-i_1} \dots \mathbf{A}_{i_k} \boldsymbol{\Upsilon}_{t-i_1-\dots-i_k} \boldsymbol{\omega}_{t-i_1-\dots-i_k},$$

which has nonnegative terms. Thus, by arguments already given,

$$E(\mathbf{S}_t - \mathbf{g}_{t,N}^*)^{\odot r} \leq \sum_{k=N+1}^{\infty} (\mathcal{A}^{\odot r} E\boldsymbol{\Upsilon}_t^{\odot r})^k \rightarrow 0, \quad \text{as } N \rightarrow \infty.$$

Similarly, the entries of $\mathbf{R}_{t,N}$ are nonnegative and, using the fact that the process \mathbf{g}_t^* is nonanticipative and satisfies $E\{(\mathbf{g}_t^*)^{\odot r}\} < \infty$, we have

$$E(\mathbf{R}_{t,N})^{\odot r} \leq (\mathcal{A}^{\odot r} E \Upsilon_t^{\odot r})^{N+1} E\{(\mathbf{g}_t^*)^{\odot r}\} \rightarrow 0, \quad \text{as } N \rightarrow \infty.$$

We thus have $\mathbf{g}_t^* = \mathbf{S}_t$ a.s. The proof of Theorem 3.1 is now complete. \square

We now turn to the proof of Proposition 3.1.

Proof of Proposition 3.1. In view of (3.7), we have

$$\boldsymbol{\ell}_t = \sum_{k=0}^{\infty} \mathbf{D}_k \mathbf{c}_{t-k},$$

where $\mathbf{c}_t = \mathbf{c}(\mathbf{v}_{t-1}, \dots, \mathbf{v}_{t-q}, \mathbf{g}_{t-1}^{1/2}, \dots, \mathbf{g}_{t-q}^{1/2}) := \mathbf{c}(\mathbf{x}_{t-1})$, and (\mathbf{D}_k) is the sequence of matrices obtained in the inversion of the matrix polynomial $\mathbf{C}(z)$. It follows that, for $r \leq 1/2$

$$E\|\boldsymbol{\ell}_t\|^{2r} = \sum_{k=0}^{\infty} \|\mathbf{D}_k\|^{2r} E\|\mathbf{c}_t\|^{2r} \leq K(E\|\mathbf{x}_t\|^{2r} + 1) < \infty,$$

and thus $E\|\mathbf{L}_t\|^{2r} < \infty$. By Hölder's inequality, we deduce

$$E\|\boldsymbol{\varepsilon}_t\|^r \leq \{E\|\mathbf{L}_t\|^{2r}\}^{1/2} \{E\|\mathbf{G}_t^{1/2}\|^{2r}\}^{1/2} E\|\boldsymbol{\eta}_t\|^r < \infty$$

which concludes the proof. \square

Before developing the proof of Theorem 3.2, we first state the following elementary lemma.

Lemma 3.1. *Let (u_t) , (x_t) and (w_t) be three sequences of random variables such that u_t is independent of $x_t w_t$, and for $\xi > 0$, $\rho \in (0, 1]$ and $a > 1 - \rho/2$, we have $\sup_t E|u_t|^\rho < \infty$, $\sup_t \| |x_t|^\rho \|_{\frac{1+\xi}{\xi}} < \infty$ and $\| |w_t|^\rho \|_{1+\xi} \leq K/t^a$. Then*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n u_t x_t w_t \rightarrow 0 \text{ in probability as } n \rightarrow \infty.$$

Proof of Lemma 3.1. The result comes from the Markov and Hölder inequalities:

$$E \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n u_t x_t w_t \right|^\rho \leq \frac{K}{n^{\rho/2}} \sum_{t=1}^n \| |x_t|^\rho \|_{\frac{1+\xi}{\xi}} \| |w_t|^\rho \|_{1+\xi} \leq K n^{-\rho/2} (n^{-a+1} + 1) \rightarrow 0.$$

\square

Proof of Theorem 3.2. It suffices to modify a few steps of the proof of Theorem 4.1 in DFL. Since we are considering volatility models with a larger memory than standard GARCH with finite orders, the main difficulty is to show that the initial values are asymptotically unimportant.

More precisely, we have to show that

$$\lim_{n \rightarrow \infty} \sup_{\boldsymbol{\theta} \in \Theta} |O_n(\boldsymbol{\theta}) - \tilde{O}_n(\boldsymbol{\theta})| = 0 \quad \text{a.s.} \quad (3.24)$$

where

$$O_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{t=1}^n q_t(\boldsymbol{\theta}), \quad q_t(\boldsymbol{\theta}) = \boldsymbol{\varepsilon}'_t \mathbf{B}'_t \mathbf{G}_t^{-1} \mathbf{B}_t \boldsymbol{\varepsilon}_t + \sum_{i=1}^m \log g_{it}.$$

Using **A1**–**A3**, DFL showed that one can replace $\tilde{\mathbf{B}}_t$ by \mathbf{B}_t in $\tilde{q}_t(\boldsymbol{\theta})$ to establish (3.24). It thus remains to prove that

$$a_n := \frac{1}{n} \sum_{t=1}^n \sup_{\boldsymbol{\theta} \in \Theta} \left| \boldsymbol{\varepsilon}'_t \mathbf{B}'_t \mathbf{G}_t^{-1} (\tilde{\mathbf{G}}_t - \mathbf{G}_t) \tilde{\mathbf{G}}_t^{-1} \mathbf{B}_t \boldsymbol{\varepsilon}_t \right| \rightarrow 0 \quad \text{a.s.} \quad (3.25)$$

Let $\varepsilon > 0$. By **A2***, almost surely there exists N such that $\sup_{\boldsymbol{\theta} \in \Theta} \|\tilde{\mathbf{G}}_t - \mathbf{G}_t\| < \varepsilon$ for all $t > N$. We thus have almost surely

$$\begin{aligned} a_n &= \frac{1}{n} \sum_{t=1}^n \sup_{\boldsymbol{\theta} \in \Theta} \left| \text{Tr} \left\{ (\tilde{\mathbf{G}}_t - \mathbf{G}_t) \tilde{\mathbf{G}}_t^{-1} \mathbf{B}_t \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t \mathbf{B}'_t \mathbf{G}_t^{-1} \right\} \right| \\ &\leq \varepsilon \frac{K}{n} \sum_{t=1}^n \sup_{\boldsymbol{\theta} \in \Theta} \left\| \tilde{\mathbf{G}}_t^{-1} \mathbf{B}_t \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t \mathbf{B}'_t \mathbf{G}_t^{-1} \right\| + \frac{K}{n} \sum_{t=1}^N \sup_{\boldsymbol{\theta} \in \Theta} \|\tilde{\mathbf{g}}_t - \mathbf{g}_t\| \left\| \tilde{\mathbf{G}}_t^{-1} \mathbf{B}_t \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t \mathbf{B}'_t \mathbf{G}_t^{-1} \right\|. \end{aligned}$$

By **A1** and **A3***, we have $c := E \sup_{\boldsymbol{\theta} \in \Theta} \left\| \tilde{\mathbf{G}}_t^{-1} \mathbf{B}_t \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t \mathbf{B}'_t \mathbf{G}_t^{-1} \right\| < \infty$. By the ergodic theorem, it follows that $\lim_{n \rightarrow \infty} a_n \leq \varepsilon K c$. Since ε can be chosen arbitrarily small, we obtain (3.25). The consistency follows by the arguments given in DFL.

To prove asymptotic normality, the main difficulty is again to show that the initial values are asymptotically unimportant. More precisely we have to show that there exists a neighborhood $V(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ such that

$$\lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \sum_{t=1}^n \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\| \frac{\partial q_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} - \frac{\partial \tilde{q}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right\| = 0 \quad \text{a.s.} \quad (3.26)$$

Using standard matrix derivative computations, we have

$$\frac{\partial}{\partial \theta_i} q_t(\boldsymbol{\theta}) = \text{Tr} \left\{ (\mathbf{I}_m - \mathbf{G}_t^{-1} \mathbf{B}_t \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t \mathbf{B}'_t) \mathbf{G}_t^{-1} \frac{\partial \mathbf{G}_t}{\partial \theta_i} \right\} + 2 \text{Tr} \left\{ \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t \mathbf{B}'_t \mathbf{G}_t^{-1} \frac{\partial \mathbf{B}_t}{\partial \theta_i} \right\}.$$

We thus have

$$\frac{\partial}{\partial \theta_i} q_t(\boldsymbol{\theta}) - \frac{\partial}{\partial \theta_i} \tilde{q}_t(\boldsymbol{\theta}) = \sum_{j=1}^6 a_{jt}(\boldsymbol{\theta}), \quad a_{jt}(\boldsymbol{\theta}) = \text{Tr} \mathbf{A}_{jt},$$

with

$$\begin{aligned}
\mathbf{A}_{1t} &= \left(\tilde{\mathbf{G}}_t^{-1} - \mathbf{G}_t^{-1} \right) \mathbf{B}_t \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t \mathbf{B}'_t \mathbf{G}_t^{-1} \frac{\partial \mathbf{G}_t}{\partial \theta_i}, & \mathbf{A}_{2t} &= \tilde{\mathbf{G}}_t^{-1} \left(\tilde{\mathbf{B}}_t - \mathbf{B}_t \right) \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t \mathbf{B}'_t \mathbf{G}_t^{-1} \frac{\partial \mathbf{G}_t}{\partial \theta_i}, \\
\mathbf{A}_{3t} &= \tilde{\mathbf{G}}_t^{-1} \tilde{\mathbf{B}}_t \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t \left(\tilde{\mathbf{B}}'_t - \mathbf{B}'_t \right) \mathbf{G}_t^{-1} \frac{\partial \mathbf{G}_t}{\partial \theta_i}, & \mathbf{A}_{4t} &= 2 \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t \left(\mathbf{B}'_t - \tilde{\mathbf{B}}'_t \right) \mathbf{G}_t^{-1} \frac{\partial \mathbf{B}_t}{\partial \theta_i}, \\
\mathbf{A}_{5t} &= 2 \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t \tilde{\mathbf{B}}'_t \left(\mathbf{G}_t^{-1} - \tilde{\mathbf{G}}_t^{-1} \right) \frac{\partial \mathbf{B}_t}{\partial \theta_i}, & \mathbf{A}_{6t} &= \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t \tilde{\mathbf{B}}'_t \tilde{\mathbf{G}}_t^{-1} \left(\frac{\partial \mathbf{B}_t}{\partial \theta_i} - \frac{\partial \tilde{\mathbf{B}}_t}{\partial \theta_i} \right).
\end{aligned}$$

Using **A1–A3** and **A8**, the elementary inequalities $(\sum_i |a_i|)^s \leq \sum_i |a_i|^s$ for $s \in (0, 1]$ and $|\text{Tr}(\mathbf{A}\mathbf{B})| \leq K \|\mathbf{A}\| \|\mathbf{B}\|$ with obvious notations, together with the Hölder inequality, we note that

$$E \left(\sum_{t=1}^{\infty} \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} |a_{4t}(\boldsymbol{\theta})| \right)^s \leq KE \|\boldsymbol{\eta}_t \boldsymbol{\eta}'_t\|^s \sum_{t=1}^{\infty} E \|\boldsymbol{\ell}_t\|^{2s} \|\mathbf{g}_t\|^{2s} \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\| \frac{\partial \boldsymbol{\beta}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\|^s \rho_t^s$$

is finite when s is sufficiently small. Therefore $\sum_{t=1}^{\infty} \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} |a_{4t}(\boldsymbol{\theta})|$ is a.s. finite and we have shown

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} |a_{it}(\boldsymbol{\theta})| = o(1) \quad \text{a.s.} \quad (3.27)$$

for $i = 4$. Similarly, it can be seen that (3.27) holds for $i = 2, 3, 6$. The case $i = 1$ in (3.27) follows from **A4***, applying Lemma 3.1 with $u_t = \|\boldsymbol{\eta}_t \boldsymbol{\eta}'_t\|$, $w_t = \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\| \tilde{\mathbf{G}}_t^{-1} - \mathbf{G}_t^{-1} \right\|$, and $x_t = \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \|\mathbf{B}_t \mathbf{L}_{0t}\|^2 \left\| \mathbf{G}_t^{-1} \frac{\partial \mathbf{G}_t}{\partial \theta_i} \right\| \|\mathbf{G}_{0t}\|^2$. The case $i = 5$ follows from **A5***. We thus have shown (3.26). The rest of the proof follows that of Theorem 4.1 in DFL. \square

3.8 Appendix B: Finite sample properties of the Wald statistics when testing for linear constraints

In this appendix, we present additional Monte Carlo experiments to assess the finite sample properties of the Wald test statistics used in Section 3.5. To conduct our experiments, we repeat the simulations presented in section 3.4. In particular, we use the monthly realized beta between the Agricultural industry and the Market portfolio as an exogenous variable in the conditional beta dynamics. We simulate a thousand samples of size 5000 of the Cholesky-ARCH(∞)-X using the value for θ_0 displayed in Table 3.1. On each sample we compute the QMLE using the realized semibetas as exogenous variables, which represents the unconstrained specification in the Cholesky-ARCH(∞)-X model defined as follows

$$\begin{aligned} g_{M,t}^2 &= \omega_M + \sum_{k=1}^{\infty} [\alpha_M \beta_M^{k-1} + \gamma_M k^{-2}] \varepsilon_{M,t-k}^2 \\ g_{i,t}^2 &= \omega_i + \sum_{k=1}^{\infty} [\alpha_i \beta_i^{k-1} + \gamma_i k^{-2}] v_{i,t-k}^2 + \delta_i \varepsilon_{M,t-1}^2 \\ \beta_{i,M,t} &= \varpi_{i,M} + \pi_{i,M}^{\mathcal{P}} \bar{\beta}_{i,M,t}^{\mathcal{P}} + \pi_{i,M}^{\mathcal{N}} \bar{\beta}_{i,M,t}^{\mathcal{N}} + \pi_{i,M}^{S^+} \bar{\beta}_{i,M,t}^{S^+} + \pi_{i,M}^{S^-} \bar{\beta}_{i,M,t}^{S^-} + c_{i,M} \beta_{i,M,t} \end{aligned}$$

where $\bar{\beta}_{i,M,t}^{\bullet}$ denotes the four realized semibetas between the Agricultural industry and the Market portfolio. We then compute the Wald test statistics for

$$H_0 : \pi_{ij}^{\mathcal{P}} = \pi_{ij}^{\mathcal{N}} \quad \text{and} \quad \pi_{ij}^{\mathcal{P}} = -\pi_{ij}^{S^+} \quad \text{and} \quad \pi_{ij}^{\mathcal{N}} = -\pi_{ij}^{S^-}.$$

Figure 3.2 shows the empirical distribution of the statistics against its theoretical χ_3^2 asymptotic distribution and illustrates the good behavior of the statistics under the null.

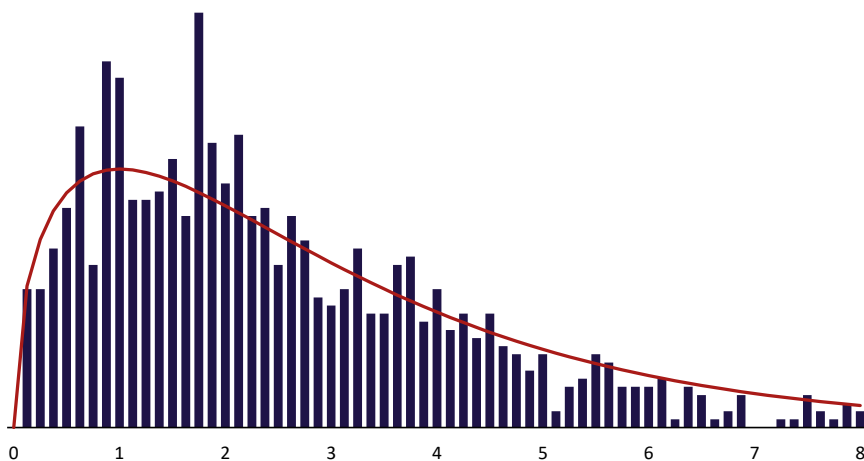


Figure 3.2: *Wald statistics empirical distribution (blue) and theoretical asymptotic distribution (red line)*

Chapter 4

Empirical Asset Pricing with Score-Driven Conditional Betas

Thomas Giroux

Julien Royer

Abstract: This paper introduces a novel empirical asset pricing framework based on the newly introduced score-driven conditional betas model. We extend the theory of the studied conditional betas by establishing the asymptotic distribution of standard tests statistics for parameter constancy in the conditional regression. In particular, these tests allow to assess the significance of a given factor in the regression. Additionally, we introduce a residual bootstrap procedure for the Wald statistic and establish its validity. We then propose a two-step estimation procedure to recover time-varying factor risk premia from individual stock returns. We illustrate the performance of our tests and risk premia estimation procedure on simulations. Finally, we present an application in which we assess the existence of a time-varying risk premium associated to a Carbon Risk factor in the cross-section of US industry portfolios.

Keywords: Asset Pricing Models, Dynamic Factor Models, Score-Driven Models, Carbon risk

4.1 Introduction

Conditional linear factor models have long been used in the financial literature to decompose the performance of assets into systematic risk factors exposures. For example, amongst popular linear factor models, the CAPM (Sharpe (1964), Lintner (1965)) assumes a single factor model using the excess market returns, while Fama and French (1993) consider two additional factors capturing size and value effects. A vast majority of the asset pricing literature has however considered time invariant estimates of the parameters in the conditional regression underlying the model. In particular, the widely used two-pass regression method of Fama and MacBeth (1973) assumes constancy of the betas and risk premia, defined as the compensation for being exposed to a systematic risk factor. However, wrongfully assuming coefficient constancy in the conditional regression may lead to erroneous results (see for example Lewellen and Nagel (2006) or Grassi and Violante (2021)).

Practitioners usually obtain time-varying estimates of the regression parameters by computing OLS estimators on rolling windows. However, this method lies on inconsistent assumptions since the obtained estimator targets constant betas on the considered time frame. Engle (2016) proposed to model conditional betas as a transformation of the conditional covariance matrix of a multivariate GARCH model. Although this method provides a direct estimation of the betas after estimating the model (usually using a Dynamic Conditional Correlation GARCH), the variable of interest is not directly specified. Indeed, the conditional betas are obtained from a transformation of the conditional covariance matrix, which makes hypothesis testing on the dynamic betas impractical. In addition, the estimated time-varying coefficients are often highly volatile.

Directly modeling betas dynamics provides numerous advantages as it allows to gain economic and financial interpretability on the time-varying dependency between assets or factors and allows for testing procedures. Gagliardini et al. (2016) proposed to introduce dynamics into the regression coefficient through instrumental variables, following ideas from Ferson and Harvey (1991) among others. The obtained factor loadings are however highly sensitive to the choice of instrumental variables as noted by Ghysels (1998). Darolles et al. (2018) proposed a new multivariate GARCH model where the conditional betas can be obtained from a Cholesky decomposition of the conditional covariance matrix. However, it requires estimating the full multivariate system even if only one equation is of interest. To circumvent this issue, Grassi and Violante (2021) propose a Block-Cholesky GARCH model, based on Block-Cholesky decomposition between the set of factors and the set of individual assets. Although better suited to an empirical asset pricing exercise, the model imposes constancy of the risk premia, which can be restrictive.

In this paper, we propose to leverage the recently introduced Autoregressive Conditional Beta model (ACB) of Blasques et al. (2022a). The ACB model builds upon the score-driven models literature introduced by Harvey (2013) and Creal et al. (2013) to allow for both time-varying

regression coefficients and conditional heteroskedasticity of the factors. We extend the theoretical results by establishing the asymptotic distribution of standard test statistics for coefficients constancy. Moreover, a particular form of the constraint under the null hypothesis can be used to test the significance of a factor in the conditional regression. Additionally, in the spirit of Gagliardini et al. (2016), we introduce a two-step estimation procedure for the time-varying risk premia.

The remainder of the paper is organized as follows. Section 2 presents the ACB model of Blasques et al. (2022a) and introduces the test statistics. In particular, it presents their asymptotic distributions under the null. Section 3 develops our two-step estimation procedure for the time-varying risk premia. Monte Carlo experiments are conducted in Section 4. Section 5 provides an application to assess the risk premium associated with a carbon-related long short portfolio. Section 6 concludes. Proofs and technical results are relegated to an appendix.

4.2 Testing in the Score-Driven Betas framework

4.2.1 A Score-Driven Conditional Regression model

Asset pricing models are built on linear regressions where the excess returns of a financial asset are explained by a set of factors. Let us denote $(r_{i,t})$ the excess returns time series of asset $i = 1, \dots, N$ in a system of N assets and $\mathbf{f}_t = (f_{1,t}, \dots, f_{m,t})'$ a set of m observable factors. We consider a conditional regression model of the form

$$r_{i,t} = \alpha_{i,t} + \beta_{i,t}' \mathbf{f}_t + \varepsilon_{i,t} \quad (4.1)$$

where each factor f_j follows a general conditional volatility model of the form

$$f_{j,t} = \mu_j + \sigma_{j,t} \eta_{j,t}$$

with $\sigma_{j,t}$ a measurable function with respect to the natural filtration and $\eta_{j,t}$ a sequence of iid centered random variable with unit variance.

To derive updating equations for the conditional regression parameters, Blasques et al. (2022a) build upon score-driven models. Such models, also known as Generalized Autoregressive Score models (GAS) were introduced by Harvey (2013) and Creal et al. (2013). They aim to generalize models with time-varying parameters where the updating is *observation-driven*. In particular, consider an observable process y_t and assume it follows the conditional density $p(y_t | \psi_t, \Omega_t, \theta)$ where ψ_t is the time varying parameter of interest, θ a constant parameter, and Ω_t the information set available at time t . A score-driven model features an updating equation of the form

$$\psi_{t+1} = \bar{\omega} + \xi \underbrace{S(\psi_t) \frac{\partial \log p(y_t | \psi_t, \Omega_t, \theta)}{\partial \psi_t}}_{\text{updating term}} + c \psi_t \quad (4.2)$$

where $\bar{\omega}$, ξ and c are unknown parameters to be estimated. The updating term is decomposed into a scaling term $S(\psi_t)$, that is usually related to the inverse of the information matrix, and the score of the likelihood with respect to the time-varying parameter. The dynamics of the parameter of interest is thus driven by the scaled score. Interestingly, numerous econometric models can be expressed as a score-driven model. For example, if we assume that the observation conditional density is Gaussian and the parameter of interest is the time-varying variance, under an appropriate choice of the scaling function, Equation (4.2) yields the standard GARCH(1,1) model.

Assuming, for all i , that the residuals in the conditional regression (4.1) follow a conditional volatility model $\varepsilon_{i,t} = g_{i,t}\zeta_{i,t}$ where $\zeta_{i,t}$ are iid Gaussian, Blasques et al. (2022a) derive the updating equations for the conditional regression coefficients

$$\begin{aligned}\alpha_{i,t+1} &= \bar{\omega}_{\alpha_i} + \xi_{\alpha_i}\varepsilon_{i,t} + c_{\alpha_i}\alpha_{i,t} \\ \beta_{i,j,t+1} &= \bar{\omega}_{i,j} + \xi_{i,j}\frac{f_{j,t}\varepsilon_{i,t}}{\mu_j^2 + \sigma_{j,t}^2} + c_{i,j}\beta_{i,j,t}\end{aligned}\quad (4.3)$$

for $i = 1, \dots, N$ and $j = 1, \dots, m$.

Note that, although derived from technical assumptions, the updating term $f_{j,t}\varepsilon_{i,t}/(\mu_j^2 + \sigma_{j,t}^2)$ is very intuitive. Indeed, if $\xi_{i,j}$ and $c_{i,j}$ are positive, the term $f_{j,t}\varepsilon_{i,t}$ implies that the update attempts to obtain beta values for which the residuals $\varepsilon_{i,t}$ are not only unconditionally orthogonal to the factor f_j as usual in linear regressions, but also conditionally orthogonal. Additionally, as the updating term is inversely proportional to the conditional volatility $\sigma_{j,t}^2$, the updating step size is less important in periods of high volatility, ensuring less noisy conditional betas. The updating equation of the intercept $\alpha_{i,t}$ is similar to the betas equation, except for the fact that since the related factor is constantly equal to one, its variance is null, thus the denominator of the updating term is equal to one.

Additionally, following Blasques et al. (2022a), we assume that both the conditional variances of the observable factors and the conditional variance of the residuals follow a GARCH(1,1) equation

$$\begin{aligned}\sigma_{j,t}^2 &= \omega_j + a_j(f_{j,t-1} - \mu_j)^2 + b_j\sigma_{j,t-1}^2 \\ g_{i,t}^2 &= \omega_{\varepsilon_i} + a_{\varepsilon_i}(r_{i,t-1} - \alpha_{i,t-1} - \beta_{i,t-1}'\mathbf{f}_{t-1})^2 + b_{\varepsilon_i}g_{i,t-1}^2.\end{aligned}\quad (4.4)$$

While this assumption is coherent with the distribution of $\zeta_{i,t}$ used to derive the dynamic of the conditional betas, score-driven models have arguably built their success on the use of different distributions for the observable process. Departing from normality and assuming a conditional volatility model $y_t = s_t\zeta_t$ where ζ_t are iid innovations following a standardized Student's t distribution with $\nu > 2$ degree of freedom, Harvey (2013) shows that the score-driven conditional variance equation is given by

$$s_t^2 = \omega + \alpha\frac{1 + \iota}{1 - 2\iota + \iota\zeta_t^2}y_t^2 + \beta s_t^2\quad (4.5)$$

where $\iota = 1/\nu$. Interestingly, (4.5) yields the GARCH(1,1) equation when $\iota = 0$. The popularity of this model, known as the Beta- t -GARCH, is due to the effect of the parameter ν on the conditional variance. Indeed, the parameter both controls the shape of the innovations density and limits the effect of large values of ζ_t on future values of s_t . While it is not considered in this manuscript, one could leverage the recently introduced quasi score-driven models of Blasques et al. (2022b) to consider both a Beta- t -GARCH equation for the volatility and derive the conditional betas updating equation using a Gaussian log-likelihood. Alternatively, deriving the dynamics of the regression coefficients in a pure score-driven model without assuming a Gaussian distribution could yield promising extensions that are beyond the scope of this chapter.

Let us now introduce the parametric form of the model. We denote for any $j \in \{1, \dots, m\}$ the true parameter $\theta_0^{(j)} = (\mu_{0j}, \omega_{0j}, a_{0j}, b_{0j})'$ driving the individual GARCH(1,1) equations of each factor f_j as defined in (4.4). The first step of the estimation consists in estimating independently $\theta_0^{(j)}$ by standard Quasi Maximum Likelihood (QML)

$$\hat{\theta}_n^{(j)} = \arg \min_{\theta \in \Theta} \tilde{O}_n^{(j)}(\theta), \quad \tilde{O}_n^{(j)}(\theta) = \frac{1}{n} \sum_{t=2}^n \tilde{l}_{jt}(\theta) \quad (4.6)$$

where θ is a generic element of the parameter space Θ such that $\theta_0^{(j)} \in \Theta$ for all $j \in \{1, \dots, m\}$ and

$$\tilde{l}_{jt}(\theta) = \frac{(f_{j,t} - \mu)^2}{\tilde{\sigma}_{j,t}^2(\theta)} + \log \tilde{\sigma}_{j,t}^2(\theta), \quad \tilde{\sigma}_{j,t}^2(\theta) = \omega + a(f_{j,t-1} - \mu)^2 + b\tilde{\sigma}_{j,t-1}^2(\theta)$$

with a given initial value $\tilde{\sigma}_{j,1}^2(\theta) = \tilde{g} > 0$. Let us denote $\hat{\theta}_n = (\hat{\theta}_n^{(1)'}, \dots, \hat{\theta}_n^{(m)'})'$.

In the second step, we estimate the parameter driving the GARCH(1,1) equation (4.4) of the residuals $\varepsilon_{i,t}$ denoted $\vartheta_0^{(\varepsilon_i)} = (\omega_{0\varepsilon_i}, a_{0\varepsilon_i}, b_{0\varepsilon_i})'$, the parameter $\vartheta_0^{(i,0)} = (\bar{\omega}_{0\alpha_i}, \xi_{0\alpha_i}, c_{0\alpha_i})'$ driving the dynamic of the time-varying intercept, and $\vartheta_0^{(i,j)} = (\bar{\omega}_{0,i,j}, \xi_{0,i,j}, c_{0,i,j})'$ the parameter driving the dynamic of the time-varying $\beta_{i,j}$, for $j \in \{1, \dots, m\}$. Let $\vartheta_0^{(i)} = (\vartheta_0^{(\varepsilon_i)'}, \vartheta_0^{(i,0)'}, \vartheta_0^{(i,1)'}, \dots, \vartheta_0^{(i,m)'})'$, $\theta_0 = (\theta_0^{(1)'}, \dots, \theta_0^{(m)'})'$ and the full parameter $\varphi_0^{(i)} = (\theta_0', \vartheta_0^{(i)'})'$. Let ϑ a generic element of Θ_ϑ such that $\vartheta_0^{(i)} \in \Theta_\vartheta$ for all $i \in \{1, \dots, N\}$ and let $\varphi = (\theta', \vartheta')'$. We estimate $\vartheta_0^{(i)}$ by

$$\hat{\vartheta}_n^{(i)} = \arg \min_{\vartheta \in \Theta_\vartheta} \tilde{O}_n(\hat{\theta}_n, \vartheta), \quad \tilde{O}_n^{(i)}(\varphi) = \frac{1}{n} \sum_{t=2}^n \tilde{l}_t^{(i)}(\varphi) \quad (4.7)$$

where

$$\begin{aligned} \tilde{l}_t^{(i)}(\varphi) &= \frac{\tilde{\varepsilon}_{i,t}^2(\varphi)}{\tilde{g}_{i,t}^2(\varphi)} + \log \tilde{g}_{i,t}^2(\varphi), \quad \tilde{\varepsilon}_{i,t}(\varphi) = r_{i,t} - \tilde{\alpha}_{i,t}(\varphi) - \sum_{j=1}^m \tilde{\beta}_{i,j,t}(\varphi) f_{j,t}, \\ \tilde{g}_{i,t}^2(\varphi) &= \omega + \alpha \tilde{\varepsilon}_{i,t-1}^2(\varphi) + \beta \tilde{g}_{i,t-1}^2(\varphi), \quad \tilde{\alpha}_{i,t}(\varphi) = \bar{\omega}_{\alpha_i} + \xi_{\alpha_i} \tilde{\varepsilon}_{i,t-1}(\varphi) + c_{\alpha_i} \tilde{\alpha}_{i,t-1}(\varphi) \\ \tilde{\beta}_{i,j,t}(\varphi) &= \bar{\omega}_{i,j} + \xi_{i,j} \frac{f_{j,t-1} \tilde{\varepsilon}_{i,t-1}(\varphi)}{\mu_j^2 + \tilde{\sigma}_{j,t-1}^2(\theta)} + c_{i,j} \tilde{\beta}_{i,j,t-1}(\varphi). \end{aligned}$$

The multistep QML estimator is thus given by $\hat{\varphi}_n^{(i)} = (\hat{\theta}_n', \hat{\vartheta}_n^{(i)'})'$.

Under a set of technical assumptions that we denote **[A-ACB]**, Blasques et al. (2022a) establish consistency and asymptotic normality of this estimator

$$\widehat{\boldsymbol{\theta}}_n \rightarrow \boldsymbol{\theta}_0 \text{ a.s.} \quad \text{and} \quad \widehat{\boldsymbol{\vartheta}}_n^{(i)} \rightarrow \boldsymbol{\vartheta}_0^{(i)} \text{ a.s. as } n \rightarrow \infty.$$

In addition,

$$\sqrt{n}(\widehat{\boldsymbol{\varphi}}_n^{(i)} - \boldsymbol{\varphi}_0^{(i)}) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_i), \quad \boldsymbol{\Omega}_i = \begin{pmatrix} \boldsymbol{\Omega}_{\boldsymbol{\theta},i} & \boldsymbol{\Omega}_{\boldsymbol{\theta}\boldsymbol{\vartheta},i} \\ \boldsymbol{\Omega}_{\boldsymbol{\vartheta}\boldsymbol{\theta},i} & \boldsymbol{\Omega}_{\boldsymbol{\vartheta},i} \end{pmatrix} \quad (4.8)$$

where $\boldsymbol{\Omega}_i$ is an invertible matrix.

4.2.2 Asymptotics of standard test statistics

Although the assumption of beta constancy is rarely backed by an economic rationale, linear regressions with constant slope parameters remain predominant in the financial literature. It is therefore useful to introduce test procedures to verify this assumption. Moreover, our general testing setup easily allows to assess factors relevancy. In the case of an unconditional regression model, a factor is said to be significant if the beta associated to this factor is significantly different from zero. We thus propose to assess a factor j significance by testing that its conditional beta is both constant and equal to zero.

Assume that we want to test the hypothesis $H_{0,ij} : \beta_{i,j,t} = \bar{\omega}_{0,ij}$ against $H_{1,ij} : \beta_{i,j,t}$ is time-varying. From (4.3), constancy of $\beta_{i,j,t}$ is obtained when $\xi_{0ij} = 0$. However, testing for this constraint on $\boldsymbol{\vartheta}_0^{(i)}$ is difficult. Indeed, under $H_{0,ij} : \xi_{0ij} = 0$, the conditional beta $\beta_{i,j,t}$ tends to $\bar{\omega}_{ij}/(1 - c_{ij})$ and there exists an infinity of pairs $(\bar{\omega}_{ij}, c_{ij})$ such that $\beta_{i,j,t} = \bar{\omega}_{0,ij}$, resulting in a non identifiable model under the null hypothesis. A similar problem occurs when testing for the significance of a factor in the conditional regression, which is equivalent to $\xi_{0ij} = 0$ and $\bar{\omega}_{i,j} = 0$. Situations where parameters are not identified under the null induce a non standard asymptotic distribution for the QML estimator and modify the behavior of classical tests statistics. Hansen (1996) in particular presents inference results while Andrews (2001) provides a testing procedure when a parameter only appears under the alternative hypothesis. To simplify the testing procedures, we rather set the parameter c_{0ij} at an arbitrary value \bar{c}_{ij} . This solves the identification issue and we can now test $H_{0,ij}(\bar{c}_{ij}) : \mathbf{R}\boldsymbol{\vartheta}_0^{(i)} = 0$ using standard techniques where \mathbf{R} is the constraint matrix¹. For example, to test the hypothesis of constancy, $H_{0,ij} : \beta_{i,j,t} = \bar{\omega}_{0,ij}$, the constraint matrix is such that $\mathbf{R}\boldsymbol{\vartheta}_0^{(i)} = \xi_{0ij}$. Alternatively, for the hypothesis of insignificance of factor f_j , $H_{0,ij} : \beta_{i,j,t} = 0$, the constraint matrix is such that $\mathbf{R}\boldsymbol{\vartheta}_0^{(i)} = (\bar{\omega}_{0ij}, \xi_{0ij})' = \mathbf{0}_2$. The triptych of the Wald, Rao-score, and Quasi Likelihood Ratio

¹For the ease of notation, we keep on denoting $\boldsymbol{\vartheta}_0^{(i)}$ the vector of parameters to estimate such that $\boldsymbol{\vartheta}_0^{(i)} = (\boldsymbol{\vartheta}_0^{(\varepsilon_i)'} , \boldsymbol{\vartheta}_0^{(i,0)'}, \boldsymbol{\vartheta}_0^{(i,1)'}, \dots, \boldsymbol{\vartheta}_0^{(i,m)'})'$ where $\boldsymbol{\vartheta}_0^{(i,m)}$ now only contains two parameters ω_{0ij} and ξ_{0ij} . Similar notation malapropisms are extended to the related estimators.

(LR) statistics is given by

$$\begin{aligned}
W_n &= n(\mathbf{R} \hat{\boldsymbol{\vartheta}}_n^{(i)})' (\mathbf{R} \hat{\boldsymbol{\Omega}}_{\boldsymbol{\vartheta}, i, n} \mathbf{R}')^{-1} (\mathbf{R} \hat{\boldsymbol{\vartheta}}_n^{(i)}) \\
R_n &= n \frac{\partial \tilde{O}_n(\hat{\boldsymbol{\varphi}}_{n|H_0}^{(i)})}{\partial \boldsymbol{\vartheta}'} \hat{\mathbf{J}}_{\boldsymbol{\vartheta}, i, n|H_0}^{-1} \mathbf{R}' (\mathbf{R} \hat{\boldsymbol{\Omega}}_{\boldsymbol{\vartheta}, i, n|H_0} \mathbf{R}')^{-1} \mathbf{R} \hat{\mathbf{J}}_{\boldsymbol{\vartheta}, i, n|H_0}^{-1} \frac{\partial \tilde{O}_n(\hat{\boldsymbol{\varphi}}_{n|H_0}^{(i)})}{\partial \boldsymbol{\vartheta}} \\
L_n &= 2n \left[\tilde{O}_n(\hat{\boldsymbol{\varphi}}_{n|H_0}^{(i)}) - \tilde{O}_n(\hat{\boldsymbol{\varphi}}_n^{(i)}) \right]
\end{aligned} \tag{4.9}$$

where $\hat{\boldsymbol{\varphi}}_{n|H_0}^{(i)}$ is the multistep QMLE restricted by $H_{0,ij}(\bar{c}_{ij})$, $\hat{\boldsymbol{\Omega}}_{\boldsymbol{\vartheta}, i, n}$ is a consistent estimator of $\boldsymbol{\Omega}_{\boldsymbol{\vartheta}, i}$, $\hat{\boldsymbol{\Omega}}_{\boldsymbol{\vartheta}, i, n|H_0}$ is a consistent estimator of $\boldsymbol{\Omega}_{\boldsymbol{\vartheta}, i}$ under $H_{0,ij}(\bar{c}_{ij})$, and

$$\hat{\mathbf{J}}_{i, n|H_0} = \frac{\partial^2 \tilde{O}_n(\hat{\boldsymbol{\varphi}}_{n|H_0}^{(i)})}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\varphi}'} = \begin{pmatrix} \hat{\mathbf{J}}_{\boldsymbol{\vartheta}, i, n|H_0} & \hat{\mathbf{J}}_{\boldsymbol{\vartheta} \boldsymbol{\vartheta}, i, n|H_0} \\ \hat{\mathbf{J}}_{\boldsymbol{\vartheta} \boldsymbol{\vartheta}, i, n|H_0} & \hat{\mathbf{J}}_{\boldsymbol{\vartheta}, i, n|H_0} \end{pmatrix}$$

is a consistent estimators of

$$\mathbf{J}_i = \mathbb{E} \frac{\partial^2 l_t^{(i)}(\boldsymbol{\varphi}_0^{(i)})}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\varphi}'} = \begin{pmatrix} \mathbf{J}_{\boldsymbol{\vartheta}, i} & \mathbf{J}_{\boldsymbol{\vartheta} \boldsymbol{\vartheta}, i} \\ \mathbf{J}_{\boldsymbol{\vartheta} \boldsymbol{\vartheta}, i} & \mathbf{J}_{\boldsymbol{\vartheta}, i} \end{pmatrix}$$

with obvious matrix partitions.

Theorem 4.1. *Let r the rank of the constraint matrix \mathbf{R} . Under the set of assumptions [A-ACB], under $H_{0,ij}(\bar{c}_{ij})$, W_n and R_n test statistics follow a χ_r^2 distribution and the critical regions at the asymptotic level ν are given by*

$$\{W_n > \chi_r^2(1-\nu)\}, \{R_n > \chi_r^2(1-\nu)\}.$$

where $\chi_r^2(1-\nu)$ is the $(1-\nu)$ -quantile of the χ^2 distribution with r degree of freedom. In addition, under the same assumptions,

$$L_n \rightarrow \bar{\chi}(\boldsymbol{\pi}) \text{ with } \bar{\chi}(\boldsymbol{\pi}) = \sum_{j=1}^{3(m+1)} \pi_j \chi_j^2$$

where $\boldsymbol{\pi} = (\pi_j)_{j=1, \dots, 3(m+1)}$ is the vector of eigenvalues of the matrix $\mathbf{J}_{\boldsymbol{\vartheta}, i}^{-1} \mathbf{R}' \boldsymbol{\Lambda}_i \mathbf{R}$ with

$$\boldsymbol{\Lambda}_i = \left[\mathbf{R} \mathbf{J}_{\boldsymbol{\vartheta}, i}^{-1} \mathbf{R}' \right]^{-1} \left[\mathbf{R} \boldsymbol{\Omega}_{\boldsymbol{\vartheta}, i} \mathbf{R}' \right] \left[\mathbf{R} \mathbf{J}_{\boldsymbol{\vartheta}, i}^{-1} \mathbf{R}' \right]^{-1'}.$$

We remark that the Quasi LR statistic has a different asymptotic distribution than the standard χ^2 distribution of the Wald and Rao-score statistics. The critical region at the asymptotic level ν for L_n can however be easily obtained following the method of Imhof (1961).

4.2.3 Residual Bootstrap approximation of the test statistics distributions

Although the theoretical asymptotic distributions of the test statistics have been derived, Monte Carlo experiments conducted in Section 4.4 show that their finite sample distributions may be distant from those. To remedy this issue, we propose to use a residual-based bootstrap procedure to approximate the statistics asymptotic distribution. Similar bootstrap schemes were developed in the case of univariate GARCH models by Beutner et al. (2018) and Cavaliere et al. (2022), while Francq et al. (2016) used a similar bootstrap procedure for multivariate volatility models. In the following, we present the algorithm to approximate the Wald statistic.

1. From a set of n observations $(r_{i,1}, \dots, r_{i,n})$, and a set of factors $(\mathbf{f}_1, \dots, \mathbf{f}_n)$, compute the constrained multistep QML estimator $\hat{\varphi}_{n|H_0}^{(i)} = (\hat{\boldsymbol{\theta}}_n', \hat{\boldsymbol{\vartheta}}_{n|H_0}^{(i)'})'$. Derive the standardized residuals $\tilde{\boldsymbol{\eta}}_{i,t} = (\tilde{\eta}_{1,t}, \dots, \tilde{\eta}_{m,t}, \tilde{\zeta}_{i,t})$ where for $j = 1, \dots, n$

$$\begin{aligned} \tilde{\eta}_{j,t} &= \frac{\hat{\eta}_{j,t}}{s_{j,n}} & \tilde{\zeta}_{i,t} &= \frac{\hat{\zeta}_{i,t}}{s_n} \\ \hat{\eta}_{j,t} &= \frac{f_{j,t} - \hat{\mu}_{j,n}}{\tilde{\sigma}_{j,t}(\hat{\boldsymbol{\theta}}_n^{(j)})} & \hat{\zeta}_{i,t} &= \frac{\tilde{\varepsilon}_{i,t}(\hat{\varphi}_{n|H_0}^{(i)})}{\tilde{g}_{i,t}(\hat{\varphi}_{n|H_0}^{(i)})} \\ s_{j,n}^2 &= n^{-1} \sum_{t=1}^n \hat{\eta}_{j,t}^2 & s_n^2 &= n^{-1} \sum_{t=1}^n \hat{\zeta}_{i,t}^2. \end{aligned}$$

Denote F_n^* the empirical distribution of the vector $\tilde{\boldsymbol{\eta}}_{i,t}$.

2. Simulate a trajectory $(r_{i,t}^*, \mathbf{f}_t^*)_{t=1, \dots, n}$ using parameters $\hat{\varphi}_{n|H_0}^{(i)}$ and with innovations $\boldsymbol{\eta}_{i,t}^* = (\eta_{1,t}^*, \dots, \eta_{m,t}^*, \zeta_{i,t}^*) \stackrel{\text{iid}}{\sim} F_n^*$. On this simulated sample, compute the unconstrained estimator $\hat{\varphi}_n^{(i)*}$ and compute the Wald statistic W_n^* .
3. From the set of observations $(r_{i,1}, \dots, r_{i,n})$, and factors $(\mathbf{f}_1, \dots, \mathbf{f}_n)$, compute the unconstrained multistep QML estimator $\hat{\varphi}_n^{(i)} = (\hat{\boldsymbol{\theta}}_n', \hat{\boldsymbol{\vartheta}}_n^{(i)'})'$ and the Wald statistic W_n .
4. Repeat B times Step 2 and denote $W_n^{*1}, \dots, W_n^{*B}$ the obtained bootstrap statistics. Approximate the p-value of the test for $H_{0,ij}(\bar{c}_{ij})$ by $\#\{W_n^{*b} \geq W_n, b = 1, \dots, B\}/B$.

From Blasques et al. (2022a) (equation (26) and arguments in the proofs of Theorem 2 and 3), we have the following Bahadur-type representation for the multistep QML estimator

$$\begin{aligned} \sqrt{n}(\hat{\varphi}_n^{(i)} - \varphi_0^{(i)}) &= \begin{bmatrix} -\mathbf{J}_*^{-1} & \mathbf{0} \\ \mathbf{J}_{\boldsymbol{\vartheta},i}^{-1} \mathbf{J}_{\boldsymbol{\vartheta}\boldsymbol{\theta},i} \mathbf{J}_*^{-1} & -\mathbf{J}_{\boldsymbol{\vartheta},i}^{-1} \\ -\mathbf{J}_*^{-1} & \mathbf{0} \\ \mathbf{J}_{\boldsymbol{\vartheta},i}^{-1} \mathbf{J}_{\boldsymbol{\vartheta}\boldsymbol{\theta},i} \mathbf{J}_*^{-1} & -\mathbf{J}_{\boldsymbol{\vartheta},i}^{-1} \end{bmatrix} \frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbf{S}_t + o_P(1) \\ &= \begin{bmatrix} -\mathbf{J}_*^{-1} & \mathbf{0} \\ \mathbf{J}_{\boldsymbol{\vartheta},i}^{-1} \mathbf{J}_{\boldsymbol{\vartheta}\boldsymbol{\theta},i} \mathbf{J}_*^{-1} & -\mathbf{J}_{\boldsymbol{\vartheta},i}^{-1} \end{bmatrix} \frac{1}{\sqrt{n}} \sum_{t=1}^n \begin{bmatrix} \mathbf{x}_t \\ \mathbf{y}_t \end{bmatrix} + o_P(1) \end{aligned} \quad (4.10)$$

where $\mathbf{S}_t = \left(\frac{\partial l_{1t}(\boldsymbol{\theta}_0^{(1)})}{\partial \boldsymbol{\theta}'}, \dots, \frac{\partial l_{mt}(\boldsymbol{\theta}_0^{(m)})}{\partial \boldsymbol{\theta}'}, \frac{\partial l_t^{(i)}(\boldsymbol{\varphi}_0^{(i)})}{\partial \boldsymbol{\vartheta}'} \right)$, $\mathbf{x}_t = (\mathbf{x}_t^{(1)}, \dots, \mathbf{x}_t^{(m)})$ with

$$\begin{aligned} \mathbf{x}_t^{(j)} &= (1 - \eta_{j,t}^2) \frac{1}{\sigma_{j,t}^2(\boldsymbol{\theta}_0^{(j)})} \frac{\partial \sigma_{j,t}^2(\boldsymbol{\theta}_0^{(j)})}{\partial \boldsymbol{\theta}} - \frac{2\eta_{j,t}}{\sigma_{j,t}(\boldsymbol{\theta}_0^{(j)})} \mathbf{e}_1 \\ \text{and } \mathbf{y}_t &= (1 - \zeta_{i,t}^2) \frac{1}{g_{i,t}^2(\boldsymbol{\varphi}_0^{(i)})} \frac{\partial g_{i,t}^2(\boldsymbol{\varphi}_0^{(i)})}{\partial \boldsymbol{\vartheta}} - \frac{2\zeta_{i,t}}{g_{i,t}(\boldsymbol{\varphi}_0^{(i)})} \frac{\partial \beta_{i,t}(\boldsymbol{\varphi}_0^{(i)})'}{\partial \boldsymbol{\vartheta}} \mathbf{f}_t \end{aligned}$$

with $\mathbf{e}_1 = (1, 0, 0, 0)'$, $\mathbf{J}_* = \text{diag}(\mathbf{J}^{(j)})_{j=1, \dots, m}$, and $\mathbf{J}^{(j)} = \lim_{n \rightarrow \infty} \frac{\partial^2 \tilde{O}_n^{(j)}(\boldsymbol{\theta}_0^{(j)})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$.

Following Kreiss et al. (2011) and Shimizu (2013), we propose to replace Step 2 of our bootstrap procedure by computing the bootstrap estimator $\hat{\boldsymbol{\varphi}}_n^{(i)*} = (\hat{\boldsymbol{\theta}}_n^{*'}, \hat{\boldsymbol{\vartheta}}_n^{(i)*'})'$ using a Newton-Raphson updating. From (4.10), we set

$$\hat{\boldsymbol{\varphi}}_n^{(i)*} = \hat{\boldsymbol{\varphi}}_{n|H_0}^{(i)} + \frac{1}{n} \begin{bmatrix} -\hat{\mathbf{J}}_*^{-1} & \mathbf{0} \\ \hat{\mathbf{J}}_{\boldsymbol{\vartheta}, i, n|H_0}^{-1} \hat{\mathbf{J}}_{\boldsymbol{\vartheta} \boldsymbol{\theta}, i, n|H_0} \hat{\mathbf{J}}_*^{-1} & -\hat{\mathbf{J}}_{\boldsymbol{\vartheta}, i, n|H_0}^{-1} \end{bmatrix} \sum_{t=1}^n \begin{bmatrix} \hat{\mathbf{x}}_t^* \\ \hat{\mathbf{y}}_t^* \end{bmatrix} \quad (4.11)$$

where $\hat{\mathbf{x}}_t^* = (\hat{\mathbf{x}}_t^{(1)*}, \dots, \hat{\mathbf{x}}_t^{(m)*})$,

$$\begin{aligned} \hat{\mathbf{x}}_t^{(j)*} &= (1 - \eta_{j,t}^{*2}) \frac{1}{\tilde{\sigma}_{j,t}^2(\hat{\boldsymbol{\theta}}_n^{(j)})} \frac{\partial \tilde{\sigma}_{j,t}^2(\hat{\boldsymbol{\theta}}_n^{(j)})}{\partial \boldsymbol{\theta}} - \frac{2\eta_{j,t}^*}{\tilde{\sigma}_{j,t}(\hat{\boldsymbol{\theta}}_n^{(j)})} \mathbf{e}_1 \\ \text{and } \hat{\mathbf{y}}_t^* &= (1 - \zeta_{i,t}^{*2}) \frac{1}{\tilde{g}_{i,t}^2(\hat{\boldsymbol{\varphi}}_{n|H_0}^{(i)})} \frac{\partial \tilde{g}_{i,t}^2(\hat{\boldsymbol{\varphi}}_{n|H_0}^{(i)})}{\partial \boldsymbol{\vartheta}} - \frac{2\zeta_{i,t}^*}{\tilde{g}_{i,t}(\hat{\boldsymbol{\varphi}}_{n|H_0}^{(i)})} \frac{\partial \tilde{\beta}_{i,t}(\hat{\boldsymbol{\varphi}}_{n|H_0}^{(i)})'}{\partial \boldsymbol{\vartheta}} \mathbf{f}_t \end{aligned}$$

where innovations $\boldsymbol{\eta}_{i,t}^* = (\eta_{1,t}^*, \dots, \eta_{m,t}^*, \zeta_{i,t}^*)' \stackrel{\text{iid}}{\sim} F_n^*$ and $\hat{\mathbf{J}}_* = \text{diag}(\hat{\mathbf{J}}^{(j)})_{j=1, \dots, m}$ with $\hat{\mathbf{J}}^{(j)} = \frac{\partial^2 \tilde{O}_n^{(j)}(\hat{\boldsymbol{\theta}}_n^{(j)})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$ a consistent estimator of \mathbf{J}_* .

In order to establish the consistency of the bootstrap procedure, we need to slightly strengthen the assumption on the distribution of $\zeta_{i,t}$.

[A+] $\zeta_{i,t}$ admits a finite fourth-order moment $\mathbb{E}\zeta_{i,t}^4 = \kappa < \infty$.

The proposed resampling algorithm is valid in the following sense.

Theorem 4.2. *Under the set of assumptions [A-ACB] and [A+], under $H_{0,ij}(\bar{c}_{ij})$, and if the distribution of $\zeta_{i,t}$ admits a bounded density with respect to the Lebesgue measure, then for almost all given realisations $(r_{i,t}, \mathbf{f}_t)_{t=1, \dots, n}$, as $n \rightarrow \infty$,*

$$W_n^* \rightarrow \chi_r^2$$

where W_n^* is computed from the unconstrained bootstrap estimator in (4.11).

This theorem ensures that the distribution of the bootstrap Wald test statistic given $(r_{i,t}, \mathbf{f}_t)$ well mimics the unconditional distribution of W_n . Additionally, in finite sample, the boot-

strap distribution of W_n^* should better approximate the distribution of W_n than the theoretical asymptotic distribution.

4.3 Conditional Risk Premia Estimation Procedure

The estimation of time invariant risk premia is usually conducted following Fama and MacBeth (1973) procedure. To capture dynamic behavior of the risk premia, a common approach is to roll short-window two-pass regressions (see for example the famous model of Fama and French (1997)). However, this method may yield erroneous results. Indeed, in the case of constant risk premia and time-varying coefficients in the conditional regression, Grassi and Violante (2021) show that risk premia estimated from two-pass regressions are usually not significantly different from 0, due to unfavorable signal-to-noise ratio of the rolling OLS estimator. We thus propose to introduce a two-step estimation procedure to recover the risk premia from a multi-factor model in a large cross-section of assets, under the assumption of absence of arbitrage opportunities.

Ross (1976) and Chamberlain and Rothschild (1983) showed that the absence of asymptotic arbitrage opportunities implies asset pricing restrictions in large cross-section of assets. We borrow from Gagliardini et al. (2016) who derive no-arbitrage restrictions for a multi-period economy with a continuum of assets^{2,3}. In particular, they show that the asset pricing restriction is such that, for any t , there exists a unique random vector $\mathbf{v}_t = (v_{1,t}, \dots, v_{m,t})'$ such that for almost all i ,

$$\alpha_{i,t} = \beta_{i,t}' \mathbf{v}_t \text{ almost surely.} \quad (4.12)$$

The dynamic asset pricing model (4.1) combined with (4.12) yields that for almost all i ,

$$\mathbb{E}[r_{i,t} | \mathcal{F}_{t-1}] = \beta_{i,t}' \boldsymbol{\lambda}_t \text{ almost surely,} \quad (4.13)$$

where $\boldsymbol{\lambda}_t$, the vector of time-varying risk premia, is given by $\boldsymbol{\lambda}_t = \mathbf{v}_t + \boldsymbol{\mu}$ with $\boldsymbol{\mu} = \mathbb{E}[\mathbf{f}_t] = (\mu_1, \dots, \mu_m)$. Note that Equation (4.13) nests favored asset pricing models with time-invariant coefficients. For example, in the CAPM where the only factor is the Market portfolio excess returns or in the three factor model of Fama and French (1993), we have $\mathbf{v}_t = 0$.

Building upon the asset pricing restriction (4.12), we propose the following two step procedure to recover time-varying risk premia from the score-driven conditional regression model.

1. For each asset i in the system, $i = 1, \dots, N$, we estimate $\hat{\alpha}_{i,t} := \hat{\alpha}_{i,t}(\hat{\varphi}_n^{(i)})$ and $\hat{\beta}_{i,t} := \hat{\beta}_{i,t}(\hat{\varphi}_n^{(i)})$ by QML
2. For each t , we estimate the vector $\hat{\mathbf{v}}_t = \arg \min_{\mathbf{v}} \left\| \hat{\beta}'_t \mathbf{v} - \hat{\boldsymbol{\alpha}}_t \right\|_2$ by OLS, where $\hat{\beta}_t = (\hat{\beta}_{i,t})_{i=1, \dots, N}$ and $\hat{\boldsymbol{\alpha}}_t = (\hat{\alpha}_{i,t})_{i=1, \dots, N}$.

²See Al-Najjar (1995, 1999) for a discussion on the advantages of working with a continuum of assets in arbitrage pricing models.

³The authors define arbitrage opportunities as zero-cost investment with nonnegative payoffs in each state of the world and positive payoff in some states of the world.

Combining \widehat{v}_t with parameter $\widehat{\mu}_n$ estimated in $\widehat{\theta}_n$ yields the estimation of the time-varying risk premia $\widehat{\lambda}_t = \widehat{\mu}_n + \widehat{v}_t$.

4.4 Monte Carlo experiments

In order to study the empirical behavior of the test statistics defined in Section 2 and to assess the finite sample properties of our risk premia estimation procedure, we carry out some Monte Carlo experiments. We assume Gaussian innovations for both the factors and the conditional regression residuals, ie $\eta_{j,t} \sim \mathcal{N}(0, 1)$ and $\zeta_{i,t} \sim \mathcal{N}(0, 1)$, where $\zeta_{i,t}$ is such that $\varepsilon_{i,t} = g_{i,t}\zeta_{i,t}$. We simulate 1000 trajectories of size 2500 with two factors and a time-varying intercept. More precisely, we consider two scenarios, one under the null assumption that the beta to the first factor is equal to 0, and one under the alternative with a dynamic conditional betas between the simulated asset returns and the simulated factors. In both cases, we set parameters driving the two factors as $\theta_0^{(1)} = (-0.05, 0.1, 0.1, 0.75)'$ and $\theta_0^{(2)} = (0, 0.1, 0.15, 0.7)'$. The parameter driving the innovations ε_t is set to $\vartheta_0^{(\varepsilon_i)} = (0.1, 0.05, 0.94)'$ while the parameter for the dynamic of the conditional constant is set to $\vartheta_0^{(i,0)} = (0.1, 0.05, 0.94)'$. Additionally, the parameter involved in the dynamic beta between the simulated asset and the second factor is set to $\vartheta_0^{(1)} = (0.05, 0.2, 0.99)'$. Finally, when simulating under the alternative, we set $\vartheta_0^{(2)} = (0.001, 0.2, 0.99)'$. For both scenarios we set $c_{0i1} = \bar{c}_{i1} = 0.9$. Results are displayed in Table 4.1.

% of rejection at 5% level	Wald	Rao	LR
Under H_0	2.3	2.2	2.4
Under H_1	100	100	100

Table 4.1: *Finite sample properties of the test statistics factor significance*

While one should reject the null assumption about 5% of the time, simulations in finite sample lead to a smaller alpha than expected following the asymptotic distribution of test statistics. This empirical weakness of the test statistics in controlling the size of the test may be due to numeric approximations in the computations of the derivatives involved in matrices $\mathbf{J}_{\vartheta,i}$ and $\mathbf{\Omega}_{\vartheta,i}$. On the other hand, the test statistics appear very powerful when the data are far from the null hypothesis as it is the case in the second scenario of our simulations.

We then turn to the empirical behavior of our proposed two-step estimation procedure of the factors risk premia. We need to simulate a system of N assets with 2 factors, where the dynamics of the conditional regression coefficients is given by Equation (4.3). However, remark that the asset pricing restriction (4.12) cannot be directly obtained from a parametric restriction on the individual assets. To evaluate the performance of the risk premia estimation, we propose to compare the *implied* simulated risk premia to the estimated one. In particular, we develop the following simulation scheme.

1. We simulate the factor's time series from the GARCH parameters $\theta_{01} = (-0.05, 0.1, 0.1, 0.75)$ and $\theta_{02} = (0, 0.1, 0.15, 0.7)$ with $\eta_{j,t} \sim \mathcal{N}(0, 1)$, $j = 1, 2$.

2. For $i = 1, \dots, N$, we randomly draw the parameters driving $g_{i,t}$, $\alpha_{i,t}$ and $\beta_{i,j,t}$, $j = 1, 2$. In particular, $\omega_{\varepsilon_i} \sim \mathcal{U}[0.01, 0.1]$, $a_{\varepsilon_i} \sim \mathcal{U}[0.01, 0.2]$, $b_{\varepsilon_i} = 1 - a_{\varepsilon_i} - u$, with $u \sim \mathcal{U}[0.01, 0.1]$. Additionally, $\bar{\omega}_{\alpha_i} \sim \mathcal{U}[-0.25, 0.25]$, $\xi_{\alpha_i} \sim \mathcal{U}[0.01, 0.2]$, $c_{\alpha_i} \sim \mathcal{U}[0.7, 0.99]$ and for $j = 1, 2$, $\bar{\omega}_{i,j} \sim \mathcal{U}[-0.25, 0.25]$, $\xi_{i,j} \sim \mathcal{U}[0.01, 0.2]$, $c_{i,j} \sim \mathcal{U}[0.7, 0.99]$.
3. From the $\varphi_0^{(i)}$ obtained in step 2, we simulate $\alpha_{i,t}$ and $\beta_{i,t}$ using $\varepsilon_{i,t} = g_{i,t}\zeta_{i,t}$, with $\zeta_{i,t} \sim \mathcal{N}(0, 1)$.
4. For all t , estimate the simulated *implied* risk premia $\mathbf{v}_t = \arg \min_{\mathbf{v}} \|\beta_t' \mathbf{v} - \alpha_t\|_2$ where $\beta_t = (\beta_{i,t})_{i=1, \dots, N}$ and $\alpha_t = (\alpha_{i,t})_{i=1, \dots, N}$.
5. Using the two-step estimation procedure, compute $\hat{\mathbf{v}}_t$ and compare it to \mathbf{v}_t .

Figure 4.1 shows the estimated risk premia for a simulated system of $N = 50$ assets. We see that the estimations, represented in dashed lines, are very close to the simulated *implied* risk premia. Note that the vector of *implied* risk premia might not verify the asset pricing restriction (4.12). This may be due to the estimation of a continuum of assets by a large discrete system of assets.

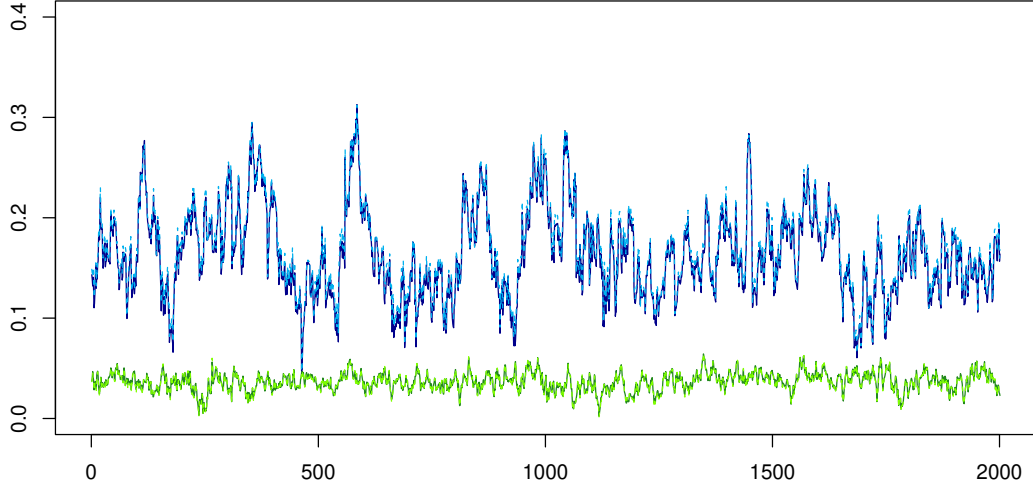


Figure 4.1: *Empirical results of the two-step risk premia estimation procedure for a two-factor system with 50 assets. Simulated implied risk premia are plotted in dark green and blue while estimates are plotted in light dashed lines*

4.5 Application: Assessing a dynamic carbon-related risk premium

The fast development of Sustainable Investing has spurred academic interest in the impact of climate-related risks on financial returns. However, findings are often contradictory. On the one hand, Pastor et al. (2021b) and De Angelis et al. (2022) argue that exposure to climate-related risks should be rewarded by a risk premium while the utility gained by sustainable investors for holding Green stocks compensates the expected lower returns. In addition, Bolton and Kacperczyk (2020) show empirically that firms' carbon intensity is linked to higher returns. On the other hand, sustainable assets have outperformed assets with high carbon exposures in recent years. Pastor et al. (2021a) argue that this out-performance is due to unexpected higher environmental concerns among investors. Therefore, the aggregated effect of climate-related risks over the cross-section of returns is yet unclear. Part of this issue lies in the fact that sustainable preferences are evolving among investors. Awareness about climate change and its consequences is growing along with new scientific findings and political will. Krueger et al. (2020) outline that institutional investors consider climate change as a growing risk in the market. Such a transition needs to be taken into account when estimating econometric models and underlines the need for dynamic parameters. While most of the literature analysing this issue relies on panel analysis (Bolton and Kacperczyk, 2020; Pastor et al., 2021a) or the standard Fama and MacBeth (1973) two steps procedure (Görge et al., 2019), we propose to compute a dynamic risk premium using the new methodology presented in this Chapter.

In the financial literature, the estimation of risk premia is commonly conducted on portfolios instead of single stocks to reduce the error variances issue⁴. Therefore, we use the thirty-eight industries portfolios to regress against well known three factors of Fama and French (1993): Market (Mkt), Size (SMB) and Value (HML). In addition to the classical three factors we introduce an additional carbon-related risk factor (hereafter referred to as Green Minus Brown - GMB). We analyse a period of 21 years from January 2000 until December 2020 which covers the increasing awareness about climate change.

4.5.1 Construction of a carbon-related risk factor

We follow the *mimicking portfolio* approach of Fama and French (1993) to build a carbon long/short risk factor as a zero-cost portfolio long on Green stocks and short on Brown ones. This approach is a standard in the financial literature and is commonly adopted to build climate-related portfolio (see for example Bolton and Kacperczyk (2020); Pastor et al. (2021a)). Because of broader data availability compared to other economic areas, we focus our analysis on the US stock market.

Daily returns for US companies are obtained from the Center for Research in Security Prices

⁴Gagliardini et al. (2016) discuss this practice and advocate the use of single stocks rather than aggregations. However their empirical results show that using industry portfolios covering a wide sectorial dispersion yields satisfactory results when estimating dynamic risk premia.

(CRSP) database. We restrain the analysis to companies that exist over the entire period and for which carbon data is available at some point. This results in a final sample of 1241 companies out of 18800 in the original CRSP database from which 3328 are matching the carbon database, but not all surviving the entire period. It accounts for 57% of the total market capitalisation on yearly average during the period and 80% of the S&P 500 constituents are present in the final sample.

Carbon emissions with scope 1, 2 and 3 are collected from *TruCost*. According to the GHG Protocol Corporate Standard ⁵ "a company's GHG emissions [are divided] into three scopes. Scope 1 emissions are direct emissions from owned or controlled sources. Scope 2 emissions are indirect emissions from the generation of purchased energy. Scope 3 emissions are all indirect emissions (not included in scope 2) that occur in the value chain of the reporting company, including both upstream and downstream emissions.". Following recommendations of TCFD (2017) we standardise carbon emissions by year-end net revenues to build carbon intensity metrics that are comparable between companies. We observe carbon intensities at a yearly frequency from 2005 to 2020. We focus on Scope 2 carbon emissions as it covers broader range than scope 1 and because the calculation methodology of scope 3, despite its ambitious aims, reaches a lower consensus due to unresolved double counting issues. Grouping companies by carbon emissions intensity deciles reveals large differences in financial profiles. Indeed, as Table 4.5 in Appendix 4.8 shows, Green companies (first deciles) distribute on average a smaller dividend yield than Brown companies (last deciles), are more profitable and expensive. Green companies are also less capital expenditure extensive and thus bear less debt on average than Brown companies.

Following standard practices in sustainable investing (Statman, 2006; Kempf and Osthoff, 2007), two possible methods may be implemented to build a carbon long-short portfolio candidate: *Best-in-Universe* or *Best-in-Class*. *Best-in-Universe* aims to select companies with best/worst carbon emissions intensities in the universe while *Best-in-Class* screens companies within each sector or industry. The first may lead to over-exposition of one class meanwhile the other requires well distributed data among classes. In order to recover robust results, we apply both approaches and additionally control for size bias following Fama and French (1993) *mimicking* portfolio approach.

As screening is applied at each date for portfolio rebalancing, data scarcity in carbon emissions intensity may be an issue. Indeed, especially prior to 2016, most sectors encounter more than 50% of missing values (See Table 4.6 in Appendix 4.8). On the one hand, it makes standard *Best-in-Class* approaches irrelevant because environmentally unaligned companies may still be selected for sectors with a lot of missing data. On the other hand, however, this is not a major issue for applying a *Best-in-Universe* screening since high-stake sectors for the climate transition are well disclosed ⁶.

⁵https://ghgprotocol.org/sites/default/files/standards_supporting/FAQ.pdf

⁶Details about high and low-stakes sectors are provided in Hoepner et al. (2019)

Figure 4.2 presents the performance of the quintile portfolios built upon Scope 2 carbon intensity. Note that prior to 2013, the bottom quintile was hardly outperformed by the Green portfolio. However it is clear than Brown stocks have experienced lower performances in more recent years.

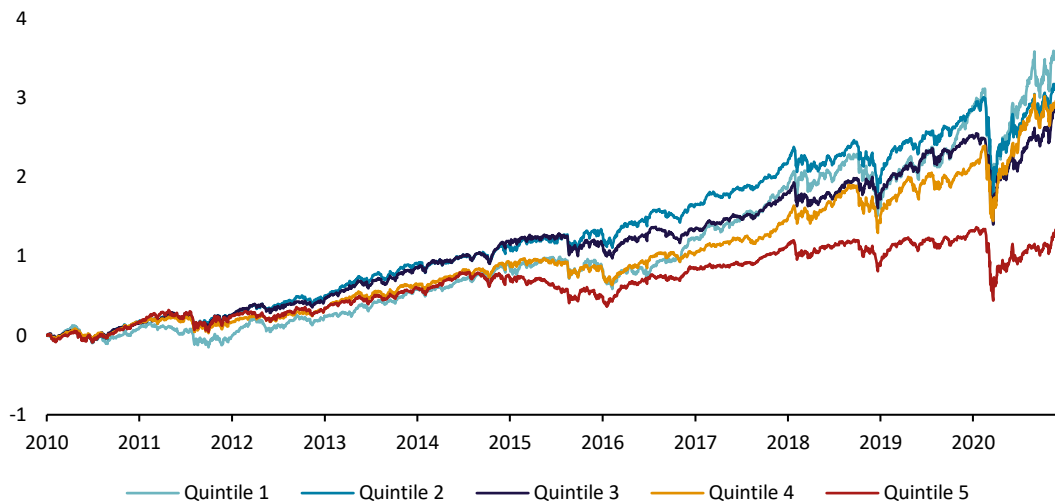


Figure 4.2: *Performance of Long-Only portfolios constructed on scope 2 carbon intensity*

4.5.2 Carbon factor significance testing

Although monthly returns are regularly chosen when estimating risk premium, we use daily frequency to increase the amount of observations as it is of critical importance to accurately estimate dynamic conditional betas. We are compelled to make this choice because of limited historical data regarding firms carbon emissions. As presented in Equation (4.1) we estimate the following model

$$r_{i,t} = \alpha_{i,t} + \beta_{i,\text{Mkt}} f_{\text{Mkt},t} + \beta_{i,\text{SMB}} f_{\text{SMB},t} + \beta_{i,\text{HML}} f_{\text{HML},t} + \beta_{i,\text{GMB}} f_{\text{GMB},t} + \varepsilon_{i,t} \quad (4.14)$$

where $i = \{1, \dots, 48\}$ denotes one of the 48 industry portfolio and assess the significance of the GMB factor using the test statistics presented in Equation (4.9). As shown in Table 4.2, the GMB factor appears significant for most of the sectors.

	Wald	Rao	LR
2% level	0.81	0.83	0.79
5% level	0.87	0.87	0.85
10% level	0.90	0.87	0.89

Table 4.2: *Proportion of industry portfolios with a significant beta to the GMB factor*

Moreover, the dynamics of betas are diversified among industries, correlated with OLS rolling-

window betas estimates while less volatile than the former as Table 4.7 in the appendix illustrates. This result holds for different window sizes and emphasises the ability of the ACB model to perform well without the need to fit exogenous parameter on subjective criteria.

4.5.3 Risk premium estimation

Applying our two-step procedure allows to recover significant risk premia for all factors at the 95% confidence level. The market portfolio delivers an average risk premium of around 17% annually, while Size and Value factors returned negative risk premium over the period. The disappearance of the Size premium is the subject of an ongoing debate in the financial literature (see for example Van Dijk (2011)) and can be related to the results presented in Chapter 2. Additionally, the under-performance of the Value factor is not surprising given the observation period. Indeed, the Value factor is often linked to interest rates and the expansionary monetary policies of the US Federal Reserve has weighted on Value stocks in the last decade. Finally, the premium associated to the GMB factor is significant and positive with an average of 7.9% annually over the period, as presented in the following Table.

	Estimate	<i>p</i> -value
Market	17.28	0.00
Size	-4.24	0.00
Value	-10.28	0.00
Green Minus Brown	7.97	0.00

Table 4.3: *Annualized average risk premia of the Fama-French and GMB factors*

Figure 4.4 in the appendix presents the obtained dynamics of the four risk premia. The time-varying behavior of the conditional risk premia are clearly observable and allow for financial interpretability of the result. For example, we can notice that the Market risk premium tends to increase in period of market turbulence such as the summer 2015 Chinese stock market crisis or the recent Covid-19 shock. Additionally, Figure 4.3 presents the dynamic of our estimated carbon-related risk premia for GMB factors constructed using scope 1 (light blue) or scope 2 (dark blue) carbon intensity data. Both risk premia appear positive and share a similar dynamic, underlining the robustness of our empirical finding to the selected carbon metric. Again, there appears to be indisputable time variation in the conditional risk premium. This might explain the inconsistency in the different contributions to the debate on the estimation of a carbon-related risk premium. Instability of the performance of climate-related portfolios was already observed by Pastor et al. (2021a) and Ardia et al. (2020) who attribute the changes in the risk-premium dynamic to modifications in the investors' attention to climate news. On that matter, it is remarkable to see the inflection of our estimated carbon risk premium around the Paris COP-21 in December 2015. Additional potential drivers behind the dynamic of the risk premium are discussed in Appendix 4.8.

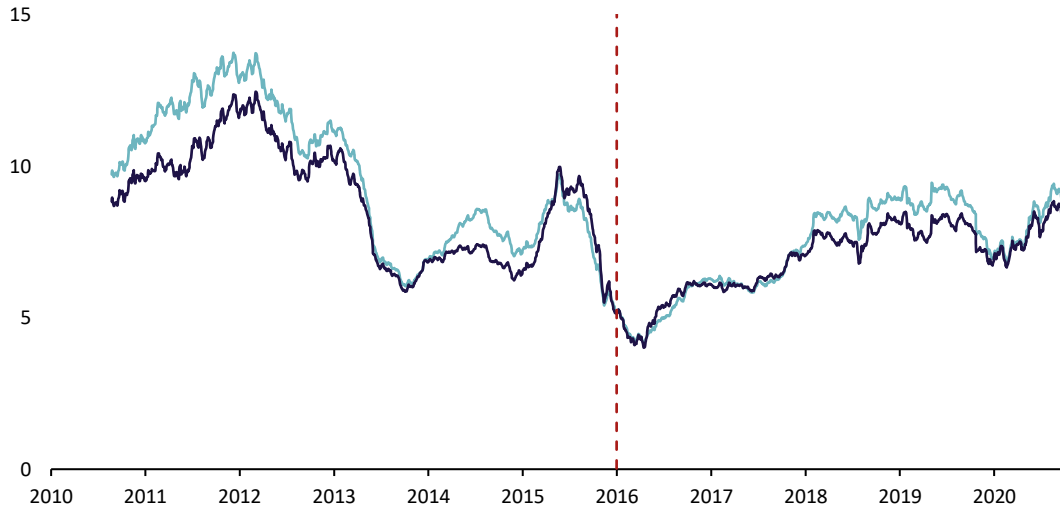


Figure 4.3: *Green Minus Brown conditional risk premium associated to portfolio formed on scope 1 (light blue) and scope 2 (dark blue) carbon intensity. The red dashed line mark the COP-21 meetings*

4.6 Conclusion

Linear asset pricing theory heavily relies on regression coefficients to assess assets exposures to systemic risk factors. For simplicity, the estimation is commonly based on OLS and time-varying coefficients are obtained by rolling the window of estimation. However, such methods may lead to spurious results on the risk premia estimates as pointed out by Grassi and Violante (2021). In this Chapter, we build upon the recently introduced ACB model of Blasques et al. (2022a) and extend it to the case of an empirical asset pricing exercise.

In particular, on the statistical front, we extend the asymptotic results obtained by the authors in two main aspects. First, we derive the asymptotic distributions of standard test statistics for testing constancy or nullity of a beta coefficient, allowing to assess the significance of a factor in the conditional regression. Additionally, we introduce a residual-based bootstrap test procedure. We establish the validity of our proposed scheme by showing that, under the null, the distribution of the bootstrap estimator of the parameter given the observations well mimics the unconditional distribution of a multi-step QMLE introduced by Blasques et al. (2022a).

Moreover, on the financial front, in the spirit of Gagliardini et al. (2016), we derive a two-step estimation procedure of the time-varying risk premia from a panel of assets and assess the validity of our method on Monte Carlo experiments. Applying this procedure to US industry portfolios, we find a significant and positive time-varying carbon-related risk premium. The dynamic of the risk premium appears to be driven by exogenous shocks, such as large events (COP21, Covid Crisis), bear economic outlooks and sustainable finance inflows. This result adds to the recent literature in theoretical sustainable finance arguing in favour of including

extra-financial criteria to explain the cross-section of returns. In addition, it gives empirical proof of a positive risk premium for holding Green stocks and contradicts equilibrium models stating that climate-concerned investors should be willing to accept lower returns to avoid extra-financial risks (see for instance Pastor et al. (2021b) and De Angelis et al. (2022)). The reason behind this out-performance may however lie on an attention-driven appetite for carbon-efficient stocks as climate risks become more material creating a demand and supply imbalance. The time-varying behavior of investors nevertheless provides a strong argument in favor of a model with dynamic risk premia.

4.7 Appendix A: Proofs and technical results

4.7.1 Proof of Theorem 4.1

Proof. We begin by studying the asymptotic distribution of the Wald statistic under the null $H_{0,ij}(\bar{c}_{ij}) : \mathbf{R}\boldsymbol{\vartheta}_0^{(i)} = 0$. From (4.8) and Slutsky lemma, we obtain

$$\sqrt{n}(\mathbf{R}\widehat{\boldsymbol{\vartheta}}_n^{(i)} - \mathbf{R}\boldsymbol{\vartheta}_0^{(i)}) = \sqrt{n}\mathbf{R}(\widehat{\boldsymbol{\vartheta}}_n^{(i)} - \boldsymbol{\vartheta}_0^{(i)}) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \mathbf{R}\boldsymbol{\Omega}_{\boldsymbol{\vartheta},i}\mathbf{R}') \quad (4.15)$$

and from the quadratic form, we thus have

$$n(\mathbf{R}\widehat{\boldsymbol{\vartheta}}_n^{(i)})' \left(\mathbf{R}\widehat{\boldsymbol{\Omega}}_{\boldsymbol{\vartheta},i,n}\mathbf{R}' \right)^{-1} (\mathbf{R}\widehat{\boldsymbol{\vartheta}}_n^{(i)}) \xrightarrow{\mathcal{L}} \chi_r^2$$

under $H_{0,ij}(\bar{c}_{ij}) : \mathbf{R}\boldsymbol{\vartheta}_0^{(i)} = 0$. Thus, the critical region of the Wald test at the asymptotic level ν is $\{W_n > \chi_r^2(1 - \nu)\}$.

To study the Rao-score statistic, we first introduce the Lagrangian function associated with the likelihood optimization problem constrained by $H_{0,ij}(\bar{c}_{ij})$, $\tilde{O}_n(\widehat{\boldsymbol{\theta}}_n, \boldsymbol{\vartheta}) + (\mathbf{R}\boldsymbol{\vartheta})'\boldsymbol{\lambda}$. The first-order condition yields

$$\frac{\partial \tilde{O}_n(\widehat{\boldsymbol{\theta}}_n, \widehat{\boldsymbol{\vartheta}}_{n|H_0}^{(i)})}{\partial \boldsymbol{\vartheta}} + \mathbf{R}'\tilde{\boldsymbol{\lambda}}_n = \frac{\partial \tilde{O}_n(\widehat{\boldsymbol{\varphi}}_{n|H_0}^{(i)})}{\partial \boldsymbol{\vartheta}} + \mathbf{R}'\tilde{\boldsymbol{\lambda}}_n = 0 \quad (4.16)$$

with $\tilde{\boldsymbol{\lambda}}_n$ the Lagrange multipliers vector.

Under $H_{0,ij}(\bar{c}_{ij})$, we have

$$\sqrt{n}(\mathbf{R}\widehat{\boldsymbol{\vartheta}}_n^{(i)}) = \mathbf{R}\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n^{(i)} - \boldsymbol{\vartheta}_0^{(i)})$$

and

$$0 = \sqrt{n}(\mathbf{R}\widehat{\boldsymbol{\vartheta}}_{n|H_0}^{(i)}) = \mathbf{R}\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_{n|H_0}^{(i)} - \boldsymbol{\vartheta}_0^{(i)})$$

since $\widehat{\boldsymbol{\vartheta}}_{n|H_0}^{(i)}$ is the constrained estimator. By subtraction, we thus obtain

$$\sqrt{n}(\mathbf{R}\widehat{\boldsymbol{\vartheta}}_n^{(i)}) = \mathbf{R}\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n^{(i)} - \widehat{\boldsymbol{\vartheta}}_{n|H_0}^{(i)}). \quad (4.17)$$

Using Taylor expansions, we can also notice that

$$0 = \sqrt{n} \frac{\partial \tilde{O}_n(\widehat{\boldsymbol{\varphi}}_{n|H_0}^{(i)})}{\partial \boldsymbol{\vartheta}} \stackrel{o_P(1)}{=} \sqrt{n} \frac{\partial \tilde{O}_n(\boldsymbol{\varphi}_0^{(i)})}{\partial \boldsymbol{\vartheta}} + \sqrt{n}\mathbf{J}_{\boldsymbol{\vartheta},i}(\widehat{\boldsymbol{\vartheta}}_n^{(i)} - \boldsymbol{\vartheta}_0^{(i)}) \quad (4.18)$$

and

$$\sqrt{n} \frac{\partial \tilde{O}_n(\widehat{\boldsymbol{\varphi}}_{n|H_0}^{(i)})}{\partial \boldsymbol{\vartheta}} \stackrel{o_P(1)}{=} \sqrt{n} \frac{\partial \tilde{O}_n(\boldsymbol{\varphi}_0^{(i)})}{\partial \boldsymbol{\vartheta}} + \sqrt{n}\mathbf{J}_{\boldsymbol{\vartheta},i}(\widehat{\boldsymbol{\vartheta}}_{n|H_0}^{(i)} - \boldsymbol{\vartheta}_0^{(i)})$$

which yields by subtraction

$$\sqrt{n} \frac{\partial \tilde{O}_n(\widehat{\boldsymbol{\varphi}}_{n|H_0}^{(i)})}{\partial \boldsymbol{\vartheta}} \stackrel{o_P(1)}{=} -\sqrt{n}\mathbf{J}_{\boldsymbol{\vartheta},i}(\widehat{\boldsymbol{\vartheta}}_n^{(i)} - \widehat{\boldsymbol{\vartheta}}_{n|H_0}^{(i)})$$

hence

$$\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n^{(i)} - \widehat{\boldsymbol{\vartheta}}_{n|H_0}^{(i)}) \stackrel{OP(1)}{=} -\sqrt{n}\mathbf{J}_{\boldsymbol{\vartheta},i}^{-1} \frac{\partial \tilde{O}_n(\widehat{\boldsymbol{\varphi}}_{n|H_0}^{(i)})}{\partial \boldsymbol{\vartheta}}. \quad (4.19)$$

From (4.16), (4.17) and (4.19), we thus obtain

$$\sqrt{n}(\mathbf{R}\widehat{\boldsymbol{\vartheta}}_n^{(i)}) \stackrel{OP(1)}{=} \mathbf{R}\mathbf{J}_{\boldsymbol{\vartheta},i}^{-1}\mathbf{R}'\sqrt{n}\tilde{\boldsymbol{\lambda}}_n \quad (4.20)$$

which yields

$$\sqrt{n}\tilde{\boldsymbol{\lambda}}_n \stackrel{OP(1)}{=} [\mathbf{R}\mathbf{J}_{\boldsymbol{\vartheta},i}^{-1}\mathbf{R}']^{-1} \sqrt{n}(\mathbf{R}\widehat{\boldsymbol{\vartheta}}_n^{(i)})$$

hence from (4.15), under $H_{0,ij}(\bar{c}_{ij})$,

$$\sqrt{n}\tilde{\boldsymbol{\lambda}}_n \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}_i) \text{ with } \boldsymbol{\Lambda}_i = [\mathbf{R}\mathbf{J}_{\boldsymbol{\vartheta},i}^{-1}\mathbf{R}']^{-1} [\mathbf{R}\boldsymbol{\Omega}_{\boldsymbol{\vartheta},i}\mathbf{R}'] [\mathbf{R}\mathbf{J}_{\boldsymbol{\vartheta},i}^{-1}\mathbf{R}']^{-1}$$

as $n \rightarrow \infty$. Taking the quadratic form, we obtain under $H_{0,ij}(\bar{c}_{ij})$,

$$n\tilde{\boldsymbol{\lambda}}_n' \mathbf{R}\hat{\mathbf{J}}_{\boldsymbol{\vartheta},i,n|H_0}^{-1} \mathbf{R}' (\mathbf{R}\hat{\boldsymbol{\Omega}}_{\boldsymbol{\vartheta},i,n|H_0} \mathbf{R}')^{-1} \mathbf{R}\hat{\mathbf{J}}_{\boldsymbol{\vartheta},i,n|H_0}^{-1} \mathbf{R}' \tilde{\boldsymbol{\lambda}}_n \xrightarrow{\mathcal{L}} \chi_r^2$$

and (4.16) yields

$$n \frac{\partial \tilde{O}_n(\widehat{\boldsymbol{\varphi}}_{n|H_0}^{(i)})}{\partial \boldsymbol{\vartheta}} \hat{\mathbf{J}}_{\boldsymbol{\vartheta},i,n|H_0}^{-1} \mathbf{R}' (\mathbf{R}\hat{\boldsymbol{\Omega}}_{\boldsymbol{\vartheta},i,n|H_0} \mathbf{R}')^{-1} \mathbf{R}\hat{\mathbf{J}}_{\boldsymbol{\vartheta},i,n|H_0}^{-1} \frac{\partial \tilde{O}_n(\widehat{\boldsymbol{\varphi}}_{n|H_0}^{(i)})}{\partial \boldsymbol{\vartheta}} \xrightarrow{\mathcal{L}} \chi_r^2. \quad (4.21)$$

It follows that the critical region of the Rao-score test at the asymptotic level ν is $\{R_n > \chi_r^2(1 - \nu)\}$.

We finally focus on the Quasi Likelihood Ratio statistic. Using Taylor expansions, we get

$$\tilde{O}_n(\widehat{\boldsymbol{\varphi}}_n^{(i)}) \stackrel{OP(1)}{=} \tilde{O}_n(\widehat{\boldsymbol{\theta}}_n, \boldsymbol{\vartheta}_0^{(i)}) + \frac{\partial \tilde{O}_n(\boldsymbol{\varphi}_0^{(i)})}{\partial \boldsymbol{\vartheta}'} (\widehat{\boldsymbol{\vartheta}}_n^{(i)} - \boldsymbol{\vartheta}_0^{(i)}) + \frac{1}{2} (\widehat{\boldsymbol{\vartheta}}_n^{(i)} - \boldsymbol{\vartheta}_0^{(i)})' \mathbf{J}_{\boldsymbol{\vartheta},i} (\widehat{\boldsymbol{\vartheta}}_n^{(i)} - \boldsymbol{\vartheta}_0^{(i)})$$

and

$$\begin{aligned} \tilde{O}_n(\widehat{\boldsymbol{\varphi}}_{n|H_0}^{(i)}) \stackrel{OP(1)}{=} & \tilde{O}_n(\widehat{\boldsymbol{\theta}}_n, \boldsymbol{\vartheta}_0^{(i)}) + \frac{\partial \tilde{O}_n(\boldsymbol{\varphi}_0^{(i)})}{\partial \boldsymbol{\vartheta}'} (\widehat{\boldsymbol{\vartheta}}_{n|H_0}^{(i)} - \boldsymbol{\vartheta}_0^{(i)}) \\ & + \frac{1}{2} (\widehat{\boldsymbol{\vartheta}}_{n|H_0}^{(i)} - \boldsymbol{\vartheta}_0^{(i)})' \mathbf{J}_{\boldsymbol{\vartheta},i} (\widehat{\boldsymbol{\vartheta}}_{n|H_0}^{(i)} - \boldsymbol{\vartheta}_0^{(i)}), \end{aligned}$$

hence, by subtraction,

$$\begin{aligned} L_n \stackrel{OP(1)}{=} & 2n \frac{\partial \tilde{O}_n(\boldsymbol{\varphi}_0^{(i)})}{\partial \boldsymbol{\vartheta}'} (\widehat{\boldsymbol{\vartheta}}_{n|H_0}^{(i)} - \widehat{\boldsymbol{\vartheta}}_n^{(i)}) + n(\widehat{\boldsymbol{\vartheta}}_{n|H_0}^{(i)} - \boldsymbol{\vartheta}_0^{(i)})' \mathbf{J}_{\boldsymbol{\vartheta},i} (\widehat{\boldsymbol{\vartheta}}_{n|H_0}^{(i)} - \boldsymbol{\vartheta}_0^{(i)}) \\ & - n(\widehat{\boldsymbol{\vartheta}}_n^{(i)} - \boldsymbol{\vartheta}_0^{(i)})' \mathbf{J}_{\boldsymbol{\vartheta},i} (\widehat{\boldsymbol{\vartheta}}_n^{(i)} - \boldsymbol{\vartheta}_0^{(i)}), \end{aligned}$$

and, from (4.18),

$$\begin{aligned} L_n \stackrel{OP(1)}{=} & 2n(\widehat{\boldsymbol{\vartheta}}_n^{(i)} - \boldsymbol{\vartheta}_0^{(i)})' \mathbf{J}_{\boldsymbol{\vartheta},i} (\widehat{\boldsymbol{\vartheta}}_{n|H_0}^{(i)} - \widehat{\boldsymbol{\vartheta}}_n^{(i)}) \\ & + n(\widehat{\boldsymbol{\vartheta}}_{n|H_0}^{(i)} - \boldsymbol{\vartheta}_0^{(i)})' \mathbf{J}_{\boldsymbol{\vartheta},i} (\widehat{\boldsymbol{\vartheta}}_{n|H_0}^{(i)} - \boldsymbol{\vartheta}_0^{(i)}) - n(\widehat{\boldsymbol{\vartheta}}_n^{(i)} - \boldsymbol{\vartheta}_0^{(i)})' \mathbf{J}_{\boldsymbol{\vartheta},i} (\widehat{\boldsymbol{\vartheta}}_n^{(i)} - \boldsymbol{\vartheta}_0^{(i)}) \\ \stackrel{OP(1)}{=} & n(\widehat{\boldsymbol{\vartheta}}_n^{(i)} - \widehat{\boldsymbol{\vartheta}}_{n|H_0}^{(i)})' \mathbf{J}_{\boldsymbol{\vartheta},i} (\widehat{\boldsymbol{\vartheta}}_n^{(i)} - \widehat{\boldsymbol{\vartheta}}_{n|H_0}^{(i)}). \end{aligned}$$

From (4.16) and (4.19), it follows that under $H_{0,ij}(\bar{c}_{ij})$,

$$L_n \stackrel{op(1)}{=} \sqrt{n} \mathbf{R}' \tilde{\boldsymbol{\lambda}}_n \mathbf{J}_{\vartheta,i}^{-1} \tilde{\boldsymbol{\lambda}}_n' \mathbf{R} \sqrt{n}. \quad (4.22)$$

Using standard results on quadratic form of Gaussian vectors (see for example Box (1954) or Scheffe (1959)), from (4.20) and Slutsky lemma, we obtain

$$L_n \rightarrow \bar{\chi}(\boldsymbol{\pi}) \text{ with } \bar{\chi}(\boldsymbol{\pi}) = \sum_{j=1}^{3(m+1)} \pi_j \lambda_j^2$$

where $\boldsymbol{\pi} = (\pi_j)_{j=1, \dots, 3(m+1)}$ is the vector of eigenvalues of the matrix $\mathbf{J}_{\vartheta,i}^{-1} \mathbf{R}' \boldsymbol{\Lambda}_i \mathbf{R}$ with

$$\boldsymbol{\Lambda}_i = \left[\mathbf{R} \mathbf{J}_{\vartheta,i}^{-1} \mathbf{R}' \right]^{-1} \left[\mathbf{R} \boldsymbol{\Omega}_{\vartheta,i} \mathbf{R}' \right] \left[\mathbf{R} \mathbf{J}_{\vartheta,i}^{-1} \mathbf{R}' \right]^{-1'}. \quad \square$$

4.7.2 Proof of Theorem 4.2

Proof. We prove Theorem 4.2 by showing that for almost all given realisations $(r_{i,t}, \mathbf{f}_t)$,

$$\sqrt{n}(\hat{\boldsymbol{\varphi}}_n^{(i)*} - \hat{\boldsymbol{\varphi}}_n^{(i)}) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_i) \text{ as } n \rightarrow \infty. \quad (4.23)$$

For any $\boldsymbol{\varphi} \in \Theta_{\boldsymbol{\varphi}}$, let

$$\hat{\boldsymbol{\eta}}_{i,t}(\boldsymbol{\varphi}) = \left(\frac{f_{1,t} - \mu_1}{\tilde{\sigma}_{1,t}(\boldsymbol{\theta}^{(1)})}, \dots, \frac{f_{m,t} - \mu_m}{\tilde{\sigma}_{m,t}(\boldsymbol{\theta}^{(m)})}, \frac{\tilde{\varepsilon}_{i,t}(\boldsymbol{\varphi})}{\tilde{g}_{i,t}(\boldsymbol{\varphi})} \right)$$

and

$$\boldsymbol{\eta}_{i,t}(\boldsymbol{\varphi}) = \left(\frac{f_{1,t} - \mu_1}{\sigma_{1,t}(\boldsymbol{\theta}^{(1)})}, \dots, \frac{f_{m,t} - \mu_m}{\sigma_{m,t}(\boldsymbol{\theta}^{(m)})}, \frac{\varepsilon_{i,t}(\boldsymbol{\varphi})}{g_{i,t}(\boldsymbol{\varphi})} \right)$$

so that $\hat{\boldsymbol{\eta}}_{i,t}(\hat{\boldsymbol{\varphi}}_{n|H_0}^{(i)}) = \hat{\boldsymbol{\eta}}_{i,t}$ and $\boldsymbol{\eta}_{i,t}(\boldsymbol{\varphi}_0^{(i)}) = \boldsymbol{\eta}_{i,t}$. From Proposition 3 in Blasques et al. (2022a), we have the asymptotic irrelevance of initial values yielding

$$\lim_{t \rightarrow \infty} \rho^{-t} \sup_{\boldsymbol{\varphi} \in \Theta_{\boldsymbol{\varphi}}} \|\boldsymbol{\eta}_{i,t}(\boldsymbol{\varphi}) - \hat{\boldsymbol{\eta}}_{i,t}(\boldsymbol{\varphi})\| = 0 \quad (4.24)$$

almost surely for some $\rho \in [0, 1)$. In addition, from a Taylor expansion we have

$$\boldsymbol{\eta}_{i,t}(\hat{\boldsymbol{\varphi}}_{n|H_0}^{(i)}) = \boldsymbol{\eta}_{i,t}(\boldsymbol{\varphi}_0^{(i)}) + \frac{\partial \boldsymbol{\eta}_{i,t}(\boldsymbol{\varphi}^{(*)})}{\partial \boldsymbol{\varphi}} (\hat{\boldsymbol{\varphi}}_{n|H_0}^{(i)} - \boldsymbol{\varphi}_0^{(i)}) \quad (4.25)$$

with $\boldsymbol{\varphi}^{(*)}$ between $\hat{\boldsymbol{\varphi}}_{n|H_0}^{(i)}$ and $\boldsymbol{\varphi}_0^{(i)}$. From Francq and Zakoian (2004), we have that, for any $k > 0$, for any $j \in \{1, \dots, m\}$, there exists a neighborhood $\mathcal{V}(\boldsymbol{\theta}_0^{(j)})$ of $\boldsymbol{\theta}_0^{(j)}$ such that

$$\mathbb{E} \sup_{\boldsymbol{\theta} \in \mathcal{V}(\boldsymbol{\theta}_0^{(j)})} \left| \frac{\sigma_{j,t}(\boldsymbol{\theta}_0^{(j)})}{\sigma_{j,t}(\boldsymbol{\theta})} \right|^k < \infty, \quad \text{and} \quad \mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta} \left\| \frac{1}{\sigma_{t,j}^2(\boldsymbol{\theta})} \frac{\partial \sigma_{t,j}^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\|^k < \infty$$

whence

$$\sup_{\boldsymbol{\theta} \in \mathcal{V}(\boldsymbol{\theta}_0^{(j)})} \left\| \frac{\partial \eta_{t,j}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\| \leq K \sup_{\boldsymbol{\theta} \in \mathcal{V}(\boldsymbol{\theta}_0^{(j)})} \left\| \frac{\sigma_{j,t}(\boldsymbol{\theta}_0^{(j)})}{\sigma_{j,t}(\boldsymbol{\theta})} \frac{1}{\sigma_{t,j}^2(\boldsymbol{\theta})} \frac{\partial \sigma_{t,j}^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\| \leq K u_{j,t} |\eta_{j,t}| \quad (4.26)$$

where $u_{j,t} \in \mathcal{F}_{t-1}$ and $\mathbb{E}|u_t|^{k/2} < \infty$ by Holder inequality. Additionally, by Assumption A10 in Blasques et al. (2022a), there exists a neighborhood $\mathcal{V}(\boldsymbol{\varphi}_0)$ of $\boldsymbol{\varphi}_0$ such that

$$\begin{aligned} \sup_{\boldsymbol{\varphi} \in \mathcal{V}(\boldsymbol{\varphi}_0)} \left\| \frac{\partial \zeta_{i,t}(\boldsymbol{\varphi})}{\partial \boldsymbol{\vartheta}} \right\| &= \sup_{\boldsymbol{\varphi} \in \mathcal{V}(\boldsymbol{\varphi}_0)} \left\| \frac{1}{g_{i,t}^2(\boldsymbol{\varphi})} \frac{\partial \varepsilon_{i,t}(\boldsymbol{\varphi})}{\partial \boldsymbol{\vartheta}} - \zeta_{i,t} \frac{g_{i,t}(\widehat{\boldsymbol{\varphi}}_{n|H_0}^{(i)})}{g_{i,t}(\boldsymbol{\varphi})} \frac{1}{g_{i,t}(\boldsymbol{\varphi})} \frac{\partial g_{i,t}(\boldsymbol{\varphi})}{\partial \boldsymbol{\vartheta}} \right\| \\ &\leq K u_t |\zeta_{i,t}| \end{aligned} \quad (4.27)$$

with $u_t \in \mathcal{F}_{t-1}$.

From (4.24), (4.25), (4.26) and (4.27), we obtain

$$\begin{aligned} |\hat{\eta}_{1,t} - \eta_{1,t}| &\leq (\rho_1^t + \|\widehat{\boldsymbol{\theta}}_n^{(1)} - \boldsymbol{\theta}_0^{(1)}\|) u_{1,t} |\eta_{1,t}| \\ &\dots \\ |\hat{\eta}_{m,t} - \eta_{m,t}| &\leq (\rho_m^t + \|\widehat{\boldsymbol{\theta}}_n^{(m)} - \boldsymbol{\theta}_0^{(m)}\|) u_{m,t} |\eta_{m,t}| \\ \text{and } |\hat{\zeta}_{i,t} - \zeta_{i,t}| &\leq (\rho^t + \|\widehat{\boldsymbol{\vartheta}}_{n|H_0} - \boldsymbol{\vartheta}_0\|) u_t |\zeta_{i,t}| \end{aligned}$$

where $\rho \in (0, 1)$ and $\rho_j \in (0, 1)$ for any $j \in \{1, \dots, m\}$. This entails

$$\begin{aligned} n^{-1} \sum_{t=1}^n |\hat{\eta}_{1,t} - \eta_{1,t}|^k &\rightarrow 0 \\ &\dots \\ n^{-1} \sum_{t=1}^n |\hat{\eta}_{m,t} - \eta_{m,t}|^k &\rightarrow 0 \\ \text{and } n^{-1} \sum_{t=1}^n |\hat{\zeta}_{i,t} - \zeta_{i,t}|^k &\rightarrow 0 \end{aligned}$$

as n tends to infinity for $k = 1, 2, 3, 4$. From assumption A5(b) in Blasques et al. (2022a), $n^{-1} \sum_{t=1}^n \eta_{j,t}^k \rightarrow \mathbb{E} \eta_{j,1}^k < \infty$ and from assumption [A+] we have $n^{-1} \sum_{t=1}^n \zeta_{i,t}^k \rightarrow \mathbb{E} \zeta_{i,1}^k < \infty$, whence

$$\frac{1}{n} \sum_{t=1}^n \hat{\boldsymbol{\eta}}_{i,t}^k \rightarrow \mathbb{E} \boldsymbol{\eta}_{i,1}^k. \quad (4.28)$$

Additionally, from arguments in the proof of Theorem 2 and 3 in Blasques et al. (2022a), we have under $H_{0,ij}(\bar{c}_{ij})$

$$\begin{aligned} &\begin{bmatrix} -\widehat{\mathbf{J}}_*^{-1} & \mathbf{0} \\ \widehat{\mathbf{J}}_{\boldsymbol{\vartheta},i,n|H_0}^{-1} \widehat{\mathbf{J}}_{\boldsymbol{\vartheta}\boldsymbol{\theta},i,n|H_0} \widehat{\mathbf{J}}_*^{-1} & -\widehat{\mathbf{J}}_{\boldsymbol{\vartheta},i,n|H_0}^{-1} \end{bmatrix} \frac{1}{\sqrt{n}} \sum_{t=1}^n \begin{bmatrix} \widehat{\mathbf{x}}_t^* \\ \widehat{\mathbf{y}}_t^* \end{bmatrix} \\ &= \begin{bmatrix} -\mathbf{J}_*^{-1} & \mathbf{0} \\ \mathbf{J}_{\boldsymbol{\vartheta},i}^{-1} \mathbf{J}_{\boldsymbol{\vartheta}\boldsymbol{\theta},i} \mathbf{J}_*^{-1} & -\mathbf{J}_{\boldsymbol{\vartheta},i}^{-1} \end{bmatrix} \frac{1}{\sqrt{n}} \sum_{t=1}^n \begin{bmatrix} \mathbf{x}_t^* \\ \mathbf{y}_t^* \end{bmatrix} + o_P(1) \end{aligned}$$

where $\mathbf{x}_t^* = (\mathbf{x}_t^{(1)*}, \dots, \mathbf{x}_t^{(m)*})$,

$$\begin{aligned} \mathbf{x}_t^{(j)*} &= (1 - \eta_{j,t}^{*2}) \frac{1}{\sigma_{j,t}^2(\boldsymbol{\theta}_0^{(j)})} \frac{\partial \sigma_{j,t}^2(\boldsymbol{\theta}_0^{(j)})}{\partial \boldsymbol{\theta}} - \frac{2\eta_{j,t}^*}{\sigma_{j,t}(\boldsymbol{\theta}_0^{(j)})} \mathbf{e}_1 \\ \text{and } \mathbf{y}_t^* &= (1 - \zeta_{i,t}^{*2}) \frac{1}{g_{i,t}^2(\boldsymbol{\varphi}_0^{(i)})} \frac{\partial g_{i,t}^2(\boldsymbol{\varphi}_0^{(i)})}{\partial \boldsymbol{\vartheta}} - \frac{2\zeta_{i,t}^*}{g_{i,t}(\boldsymbol{\varphi}_0^{(i)})} \frac{\partial \boldsymbol{\beta}_{i,t}(\boldsymbol{\varphi}_0^{(i)})}{\partial \boldsymbol{\vartheta}} \mathbf{f}_t. \end{aligned}$$

Since

$$\boldsymbol{\Omega}_i = \begin{bmatrix} -\mathbf{J}_*^{-1} & \mathbf{0} \\ \mathbf{J}_{\boldsymbol{\vartheta},i}^{-1} \mathbf{J}_{\boldsymbol{\vartheta}\boldsymbol{\theta},i} \mathbf{J}_*^{-1} & -\mathbf{J}_{\boldsymbol{\vartheta},i}^{-1} \end{bmatrix} \mathbf{I}_i \begin{bmatrix} -\mathbf{J}_*^{-1} & \mathbf{J}_*^{-1} \mathbf{J}_{\boldsymbol{\theta}\boldsymbol{\vartheta},i} \mathbf{J}_{\boldsymbol{\vartheta},i}^{-1} \\ \mathbf{0} & -\mathbf{J}_{\boldsymbol{\vartheta},i}^{-1} \end{bmatrix}$$

where $\mathbf{I}_i = \mathbb{E} \mathbf{S}_t \mathbf{S}_t'$, it suffices to show that conditionally on $(r_{i,t}, \mathbf{f}_t)$,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbf{z}_t^* := \frac{1}{\sqrt{n}} \sum_{t=1}^n \begin{bmatrix} \mathbf{x}_t^* \\ \mathbf{y}_t^* \end{bmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \mathbf{I}_i) \quad \text{as } n \rightarrow \infty. \quad (4.29)$$

Since, conditionally on $(r_{i,t})$ and (\mathbf{f}_t) , for each n , the random vectors $\mathbf{z}_{1,n}^*, \mathbf{z}_{2,n}^*, \dots$ are independent and centered with finite second-order moments, from Lindeberg's CLT for triangular arrays and the Wold–Cramer device, it suffices to show that for any $\boldsymbol{\lambda} \in \mathbb{R}^{7m+5}$

$$\frac{1}{n} \sum_{t=1}^n \text{Var}[\boldsymbol{\lambda}' \mathbf{z}_t^*] \rightarrow \boldsymbol{\lambda}' \mathbf{I}_i \boldsymbol{\lambda} \quad \text{as } n \rightarrow \infty \quad (4.30)$$

and for all $\nu > 0$

$$\frac{1}{n} \sum_{t=1}^n \mathbb{E}[(\boldsymbol{\lambda}' \mathbf{z}_t^*)^2 \mathbf{1}_{\{|\boldsymbol{\lambda}' \mathbf{z}_t^*| > \sqrt{n\nu}\}}] \rightarrow 0 \quad \text{as } n \rightarrow \infty \quad (4.31)$$

to prove (4.29). Note that conditionally on $(r_{i,t})$ and (\mathbf{f}_t) , we can rewrite for any $j = 1, \dots, m$,

$$\begin{aligned} \mathbf{x}_t^{(j)*} &= (1 - \eta_{j,t}^{*2}) \boldsymbol{\phi}_t^{(j)}(\boldsymbol{\theta}_0^{(j)}) + \eta_{j,t}^* \boldsymbol{\psi}_t^{(j)}(\boldsymbol{\theta}_0^{(j)}) \\ \text{and } \mathbf{y}_t^* &= (1 - \zeta_{i,t}^{*2}) \boldsymbol{\phi}_t(\boldsymbol{\varphi}_0^{(i)}) + \zeta_{i,t}^* \boldsymbol{\psi}_t(\boldsymbol{\varphi}_0^{(i)}) \end{aligned}$$

where $\boldsymbol{\phi}_t^{(j)}(\boldsymbol{\theta}_0^{(j)})$, $\boldsymbol{\psi}_t^{(j)}(\boldsymbol{\theta}_0^{(j)})$, $\boldsymbol{\phi}_t(\boldsymbol{\varphi}_0^{(i)})$ and $\boldsymbol{\psi}_t(\boldsymbol{\varphi}_0^{(i)})$ are non random vectors. Denoting, for $j = 1, \dots, m$,

$$\begin{aligned} \hat{\boldsymbol{x}}_{t,n}^{(j)} &= (1 - \hat{\eta}_{j,t}^{*2}) \boldsymbol{\phi}_t^{(j)}(\hat{\boldsymbol{\theta}}_n^{(j)}) + \hat{\eta}_{j,t}^* \boldsymbol{\psi}_t^{(j)}(\hat{\boldsymbol{\theta}}_n^{(j)}) \\ \text{and } \hat{\boldsymbol{y}}_{t,n} &= (1 - \hat{\zeta}_{i,t}^2) \boldsymbol{\phi}_t(\hat{\boldsymbol{\varphi}}_{n|H_0}^{(i)}) + \hat{\zeta}_{i,t} \boldsymbol{\psi}_t(\hat{\boldsymbol{\varphi}}_{n|H_0}^{(i)}) \end{aligned}$$

and $\hat{\mathbf{z}}_{t,n} = (\hat{\boldsymbol{x}}_{t,n}^{(1)'}, \dots, \hat{\boldsymbol{x}}_{t,n}^{(m)'})', \hat{\boldsymbol{y}}_{t,n}'$, from (4.10), (4.28) and an application of the continuous mapping theorem, we have

$$\text{Var}[\boldsymbol{\lambda}' \hat{\mathbf{z}}_{t,n}^*] = \frac{1}{n} \sum_{t=1}^n (\boldsymbol{\lambda}' \hat{\mathbf{z}}_{t,n}^*)^2 \rightarrow \mathbb{E}[(\boldsymbol{\lambda}' \mathbf{S}_t)^2] \quad \text{as } n \rightarrow \infty.$$

Cesaro mean convergence theorem and the fact that $\mathbb{E} \mathbf{S}_t \mathbf{S}_t' = \mathbf{I}_i$, yield (4.30).

We now turn to the proof of (4.30). Consider first that $\boldsymbol{\lambda} = (\mathbf{e}'_4, \mathbf{0}'_{7m+1})'$. We then obtain

$$\begin{aligned} & \mathbb{E}[(\boldsymbol{\lambda}' \mathbf{z}_t^*)^2 \mathbb{1}_{\{|\boldsymbol{\lambda}' \mathbf{z}_t^*| > \sqrt{n\nu}\}}] \\ &= \mathbb{E}[(\mathbf{e}'_4 \mathbf{x}_t^{(1)*})^2 \mathbb{1}_{\{|\mathbf{e}'_4 \mathbf{x}_t^{(1)*}| > \sqrt{n\nu}\}}] \\ &\leq \mathbb{E}[(1 - \eta_{j,t}^{*2})a + \eta_{j,t}^* b]^2 \mathbb{1}_{\{|(1 - \eta_{j,t}^{*2})a + \eta_{j,t}^* b| > \sqrt{n\nu}\}} \end{aligned}$$

where $a = \mathbf{e}'_4 \boldsymbol{\phi}_t^{(j)}(\boldsymbol{\theta}_0^{(j)})$ and $b = \mathbf{e}'_4 \boldsymbol{\psi}_t^{(j)}(\boldsymbol{\theta}_0^{(j)})$ are scalars.

For any A there exists n_A such that if $n > n_A$, then the previous expectation is bounded by

$$\mathbb{E}[(1 - \eta_{j,t}^{*2})a + \eta_{j,t}^* b]^2 \mathbb{1}_{\{|(1 - \eta_{j,t}^{*2})a + \eta_{j,t}^* b| > A\}}$$

which tends to

$$\int_{|(1-u^2)a+ub|>A} |(1-u^2)a+ub|^2 F^{(1)}(x) \quad (4.32)$$

by Lemma A1 in Francq and Zakoian (2022), with $F^{(1)}$ the density of $\eta_{1,t}$. Since (4.32) is arbitrarily small when A is large, we have proved (4.31) for $\boldsymbol{\lambda} = (\mathbf{e}'_4, \mathbf{0}'_{7m+1})'$. Iterating the argument for any $\boldsymbol{\lambda}$, such that $\boldsymbol{\lambda}' \mathbf{z}_t^* = \mathbf{e}'_4 \mathbf{x}_t^{(j)*}$ for $j = 2, \dots, m$, and $\boldsymbol{\lambda} = (\mathbf{0}'_{4m}, \mathbf{e}'_{3m+5})'$, yields (4.31) which concludes the proof. □

4.8 Appendix B: Details on empirical results

Companies profiles by carbon emissions

	2000-2007	2008-2012	2013-2016	2016-2020	
Communication Services	1.02	84.15	82.61	78.36	87.08
Consumer Discretionary	15.67	254.89	225.81	207.05	222.61
Consumer Staples	5.35	687.04	617.75	574.11	567.85
Energy	4.02	831.57	751.94	782.09	753.79
Financials	18.19	49.49	46.92	54.16	45.85
Health Care	7.87	169.89	139.96	127.00	129.71
Industrials	17.17	420.29	376.01	365.28	350.51
Information Technology	13.94	181.70	171.54	158.03	151.29
Materials	6.22	1202.59	1164.24	1013.76	979.61
Real Estate	6.22	166.06	137.79	133.16	130.61
Utilities	4.33	4413.52	4089.86	3511.01	2682.21

Table 4.4: Average scope 2 carbon intensity by sector

carbon Scope 2 decile	1	2	3	4	5	6	7	8	9	10
Dividends/Net Income mean	0.31	0.23	0.24	0.26	0.19	0.21	0.28	0.20	0.30	0.38
Dividends/Net Income median	0.32	0.15	0.14	0.17	0.07	0.12	0.16	0.08	0.19	0.34
Dividends/Net Income std	0.05	0.25	0.30	0.32	0.25	0.26	0.35	0.29	0.34	0.34
Debt/Assets mean 0.07	0.21	0.28	0.32	0.26	0.29	0.32	0.37	0.31	0.34	
Debt/Assets median	0.04	0.18	0.25	0.30	0.24	0.28	0.33	0.42	0.31	0.36
Debt/Assets std	0.09	0.17	0.19	0.18	0.17	0.19	0.18	0.21	0.16	0.13
EBIT/Assets mean	0.11	0.12	0.10	0.10	0.08	0.09	0.08	0.10	0.09	0.07
EBIT/Assets median	0.11	0.10	0.09	0.09	0.09	0.09	0.07	0.10	0.07	0.06
EBIT/Assets std	0.06	0.07	0.09	0.07	0.08	0.09	0.08	0.08	0.07	0.07
CapEx/Assets mean	0.01	0.01	0.02	0.03	0.02	0.02	0.03	0.03	0.05	0.06
CapEx/Assets median	0.01	0.01	0.01	0.02	0.02	0.02	0.02	0.03	0.05	0.06
CapEx/Assets std	0.00	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.03	0.03
Book-to-Market mean 115.81	116.46	88.50	56.13	33.57	43.75	30.85	16.34	61.77	80.80	
Book-to-Market median	9.91	2.93	3.25	1.99	1.88	1.49	0.89	0.80	2.95	8.72
Book-to-Market std	216.99	231.32	194.27	147.22	94.28	127.73	103.05	87.78	152.15	172.94
Cost-of-capital mean	1.30	0.55	1.21	0.96	0.88	0.97	1.01	1.21	0.99	1.49
Cost-of-capital median	0.41	0.32	0.59	0.38	0.33	0.36	0.46	0.45	0.40	0.67
Cost-of-capital std	1.98	0.69	1.57	1.49	1.33	1.44	1.57	1.81	1.50	1.88

Table 4.5: Financial ratios summary statistics by scope 2 carbon intensity deciles

	2008	2011	2014	2017	2020
Communication Services	0.15	0.15	0.08	0.00	0.00
Consumer Discretionary	0.45	0.43	0.41	0.01	0.00
Consumer Staples	0.46	0.44	0.43	0.01	0.00
Energy	0.33	0.31	0.29	0.10	0.00
Financials	0.58	0.58	0.55	0.00	0.00
Health Care	0.50	0.47	0.47	0.05	0.00
Industrials	0.53	0.52	0.49	0.01	0.00
Information Technology	0.55	0.54	0.50	0.04	0.00
Materials	0.37	0.34	0.30	0.01	0.00
Real Estate	0.56	0.56	0.54	0.01	0.00
Utilities	0.18	0.18	0.13	0.00	0.00

Table 4.6: Proportion of missing values by sector for scope 2 carbon intensity

Comparison to rolling windows estimators

	Market	GMB
Aero	1.10	–
Agric	0.90	3.60
Autos	0.90	0.60
Banks	0.80	0.90
Beer	1.40	1.10
BldMt	0.90	0.80
Books	1.10	0.90
Boxes	3.90	1.60
BusSv	1.00	2.10
Chems	1.10	0.90
Chips	0.80	2.60
Clths	1.20	1.30
Cnstr	0.80	0.60
Coal	1.00	1.10
Comps	1.00	1.30
Drugs	0.80	1.20
EleEq	1.20	1.00
FabPr	1.80	0.80
Fin	0.90	1.10
Food	1.30	1.70
Fun	0.90	2.3
Gold	0.80	0.90
Guns	2.80	2.10
Hlth	–	–
Hshld	0.90	3.30
Insur	1.10	1.00
LabEq	1.00	0.90
Mach	0.80	0.80
Meals	1.10	5.40
MedEq	1.40	4.40
Mines	0.90	1.00
Oil	0.70	0.60
Other	1.00	1.50
Paper	0.90	3.30
PerSv	1.00	0.70
REst	0.80	4.00
Rtail	1.10	1.00
Rubbr	1.00	–
Ships	0.90	1.10
Smoke	0.90	–
Soda	1.00	–
Steel	0.80	0.90
Telcm	0.90	–
Toys	–	0.50
Trans	0.60	0.80
Txtls	–	0.70
Util	0.80	2.3
Whsl	0.80	0.80

Table 4.7: *Factor loadings volatility ratios between 2-year fixed-rolling windows regressions and ACB regressions*

Estimated dynamic risk premia

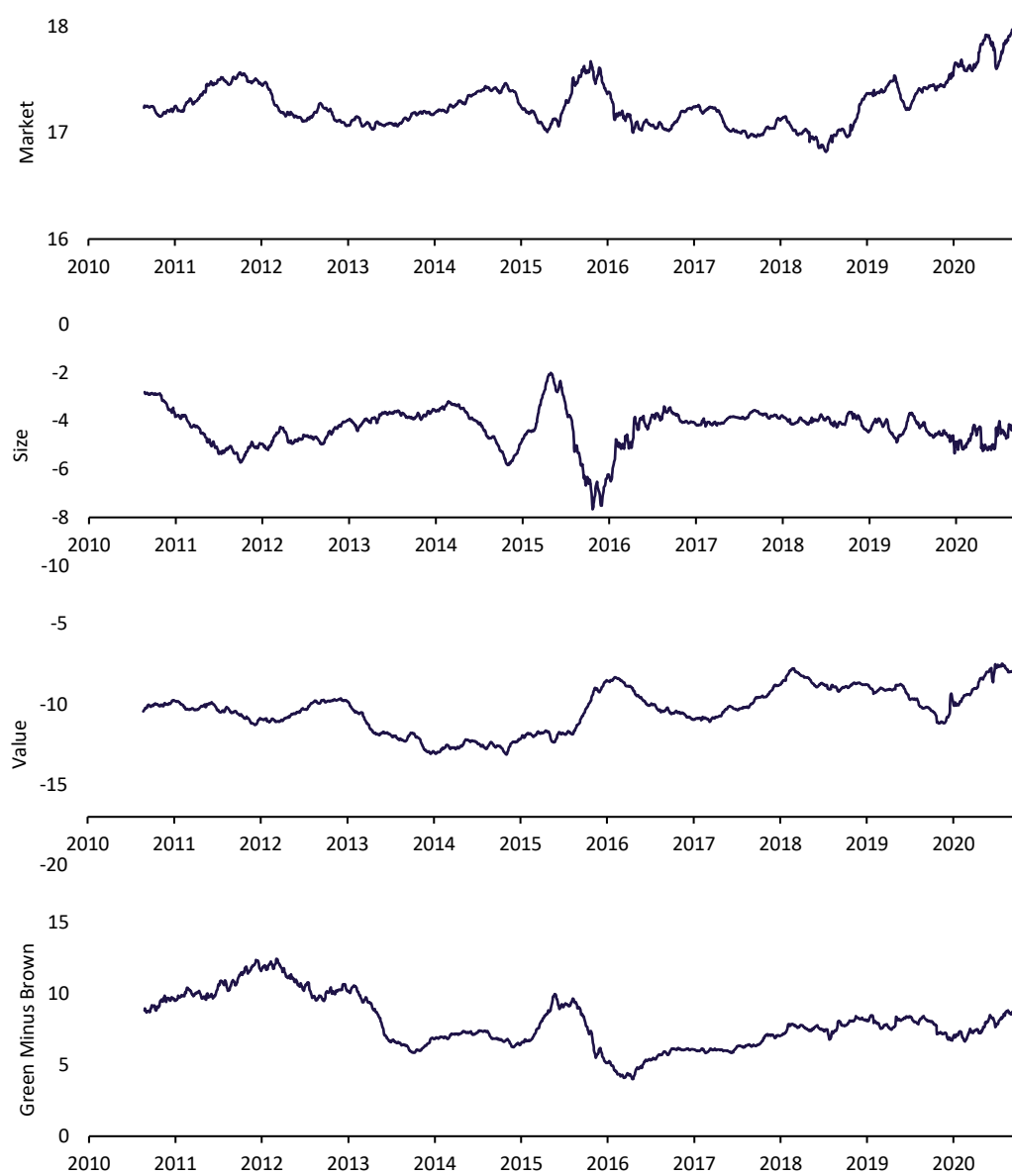


Figure 4.4: *Conditional risk premia of the Fama French and GMB factors*

Explaining the Green premium

Variable name	Source	Estimate	Se	Pval
Climate Change Attention	MCCC Sentometrics-research	0.23	0.13	0.092
Total Sustainable Funds Flows	Morningstar & Factset	0.96	0.58	0.108
Bond Spread	FRED-MD Dataset	-1.15	0.52	0.027
Oil Price	FRED-MD Dataset	0.03	0.23	0.89
Economic State	FRED-MD Dataset	-0.05	0.03	0.095

Table 4.8: *Potential explanatory variables of the conditional carbon-related risk premium*

The regression is estimated on a sample from 2010 to 2019 because of the lack of data for the most recent period. Transformations are applied to raw variables following literature and standard practices:

- "*Attention*" is derived following Ardia et al. (2020).
- "*Total Sustainable Funds Flows*" is computed following a list of sustainable funds based on Morningstar rating and using Factset quarterly holdings.
- "*Economic state*" is obtained from a two-states Hidden-Markov Model on the S&P 500 monthly returns.

Chapter 5

Conclusion and Perspectives

5.1 Conclusion

This thesis focuses on conditional volatility models and their statistical and probabilistic properties. More precisely, this manuscript addresses two extensions of the standard financial econometrics literature on heteroskedastic time series: the memory property of power transformations of financial returns, and the modeling of dynamic regression coefficients.

It is well-known that, empirically, autocorrelations of squared returns are decaying slowly toward zero, a stylized fact that is incompatible with short memory models such as the prominent GARCH(p, q) framework. On the contrary, ARCH(∞) models allow for a flexible representation of the sequence of coefficients driving the conditional volatility process. We introduce a novel ARCH(∞) specification taking into account the asymmetry of the volatility response to a negative and a positive shock. We give a condition for the existence of a unique strictly stationary solution to our model which is equivalent to the one of Giraitis et al. (2000), Kazakevičius and Leipus (2002) and Douc et al. (2008) in the case of a symmetric process. Additionally, we extend the asymptotic results on QML estimation derived by Robinson and Zaffaroni (2006). In particular, we are able to derive the consistency and asymptotic distribution of the QMLE using milder assumptions, noticeably without restricting the parameter to the interior of the parameter space. A portmanteau goodness-of-fit test is introduced as an alternative to the omnibus test of Hidalgo and Zaffaroni (2007) as well as standard statistics for testing a linear restriction on the parameter. For the latter, we extend the results of Francq and Zakoïan (2009) and derive the asymptotic distributions of the Wald, Rao-score and quasi-likelihood ratio test statistics when the parameter may be on the boundary of the parameter space. Additionally, a multivariate ARCH(∞) specification is considered. The conditions for the existence of a stationary solution extend the results from Doukhan et al. (2006) while asymptotic properties of the QMLE are established. Moreover, the specification bridges the gap between moderate memory volatility and the modeling of the time-varying parameters in a conditional regression.

Bollerslev et al. (1988) were the first to use a multivariate conditional volatility model to obtain

dynamic slope coefficients in a bivariate system. Since in this case, the unconditional beta is the ratio between the unconditional covariance between an asset and a factor, and the unconditional variance of the factor, they proposed to obtain the conditional beta as the ratio of the conditional covariance and variance. Building upon this idea, Engle (2016) generalized this indirect modeling of the slope coefficients with multiple regressors and a dynamic conditional correlation matrix. We consider a different approach and extend the Cholesky-GARCH model of Darolles et al. (2018), which allows to directly model the conditional betas. Relaxing the GARCH-type specification of the model allows for a more flexible modeling of the memory pattern exhibited by individual assets. Additionally, we introduce exogenous variables in the dynamic of both the conditional betas and the conditional variances. Such extension has proved useful in the modeling of individual stock variance (Francq and Thieu (2019)) and multivariate system (Bollerslev et al. (2020b)).

However, the Cholesky GARCH framework, as well as its ARCH(∞) extension, is ill-suited for the modeling of large asset systems. We rather propose to leverage the recently introduced Autoregressive Conditional Beta model of Blasques et al. (2022a). In particular, we introduce test statistics to assess the significance of a factor in the conditional regression and establish their asymptotic distributions. A bootstrap procedure to estimate the Wald statistic is also considered and the validity of the proposed recursive bootstrap scheme is established, extending the results for the multi-step QMLE considered in the seminal article of Blasques et al. (2022a). Finally, in the spirit of the two-step estimation of Fama and MacBeth (1973) and building upon Gagliardini et al. (2016), we propose a two-step procedure to estimate time-varying risk premia from the conditional betas obtained in the ACB model.

Although theoretical results are central to this manuscript, a significant attention has been devoted to empirical applications. In the univariate case, they provide unequivocal arguments in favor of ARCH(∞) models over standard GARCH model when modeling peripheral assets. In a multivariate exercise, they illustrate the need to account for the time varying behavior of linear regressions coefficients.

5.2 Perspectives

The results presented in the previous chapter have opened several potential future lines of research, either on applications or theoretical aspects of the studied models.

In particular, this dissertation has provided numerous arguments in favor of volatility models exhibiting more persistence than the prominent GARCH formulations. However, such models are sometimes faced with scepticism and a recurrent critique stems from the neglect of potential breaks in the sample. Indeed, neglecting structural breaks generates *spurious* memory effects when applying standard econometric tools without accounting for the implied non-stationarity. This effect was for example established on linear time series with infrequent breaks (see Gouriéroux and Jasiak (2001) or Diebold and Inoue (2001)) and extended to conditional volatility mod-

els. Both Mikosch and Stărică (2004) and Hillebrand (2005) show that ignoring non-stationarities in the data sample when estimating a GARCH(1,1) will bias respectively the Whittle estimator and maximum likelihood estimator towards an integrated GARCH (IGARCH) effect where $\alpha + \beta = 1$. In addition, Andreou and Ghysels (2002) show that there exist multiple change points in the dynamics of conditional variance when considering data from foreign exchange and stock markets. Most studies conclude that the strong memory pattern observed empirically is only a consequence of the presence of breaks. However, the existence of infrequent structural breaks might not contradict the presence of long memory on the subsamples where no break occurs. Empirical results presented in Chapters 2 and 3 show that there might exist a relationship between liquidity or idiosyncrasy and high persistence. Therefore, there might be intrinsic market components that justify memory effects beyond potential breaks in the data. Testing this hypothesis is however challenging and it would be interesting to extend the testing procedure in Chapter 2 to allow for potential breaks. Change points test procedures have mostly been studied in the case of GARCH processes, leaving out the case of ARCH(∞) models. A notable exception is the CUSUM (for cumulative sum) test of Kokoszka and Leipus (2000) who extend the seminal work of Brown et al. (1975), however the authors rely on a higher order moment assumption to derive their asymptotic results. Moreover, their test statistic is constructed in terms of sums of squares of the observed data although it is well known from regression and linear time series, that diagnostics based on residuals prove often to be better tools. To that extent, Kulperger and Yu (2005) proposed a CUSUM test based on sum of squares of the residuals of a GARCH model. As an alternative to tests focusing on identifying structural breaks *in sample*, Berkes et al. (2004) introduce a sequential testing methodology to identify change points in a flow of data. These results have not yet been extended to the ARCH(∞) framework.

While Chapter 4 does not focus on memory properties of conditional variance, it contributes to the fast growing literature on *Sustainable Finance*. In particular, it gives arguments in favor of the existence of a *Green Minus Brown* premium. However, such long-short portfolios might be difficult to implement (see for example Brière and Szafarz (2017) or Brière et al. (2019)) and many investors have investment constraints precluding short selling. Although a long only position in a *Green* portfolio might still yield a premium, this portfolio may not be hedged against carbon-risk related shocks. Recently Engle et al. (2020) proposed a methodology to construct a portfolio that is not linearly exposed to climate news shock. However, there might be volatility spillovers from *Brown* stock to *Green* stocks that may not be captured by linear dependence. Studies of volatilities in the *Sustainable Finance* literature are regrettably scarce as the main focus has been the cross-section of financial assets. Nevertheless, Campos-Martins and Engle (2022) and Campos-Martins and Hendry (2020) considered volatility shocks in a climate finance framework. Their work however aims at identifying a global volatility shock stemming from climate risk rather than studying *Brown* stocks specific shocks and their contagion to *Green* stocks. Building upon the two portfolios constructed in the application of Chapter 4, one could use a multivariate GARCH model to study these spillovers. More precisely, consider the *unrestricted* ECCC (uECCC) GARCH model of Conrad and Karanasos (2010) relaxing the hypothesis of positivity of the GARCH spillovers. This model is particularly suited for the

analysis of volatility contagion as spillovers can easily be derived from the off-diagonal elements of the matrices \mathbf{A}_i and \mathbf{B}_j in equation (3.4). For example, from a bivariate uECCC-GARCH(1,1) model, we can obtain the following equation for the conditional variance vector

$$\begin{cases} g_{B,t} = \omega_1 + \overbrace{a_{11}r_{B,t-1}^2 + b_{11}g_{B,t-1}}^{\text{Brown own effect}} + \overbrace{a_{12}r_{G,t-1}^2 + b_{12}g_{G,t-1}}^{\text{Green spillover}} \\ g_{G,t} = \omega_2 + \overbrace{a_{22}r_{G,t-1}^2 + b_{22}g_{G,t-1}}^{\text{Green own effect}} + \overbrace{a_{21}r_{B,t-1}^2 + b_{21}g_{B,t-1}}^{\text{Brown spillover}} \end{cases}$$

where $r_{B,t}$ and $r_{G,t}$ are returns of the Brown and Green portfolios and $g_{B,t}$ and $g_{G,t}$ their respective conditional variances. Studying the empirical properties of these spillovers might help building portfolio less exposed to carbon-related risks.

Bibliography

- Al-Najjar, N. I. (1995). Decomposition and characterization of risk with a continuum of random variables. *Econometrica*, 63(5):1195–1224.
- Al-Najjar, N. I. (1999). Decomposition and characterization of risk with a continuum of random variables: Corrigendum. *Econometrica*, 67(4):919–920.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., and Wu, G. (2006). Realized beta: Persistence and predictability. *Advances in Econometrics*, 20:1–40.
- Andreou, E. and Ghysels, E. (2002). Detecting multiple breaks in financial market volatility dynamics. *Journal of Applied Econometrics*, 17(5):579–600.
- Andrews, D. W. (2001). Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica*, 69(3):683–734.
- Ang, A. and Kristensen, D. (2012). Testing conditional factor models. *Journal of Financial Economics*, 106(1):132–156.
- Ardia, D., Bluteau, K., Boudt, K., and Inghelbrecht, K. (2020). Climate change concerns and the performance of green versus brown stocks. *National Bank of Belgium, Working Paper Research*, (395).
- Aue, A., Hörmann, S., Horváth, L., and Reimherr, M. (2009). Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics*, 37(6B):4046–4087.
- Baillie, R. T., Bollerslev, T., and Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74(1):3–30.
- Bardet, J.-M. and Wintenberger, O. (2009). Asymptotic normality of the quasi-maximum likelihood estimator for multidimensional causal processes. *The Annals of Statistics*, 37(5B):2730–2759.
- Barndorff-Nielsen, O. E., Kinnebroek, S., and Shephard, N. (2010). Measuring downside risk: realised semivariance. In T. Bollerslev, J. R. and Watson, M., editors, *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*, pages 117–136. Oxford University Press.

- Barndorff-Nielsen, O. E. and Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2):253–280.
- Barndorff-Nielsen, O. E. and Shephard, N. (2004). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica*, 72(3):885–925.
- Bauwens, L., Laurent, S., and Rombouts, J. V. K. (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics*, 21(1):79–109.
- Bekaert, G., Harvey, C., and Ng, A. (2005). Market integration and contagion. *The Journal of Business*, 78(1):39–69.
- Bekaert, G. and Harvey, C. R. (1995). Time-varying world market integration. *The Journal of Finance*, 50(2):403–444.
- Bekaert, G. and Wu, G. (2000). Asymmetric volatility and risk in equity markets. *The Review of Financial Studies*, 13(1):1–42.
- Berkes, I., Gombay, E., Horváth, L., and Kokoszka, P. (2004). Sequential change-point detection in GARCH(p, q) models. *Econometric Theory*, 20(6):1140–1167.
- Berkes, I., Horváth, L., and Kokoszka, P. (2003a). Asymptotics for GARCH squared residual correlations. *Econometric Theory*, 19(4):515–540.
- Berkes, I., Horváth, L., and Kokoszka, P. (2003b). GARCH processes: Structure and estimation. *Bernoulli*, 9(2):201–227.
- Bernardi, M. and Catania, L. (2018). The model confidence set package for R. *International Journal of Computational Economics and Econometrics*, 8(2):144–158.
- Beutner, E., Heinemann, A., and Smeekes, S. (2018). A residual bootstrap for conditional value-at-risk. *arXiv preprint arXiv:1808.09125*.
- Billingsley, P. (1961). The Lindeberg-Lévy theorem for martingales. *Proceedings of the American Mathematical Society*, 12(5):788–792.
- Billingsley, P. (1995). *Probability and measure*. John Wiley & Sons, 3rd edition.
- Black, F. (1976). Studies of stock price volatility changes. In *Proceedings from the American Statistical Association, Business and Economic Statistics Section*, pages 245–258. American Statistical Association.
- Blasques, F., Francq, C., and Laurent, S. (2022a). Autoregressive conditional betas. *Working paper*.
- Blasques, F., Francq, C., and Laurent, S. (2022b). Quasi score-driven models. *Journal of Econometrics, Forthcoming*.

- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.
- Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate Generalized ARCH model. *The Review of Economics and Statistics*, 72(3):498–505.
- Bollerslev, T., Engle, R. F., and Wooldridge, J. M. (1988). A capital asset pricing model with time-varying covariances. *Journal of Political Economy*, 96(1):116–131.
- Bollerslev, T., Li, J., Patton, A. J., and Quaedvlieg, R. (2020a). Realized semicovariances. *Econometrica*, 88(4):1515–1551.
- Bollerslev, T. and Mikkelsen, H. O. (1996). Modeling and pricing long memory in stock market volatility. *Journal of Econometrics*, 73(1):151–184.
- Bollerslev, T., Patton, A. J., and Quaedvlieg, R. (2020b). Multivariate leverage effects and realized semicovariance GARCH models. *Journal of Econometrics*, 217(2):411–430.
- Bollerslev, T., Patton, A. J., and Quaedvlieg, R. (2022). Realized semibetas: Disentangling “good” and “bad” downside risks. *Journal of Financial Economics*, 144(1):227–246.
- Bolton, P. and Kacperczyk, M. (2020). Do investors care about carbon risk? Technical report, National Bureau of Economic Research.
- Boudt, K., Laurent, S., Lunde, A., Quaedvlieg, R., and Sauri, O. (2017). Positive semidefinite integrated covariance estimation, factorizations and asynchronicity. *Journal of Econometrics*, 196(2):347–367.
- Box, G. E. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, i. effect of inequality of variance in the one-way classification. *The Annals of Mathematical Statistics*, pages 290–302.
- Box, G. E. and Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65(332):1509–1526.
- Brière, M., Lehalle, C.-A., Nefedova, T., and Raboun, A. (2019). Stock market liquidity and the trading costs of asset pricing anomalies. *Université Paris-Dauphine Research Paper:3380239*.
- Brière, M. and Szafarz, A. (2017). Factor investing: The rocky road from long-only to long-short. In Jurczenko, E., editor, *Factor Investing*, pages 25–45. Elsevier.
- Brockwell, P. J. and Davis, R. A. (2009). *Time series: theory and methods*. Springer science & business media.
- Brown, R. L., Durbin, J., and Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society. Series B (Methodological)*, 37(2):149–192.

- Campbell, J. Y. and Hentschel, L. (1992). No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of Financial Economics*, 31(3):281–318.
- Campos-Martins, S. and Engle, R. F. (2022). What are the events that shake our world? measuring and hedging global COVOL. *Journal of Financial Economics*, *Forthcoming*.
- Campos-Martins, S. and Hendry, D. F. (2020). Identifying climate global common volatility shocks. Technical report, Working Paper, Nuffield College, Oxford University.
- Carbon, M. and Francq, C. (2011). Portmanteau goodness-of-fit test for asymmetric power GARCH models. *Austrian Journal of Statistics*, 40(1&2):55–64.
- Cavaliere, G., Nielsen, H. B., Pedersen, R. S., and Rahbek, A. (2022). Bootstrap inference on the boundary of the parameter space, with application to conditional volatility models. *Journal of Econometrics*, 227(1):241–263.
- Chaieb, I., Langlois, H., and Scaillet, O. (2021). Factors and risk premia in individual international stock returns. *Journal of Financial Economics*, 141(2):669–692.
- Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, 51(5):1281–1304.
- Christie, A. A. (1982). The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of Financial Economics*, 10(4):407–432.
- Comte, F. and Lieberman, O. (2003). Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis*, 84(1):61–84.
- Conrad, C. and Karanasos, M. (2010). Negative volatility spillovers in the unrestricted ECCC-GARCH model. *Econometric Theory*, 26(3):838–862.
- Conrad, C., Karanasos, M., and Zeng, N. (2011). Multivariate fractionally integrated APARCH modeling of stock market volatility: A multi-country study. *Journal of Empirical Finance*, 18(1):147–159.
- Corsi, F. (2009). A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics*, 7(2):174–196.
- Creal, D., Koopman, S. J., and Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28(5):777–795.
- Darolles, S., Francq, C., and Laurent, S. (2018). Asymptotics of Cholesky GARCH models and time-varying conditional betas. *Journal of Econometrics*, 204(2):223–247.
- De Angelis, T., Tankov, P., and Zerbib, O. D. (2022). Climate impact investing. *Management Science*, *Forthcoming*.
- Diebold, F. X. and Inoue, A. (2001). Long memory and regime switching. *Journal of Econometrics*, 105(1):131–159.

- Ding, Z. and Granger, C. W. (1996). Modeling volatility persistence of speculative returns: A new approach. *Journal of Econometrics*, 73(1):185–215.
- Ding, Z., Granger, C. W., and Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1(1):83–106.
- Douc, R., Roueff, F., and Soulier, P. (2008). On the existence of some ARCH(∞) processes. *Stochastic Processes and their Applications*, 118(5):755–761.
- Doukhan, P., Teyssiere, G., and Winant, P. (2006). A LARCH(∞) vector valued process. In Bertail, P., Doukhan, P., and Soulier, P., editors, *Dependence in probability and statistics*, pages 245–258. Springer.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007.
- Engle, R. F. (2002). Dynamic conditional correlation: A simple class of multivariate GARCH models. *Journal of Business & Economic Statistics*, 20(3):339–350.
- Engle, R. F. (2016). Dynamic conditional beta. *Journal of Financial Econometrics*, 14(4):643–667.
- Engle, R. F., Giglio, S., Kelly, B., Lee, H., and Stroebel, J. (2020). Hedging climate change news. *The Review of Financial Studies*, 33(3):1184–1216.
- Engle, R. F., Granger, C., and Kraft, D. (1984). Combining competing forecasts of inflation using a bivariate arch model. *Journal of Economic Dynamics and Control*, 8(2):151–165.
- Engle, R. F., Hendry, D. F., and Richard, J.-F. (1983). Exogeneity. *Econometrica*, 51(2):277–304.
- Engle, R. F. and Kroner, K. F. (1995). Multivariate simultaneous Generalized ARCH. *Econometric Theory*, 11(1):122–150.
- Engle, R. F. and Patton, A. J. (2001). What good is a volatility model? *Quantitative Finance*, 1(2):237–245.
- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2):383–417.
- Fama, E. F. (1991). Efficient capital markets: II. *The Journal of Finance*, 46(5):1575–1617.
- Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.
- Fama, E. F. and French, K. R. (1997). Industry costs of equity. *Journal of Financial Economics*, 43(2):153–193.
- Fama, E. F. and MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81(3):607–636.

- Fermanian, J.-D. and Malongo, H. (2017). On the stationarity of dynamic conditional correlation models. *Econometric Theory*, 33(3):636–663.
- Ferson, W. E. and Harvey, C. R. (1991). The variation of economic risk premiums. *Journal of Political Economy*, 99(2):385–415.
- Francq, C., Horváth, L., and Zakoïan, J.-M. (2016). Variance targeting estimation of multivariate GARCH models. *Journal of Financial Econometrics*, 14(2):353–382.
- Francq, C. and Thieu, L. Q. (2019). QML inference for volatility models with covariates. *Econometric Theory*, 35(1):37–72.
- Francq, C. and Zakoïan, J.-M. (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli*, 10(4):605–637.
- Francq, C. and Zakoïan, J.-M. (2007). Quasi-maximum likelihood estimation in GARCH processes when some coefficients are equal to zero. *Stochastic Processes and their Applications*, 117(9):1265–1284.
- Francq, C. and Zakoïan, J.-M. (2009). Testing the nullity of GARCH coefficients: correction of the standard tests and relative efficiency comparisons. *Journal of the American Statistical Association*, 104(485):313–324.
- Francq, C. and Zakoïan, J.-M. (2012). QML estimation of a class of multivariate asymmetric GARCH models. *Econometric Theory*, 28(1):179–206.
- Francq, C. and Zakoïan, J.-M. (2019). *GARCH models: structure, statistical inference and financial applications*. John Wiley & Sons, 2nd edition.
- Francq, C. and Zakoïan, J.-M. (2022). Testing the existence of moments for GARCH processes. *Journal of Econometrics*, 227(1):47–64.
- French, K. R., Schwert, G. W., and Stambaugh, R. F. (1987). Expected stock returns and volatility. *Journal of Financial Economics*, 19(1):3–29.
- Gagliardini, P., Ossola, E., and Scaillet, O. (2016). Time-varying risk premium in large cross-sectional equity data sets. *Econometrica*, 84(3):985–1046.
- Ghysels, E. (1998). On stable factor structures in the pricing of risk: do time-varying betas help or hurt? *The Journal of Finance*, 53(2):549–573.
- Ghysels, E., Santa-Clara, P., and Valkanov, R. (2006). Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics*, 131(1):59–95.
- Giacomini, R., Politis, D. N., and White, H. (2013). A warp-speed method for conducting monte carlo experiments involving bootstrap estimators. *Econometric Theory*, 29(3):567–589.
- Giraitis, L., Kokoszka, P., and Leipus, R. (2000). Stationary ARCH models: Dependence structure and central limit theorem. *Econometric Theory*, 16(1):3–22.

- Giraitis, L., Koul, H. L., and Surgailis, D. (2012). *Large sample inference for long memory processes*. World Scientific.
- Giraitis, L. and Robinson, P. M. (2001). Whittle estimation of ARCH models. *Econometric Theory*, 17(3):608–631.
- Giraitis, L., Surgailis, D., and Škarnulis, A. (2018). Stationary integrated ARCH(∞) and AR(∞) processes with finite variance. *Econometric Theory*, 34(6):1159–1179.
- Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5):1779–1801.
- Goldfeld, S. M. and Quandt, R. E. (1973). A Markov model for switching regressions. *Journal of Econometrics*, 1(1):3–15.
- González-Rivera, G., Lee, T.-H., and Mishra, S. (2004). Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood. *International Journal of Forecasting*, 20(4):629–645.
- Görgen, M., Jacob, A., Nerlinger, M., Riordan, R., Rohleder, M., and Wilkens, M. (2019). Carbon risk. *Available at SSRN 2930897*.
- Gourieroux, C. and Jasiak, J. (2001). Memory and infrequent breaks. *Economics Letters*, 70(1):29–41.
- Gouriéroux, C. and Monfort, A. (1995). *Statistics and Econometric Models*, volume 2 of *Themes in Modern Econometrics*. Cambridge University Press.
- Granger, C. W. J. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1(1):15–29.
- Grassi, S. and Violante, F. (2021). Asset Pricing Using Block-Cholesky GARCH and Time-Varying Betas. Working Papers 2021-05, Center for Research in Economics and Statistics.
- Hafner, C. M. and Preminger, A. (2017). On asymptotic theory for ARCH (∞) models. *Journal of Time Series Analysis*, 38(6):865–879.
- Hamadeh, T. and Zakoïan, J.-M. (2011). Asymptotic properties of LS and QML estimators for a class of nonlinear GARCH processes. *Journal of Statistical Planning and Inference*, 141(1):488–507.
- Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384.
- Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica*, 64(2):413–430.
- Hansen, P. R., Lunde, A., and Nason, J. M. (2011). The model confidence set. *Econometrica*, 79(2):453–497.

- Hansen, P. R., Lunde, A., and Voev, V. (2014). Realized beta GARCH: A multivariate GARCH model with realized measures of volatility. *Journal of Applied Econometrics*, 29(5):774–799.
- Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*. Econometric Society Monographs. Cambridge University Press.
- Harvey, A. C. and Chakravarty, T. (2008). Beta-t-(e) GARCH. *University of Cambridge Discussion Paper*, 08340.
- Hidalgo, J. and Zaffaroni, P. (2007). A goodness-of-fit test for ARCH(∞) models. *Journal of Econometrics*, 141(2):973–1013.
- Hillebrand, E. (2005). Neglecting parameter changes in GARCH models. *Journal of Econometrics*, 129(1):121–138.
- Hoepner, A., Masoni, P., and Kramer, B. (2019). Handbook of climate transition benchmarks, Paris-aligned benchmark and benchmarks’ ESG disclosure. Technical report.
- Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, 68(1):165–176.
- Imhof, J. P. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika*, 48(3/4):419–426.
- Jeantheau, T. (1998). Strong consistency of estimators for multivariate arch models. *Econometric Theory*, 14(1):70–86.
- Kazakevičius, V. and Leipus, R. (2002). On stationarity in the ARCH(∞) model. *Econometric Theory*, 18(1):1–16.
- Kempf, A. and Osthoff, P. (2007). The effect of socially responsible investing on portfolio performance. *European Financial Management*, 13(5):908–922.
- Kim, C.-J., Piger, J., and Startz, R. (2008). Estimation of Markov regime-switching regression models with endogenous switching. *Journal of Econometrics*, 143(2):263–273.
- Kokoszka, P. and Leipus, R. (2000). Change-point estimation in ARCH Models. *Bernoulli*, 6(3):513–539.
- Kreiss, J.-P., Paparoditis, E., and Politis, D. N. (2011). On the range of validity of the autoregressive sieve bootstrap. *The Annals of Statistics*, 39(4):2103–2130.
- Krueger, P., Sautner, Z., and Starks, L. T. (2020). The importance of climate risks for institutional investors. *The Review of Financial Studies*, 33(3):1067–1111.
- Kulperger, R. and Yu, H. (2005). High moment partial sum processes of residuals in GARCH models and their applications. *The Annals of Statistics*, 33(5):2395–2422.
- Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives*, 3(2).

- Lewellen, J. and Nagel, S. (2006). The conditional CAPM does not explain asset-pricing anomalies. *Journal of Financial Economics*, 82(2):289–314.
- Li, W. K. and Mak, T. (1994). On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity. *Journal of Time Series Analysis*, 15(6):627–636.
- Ling, S. and McAleer, M. (2003). Asymptotic theory for a Vector ARMA-GARCH model. *Econometric Theory*, 19(2):280–310.
- Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics*, 47(1):13–37.
- Linton, O. and Mammen, E. (2005). Estimating semiparametric ARCH(∞) models by kernel smoothing methods. *Econometrica*, 73(3):771–836.
- Lo, A. W. and MacKinlay, A. C. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *The Review of Financial Studies*, 1(1):41–66.
- Luati, A. and Proietti, T. (2015). The generalised autocovariance function. *Journal of Econometrics*, 186(1):245–257.
- Luati, A., Proietti, T., and Reale, M. (2012). The variance profile. *Journal of the American Statistical Association*, 107(498):607–621.
- Meyer, C. (2000). *Matrix Analysis and Applied Linear Algebra*. Other Titles in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM).
- Mikosch, T. and Stărică, C. (2004). Nonstationarities in Financial Time Series, the Long-Range Dependence, and the IGARCH Effects. *The Review of Economics and Statistics*, 86(1):378–390.
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2):347–370.
- Pastor, L., Stambaugh, R. F., and Taylor, L. A. (2021a). Dissecting green returns. Technical report, National Bureau of Economic Research.
- Pastor, L., Stambaugh, R. F., and Taylor, L. A. (2021b). Sustainable investing in equilibrium. *Journal of Financial Economics*, 142(2):550–571.
- Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika*, 86(3):677–690.
- Robinson, P. M. (1991). Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *Journal of Econometrics*, 47(1):67–84.
- Robinson, P. M. and Zaffaroni, P. (2006). Pseudo-maximum likelihood estimation of ARCH(∞) models. *The Annals of Statistics*, 34(3):1049–1074.

- Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3):341–360.
- Royer, J. (2022). Conditional asymmetry in Power ARCH(∞) models. *Journal of Econometrics*, *Forthcoming*.
- Scheffe, H. (1959). *The analysis of variance*. John Wiley & Sons.
- Schwert, G. W. (1989). Why does stock market volatility change over time? *The Journal of Finance*, 44(5):1115–1153.
- Shanken, J. (1990). Intertemporal asset pricing: An empirical investigation. *Journal of Econometrics*, 45(1-2):99–120.
- Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3):425–442.
- Shimizu, K. (2013). The bootstrap does not always work for heteroscedastic models. *Statistics and Risk Modeling*, 30(3):189–204.
- Statman, M. (2006). Socially responsible indexes. *The Journal of Portfolio Management*, 32(3):100–109.
- TCFD (2017). Implementing the recommendations of the task force on climate- related financial disclosure. Technical report, TCFD.
- Tse, Y. K. and Tsui, A. K. C. (2002). A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations. *Journal of Business & Economic Statistics*, 20(3):351–362.
- Van Dijk, M. A. (2011). Is Size dead? A review of the size effect in equity returns. *Journal of Banking & Finance*, 35(12):3263–3274.
- Yule, G. U. (1927). On a method of investigating periodicities disturbed series, with special reference to Wolfer’s sunspot numbers. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 226(636-646):267–298.
- Zaffaroni, P. (2004). Stationarity and memory of ARCH(∞) models. *Econometric Theory*, 20(1):147–160.
- Zaffaroni, P. (2009). Whittle estimation of EGARCH and other exponential volatility models. *Journal of Econometrics*, 151(2):190–200.
- Zakoïan, J.-M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, 18(5):931–955.
- Zarantonello, E. H. (1971). Projections on convex sets in Hilbert space and spectral theory. In *Contributions to nonlinear functional analysis*, pages 237–424. Elsevier.

Titre : Processus ARCH d'ordre infini, Bêtas dynamiques et applications financières

Mots clés : Hétéroscédasticité, Modèles à facteurs dynamiques, Modèles GAS, Quasi maximum de vraisemblance, Valorisation d'actifs

Résumé : La modélisation des séries temporelles financières est rendue difficile par la présence de faits stylisés. Ces propriétés statistiques empiriques rendent nécessaires l'utilisation de modèles non-linéaires hétéroscédastiques. Les modèles ARCH d'ordre infini ont été introduits afin de permettre une modélisation plus fine de ces faits stylisés, et en particulier le phénomène de forte persistance des chocs de volatilités. Nous présentons de nouvelles extensions à ces modèles flexibles et étudions leur inférence. En premier lieu, nous considérons un modèle ARCH(∞) asymétrique. Nous démontrons l'existence d'une solution stationnaire et nous établissons les propriétés asymptotiques de l'estimateur de quasi-maximum de vraisemblance dans le cadre de ce modèle. En particulier, nous permettons au paramètre d'être sur le bord de l'espace des paramètres, empêchant la normalité asymptotique. De plus, nous introduisons un test portmanteau vérifiant l'adéquation du modèle aux données ainsi qu'un test statistique pour la présence de mémoire

et d'asymétrie. Dans un second temps, nous nous intéressons à la modélisation des coefficients d'une régression linéaire conditionnelle. Les modèles à facteurs linéaires sont au cœur de nombreux modèles financiers et souvent, les coefficients de régression sont supposés constants. Nous proposons un modèle permettant la mesure de la dynamique de ces coefficients bêtas dans le cadre des modèles ARCH(∞) multivariés. En particulier, nous permettons l'ajout de variables exogènes dans la dynamique des bêtas conditionnels et discutons de potentiels candidats. Nous établissons les conditions d'existence d'une solution stationnaire et discutons l'existence de moments de cette dernière. Enfin, nous considérons un exercice de valorisation d'actifs basé sur des bêtas dynamiques. A cet effet, nous étendons les résultats de tests statistiques dans le cas de *score-driven* bêtas et proposons une procédure de rééchantillonnage. Nous introduisons une méthode d'estimation en deux étapes pour mesurer les primes de risque dynamiques sous-jacentes au modèle d'évaluation.

Title : Infinite ARCH processes, dynamic betas, and financial applications

Keywords : Asset Pricing, Conditional heteroskedasticity, Dynamic Factor Models, QML asymptotic results, Score-Driven Models

Abstract : The modeling of financial time series is made difficult by the presence of stylized facts. These empirical statistical properties led to the development of heteroskedastic nonlinear models. Infinite ARCH specifications have been introduced to allow finer modeling of these stylized facts, and in particular the phenomenon of strong persistence of volatility shocks. We present new extensions to these flexible models and study their inference. First, we consider an asymmetric ARCH(∞) model. We prove the existence of a stationary solution and establish the asymptotic properties of the quasi-maximum likelihood estimator in this framework. In particular, we allow the parameter to lie on the boundary of the parameter space, precluding asymptotic normality. Moreover, we introduce a portmanteau test assessing the goodness-of-fit of the model on data. We also propose a test for the presence of memory and asymmetry. In a second

time, we consider the modeling of the coefficients of a conditional linear regression. Linear factor models are key to many financial models and regression coefficients are often wrongfully assumed constant. We propose a model allowing for dynamic beta coefficients within the framework of multivariate ARCH(∞) models. In particular, we allow the addition of exogenous variables in the dynamics of conditional betas and discuss potential candidates. We establish the conditions of existence of a stationary solution and discuss the existence of its moments. Finally, we consider an asset pricing exercise based on dynamic betas. To this end, we extend the results of statistical tests in the case of *score-driven* betas and propose a bootstrap procedure. Additionally, we introduce a two-step estimation method to measure the dynamic risk premia underlying the asset pricing model.