



HAL
open science

Inférence de modèles conditionnellement hétéroscédastiques avec variables exogènes

Le Quyen Thieu

► **To cite this version:**

Le Quyen Thieu. Inférence de modèles conditionnellement hétéroscédastiques avec variables exogènes. Statistiques [math.ST]. Université Pierre et Marie Curie - Paris VI, 2016. Français. NNT : 2016PA066260 . tel-01527888

HAL Id: tel-01527888

<https://theses.hal.science/tel-01527888>

Submitted on 26 May 2017

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.



Université Pierre et Marie Curie

Ecole doctorale 386

Sciences mathématiques de Paris Centre

Laboratoire de Statistiques Théoriques et Appliquées

**Inférence de modèles conditionnellement
hétéroscédastiques avec variables exogènes.**

Par Le Quyen THIEU

Thèse de Doctorat de Mathématiques

Spécialité : Statistiques

Dirigée par Denis BOSQ et Christian FRANCO

Devant un jury composé de :

M. BIAU Gérard	Professeur, Université Paris 6	Président
Mme. BILLIO Monica	Professeur, Université Ca' Foscari Venice	Examineur
M. BOSQ Denis	Professeur, Université Paris 6	Directeur
M. FRANCO Christian	Professeur, Université Lille 3	Co-Directeur
Mme. PHILIPPE Anne	Professeur, Université de Nantes	Rapporteur
M. ROMBOUTS Jeroen	Professeur, ESSEC Business School	Examineur

Thèse préparée au
Laboratoire de Statistique
Théorique et Appliquée
Bâtiments 15-25 et 15-16, 2^{ème} étage
4, Place Jussieu
75005 Paris

UPMC
École Doctorale de Sciences
Mathématiques de Paris Centre
4 place Jussieu
75252 Paris Cedex 05
Boite courrier 290

Remerciements

Ces trois années de thèse ont été pour moi l'occasion de rencontrer et côtoyer des personnes qui ont compté dans l'aboutissement de ce projet. Les quelques lignes suivantes me permettent de leur exprimer mes remerciements.

Je tiens tout d'abord à remercier chaleureusement le Professeur Denis Bosq, mon directeur de thèse, pour m'avoir donné l'occasion de poursuivre mes études de doctorat au Laboratoire de Statistique Théorique et Appliquée (LSTA). Je tiens à exprimer ma profonde gratitude au Professeur Christian Francq qui est mon co-directeur de thèse. Vos qualités scientifiques et humaines, votre grande disponibilité, votre expérience et votre méthode de travail m'ont permis d'apprendre beaucoup à vos côtés et m'ont permis de mener à bien cette thèse.

Je tiens à remercier la Professeure Anne Philippe et le Professeur Anders Rahbek qui ont accepté d'être les deux rapporteurs de cette thèse. Je suis très honorée par leur présence et les remercie du temps qu'ils ont consacré à l'analyse du travail présenté ici. De plus, je manifeste une grande gratitude aux professeurs Gérard Biau, Monica Billio et Jeroen Rombouts pour leur participation au jury de cette thèse.

Un grand merci également pour les membres de l'administration du laboratoire LSTA, Madame Louise Lamart pour ses grandes qualités professionnelles et humaines, Madame Corine Van Vliberghe pour sa gentillesse.

Trois années de thèse, ce sont aussi trois années au cours desquelles j'ai eu l'occasion de faire la connaissance de collègues qui m'ont beaucoup appris humainement et intellectuellement. Je pense à Assia, Baptise, Cécile, Erwan, Lucie, Matthieu, Mokhtar, Nazih, Roxane, Thibaut.

Un merci particulier à madame Christiane Francq pour votre gentillesse, votre accueil si sympathique et attentionné pendant les conférences à Lille. Je vous remercie sincèrement d'avoir relu la partie en français de ma thèse.

Vient maintenant le moment de remercier ma famille qui a joué un rôle tellement essentiel durant ces trois années. Tout d'abord je souhaiterais remercier mes parents pour leur affection

et leur soutien moral. J'espère qu'au moment où je finis cette thèse, de là où vous êtes, vous éprouvez de la fierté. Je voudrais également remercier ma petite princesse, Cún qui n'a pas eu peur de partir avec maman à l'autre bout du monde où les gens parlent des langues "bizarres", où tout semblait totalement étrange pour elle. Grâce à elle, j'ai passé de bons moments et oublié la fatigue, le stress du travail. Pour finir je remercie également avec une mention spéciale celui qui occupe une place particulière dans ma vie : Mikaël. Je n'ai pas de mots assez forts pour t'exprimer ma gratitude. Tu m'as toujours soutenue au quotidien, tu t'es bien occupé de moi tout le temps. Tu as toujours été à mes côtés malgré les longues soirées passées à m'écouter parler des modèles "hétéroskédastiques". En écrivant ces mots, je mesure plus que jamais l'équilibre que tu m'apportes. Sache que l'achèvement de cette thèse est aussi ton succès !

Résumé

Cette thèse de doctorat a pour objectif principal d'étudier certaines propriétés probabilistes et statistiques de modèles de volatilité contenant des variables explicatives exogènes. Elle comporte deux parties.

Dans une première partie, nous étudions le comportement asymptotique de l'estimation du quasi-maximum de vraisemblance (QMV) pour la classe polyvalente des modèles PGARCH semi-forts augmentés avec des covariables. Les hypothèses principales sur les variables exogènes sont la stationnarité et la non-colinéarité avec les autres variables explicatives de la volatilité. Pour la distribution asymptotique du QMV, nous étudions quatre situations différentes correspondant à des modèles forts ou semi-forts, et des paramètres à l'intérieur ou au bord de l'espace des paramètres. Nous montrons la normalité asymptotique du QMV sans imposer aucune condition de moment sur le processus observé lorsque le paramètre GARCH-X appartient à l'intérieur de l'espace des paramètres. Par contre, quand un ou plusieurs coefficients sont égaux à zéro, les conditions de moment d'ordre 4 sont requises pour que la matrice d'information soit finie et la loi asymptotique est alors la projection d'une loi normale sur un cône convexe. Comme la vraie valeur du paramètre n'est pas contrainte à appartenir à l'intérieur de l'espace des paramètres, nous proposons des tests pour déterminer l'ordre du modèle et vérifier la signification des variables exogènes.

La deuxième partie est consacrée à l'étude de l'influence des variables exogènes sur les matrices de covariance conditionnelle de rendements d'actifs. Plus précisément, nous considérons des modèles BEKK avec variables exogènes. Les paramètres sont estimés par deux méthodes qui s'appellent l'estimation par ciblage de la variance et l'estimation équation par équation. Ces deux méthodes nous permettent de réduire la complexité numérique liée à l'estimation d'un nombre élevé des paramètres des modèles GARCH multivariés, en particulier, en présence de variables exogènes. La consistance ainsi que la loi limite de ces estimateurs sont établies pour des hypothèses relativement peu restrictives. En particulier, les innovations sont supposées être une différence de martingales au lieu d'être iid.

Nos résultats sont illustrés par des expériences de Monte Carlo et des applications sur séries réelles.

Mots-clefs

Modèles GARCH-X, variables exogènes, covariables, modèles conditionnellement hétéroscédastiques.

Abstract

This PhD Dissertation is dedicated to the study of probabilistic and statistical properties of volatility models augmented with exogenous variables. It consists of two parts which are summarized below. In the first part of this work, we study asymptotic behavior of the QMLE for the versatile class of the semi-strong PGARCH models augmented with exogenous variables. The main assumptions on the exogenous variables are the stationarity and the non-colinearity with the other explanatory variables of the volatility. For the asymptotic distribution of the QMLE, we investigated four different situations corresponding to strong or semi-strong models, and to parameters inside or at the boundary of the parameter space. When the GARCH-X parameter belongs to the interior of the parameter space, the asymptotic distribution of the QMLE is normal, whereas it is the projection of a normal distribution on a convex cone when one or several coefficients are equal to zero. For models with positive GARCH coefficients, the asymptotic distribution is obtained under very mild conditions, in particular, without any moment condition on the observed process. When the GARCH parameter stands at the boundary, fourth-order moment conditions are required for the information matrix to be finite. Our asymptotic results are obtained under conditions that are only marginally stronger than these optimal moment conditions, which extends and improves the results that existed for GARCH models without covariables.

The second part is devoted to studying the influence of exogenous variables on the conditional covariance matrix of asset returns. Specifically, we consider BEKK models augmented with exogenous variables. The parameters are estimated by two methods which are called the variance targeting estimation and equation by equation estimation. Both methods allow us to reduce the curse of dimensionality which appears when modeling a conditional covariance matrix, particularly in the presence of exogenous variables. The consistency and the asymptotic distribution of these estimators are established under mild assumptions. In particular, the innovation is assumed to be a martingale difference instead of iid.

Our results are illustrated by Monte Carlo experiences and the applications on real series.

Keywords

GARCH-X models, exogenous variables, covariates, conditional heteroskedasticity models

Table des matières

Résumé	5
Abstract	6
1 Introduction	11
1.1 Variables exogènes pour modéliser la volatilité des series financières	11
1.1.1 Description et contribution de cette thèse	14
1.2 Résultats du chapitre 2	15
1.2.1 Convergence forte du QMV	16
1.2.2 Loi asymptotique du QMV	19
1.2.3 Test de pertinence de covariables	22
1.3 Résultats du chapitre 3	24
1.3.1 Modèle et l'estimation par ciblage de la variance	24
1.3.2 Consistance forte et Normalité asymptotique	26
1.4 Résultats du chapitre 4	31
1.4.1 Modèle et estimation équation par équation	31
1.4.2 L'inférence de l'estimation EbE	34
2 QML inference for volatility models with covariates	43
2.1 Introduction	43
2.1.1 The model	45
2.1.2 The objectives	45
2.2 Main results	47
2.2.1 Strict stationarity	47
2.2.2 Strong consistency of the QMLE	49
2.2.3 Asymptotic distribution of the QMLE	51
2.2.4 Testing the relevance of the covariates	57
2.3 Numerical illustrations	58

2.3.1	Simulation experiments	58
2.3.2	SP500 with realized range, volume and other indices	61
2.3.3	US stocks with realized volatility	64
2.4	Conclusion	65
2.5	Proofs and technical lemmas	66
3	Variance targeting estimation of the BEKK-X model	81
3.1	Introduction	81
3.2	The model and variance targeting estimation	83
3.2.1	The model	83
3.2.2	Variance targeting estimation	84
3.3	VTE inference	86
3.3.1	Consistency and asymptotic normality	86
3.3.2	Estimating the asymptotic covariance matrix	89
3.4	Numerical illustration	90
3.4.1	A Monte Carlo experiment	90
3.4.2	An application to stocks US	91
3.5	Conclusion	93
3.6	Proofs	94
4	Equation by equation estimation of the semi-diagonal BEKK model with covariates	112
4.1	Introduction	112
4.2	The model and EbE estimation	114
4.2.1	The model	114
4.2.2	Equation-by-equation estimation of parameters	116
4.3	EbE estimation inference	117
4.4	Numerical illustrations	120
4.5	Conclusion	123
4.6	Proofs	125
	Bibliographie	144

Table des figures

2.1	Return, range and volume of the SP500 index from January 4, 1985 to August 26, 2011 (October 19, 1987 corresponds to the black Monday, and October 27, 1997 is the date of a mini-crash that was caused by an economic crisis in Asia).	63
2.2	Autocorrelations of the SP500 returns and of the squared residuals of model (2.28), with the significance bands obtained from the standard Bartlett's formula (in dotted lines) and the generalized Bartlett's formula (in full lines).	64
4.1	Kernel density estimator (in full line) of the distribution of the EbEE errors for the estimation of the parameters involved $vech(\mathbf{\Omega})$.	121
4.2	Kernel density estimator (in full line) of the distribution of the EbEE errors for the estimation of the parameters involved \mathbf{A} .	123
4.3	Kernel density estimator (in full line) of the distribution of the EbEE errors for the estimation of the parameters involved $diag(\mathbf{B})$.	124
4.4	Kernel density estimator (in full line) of the distribution of the EbEE errors for the estimation of the parameters involved \mathbf{C} .	125
4.5	Boxplots of 500 estimation errors for the EbEE and VTE.	127

Liste des tableaux

2.1	Sampling distribution of the QMLE of ϑ_0 over 1000 replications for the TARCH-X(1,1) model in Case A	59
2.2	Sampling distribution of the QMLE of ϑ_0 over 1000 replications for the TARCH-X(1,1) model in Case B	60
2.3	Sampling distribution of the QMLE of ϑ_0 over 1000 replications for the TARCH-X(1,1) model in Case C	60
2.4	Sampling distribution of the QMLE of ϑ_0 over 1000 replications for the TARCH-X(1,1) model in Case D	61
2.5	Relative frequencies (in %) of rejection of the assumptions that the first and second lagged values of the exogenous variable do not appear in the conditional variance	62
2.6	APARCH-X(1,1) models (2.29) fitted by QMLE on daily returns of US stock with two lagged values of realized volatilities as covariates. The estimated standard deviations are displayed into parentheses. For the estimated values of π_1 and π_2 , one star (*) means a p -value $p \in [0.01, 0.05)$ for testing the nullity of the coefficient, two stars (**) means $p \in [0.001, 0.01)$, and three stars (***) means $p < 0.001$. The last column gives the selected value of the power δ	66
2.6	(continued)	67
3.1	Sampling distribution of the VTE of ϑ_0 over 500 replications for the BEKK(1, 1)-X model	92
4.1	Sampling distribution of the EbEE of ϑ_0 over 500 replications for the semi-diagonal BEKK(1, 1)-X model	122
4.2	Sampling distribution of the EbEE and VTE of ϑ_0 over 500 replications for the BEKK-X(1,1) model	126

Chapitre 1

Introduction

1.1 Variables exogènes pour modéliser la volatilité des series financières

Relier les variations des prix des actifs et la volatilité des marchés aux variables économiques fondamentales est d'un intérêt majeur pour les activités financières et économiques comme la gestion des risques, l'évaluation des actifs, et le choix de portefeuilles. Il est bien connu que les marchés financiers et les investisseurs réagissent rapidement aux nouvelles importantes, aux crises économiques, guerres, troubles politiques ou catastrophes naturelles. Dans de telles situations, les prix des actifs financiers peuvent subir des fluctuations substantielles. Cela signifie que la variance conditionnelle d'une série chronologique de prix n'est généralement pas constante dans le temps. Ce phénomène est connu sous le nom d'hétéroscédasticité conditionnelle. Depuis les trente dernières années, plusieurs modèles ont été élaborés pour prendre en considération la dynamique temporelle de la volatilité. Les premiers modèles économétriques pour la volatilité des séries financières sont les modèles autorégressifs à volatilité conditionnelle hétéroscédastique (ARCH) proposés par [Engle \(1982\)](#) et leur généralisation (GARCH) développée par [Bollerslev \(1986\)](#). Ces modèles ont le principal avantage de prendre en compte la dynamique de la volatilité et également la leptokurticité des rendements. Ils sont un outil commode pour l'industrie financière. Depuis l'apparition du modèle (G)ARCH, plusieurs développements ont été introduits, pour améliorer la prise en compte des faits stylisés des rendements financiers (queues de distribution épaisses, autocorrélation des carrés des variations de prix, regroupement des fortes variations de prix, asymétrie, effets de levier...). Des extensions importantes du modèle (G)ARCH, comme le modèle GARCH

exponentiel (EGARCH) de [Nelson \(1991\)](#), le modèle GARCH en puissance asymétrique de [Ding et al. \(1993\)](#), le modèle GJR-GARCH de [Glosten et al. \(1993\)](#), la classe des modèles GARCH à seuil (TGARCH) de [Zakoïan \(1994\)](#), le modèle GARCH fractionnellement intégré (FIGARCH) de [Baillie \(1996\)](#)... ont été par la suite proposés, afin de mieux répondre aux spécificités des rendements financiers. Les améliorations proposées dans tous ces modèles se font en changeant l'équation de la volatilité conditionnelle, mais en gardant les mêmes variables.

Les modèles GARCH que nous avons mentionnés ci-dessus sont des modèles univariés. Ils permettent de décrire et de prévoir le comportement de la volatilité d'une seule série monétaire. Lorsque l'on dispose de plusieurs séries présentant des dépendances temporelles ou instantanées, il est utile de les étudier conjointement, de cerner les interdépendances entre les marchés, et d'analyser les mécanismes de transmission de la volatilité d'un marché à un autre, surtout en période de turbulences financières. Or, les modèles univariés ne tiennent pas compte de la corrélation entre les actifs. Pour pallier à cela, des classes de modèles GARCH multivariés (MGARCH) ont été développées afin de capter les liens dynamiques entre les rendements des actifs. Les premières formulations conçues à partir de la fin des années 80 ont cherché à généraliser directement le modèle GARCH univarié. Nous y trouvons les modèles VEC de [Bollerslev and Wooldridge \(1988\)](#), le modèle BEKK (Baba, Engle, Kraft et Kroner) de [Engle and Kroner \(1995\)](#), le modèle à corrélations conditionnelles constantes (CCC) de [Bollerslev \(1990\)](#), le modèle GARCH à facteurs de [Engle and Rothschild \(1990\)](#), le modèle à corrélations conditionnelles dynamiques (DCC) proposé par [Engle \(2002\)](#).

Le succès des modèles GARCH capturent avec succès certains traits saillants de la volatilité conditionnelle, mais ils présentent quelques caractéristiques indésirables. Notamment ces modèles expriment la variance conditionnelle des rendements en fonction uniquement de leurs valeurs passées, alors qu'il y a des informations supplémentaires sous forme de variables exogènes, par exemple du volume journalier de transactions, des données intrajournalières à haute fréquence, ou même des séries d'autres rendements, qui peuvent aider à améliorer la prédiction de volatilité. Il est donc naturel d'ajouter les variables explicatives dans les représentations de la variance conditionnelle ou de la matrice de covariance conditionnelle.

Du point de vue pratique, nous trouvons dans la littérature des exemples d'utilisation de variables exogènes afin d'expliquer et de prévoir la variabilité des séries financières. [Engle \(2009\)](#) a fourni la preuve que les fondamentaux économiques, tels que l'inflation et la production industrielle, sont des facteurs importants pour la volatilité des marchés. [Cakmakli and Dijk \(2010\)](#) ont démontré qu'un certain nombre de variables macroéconomiques peuvent

nous aider à prédire la volatilité d'actions. [Christiansen and Schrimpf \(2012\)](#) ont obtenu des résultats similaires pour le change, le marché des matières premières et le marché obligataire. [Kanas \(2012\)](#) a étudié les rendements de l'indice SP100 de 1989 à 2007 dans le cadre du modèle GARCH-en-moyenne (GARCH-M) et obtient une relation rendement-volatilité positive et significative lorsque l'indice de volatilité implicite (VIX) est ajouté au processus de variance conditionnelle en tant que variable exogène. Malgré le fait que ces informations supplémentaires, dans les variables financières et macroéconomiques, sont largement utilisées pour expliquer et prédire la volatilité des actifs financiers, il y a relativement peu de résultats sur le comportement asymptotique de l'estimation des modèles GARCH en présence de variables exogènes. Dans le cas univarié, [Han and Kristensen \(2014\)](#) ont donné les conditions pour la consistance et la loi asymptotique (the Consistency and Asymptotic Normality (CAN)) du modèle standard GARCH(1, 1) augmenté par une covariable supplémentaire. Toutefois, leur travail est restrictif et, en particulier, ne permet pas de prendre en compte l'effet de levier qui est habituellement observé dans les séries financières. Dans le cas multivarié, [Engle and Kroner \(1995\)](#) ont suggéré le modèle BEKK augmenté par des variables exogènes (BEKK-X). Dans leur modèle, les covariables peuvent affecter toutes les volatilités et les co-volatilités des rendements. Cependant, ils n'ont pas considéré l'estimation du modèle avec variables exogènes. [Sucarrat and Escribano \(2016\)](#) ont proposé un cadre général pour l'estimation et l'inférence des modèles log-GARCH-X univariés et multivariés via la représentation (V)ARMA-X. Cependant, ils n'ont pas prouvé la CAN, ni dans le cas univarié, ni dans le cas multivarié. [Francq and Sucarrat \(2015\)](#) ont développé les travaux de [Sucarrat and Escribano \(2016\)](#) en fournissant la preuve de la CAN des estimateurs des volatilités pour les composantes du modèle log-GARCH vectoriel avec covariables. Cependant leur cadre ne précise pas directement la covariance conditionnelle.

En résumé, la modélisation du second moment conditionnel des rendements des actifs financiers a des implications importantes dans de nombreux domaines de recherche tels que la gestion des risques, la tarification de produits dérivés, la couverture de position, l'optimisation de portefeuille. Malgré l'utilisation répandue et l'utilité indubitable d'informations supplémentaires pour expliquer et prédire la volatilité, les recherches théoriques sur les modèles conditionnels hétéroscédastiques avec variables exogènes sont peu fréquents. Ces champs de recherche sont porteurs à la fois de développements théoriques intéressants et de multiples possibilités d'applications.

1.1.1 Description et contribution de cette thèse

Cette thèse se situe dans la continuité des travaux de recherche cités précédemment. Son objectif principal est d'étudier, dans les modèles conditionnellement hétéroscédastiques, des effets exogènes mesurés à l'aide de variables additionnelles. Plus précisément, elle est consacrée à proposer des estimations de modèles du type GARCH avec covariables dans le cas univarié ainsi le cas multivarié, d'établir le comportement asymptotique des estimateurs et de construire des tests de la persistance des variables exogènes. Dans le cadre de la thèse, les variables additionnelles sont supposées déterministes. Toutefois, même si certaines variables utilisées pour quantifier l'influence d'un facteur exogène sur la volatilité des rendements étaient considérées comme aléatoires, elles ne seraient pas considérées comme causales, *i.e.* elles déterminent le processus endogène à modéliser, et ne sont pas déterminées par lui. En ce sens, nous les qualifions indifféremment d'explicatives ou d'exogènes (exogénéité forte)

Le deuxième chapitre est consacré à étendre la théorie existante en étudiant en détail le comportement du QMV pour une large classe de modèles PGARCH augmentés avec des covariables. Nous imposons des conditions très douces sur les covariables (qui peuvent être en corrélation avec les innovations et/ou les rendements passés) et sur les innovations (qui peuvent être des bruits blancs demi-forts de sorte que plusieurs GARCH-X spécifications peuvent être compatibles). Pour la distribution asymptotique du QMV, nous étudions quatre situations différentes correspondant à des modèles forts ou demi-forts, et des paramètres à l'intérieur ou au bord de l'espace des paramètres. En se basant sur la distribution asymptotique du QMV lorsque le paramètre se situe sur le bord de l'espace de paramètres, nous proposons des tests pour déterminer l'ordre du modèle et vérifier la signification des variables exogènes. Nos résultats théoriques sont évidemment illustrés par simulations de Monte Carlo et des applications aux données financières. Ce chapitre a fait l'objet d'un article intitulé "*QLM inference for volatility models with covariables*", co-écrit avec Christian FRANCO et en révision.

Dans le troisième chapitre, nous nous intéressons à étudier le modèle BEKK(1, 1, 1), ci-après dénommé BEKK(1, 1), avec variables exogènes. Comme dans les modèles APARCH-X du deuxième chapitre, le processus des innovations du modèle est supposé être un bruit blanc faible qui est moins fort et plus flexible que l'hypothèse iid. En supposant l'existence du second moment du processus observé et des variables explicatives, nous considérons l'estimation des paramètres par la méthode du ciblage de la variance. Cette méthode est numériquement plus efficace que la méthode classique du QMV, en particulier, en présence de covariables puisqu'elle requiert une optimisation de dimension inférieure. Sous certaines

conditions de régularité, nous étudions les propriétés asymptotiques des estimateurs du VT pour les processus stationnaires et mélangeants. Nous obtenons la convergence et la normalité asymptotique des estimateurs. Ainsi, la matrice de covariance asymptotique de l'estimation est calculée par la méthode HAC. Nous illustrons enfin nos résultats par une étude empirique basée sur des expériences de Monte Carlo et d'une application sur des séries réelles

Enfin, le quatrième chapitre expose la méthode d'estimation équation par équation du modèle BEKK-X présenté dans le troisième chapitre. Un des défis dans la modélisation d'une matrice de covariance est la "malédiction de la dimension", en particulier, en présence de variables exogènes. La méthode d'estimation du VT présentée dans le deuxième chapitre est une façon de réduire cette malédiction de la dimension en imposant une structure sur l'interception de modèle basé sur l'information échantillon. Malgré les avantages potentiels de l'estimation du VT, lorsque le nombre de rendements ou celui de variables augmentent, la malédiction de la dimension est toujours problématique. La motivation principale pour l'utilisation d'une approche EbE dans les applications est le gain important en temps de calcul. Nous avons montré le CAN de l'estimation EbE et comparé l'efficacité entre deux méthodes EbE et VT en faisant des expériences de Monte Carlo.

Nous présentons maintenant les résultats obtenus de façon plus détaillée.

1.2 Résultats du chapitre 2

Dans ce chapitre, nous considérons la vaste classe des modèles GARCH en puissance asymétrique (Asymmetric Power GARCH (APARCH)) semi-forts augmentés avec des variables exogènes (APARCH-X). Le premier objectif est d'établir le comportement asymptotique de l'estimateur du quasi-maximum du vraisemblance (QMV) du modèle APARCH-X en relâchant l'hypothèse que le paramètre se situe à l'intérieur de l'espace des paramètres. Les covariables sont supposées être positives et stationnaires, mais elles sont autorisées à être corrélées, et aussi en corrélation avec les innovations. Nous présentons ensuite les tests utilisés pour éprouver la nullité de certains coefficients, ce qui permet de déterminer l'ordre du modèle et d'éprouver la signification des variables exogènes.

Soit $x^+ = \max(x, 0)$ et $x^- = \max(-x, 0)$. Nous considérons le modèle APARCH-X défini par

$$\begin{cases} \varepsilon_t = h_t^{1/\delta} \eta_t \\ h_t = \omega_0 + \sum_{i=1}^q [\alpha_{0i+} (\varepsilon_{t-i}^+)^{\delta} + \alpha_{0i-} (\varepsilon_{t-i}^-)^{\delta}] + \sum_{j=1}^p \beta_{0j} h_{t-j} + \boldsymbol{\pi}'_0 \mathbf{x}_{t-1}, \end{cases} \quad (1.1)$$

où $\mathbf{x}_t = (x_{1,t}, \dots, x_{r,t})'$ est un vecteur de r variables exogènes, $\alpha_{0i+} \geq 0$, $\alpha_{0i-} \geq 0$, $\beta_{0j} \geq 0$,

$\omega_0 > 0$, $\delta > 0$ et $\boldsymbol{\pi}_0 = (\pi_{01}, \dots, \pi_{0r})' \geq 0$ composante par composante.

Nous discutons d'abord la stationnarité stricte qui sera la principale condition pour la convergence forte du QMV.

En supposant que $p \geq 2$ et $q \geq 2$, soit un vecteur de dimension $2q + p - 2$

$$\mathbf{Y}_t = \left(h_{t+1}, \dots, h_{t-p+2}, (\varepsilon_t^+)^{\delta}, (\varepsilon_t^-)^{\delta}, \dots, (\varepsilon_{t-q+2}^+)^{\delta}, (\varepsilon_{t-q+2}^-)^{\delta} \right)'.$$

Il est facile à voir que (ε_t) vérifie (1.1) si et seulement si

$$\mathbf{Y}_t = \mathbf{C}_{0t} \mathbf{Y}_{t-1} + \mathbf{B}_{0t}, \quad (1.2)$$

où $\mathbf{B}_{0t} = (\omega_0 + \boldsymbol{\pi}'_0 \mathbf{x}_t, 0, \dots, 0)'$ est un vecteur de dimension $2q + p - 2$ et \mathbf{C}_{0t} est une matrice dépendant de $(\eta_t^+)^{\delta}$, $(\eta_t^-)^{\delta}$ et

$$\boldsymbol{\vartheta}_0 = (\boldsymbol{\theta}'_0, \boldsymbol{\pi}'_0)', \quad \boldsymbol{\theta}_0 = (\omega_0, \alpha_{01+}, \alpha_{01-}, \dots, \alpha_{0q+}, \alpha_{0q-}, \beta_{01}, \dots, \beta_{0p})'.$$

Nous faisons l'hypothèse suivante

A1 : (η_t, \mathbf{x}'_t) est un processus strictement stationnaire et ergodique, et il existe $s > 0$ tel que $E|\eta_1|^s < \infty$ et $E\|\mathbf{x}_1\|^s < \infty$.

Lemme 1.1. *Supposons que A1 est satisfaite. Si $\gamma < 0$, alors l'équation APARCH-X (1.1) (ou équivalente (1.2)) admet une unique solution strictement stationnaire, non anticipative et ergodique. Cette solution de (1.2) est donnée par*

$$\mathbf{Y}_t = \mathbf{B}_{0t} + \sum_{k=1}^{\infty} \left(\prod_{i=1}^k \mathbf{C}_{0,t-i-1} \right) \mathbf{B}_{0,t-k}.$$

Lorsque $\gamma \geq 0$, il n'existe pas de solution pour (1.1) et pour (1.2).

Remarque 1.1. *Il faut noter que la condition de stationnarité stricte $\gamma < 0$ donnée dans Lemme 1.1 ne comporte pas de variables exogènes \mathbf{x}_t . Prendre $\mathbf{x}_t = \varepsilon_t$ n'est pas interdit, cependant l'hypothèse A1 implique que (\mathbf{x}_t) est stationnaire, et dans ce cas, le lemme devient trivial.*

1.2.1 Convergence forte du QMV

Hamadeh and Zakoïan (2011) montrent que, pour les modèles APARCH, l'estimation du paramètre de puissance δ est difficile par le fait que dans la pratique la vraisemblance soit

très plate en direction de δ . Estimer ce paramètre conduit à des résultats imprécis et ralentit considérablement la routine d'optimisation. Nous considérons donc que δ est fixé.

Le vecteur des paramètres restant à estimer

$$\boldsymbol{\vartheta} = (\omega, \alpha_{1+}, \alpha_{1-}, \dots, \alpha_{q+}, \alpha_{q-}, \beta_1, \dots, \beta_p, \boldsymbol{\pi}')'$$

appartient à un espace des paramètres $\Theta \subseteq (0, +\infty) \times [0, +\infty)^{d-1}$, où $d = 2q + p + r + 1$.

Soient les observations $(\varepsilon_1, \dots, \varepsilon_n)$ d'un processus stationnaire (ε_t) et les observations $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ des variables exogènes. Étant données des valeurs initiales $\varepsilon_{1-q}, \dots, \varepsilon_0, \tilde{\sigma}_{1-p} \geq 0, \dots, \tilde{\sigma}_0 \geq 0, \mathbf{x}_0 \geq 0$, les $\tilde{\sigma}_t$ sont définis récursivement, pour $t \geq 1$, par

$$\tilde{\sigma}_t^\delta = \tilde{\sigma}_t^\delta(\boldsymbol{\vartheta}) = \omega + \sum_{i=1}^q \alpha_{i+} (\varepsilon_{t-i}^+)^{\delta} + \alpha_{i-} (\varepsilon_{t-i}^-)^{\delta} + \sum_{j=1}^p \beta_j \tilde{\sigma}_{t-j}^\delta + \boldsymbol{\pi}' \mathbf{x}_{t-1}.$$

Un estimateur du QMV de $\boldsymbol{\vartheta}_0$ est défini comme toute solution mesurable $\hat{\boldsymbol{\vartheta}}_n$ de

$$\hat{\boldsymbol{\vartheta}}_n = \arg \min_{\boldsymbol{\vartheta} \in \Theta} \tilde{Q}_n(\boldsymbol{\vartheta}), \quad (1.3)$$

où

$$\tilde{Q}_n(\boldsymbol{\vartheta}) = \frac{1}{n} \sum_{t=1}^n \tilde{\ell}_t, \quad \tilde{\ell}_t = \tilde{\ell}_t(\boldsymbol{\vartheta}) = \frac{\varepsilon_t^2}{\tilde{\sigma}_t^2} + \ln \tilde{\sigma}_t^2. \quad (1.4)$$

Notons $\mathcal{A}_{\boldsymbol{\vartheta}+}(z) = \sum_{i=1}^q \alpha_{i+} z^i$, $\mathcal{A}_{\boldsymbol{\vartheta}-}(z) = \sum_{i=1}^q \alpha_{i-} z^i$ et $\mathcal{B}_{\boldsymbol{\vartheta}}(z) = 1 - \sum_{j=1}^p \beta_j z^j$. Soient \mathcal{F}_{t-1} une tribu engendrée par $\{\varepsilon_u, \mathbf{x}_u, u < t\}$ et $\mathcal{F}_{t,i}$ une tribu engendrée par $\{\eta_{t-j}, j > i, \mathbf{x}_{t-k}, k > 0\}$ (telles que $\mathcal{F}_{t-1} = \mathcal{F}_{t,0}$). Pour montrer la convergence forte, les hypothèses suivantes seront faites

A2 : $E(\eta_t | \mathcal{F}_{t-1}) = 0$ et $E(\eta_t^2 | \mathcal{F}_{t-1}) = 1$.

A3 : $\boldsymbol{\vartheta}_0 \in \Theta$, Θ est compact.

A4 : Pour tout $i \geq 1$ (et tout $t \in \mathbb{Z}$), le support de la distribution de η_{t-i} sachant $\mathcal{F}_{t,i}$ n'est pas inclus dans $[0, \infty)$ ou dans $(-\infty, 0]$ et contient au moins trois points.

A5 : $\gamma < 0$ et $\sum_{j=1}^p \beta_j < 1$ pour tout $\boldsymbol{\vartheta} \in \Theta$.

A6 : Il existe $s > 0$, tel que $Eh_t^s < \infty$ et $E|\varepsilon_t|^s < \infty$.

A7 : Si $p > 0$, $\mathcal{B}_{\boldsymbol{\vartheta}_0}(z)$, $\mathcal{A}_{\boldsymbol{\vartheta}_0+}(z)$ et $\mathcal{A}_{\boldsymbol{\vartheta}_0-}(z)$ n'ont pas de racine commune; $\mathcal{A}_{\boldsymbol{\vartheta}_0+}(1) + \mathcal{A}_{\boldsymbol{\vartheta}_0-}(1) \neq 0$ et $\alpha_{0q+} + \alpha_{0q-} + \beta_{0p} \neq 0$ (avec la notation $\alpha_{00+} = \alpha_{00-} = \beta_{00} = 1$).

A8 : Si \mathbf{c} est un vecteur non nul de \mathbb{R}^r , $\mathbf{c}'\mathbf{x}_1$ a une loi non dégénérée.

Remarque 1.2. Les hypothèses **A1** et **A2** impliquent que (η_t, \mathcal{F}_t) est une différence de martingale homoscédastique. Pour les modèles de type GARCH, il est habituel de supposer l'hypothèse plus forte que (η_t) est iid $(0, 1)$. Notons, cependant, que [Escanciano \(2009\)](#) et [Han and Kristensen \(2014\)](#) ont employé **A2**. L'avantage d'utiliser **A2** est que (1.1) devient un modèle semi-fort, qui peut être satisfait pour différentes tribus \mathcal{F}_t , correspondant par exemple à différentes séquences de variables exogènes (\mathbf{x}_t) . Pour les modèles APARCH-X, l'hypothèse **A2** semble plus flexible que l'hypothèse iid.

Remarque 1.3. Prenons un exemple d'un processus de génération des données pour lequel plusieurs modèles GARCH-X de la forme (1.1) coexistent sous l'hypothèse **A2**. Supposons que $\mathbf{X}_t = (\varepsilon_t, y_t)'$ suit le modèle GARCH bivarié $\mathbf{X}_t = \Sigma_t^{1/2} \boldsymbol{\eta}_t$, où $\Sigma_t = \text{diag}(\sigma_{1,t}^2, \sigma_{2,t}^2)$ avec $\boldsymbol{\eta}_t$ iid $\mathcal{N}(0, \mathbf{I}_2)$, et $\sigma_{i,t}^2 = \omega_i + \alpha_i \varepsilon_{t-1}^2 + \beta_i \sigma_{i,t-1}^2 + \pi_i y_{t-1}^2$ avec $i = 1, 2$. Le processus (ε_t) vérifie donc un modèle (fort) GARCH-X(1,1) avec une variable exogène $x_t = y_t^2$. [Nijman and Sentana \(1996\)](#) montrent que (ε_t) vérifie également un modèle GARCH(2,2) sans variables exogènes, mais avec un bruit demi-fort satisfaisant **A2** qui n'est pas indépendant en général.

Remarque 1.4. Notons que lorsqu'il n'y a pas de covariables et quand (η_t) est iid, l'hypothèse **A4** réduit à

$$P[\eta_1 > 0] \in (0, 1) \text{ et le support de la distribution de } \eta_1 \text{ contient au moins trois points,}$$

ce qui est exactement la condition d'identifiabilité **A2** de [Hamadeh and Zakoïan \(2011\)](#). En présence de covariables, **A4** exclut l'existence de colinéarités entre les variables exogènes et les fonctions des rendements passés impliqués dans la volatilité. Par exemple, l'hypothèse exclut le cas où $\mathbf{c}'\mathbf{x}_{t-1} = (\varepsilon_{t-i}^+)^{\delta}$ avec $\mathbf{c} \in \mathbb{R}^r$ (sinon la variable $(\eta_{t-i}^+)^{\delta}$ sachant $\mathcal{F}_{t,i}$ serait dégénérée, ce qui n'est pas possible sous l'hypothèse **A4**).

Le lemme suivant montre que **A6** peut être enlevée lorsque (η_t) est iid.

Lemme 1.2. Si $\gamma < 0$ et les hypothèses **A1-A2** tiennent avec (η_t) iid $(0, 1)$, l'hypothèse **A6** est satisfaite.

Nous pouvons maintenant énoncer le théorème de la convergence forte du QMV.

Théorème 1.1. Soit $\widehat{\boldsymbol{\vartheta}}_n$ une suite d'estimateurs du QMV satisfaisant (1.3). Sous les hypothèses **A1-A8**,

$$\widehat{\boldsymbol{\vartheta}}_n \rightarrow \boldsymbol{\vartheta}_0 \text{ p.s. lorsque } n \rightarrow \infty.$$

1.2.2 Loi asymptotique du QMV

Pour les modèles GARCH standards, sans covariables et avec (η_t) iid, notons que η_t est indépendant de \mathcal{F}_{t-1} . Dans cette situation, il est connu qu'aucune condition de moment de ε_t n'est requise pour la convergence et la normalité asymptotique du QMV lorsque le paramètre GARCH appartient à l'intérieur de l'espace des paramètres (notée par $\overset{\circ}{\Theta}$), alors que des conditions de moment sont imposées lorsque le paramètre se situe sur le bord de l'espace des paramètres (voir l'exemple donné dans Section 3.1 de [Francq and Zakoïan \(2007\)](#)). Lorsque η_t n'est pas indépendant de \mathcal{F}_{t-1} , des conditions de moment qui sont plus fortes seront requises. Nous établissons donc la loi asymptotique des estimateurs dans 4 cas suivants

- Cas A : η_t est indépendant de \mathcal{F}_{t-1} et tous les composants de $\boldsymbol{\vartheta}_0$ sont strictement positifs ;
 Cas B : η_t est indépendant de \mathcal{F}_{t-1} et au moins un composant de $\boldsymbol{\vartheta}_0$ est égal à zéro ;
 Cas C : η_t n'est pas indépendant de \mathcal{F}_{t-1} et tous les composants de $\boldsymbol{\vartheta}_0$ sont strictement positifs ;
 Cas D : η_t n'est pas indépendant de \mathcal{F}_{t-1} et au moins un composant de $\boldsymbol{\vartheta}_0$ est égal à zéro.

Nous considérons les hypothèses supplémentaires suivantes

A9 : $\mathcal{C} := \lim_{n \rightarrow \infty} \sqrt{n}(\Theta - \boldsymbol{\vartheta}_0) = \prod_{i=1}^d \mathcal{C}_i$, où $\mathcal{C}_i = [0, +\infty)$ lorsque $\boldsymbol{\vartheta}_{0i} = 0$ et $\mathcal{C}_i = \mathbb{R}$ autrement.

A10 : $E\eta_t^4 < \infty$ dans les cas A et B, et $E|\eta_t|^{4+\nu} < \infty$ pour quelque $\nu > 0$ dans les cas C et D.

A11 : $E|\varepsilon_t|^{2\delta+\nu_1} < \infty$ et $E\|\mathbf{x}_t\|^{2+\nu_1} < \infty$ avec $\nu_1 > 0$ dans le cas B, et $E|\varepsilon_t|^{2\delta+8\delta/\nu} < \infty$ et $E\|\mathbf{x}_t\|^{2+8/\nu} < \infty$ avec ν satisfaisant **A10** dans le cas D.

Sous les hypothèses précédentes, nous montrons que la matrice

$$\mathbf{J} := E \left(\frac{\partial^2 \ell_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \right) = \frac{4}{\delta^2} E \left(\frac{1}{\sigma_t^{2\delta}(\boldsymbol{\vartheta}_0)} \frac{\partial \sigma_t^\delta(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}} \frac{\partial \sigma_t^\delta(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}'} \right) \quad (1.5)$$

est définie positive. Soit $\mathbf{Z}^{\mathcal{C}}$ la projection d'un vecteur $\mathbf{Z} \in \mathbb{R}^d$ sur \mathcal{C} définie par

$$\mathbf{Z}^{\mathcal{C}} = \arg \inf_{\mathbf{C} \in \mathcal{C}} \|\mathbf{C} - \mathbf{Z}\|_{\mathbf{J}},$$

par rapport à la norme $\|\mathbf{x}\|_{\mathbf{J}}^2 = \mathbf{x}' \mathbf{J} \mathbf{x}$.

La loi limite du QMV est donnée par le résultat suivant

Théorème 1.2. *Sous les hypothèses du Théorème 1.1 et A9–A11, lorsque $n \rightarrow \infty$,*

$$\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) \xrightarrow{d} \mathbf{Z}^C, \text{ où } \mathbf{Z} \sim \mathcal{N}\{0, \mathbf{J}^{-1} \mathbf{I} \mathbf{J}^{-1}\}, \quad (1.6)$$

\mathbf{J} est défini par (1.5) et

$$\mathbf{I} = \frac{4}{\delta^2} E \left[\left\{ E(\eta_t^4 | \mathcal{F}_{t-1}) - 1 \right\} \frac{1}{\sigma_t^{2\delta}(\boldsymbol{\vartheta}_0)} \frac{\partial \sigma_t^\delta(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}} \frac{\partial \sigma_t^\delta(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}'} \right].$$

Remarque 1.5. *Le théorème précédent fournit la distribution asymptotique du QMV correspondant à chacun des cas A-D. Dans tous les cas, les hypothèses A1 - A9 sont obligatoires. Remarquons que, dans les cas A et B, nous avons $\mathbf{I} = (E\eta_1^4 - 1)\mathbf{J}$. Dans les cas A et C, l'espace local des paramètres est $\mathcal{C} = \mathbb{R}^d$, et la loi asymptotique du QMV est donc normale :*

$$\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) \xrightarrow{d} \mathcal{N}\{0, (E\eta_1^4 - 1)\mathbf{J}^{-1}\} \quad \text{dans le cas A} \quad (1.7)$$

et

$$\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) \xrightarrow{d} \mathcal{N}\{0, \mathbf{J}^{-1} \mathbf{I} \mathbf{J}^{-1}\} \quad \text{dans le cas C.} \quad (1.8)$$

Ce résultat est obtenu sous l'hypothèse que $E\eta_t^4 < \infty$ dans le cas A et une condition légèrement plus forte dans le cas C (voir A10), mais sans imposer des conditions de moment sur le processus observé ε_t .

Sous la condition de moment $E|\varepsilon_t|^6 < \infty$, et d'autres conditions de régularité ressemblantes aux nôtres, [Francq and Zakoïan \(2007\)](#) ont obtenu (1.6) dans le cas particulier des modèles GARCH avec (η_t) iid et $r = 0$ (voir aussi [Andrews \(1999\)](#) et les références qui y sont). Notons que A11 dans le cas B est une condition plus douce que $E|\varepsilon_t|^6 < \infty$, donc le théorème 1.2 améliore légèrement les résultats disponibles pour les modèles GARCH standard. Même quand $r = 0$, les auteurs ne sont pas conscients de l'existence de conditions impliquant (1.6) pour le modèle général APARCH, ou même pour la sous-classe du modèle GARCH dans le cas D. Au meilleur de notre connaissance, lorsque $\boldsymbol{\vartheta}_0$ se situe sur le bord de l'espace des paramètres (les cas B et D), il n'existait pas de résultat similaire à (1.6) pour les modèles GARCH avec covariables dans la littérature.

La proposition suivante fournit des estimations pour les matrices \mathbf{I} et \mathbf{J} requises afin d'appliquer le Théorème 1.2.

Proposition 1.1. *Sous les hypothèses du Théorème 1.2, les estimateurs fortement conver-*

gents des \mathbf{J} and \mathbf{I} sont donnés par

$$\widehat{\mathbf{J}}_n = \frac{4}{\delta^2} \frac{1}{n} \sum_{t=1}^n \frac{1}{\widetilde{\sigma}_t^{2\delta}(\widehat{\boldsymbol{\vartheta}}_n)} \frac{\partial \widetilde{\sigma}_t^\delta(\widehat{\boldsymbol{\vartheta}}_n)}{\partial \boldsymbol{\vartheta}} \frac{\partial \widetilde{\sigma}_t^\delta(\widehat{\boldsymbol{\vartheta}}_n)}{\partial \boldsymbol{\vartheta}'}, \quad (1.9)$$

et

$$\widehat{\mathbf{I}}_n = \frac{4}{\delta^2} \frac{1}{n} \sum_{t=1}^n (\widehat{\eta}_t^4 - 1) \frac{1}{\widetilde{\sigma}_t^{2\delta}(\widehat{\boldsymbol{\vartheta}}_n)} \frac{\partial \widetilde{\sigma}_t^\delta(\widehat{\boldsymbol{\vartheta}}_n)}{\partial \boldsymbol{\vartheta}} \frac{\partial \widetilde{\sigma}_t^\delta(\widehat{\boldsymbol{\vartheta}}_n)}{\partial \boldsymbol{\vartheta}'}, \quad (1.10)$$

avec $\widehat{\eta}_t = \varepsilon_t / \widetilde{\sigma}_t(\widehat{\boldsymbol{\vartheta}}_n)$.

Remarque 1.6. Notons que, pour le Théorème 1.2 et la Proposition 1.1, une condition de moment est imposée sur le processus observé dans les cas B et D. Nous prenons un exemple afin de montrer que la matrice d'information \mathbf{I} n'est pas finie si cette condition n'est pas satisfaite. Dans ce cas là, la loi asymptotique du QMV n'est peut-être pas normale (Hall and Yao, 2003), bien qu'il soit toujours consistant (le Théorème 1.1 n'a pas besoin de conditions de moment).

Remarque 1.7. Dans les cas A et B, en vue de la Remarque 1.5, l'estimateur défini par (1.10) peut être remplacé par

$$\widehat{\mathbf{I}}_n = \left(\frac{1}{n} \sum_{t=1}^n \widehat{\eta}_t^4 - 1 \right) \widehat{\mathbf{J}}_n. \quad (1.11)$$

Le Théorème 1.2 et la Proposition 1.1 permettent de tester la nullité d'un ou de plusieurs coefficients, ce qui est important pour identifier les ordres du modèle et les covariables pertinentes.

Nous considérons tout d'abord le cas où nous testons la nullité d'un seul coefficient. Soit \mathbf{e}_k le k -ème élément de la base canonique de \mathbb{R}^d . Tester la significativité de la k -ème composante de $\boldsymbol{\vartheta}_0$ en supposant que les autres composantes sont positives consiste à tester

$$H_0 : \mathbf{e}'_k \boldsymbol{\vartheta}_0 = 0 \text{ et } \mathbf{e}'_\ell \boldsymbol{\vartheta}_0 > 0 \quad \forall \ell \neq k \quad \text{contre} \quad H_1 : \mathbf{e}'_k \boldsymbol{\vartheta}_0 > 0. \quad (1.12)$$

Pour ce problème, nous considérons le test t de Student défini par

$$t_n(k) = \sqrt{n} \frac{\mathbf{e}'_k \widehat{\boldsymbol{\vartheta}}_n}{\sqrt{\mathbf{e}'_k \widehat{\boldsymbol{\Sigma}} \mathbf{e}_k}}, \quad \widehat{\boldsymbol{\Sigma}} = \widehat{\mathbf{J}}_n^{-1} \widehat{\mathbf{I}}_n \widehat{\mathbf{J}}_n^{-1}.$$

En tant que corollaire du Théorème 1.2 et de la Proposition 1.1, nous montrons le résultat suivant qui nous donne la région de rejet.

Corollaire 1.1. *Sous les hypothèses de la Proposition 1.1, la région de rejete du test t de Student*

$$\{t_n(k) > \Phi^{-1}(1 - \alpha)\} \Leftrightarrow \{t_n^2(k) > \chi_1^2(1 - 2\alpha)\}$$

est asymptotiquement de niveau α sous H_0 et consistante sous H_1 définie dans (1.12).

1.2.3 Test de pertinence de covariables

Pour les modèles GARCH-X, les tests les plus intéressants concernent la pertinence des covariables. Lorsque les ordres GARCH sont petits, par exemple dans le cas d'un GARCH-X (1,1), les coefficients GARCH sont généralement loins d'être nul. Dans ce cas, il est raisonnable de supposer que le paramètre GARCH θ_0 est strictement positif composante par composante. Sous cette hypothèse, aucune condition de moment sur le processus observé ε_t n'est requise pour le QMV pour avoir la distribution asymptotique donnée par (1.6). Nous ne supposons donc que

A12 : dans le cas B, nous avons $E \|\mathbf{x}_t\|^{2+\nu_1} < \infty$ avec $\nu_1 > 0$, et dans le cas D, nous avons $E \|\mathbf{x}_t\|^{2+8/\nu} < \infty$ avec ν comme dans A10.

Théorème 1.3. *Sous les hypothèses A1–A10 et A12, quand $\theta_0 > 0$, la conclusion du Théorème 1.2 est vérifiée.*

Sans perte de généralité, nous pouvons supposer que nous testons que les $r_0 \geq 1$ dernières variables exogènes ne sont pas pertinentes et les autres le sont.

Le vecteur $\boldsymbol{\pi}_0$ peut donc s'écrire

$$\boldsymbol{\pi}_0 = (\boldsymbol{\pi}'_{01}, \boldsymbol{\pi}'_{02})', \quad \boldsymbol{\pi}_{01} = (\pi_{01}, \dots, \pi_{0,r-r_0})', \quad \boldsymbol{\pi}_{02} = (\pi_{0,r-r_0+1}, \dots, \pi_{0r})'. \quad (1.13)$$

Sous les hypothèses $\theta_0 > 0$ et $\boldsymbol{\pi}_{01} > 0$ composante par composante, nous faisons le test

$$H_0^\pi : \boldsymbol{\pi}_{02} = \mathbf{0}_{r_0} \quad \text{contre} \quad H_1^\pi : \boldsymbol{\pi}_{02} \neq \mathbf{0}_{r_0} \quad (1.14)$$

en utilisant le test de Wald

$$W_n = n \|\hat{\boldsymbol{\pi}}_{2n}\|_{(K\hat{\Sigma}K')^{-1}}^2 \quad (1.15)$$

où $K = (0_{r_0 \times (d-r_0)} : I_{r_0})$ où I_k désigne la matrice identité de taille k et $0_{k_1 \times k_2}$ désigne la matrice nulle de taille $k_1 \times k_2$.

Corollaire 1.2. *Sous les hypothèses du Théorème 1.3 et H_0^π ,*

$$W_n \xrightarrow{d} W = \|K\mathbf{Z}^C\|_{(K\Sigma K')^{-1}}^2, \quad (1.16)$$

où $\Sigma = \mathbf{J}^{-1}\mathbf{I}\mathbf{J}^{-1}$.

Au niveau asymptotique $\alpha \in (0, 1)$, la valeur critique q_α du test de Wald $\{W_n \geq q_\alpha\}$ peut être obtenue par l'algorithme récemment proposé par Pedersen (2015).

Des expériences de type Monte Carlo et des applications sur séries financières réalisées au chapitre 2 illustrent les résultats asymptotiques. En particulier, nous avons construit un modèle qui vise à expliquer la volatilité des rendements quotidiens de l'indice SP500 par ses valeurs passées, la gamme réalisée, rr_t , le volume, $v - t$ et les rendements du Nikkei, Nik_t , Nik_t , et du FTSE, Nik_t . Nous avons testé la nullité de chaque coefficient et la pertinence des variables. Le modèle avec la plus grande vraisemblance est obtenu pour $\delta = 1$, et est donné par $\varepsilon_t = h_t\eta_t$ avec

$$\begin{aligned} h_t = & 0.004 + \underset{(0.022) 0.500}{0.000} \varepsilon_{t-1}^+ + \underset{(0.043) 0.004}{0.113} \varepsilon_{t-1}^- + \underset{(0.027) 0.000}{0.880} h_{t-1} + \underset{(1.642) 0.036}{2.964} rr_{t-1} \\ & + \underset{(0.031) 0.044}{0.053} v_{t-1} + \underset{(0.008) 0.500}{0.000} Nik_{t-1}^2 + \underset{(0.018) 0.018}{0.037} Ft_{t-1}^2. \end{aligned}$$

Nous avons testé l'hypothèse

$$\alpha_{0+} = \pi_{04} = 0 \quad \text{against} \quad \alpha_{0+} \neq 0 \text{ or } \pi_{04} \neq 0,$$

où π_{04} est le coefficient du Nikkei, sous l'hypothèse maintenue que les autres coefficients sont non nuls. Le modèle a été estimé de nouveau sous la nulle, et nous avons obtenu

$$\begin{aligned} h_t = & 0.004 + \underset{(0.034) 0.000}{0.113} \varepsilon_{t-1}^- + \underset{(0.026) 0.000}{0.880} h_{t-1} + \underset{(1.146) 0.005}{2.964} rr_{t-1} \\ & + \underset{(0.033) 0.053}{0.053} v_{t-1} + \underset{(0.017) 0.014}{0.037} Ft_{t-1}^2. \end{aligned} \quad (1.17)$$

Le résultat montre que la gamme réalisée rr_{t-1} et la série Ft_{t-1}^2 sont des covariables significatives, alors qu'il existe un doute quant à la signification du volume v_{t-1} et que le Nikkei n'as pas d'impact sur l'indice SP500.

Une autre étude empirique montre que la volatilité réalisée est une covariable utile pour prévoir les carrés de rendements.

1.3 Résultats du chapitre 3

Ce chapitre est dédié à l'étude du modèle BEKK avec des variables exogènes (BEKK-X), qui envisage de prendre en compte l'influence de variables explicatives sur la covariance conditionnelle des rendements des actifs. La motivation de ce chapitre est de présenter l'estimation du modèle BEKK-X par ciblage de la variance (variance targeting -VT) et d'établir les résultats asymptotiques d'estimateurs.

1.3.1 Modèle et l'estimation par ciblage de la variance

Nous considérons un processus vectoriel $(\boldsymbol{\varepsilon}_t)_{t \in \mathbb{Z}}$ de dimension m , $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \dots, \varepsilon_{mt})'$ et un processus de r variables exogènes $\mathbf{x}_t = (x_{1t}, \dots, x_{rt})'$. Soit \mathcal{F}_{t-1} la tribu engendrée par le passé de $\boldsymbol{\varepsilon}_t$ et de \mathbf{x}_t ; *i.e.* $\mathcal{F}_{t-1} = \sigma\{\boldsymbol{\varepsilon}_u, \mathbf{x}_v; u < t, v < t\}$. Supposons que le processus $\boldsymbol{\varepsilon}_t$ vérifie pour tout $t \in \mathbb{Z}$

$$E(\boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1}) = \mathbf{0}, \quad \text{Var}(\boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1}) = \mathbf{H}_t \quad (1.18)$$

où la matrice \mathbf{H}_t existe et est définie positive.

Considérons le modèle BEKK-X(1,1) suivant

$$\begin{cases} \boldsymbol{\varepsilon}_t = \mathbf{H}_t^{1/2} \boldsymbol{\eta}_t \\ \mathbf{H}_t = \boldsymbol{\Omega} + \mathbf{A} \boldsymbol{\varepsilon}_{t-1} \boldsymbol{\varepsilon}'_{t-1} \mathbf{A}' + \mathbf{B} \mathbf{H}_{t-1} \mathbf{B}' + \mathbf{C} \mathbf{x}_{t-1} \mathbf{x}'_{t-1} \mathbf{C}', \end{cases} \quad (1.19)$$

où les matrices $\boldsymbol{\Omega}$, \mathbf{A} et \mathbf{B} sont carrées ($m \times m$), \mathbf{C} est de taille ($m \times r$), $\boldsymbol{\Omega}$ est symétrique définie positive.

Nous faisons les hypothèses suivantes

B1 : $E(\boldsymbol{\eta}_t | \mathcal{F}_{t-1}) = \mathbf{0}$ et $E(\boldsymbol{\eta}_t \boldsymbol{\eta}'_t | \mathcal{F}_{t-1}) = \mathbf{I}_m$.

B2 : $(\boldsymbol{\varepsilon}_t, \mathbf{x}_t)$ est un processus strictement stationnaire et ergodique.

B3 : $E\|\mathbf{x}_t\|^2 < \infty$ et $E\|\boldsymbol{\varepsilon}_t\|^2 < \infty$.

Soient $\boldsymbol{\Sigma}_\varepsilon := \text{Var}(\boldsymbol{\varepsilon}_t) = E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t) = E(\mathbf{H}_t)$ et $\boldsymbol{\Sigma}_x := E(\mathbf{x}_t \mathbf{x}'_t)$. En prenant les espérances des deux côtés de (1.19), on obtient

$$\boldsymbol{\Sigma}_\varepsilon = \boldsymbol{\Omega} + \mathbf{A} \boldsymbol{\Sigma}_\varepsilon \mathbf{A}' + \mathbf{B} \boldsymbol{\Sigma}_\varepsilon \mathbf{B}' + \mathbf{C} \boldsymbol{\Sigma}_x \mathbf{C}'. \quad (1.20)$$

Alors (1.19) peut être paramétré de nouveau par

$$\begin{cases} \varepsilon_t = \mathbf{H}_t^{1/2} \eta_t, \\ \mathbf{H}_t = (\boldsymbol{\Sigma}_\varepsilon - \mathbf{A}\boldsymbol{\Sigma}_\varepsilon\mathbf{A}' - \mathbf{B}\boldsymbol{\Sigma}_\varepsilon\mathbf{B}' - \mathbf{C}\boldsymbol{\Sigma}_x\mathbf{C}') + \mathbf{A}\varepsilon_{t-1}\varepsilon'_{t-1}\mathbf{A}' + \mathbf{B}\mathbf{H}_{t-1}\mathbf{B}' \\ \quad + \mathbf{C}\mathbf{x}_{t-1}\mathbf{x}'_{t-1}\mathbf{C}'. \end{cases} \quad (1.21)$$

Notons que dans cette représentation, la contrainte de la positivité de $\boldsymbol{\Omega}$ dans (1.19) devient

$$\boldsymbol{\Sigma}_\varepsilon - \mathbf{A}\boldsymbol{\Sigma}_\varepsilon\mathbf{A}' - \mathbf{B}\boldsymbol{\Sigma}_\varepsilon\mathbf{B}' - \mathbf{C}\boldsymbol{\Sigma}_x\mathbf{C}' \text{ est défini positif.} \quad (1.22)$$

Les paramètres sont donc les coefficients des matrices \mathbf{A} , \mathbf{B} et \mathbf{C} ainsi que les coefficients des parties triangulaires inférieures y compris la diagonale des matrices $\boldsymbol{\Sigma}_\varepsilon$, $\boldsymbol{\Sigma}_x$.

Le vecteur des paramètres inconnus est noté $\boldsymbol{\vartheta}_0 = (\boldsymbol{\gamma}'_0, \boldsymbol{\theta}'_0)' \in \mathbb{R}^d$ où

$$\begin{aligned} \boldsymbol{\gamma}_0 &= (\boldsymbol{\gamma}'_{x_0}, \boldsymbol{\gamma}'_{\varepsilon_0})' = (\text{vech}'(\boldsymbol{\Sigma}_x), \text{vech}'(\boldsymbol{\Sigma}_\varepsilon))' \in \mathbb{R}^{d_1}, \quad d_1 = \frac{r(r+1)}{2} + \frac{m(m+1)}{2}, \\ \boldsymbol{\theta}_0 &= (\text{vec}'(\mathbf{A}_0), \text{vec}'(\mathbf{B}_0), \text{vec}'(\mathbf{C}_0))' \in \mathbb{R}^{d_2}, \quad d_2 = 2m^2 + mr \end{aligned}$$

et $d = d_1 + d_2$ est le nombre des paramètres à estimer. L'espace des paramètres est

$$\boldsymbol{\Theta} := \boldsymbol{\Theta}_\gamma \times \boldsymbol{\Theta}_\theta \subset \mathbb{R}^{d_1} \times \mathbb{R}^{d_2},$$

dont un vecteur générique de paramètres est désigné par

$$\boldsymbol{\vartheta} = (\boldsymbol{\gamma}', \boldsymbol{\theta}')' = (\boldsymbol{\gamma}'_x, \boldsymbol{\gamma}'_\varepsilon, \boldsymbol{\theta}')' = (\text{vech}'(\boldsymbol{\Sigma}_x), \text{vech}'(\boldsymbol{\Sigma}_\varepsilon), \text{vec}'(\mathbf{A}), \text{vec}'(\mathbf{B}), \text{vec}'(\mathbf{C}))'.$$

Soient $(\varepsilon_1, \dots, \varepsilon_n)$ une réalisation de longueur n du processus stationnaire et ergodique (ε_t) et $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ les observations des variables exogènes (\mathbf{x}_t) . Conditionnellement à des valeurs initiales $\varepsilon_0, \mathbf{x}_0$ et $\widetilde{\mathbf{H}}_0$ défini positif, le *logarithme* de la quasi-vraisemblance gaussienne s'écrit

$$\widetilde{Q}_n(\boldsymbol{\gamma}, \boldsymbol{\theta}) = \frac{1}{n} \sum_{t=1}^n \widetilde{\ell}_t(\boldsymbol{\gamma}, \boldsymbol{\theta}), \quad \widetilde{\ell}_t(\boldsymbol{\gamma}, \boldsymbol{\theta}) = \varepsilon'_t \widetilde{\mathbf{H}}_t^{-1}(\boldsymbol{\gamma}, \boldsymbol{\theta}) \varepsilon_t + \log \det \left(\widetilde{\mathbf{H}}_t(\boldsymbol{\gamma}, \boldsymbol{\theta}) \right). \quad (1.23)$$

où les $\widetilde{\mathbf{H}}_t$ sont définis récursivement, pour $t \geq 1$, par

$$\begin{aligned} \widetilde{\mathbf{H}}_t(\boldsymbol{\gamma}, \boldsymbol{\theta}) &= (\boldsymbol{\Sigma}_\varepsilon - \mathbf{A}\boldsymbol{\Sigma}_\varepsilon\mathbf{A}' - \mathbf{B}\boldsymbol{\Sigma}_\varepsilon\mathbf{B}' - \mathbf{C}\boldsymbol{\Sigma}_x\mathbf{C}') + \mathbf{A}\varepsilon_{t-1}\varepsilon'_{t-1}\mathbf{A}' + \mathbf{B}\widetilde{\mathbf{H}}_{t-1}(\boldsymbol{\gamma}, \boldsymbol{\theta})\mathbf{B}' \\ &\quad + \mathbf{C}\mathbf{x}_{t-1}\mathbf{x}'_{t-1}\mathbf{C}'. \end{aligned} \quad (1.24)$$

L'estimation du VT consiste deux étapes suivantes

Étape 1 : Estimer empiriquement le paramètre $\boldsymbol{\gamma} = (\boldsymbol{\gamma}'_{\mathbf{x}}, \boldsymbol{\gamma}'_{\boldsymbol{\varepsilon}})'$ comme suit

$$\widehat{\boldsymbol{\gamma}}_n = (\widehat{\boldsymbol{\gamma}}'_{\mathbf{x}n}, \widehat{\boldsymbol{\gamma}}'_{\boldsymbol{\varepsilon}n})' = \left(\text{vech}'(\widehat{\boldsymbol{\Sigma}}_{\mathbf{x}n}), \text{vech}'(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}n}) \right)', \quad (1.25)$$

$$\text{où } \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}n} = \frac{1}{n} \sum_{t=1}^n \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' \text{ and } \widehat{\boldsymbol{\Sigma}}_{\mathbf{x}n} = \frac{1}{n} \sum_{t=1}^n \mathbf{x}_t \mathbf{x}_t'.$$

Étape 2 : Un estimateur du QMV de $\boldsymbol{\theta}$ est défini comme toute solution mesurable $\widehat{\boldsymbol{\theta}}_n$ de

$$\widehat{\boldsymbol{\theta}}_n = \arg \min_{\boldsymbol{\theta} \in \Theta_{\boldsymbol{\theta}}} \widetilde{Q}_n(\widehat{\boldsymbol{\gamma}}_n, \boldsymbol{\theta}). \quad (1.26)$$

Un estimateur du VT de $\boldsymbol{\vartheta}_0$ est donc défini par $\widehat{\boldsymbol{\vartheta}}_n = (\widehat{\boldsymbol{\gamma}}'_n, \widehat{\boldsymbol{\theta}}'_n)'$. La matrice $\boldsymbol{\Omega}$ dans le modèle original BEKK-X (1.19) peut être estimée par

$$\widehat{\boldsymbol{\Omega}}_n = \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}n} - \widehat{\mathbf{A}}_n \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}n} \widehat{\mathbf{A}}_n' - \widehat{\mathbf{B}}_n \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}n} \widehat{\mathbf{B}}_n' - \widehat{\mathbf{C}}_n \widehat{\boldsymbol{\Sigma}}_{\mathbf{x}n} \widehat{\mathbf{C}}_n', \quad (1.27)$$

où $\widehat{\mathbf{A}}_n, \widehat{\mathbf{B}}_n$ et $\widehat{\mathbf{C}}_n$ sont les estimateurs du QMV de \mathbf{A}, \mathbf{B} et \mathbf{C} respectivement. Alors un estimateur du vecteur des paramètres originaux, noté par $\boldsymbol{\xi}_0 = (\text{vech}(\boldsymbol{\Omega}_0)', \boldsymbol{\theta}'_0)'$, de (1.19) peut être défini par $\widehat{\boldsymbol{\xi}}_n = (\text{vech}'(\widehat{\boldsymbol{\Omega}}_n), \widehat{\boldsymbol{\theta}}'_n)'$.

1.3.2 Consistance forte et Normalité asymptotique

Nous ferons les hypothèses suivantes pour montrer la convergence forte de l'estimateur du VT

B4 : $\boldsymbol{\vartheta}_0 \in \Theta$ et Θ est compact.

B5 : $\rho(\mathbf{B}) < 1$ et $\rho(\mathbf{A} \otimes \mathbf{A} + \mathbf{B} \otimes \mathbf{B}) < 1, \forall \boldsymbol{\vartheta} \in \Theta$.

B6 : Si , pour quelconque $\boldsymbol{\theta} \in \Theta_{\boldsymbol{\theta}}, \mathbf{H}_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta}) = \mathbf{H}_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)$ p.s., $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

B7 : Si $\boldsymbol{\pi}$ est un vecteur non nul de \mathbb{R}^r , $\boldsymbol{\pi}' \mathbf{x}_1$ a une loi non dégénérée.

Nous pouvons maintenant énoncer le théorème de convergence suivant

Théorème 1.4. *Sous les hypothèses B1 - B7, presque sûrement*

$$\widehat{\boldsymbol{\vartheta}}_n \rightarrow \boldsymbol{\vartheta}_0, \text{ quand } n \rightarrow \infty. \quad (1.28)$$

Pour montrer la normalité asymptotique nous avons besoin des hypothèses supplémentaires suivantes.

B8 : $\boldsymbol{\theta}_0 \in \overset{\circ}{\Theta}_{\boldsymbol{\theta}}$.

B9 : $E \|\boldsymbol{\varepsilon}_t\|^6 < \infty$ et $E \|\mathbf{x}_t\|^6 < \infty$.

Dans le chapitre 3 le comportement asymptotique des estimateurs du VT est étudié dans le cas où les erreurs sont non corrélées mais pas nécessairement iid et les variables exogènes ne sont pas forcément indépendantes du $(\boldsymbol{\eta}_t)$. Ainsi nous introduisons les coefficients de mélange $\alpha_X(h)$ pour un processus stationnaire $X = (X_t)$.

$$\alpha_X(h) = \sup_{A \in \sigma(X_u, u \leq t), B \in \sigma(X_u, u \geq t+h)} |P(A \cap B) - P(A)P(B)|.$$

Nous considérons l'hypothèse suivante qui nous permet de contrôler la dépendance dans le temps du processus $\mathbf{z}_t = (\mathbf{x}'_t, \boldsymbol{\varepsilon}'_t, \boldsymbol{\eta}'_t)'$.

B10 : Le processus (\mathbf{z}_t) est tel que

$$E \|\boldsymbol{\varepsilon}_t\|^{(4+2\nu)(1+1/\delta)} < \infty, \quad E \|\mathbf{x}_t\|^{(4+2\nu)(1+1/\delta)} < \infty, \quad E \|\boldsymbol{\eta}_t\|^{(4+2\nu)(1+\delta)} < \infty$$

et les coefficients de mélange du processus (\mathbf{z}_t) vérifient $\sum_{h=0}^{\infty} \{\alpha_{\mathbf{z}}(h)\}^{\nu/(2+\nu)} < \infty$ avec $\nu > 0$ et $\delta > 0$.

Soit $\underline{\mathbf{H}}_{t,s}(\boldsymbol{\vartheta})$ tel que, avec $s > 0$,

$$\text{vec}(\underline{\mathbf{H}}_{t,s}(\boldsymbol{\vartheta})) = \sum_{k=0}^s (\mathbf{B}^{\otimes 2})^k \left(\text{vec}(\boldsymbol{\Omega}) + \mathbf{A}^{\otimes 2} \text{vec}(\boldsymbol{\varepsilon}_{t-k-1} \boldsymbol{\varepsilon}'_{t-k-1}) + \mathbf{C}^{\otimes 2} \text{vec}(\mathbf{x}_{t-k-1} \mathbf{x}'_{t-k-1}) \right),$$

où $A^{\otimes 2}$ dénote le produit de Kronecker de la matrice A et elle même. Notons \mathcal{S} un sous-espace tel que pour tout $\boldsymbol{\vartheta} \in \Theta$, $\mathbf{H}_t(\boldsymbol{\vartheta}) \in \mathcal{S}$ et pour tout $s > 0$, $\underline{\mathbf{H}}_{t,s}(\boldsymbol{\vartheta}) \in \mathcal{S}$.

B11 : Il existe $K > 0$ tel que

$$\left\| \mathbf{H}_t^{1/2}(\boldsymbol{\vartheta}) - \mathbf{H}_t^{*1/2}(\boldsymbol{\vartheta}) \right\| \leq K \|\mathbf{H}_t(\boldsymbol{\vartheta}) - \mathbf{H}_t^*(\boldsymbol{\vartheta})\| \quad \text{pour tout } \mathbf{H}_t(\boldsymbol{\vartheta}), \mathbf{H}_t^*(\boldsymbol{\vartheta}) \in \mathcal{S}.$$

Remarque 1.8. *L'hypothèse B9 impose l'existence du moment d'ordre 6 sur le processus observé et sur les variables exogènes qui est nécessaire pour montrer l'existence du moment d'ordre 2 du critère.*

Remarque 1.9. *En l'absence de variables exogènes [Comte and Lieberman \(2003\)](#) et [Pedersen and Rahbek \(2014\)](#) ont établi la normalité asymptotique du QMV et du VT, respectivement, en appliquant le théorème central limite à une différence de martingale sous hypothèse iid. Dans ce papier, nous ne supposons pas que $\boldsymbol{\eta}_t$ est indépendant. Les variables exogènes peuvent même être corrélées avec $\boldsymbol{\eta}_t$. Ainsi les hypothèses **B1** et **B10** sont clairement plus faibles que l'hypothèse de bruit blanc fort du point de vue de la dépendance.*

*Si l'hypothèse **A1** est remplacée par une hypothèse iid, les conditions de moment dans **B10** peut être affaiblie par*

A11* : $\mathbf{z}_t = (\mathbf{x}'_t, \boldsymbol{\varepsilon}'_t, \boldsymbol{\eta}'_t)'$ est tel que $E\|\mathbf{z}_t\|^{4+2\nu} < \infty$ et les coefficients de mélange du processus vérifient $\sum_{h=0}^{\infty} \{\alpha_{\mathbf{z}}(h)\}^{\nu/(2+\nu)} < \infty$, avec $\nu > 0$.

Remarque 1.10. *Dans le cas univarié, l'hypothèse **B11** est toujours vérifiée. En effet, pour la simplicité, nous considérons le modèle GARH(1,1) univarié $\varepsilon_t = \sigma_t^2(\theta)\eta_t$, où $\sigma_t^2(\theta) = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2(\theta)$, avec $\omega \geq \underline{\omega} > 0$. Pour quelconque $\sigma_t^2(\theta)$ et $\sigma_t^{*2}(\theta)$, on a*

$$|\sigma_t(\theta) - \sigma_t^*(\theta)| = \left| \sqrt{\sigma_t^{*2}(\theta) + (\sigma_t^2(\theta) - \sigma_t^{*2}(\theta)) \frac{1}{2\sqrt{\bar{\sigma}_t^2(\theta)}}} - \sqrt{\sigma_t^{*2}(\theta)} \right| \leq K |\sigma_t^2(\theta) - \sigma_t^{*2}(\theta)|,$$

où $\bar{\sigma}_t^2(\theta)$ est entre $\sigma_t^2(\theta)$ et $\sigma_t^{*2}(\theta)$ et la dernière inégalité est obtenue de $\bar{\sigma}_t^2(\theta) \geq \underline{\omega}$ pour θ quelconque.

Nous définissons $Q_n(\boldsymbol{\gamma}, \boldsymbol{\theta})$ et $\ell_t(\boldsymbol{\gamma}, \boldsymbol{\theta})$ en remplaçant $\widetilde{\mathbf{H}}_t(\boldsymbol{\gamma}, \boldsymbol{\theta})$ par $\mathbf{H}_t(\boldsymbol{\gamma}, \boldsymbol{\theta})$ dans $\widetilde{Q}_n(\boldsymbol{\gamma}, \boldsymbol{\theta})$ et $\widetilde{\ell}_t(\boldsymbol{\gamma}, \boldsymbol{\theta})$. Nous introduisons les matrices suivantes

$$\mathbf{J} = E \left(\frac{\partial^2 \ell_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right), \quad \mathbf{K}_\varepsilon = E \left(\frac{\partial^2 \ell_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\gamma}'_\varepsilon} \right), \quad \mathbf{K}_x = E \left(\frac{\partial^2 \ell_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\gamma}'_x} \right) \quad (1.29)$$

et

$$\boldsymbol{\Sigma}_{11} = \sum_{h=-\infty}^{\infty} \text{cov}(\text{vech}(\mathbf{x}_t \mathbf{x}'_t), \text{vech}(\mathbf{x}_{t-h} \mathbf{x}'_{t-h})), \quad (1.30)$$

$$\boldsymbol{\Sigma}_{22} = \sum_{h=-\infty}^{\infty} \text{cov}(\boldsymbol{\Upsilon}_{0t} \text{vec}(\boldsymbol{\eta}_t \boldsymbol{\eta}'_t), \boldsymbol{\Upsilon}_{0,t-h} \text{vec}(\boldsymbol{\eta}_{t-h} \boldsymbol{\eta}'_{t-h})), \quad (1.31)$$

$$\boldsymbol{\Sigma}_{12} = \sum_{h=-\infty}^{\infty} \text{cov}(\text{vech}(\mathbf{x}_t \mathbf{x}'_t), \boldsymbol{\Upsilon}_{0,t-h} \text{vec}(\boldsymbol{\eta}_{t-h} \boldsymbol{\eta}'_{t-h})), \quad (1.32)$$

où

$$\Upsilon_{0t} = \begin{pmatrix} \mathbf{H}_{0t}^{1/2} \otimes \mathbf{H}_{0t}^{1/2} \\ -\frac{\partial \text{vec}'(\mathbf{H}_{0t})}{\partial \boldsymbol{\theta}} \left(\mathbf{H}_{0t}^{-1/2} \otimes \mathbf{H}_{0t}^{-1/2} \right)' \end{pmatrix}. \quad (1.33)$$

Notons D_m et L_m deux matrices telles que $\text{vec}(A) = D_m \text{vech}(A)$ et $\text{vech}(A) = L_m \text{vec}(A)$ où A est une matrice symétrique ($m \times m$) quelconque.

Nous montrons le théorème suivant qui nous donne la normalité asymptotique de l'estimateur du VT.

Théorème 1.5. *Sous les hypothèses B1 - B11, lorsque $n \rightarrow \infty$,*

$$\sqrt{n} \begin{pmatrix} \widehat{\gamma}_{\mathbf{x}n} - \gamma_{\mathbf{x}0} \\ \widehat{\gamma}_{\boldsymbol{\varepsilon}n} - \gamma_{\boldsymbol{\varepsilon}0} \\ \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Gamma} \boldsymbol{\Phi} \boldsymbol{\Sigma} \boldsymbol{\Phi}' \boldsymbol{\Gamma}'), \quad (1.34)$$

où

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}'_{12} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \quad \boldsymbol{\Gamma} = \begin{pmatrix} I_{r(r+1)/2} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I_{m(m+1)/2} & \mathbf{0} \\ -\mathbf{J}^{-1} \mathbf{K}_x & -\mathbf{J}^{-1} \mathbf{K}_\varepsilon & -\mathbf{J}^{-1} \end{pmatrix} \quad (1.35)$$

et

$$\boldsymbol{\Phi} = \begin{pmatrix} I_{r(r+1)/2} & \mathbf{0} & \mathbf{0} \\ L_m(I_{m^2} - \mathbf{A}_0^{\otimes 2} - \mathbf{B}_0^{\otimes 2})^{-1} \mathbf{C}_0^{\otimes 2} D_r & L_m(I_{m^2} - \mathbf{A}_0^{\otimes 2} - \mathbf{B}_0^{\otimes 2})^{-1} (I_{m^2} - \mathbf{B}_0^{\otimes 2}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -I_{2m^2+mr} \end{pmatrix}. \quad (1.36)$$

En utilisant le théorème précédent nous obtenons dans le corollaire ci-dessous la normalité asymptotique des estimateurs des paramètres originaux.

Corollaire 1.3. *Sous les hypothèses du Théorème 1.5, on a*

$$\sqrt{n} \left(\widehat{\boldsymbol{\xi}}_n - \boldsymbol{\xi}_0 \right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Delta \boldsymbol{\Gamma} \boldsymbol{\Phi} \boldsymbol{\Sigma} \boldsymbol{\Phi}' \boldsymbol{\Gamma}' \Delta'), \quad (1.37)$$

où

$$\Delta = \begin{pmatrix} \Delta_1 & \Delta_2 \\ \mathbf{0}_{(2m^2+mr) \times (m^2+r^2)} & I_{(2m^2+mr) \times (2m^2+mr)} \end{pmatrix} \quad (1.38)$$

avec

$$\begin{aligned}\Delta_1 &= \begin{pmatrix} -L_m(\mathbf{C}_0 \otimes \mathbf{C}_0) D_d & L_m(I_{m^2} - \mathbf{A}_0 \otimes \mathbf{A}_0 - \mathbf{B}_0 \otimes \mathbf{B}_0) D_m \end{pmatrix}, \\ \Delta_2 &= -L_m(I_{m^2} + M_{mm}) \begin{pmatrix} (\mathbf{A}_0 \Sigma_{\varepsilon 0} \otimes I_m) & (\mathbf{B}_0 \Sigma_{\varepsilon 0} \otimes I_m) & (\mathbf{C}_0 \Sigma_{x0} \otimes I_m) \end{pmatrix}\end{aligned}$$

et la matrice M_{pq} vérifie $M_{pq} \text{vec}(A) = \text{vec}(A')$ pour une matrice A quelconque de taille $(p \times q)$.

Nous donnons la matrice de covariance asymptotique du Théorème 1.5. Les matrices Σ_{11} , Σ_{22} et Σ_{12} sont estimées par la méthode d'estimation non paramétrique du noyau aussi appelée HAC (Heteroscedasticity and Autocorrelation Consistent)

$$\begin{aligned}\widehat{\Sigma}_{11n} &= \frac{1}{n} \sum_{t,s=1}^n w_{|t-s|} \text{vech}(\mathbf{x}_t \mathbf{x}_t') \text{vech}'(\mathbf{x}_s \mathbf{x}_s'), \\ \widehat{\Sigma}_{22n} &= \frac{1}{n} \sum_{t,s=1}^n w_{|t-s|} \widehat{\Upsilon}_t \text{vec}(\boldsymbol{\eta}_t \boldsymbol{\eta}_t') \text{vec}'(\boldsymbol{\eta}_s \boldsymbol{\eta}_s') \widehat{\Upsilon}_s', \\ \widehat{\Sigma}_{12n} &= \frac{1}{n} \sum_{t,s=1}^n w_{|t-s|} \text{vech}(\mathbf{x}_t \mathbf{x}_t') \text{vec}'(\boldsymbol{\eta}_s \boldsymbol{\eta}_s') \widehat{\Upsilon}_s',\end{aligned}$$

où w_0, \dots, w_{n-1} est une séquence de poids (voir Newey and West (1987), Andrews (1991) et Phillips and Jin (2003) pour le problème du choix des poids et

$$\widehat{\Upsilon}_t = \begin{pmatrix} \widetilde{\mathbf{H}}_t^{1/2}(\widehat{\boldsymbol{\vartheta}}_n) \otimes \widetilde{\mathbf{H}}_t^{1/2}(\widehat{\boldsymbol{\vartheta}}_n) \\ -\frac{\partial \text{vec}'(\widetilde{\mathbf{H}}_t(\widehat{\boldsymbol{\vartheta}}_n))}{\partial \boldsymbol{\theta}} \left(\widetilde{\mathbf{H}}_t^{-1/2}(\widehat{\boldsymbol{\vartheta}}_n) \otimes \widetilde{\mathbf{H}}_t^{-1/2}(\widehat{\boldsymbol{\vartheta}}_n) \right)'\end{pmatrix}.$$

Soient

$$\widehat{\mathbf{J}}_n = \frac{1}{n} \sum_{t=1}^n \frac{\partial^2 \widetilde{\ell}_t(\widehat{\boldsymbol{\vartheta}}_n)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}, \quad \widehat{\mathbf{K}}_{x_n} = \frac{1}{n} \sum_{t=1}^n \frac{\partial^2 \widetilde{\ell}_t(\widehat{\boldsymbol{\vartheta}}_n)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\gamma}'_x}, \quad \widehat{\mathbf{K}}_{\varepsilon_n} = \frac{1}{n} \sum_{t=1}^n \frac{\partial^2 \widetilde{\ell}_t(\widehat{\boldsymbol{\vartheta}}_n)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\gamma}'_\varepsilon}. \quad (1.39)$$

Sous les hypothèses du Théorème 1.5, les estimateurs fortement convergents des $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ sont donnés par

$$\widehat{\boldsymbol{\Gamma}}_n = \begin{pmatrix} I_{r(r+1)/2} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I_{m(m+1)/2} & \mathbf{0} \\ -\widehat{\mathbf{J}}_n^{-1} \widehat{\mathbf{K}}_{x_n} & -\widehat{\mathbf{J}}_n^{-1} \widehat{\mathbf{K}}_{\varepsilon_n} & -\widehat{\mathbf{J}}_n^{-1} \end{pmatrix} \quad \text{and} \quad \widehat{\boldsymbol{\Sigma}}_n = \begin{pmatrix} \widehat{\Sigma}_{11n} & \widehat{\Sigma}_{12n} \\ \widehat{\Sigma}'_{12n} & \widehat{\Sigma}_{22n} \end{pmatrix} \quad (1.40)$$

Des expériences de Monte Carlo réalisées au chapitre 3 montrent que les résultats théoriques sont valides.

1.4 Résultats du chapitre 4

Dans ce chapitre nous étudions le modèle BEKK demi-diagonal avec variables exogènes. La méthode d'estimation est équation par équation.

1.4.1 Modèle et estimation équation par équation

Nous considérons un processus vectoriel $(\boldsymbol{\varepsilon}_t)_{t \in \mathbb{Z}}$ de dimension m , $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \dots, \varepsilon_{mt})'$ et un processus de r variables exogènes $\mathbf{x}_t = (x_{1t}, \dots, x_{rt})'$. Soit \mathcal{F}_{t-1} la tribu engendrée par le passé de $\boldsymbol{\varepsilon}_t$ et de \mathbf{x}_t ; *i.e.* $\mathcal{F}_{t-1} = \sigma\{\boldsymbol{\varepsilon}_u, \mathbf{x}_v; u < t, v < t\}$. Supposons que le processus $\boldsymbol{\varepsilon}_t$ vérifie pour tout $t \in \mathbb{Z}$

$$E(\boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1}) = \mathbf{0}, \quad \text{Var}(\boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1}) = \mathbf{H}_t \quad (1.41)$$

où la matrice \mathbf{H}_t existe et est définie positive.

Considérons le modèle BEKK-X(1,1) demi-diagonal suivant

$$\begin{cases} \boldsymbol{\varepsilon}_t = \mathbf{H}_t^{1/2} \boldsymbol{\eta}_t \\ \mathbf{H}_t = \boldsymbol{\Omega} + \mathbf{A} \boldsymbol{\varepsilon}_{t-1} \boldsymbol{\varepsilon}'_{t-1} \mathbf{A}' + \mathbf{B} \mathbf{H}_{t-1} \mathbf{B}' + \mathbf{C} \mathbf{x}_{t-1} \mathbf{x}'_{t-1} \mathbf{C}', \end{cases} \quad (1.42)$$

où $\mathbf{A} = (a_{k\ell})_{1 \leq k, \ell \leq m}$, $\mathbf{B} = \text{diag}(b_1, \dots, b_m)$, $\mathbf{C} = (c_{k\ell})_{1 \leq k \leq m, 1 \leq \ell \leq r}$ et $\boldsymbol{\Omega} = (\omega_{k\ell})_{1 \leq k, \ell \leq m}$ est symétrique définie positive.

Nous faisons les hypothèses suivantes

C1 : $E(\boldsymbol{\eta}_t | \mathcal{F}_{t-1}) = \mathbf{0}$, $\text{Var}(\boldsymbol{\eta}_t | \mathcal{F}_{t-1}) = \mathbf{I}_m$.

C2 : Le processus $(\boldsymbol{\varepsilon}_t, \mathbf{x}_t)$ est strictement stationnaire et ergodique.

C3 : $E(\|\boldsymbol{\varepsilon}_t\|^2) < \infty$ et $E(\|\mathbf{x}_t\|^2) < \infty$.

Supposons que toute l'information passée sur ε_{kt} soit résumée par la variable σ_{kt}^2 avec

$$\text{Var}(\varepsilon_{kt} | \mathcal{F}_{t-1}) = \sigma_{kt}^2.$$

Notons que σ_{kt}^2 est le k -ème élément diagonal de \mathbf{H}_t .

Posons

$$\mathbf{D}_t = \begin{pmatrix} \sigma_{1t} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_{mt} \end{pmatrix}, \quad \boldsymbol{\eta}_t^* = \mathbf{D}_t^{-1} \boldsymbol{\varepsilon}_t = \begin{pmatrix} \frac{\varepsilon_{1t}}{\sigma_{1t}} \\ \vdots \\ \frac{\varepsilon_{mt}}{\sigma_{mt}} \end{pmatrix}$$

Si $\boldsymbol{\varepsilon}_t$ vérifie (1.41), les rendements standardisés $\boldsymbol{\eta}_t^*$ vérifient

$$E(\boldsymbol{\eta}_t^* | \mathcal{F}_{t-1}) = 0, \quad \text{Var}(\boldsymbol{\eta}_t^* | \mathcal{F}_{t-1}) = \mathbf{D}_t^{-1} \mathbf{H}_t \mathbf{D}_t^{-1},$$

qui impliquent

$$E(\eta_{kt}^* | \mathcal{F}_{t-1}) = 0, \quad \text{Var}(\eta_{kt}^* | \mathcal{F}_{t-1}) = 1. \quad (1.43)$$

pour $k = 1, \dots, m$.

Les volatilités individuelles s'expriment linéairement en fonction de ses valeurs passées, des valeurs passées de $\varepsilon_{\ell,t} \varepsilon_{\ell',t}$, $\ell, \ell' = 1, \dots, m$ et des valeurs passées de $x_{s,t} x_{s',t}$, $s, s' = 1, \dots, r$.

$$\begin{cases} \varepsilon_{kt} = \sigma_{kt} \eta_{kt}^*, \\ \sigma_{kt}^2 = \omega_{kk} + \left(\sum_{\ell=1}^m a_{k\ell} \varepsilon_{\ell,t-1} \right)^2 + b_k^2 \sigma_{k,t-1}^2 + \left(\sum_{s=1}^r c_{ks} x_{s,t-1} \right)^2, \end{cases} \quad (1.44)$$

où $\omega_{kk} > 0$, $k = 1, \dots, m$.

Le vecteur des paramètres est noté

$$\boldsymbol{\theta}^{(k)} = (\omega_{kk}, \underline{\mathbf{a}}_k', b_k^2, \underline{\mathbf{c}}_k')' = (\omega_{kk}, a_{k1}, \dots, a_{km}, b_k, c_{k1}, \dots, c_{kr})' \in \mathbb{R}^d,$$

où $\underline{\mathbf{a}}_k = (a_{k1}, \dots, a_{km})'$ et $\underline{\mathbf{c}}_k = (c_{k1}, \dots, c_{kr})'$ dénotent respectivement les k -ème colonnes des matrices \mathbf{A} et \mathbf{C} et $d = m + r + 2$ est le nombre des paramètres à estimer. L'espace des paramètres est

$$\Theta^{(k)} \subset (0, +\infty)^2 \times \mathbb{R}^{m-1} \times [0, 1] \times (0, +\infty) \times \mathbb{R}^{r-1}.$$

La vraie valeur du paramètre est noté

$$\boldsymbol{\theta}_0^{(k)} = (\omega_{kk}^0, \underline{\mathbf{a}}_k^{0'}, (b_k^0)^2, \underline{\mathbf{c}}_k^{0'})', \quad \underline{\mathbf{a}}_k^0 = (a_{k1}^0, \dots, a_{km}^0)', \quad \underline{\mathbf{c}}_k^0 = (c_{k1}^0, \dots, c_{kr}^0)'.$$

Il faut clairement imposer des contraintes supplémentaires pour que σ_{kt}^2 soit identifiable. En effet on obtient la même représentation si a_{k1}^0 ou b_k^0 ou c_{k1}^0 sont remplacés par leurs opposés. Sans perte de généralité, nous supposons que $a_{k1}^0 > 0$, $b_k^0 > 0$ et $c_{k1}^0 > 0$, avec $k = 1, \dots, m$.

Soient $\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_n$ une réalisation de longueur n du processus stationnaire et ergodique $(\boldsymbol{\varepsilon}_t)$ vérifiant (1.42) et $\mathbf{x}_1, \dots, \mathbf{x}_n$ les observations des variables exogènes (\mathbf{x}_t) . Conditionnellement à des valeurs initiales $\tilde{\boldsymbol{\varepsilon}}_0, \tilde{\sigma}_0$ et \mathbf{x}_0 , le *logarithme* de la quasi-vraisemblance gaussienne s'écrit

$$\tilde{Q}_n^{(k)}(\boldsymbol{\theta}^{(k)}) = \frac{1}{n} \sum_{t=1}^n \tilde{\ell}_{kt}(\boldsymbol{\theta}^{(k)}), \quad \tilde{\ell}_{kt}(\boldsymbol{\theta}^{(k)}) = \log \tilde{\sigma}_{kt}^2(\boldsymbol{\theta}^{(k)}) + \frac{\boldsymbol{\varepsilon}_{kt}^2}{\tilde{\sigma}_{kt}^2(\boldsymbol{\theta}^{(k)})},$$

où les $\tilde{\sigma}_{kt}^2(\boldsymbol{\theta}^{(k)})$ sont définis récursivement, pour $t \geq 1$, par

$$\tilde{\sigma}_{kt}^2(\boldsymbol{\theta}^{(k)}) = \omega_{kk} + \left(\sum_{\ell=1}^m a_{k\ell} \varepsilon_{\ell, t-1} \right)^2 + b_k^2 \tilde{\sigma}_{k, t-1}^2(\boldsymbol{\theta}^{(k)}) + \left(\sum_{s=1}^r c_{ks} x_{s, t-1} \right)^2 \quad (1.45)$$

Un estimateur du EbE de $\boldsymbol{\theta}_0^{(k)}$ est défini comme toute solution mesurable $\hat{\boldsymbol{\theta}}_n^{(k)}$ de

$$\hat{\boldsymbol{\theta}}_n^{(k)} = \arg \min_{\boldsymbol{\theta}^{(k)} \in \Theta^{(k)}} \tilde{Q}_n^{(k)}(\boldsymbol{\theta}^{(k)}). \quad (1.46)$$

Soit $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}_0^{(1)'}, \dots, \boldsymbol{\theta}_0^{(m)'})'$. Notons que $\boldsymbol{\theta}_0$ inclut les éléments de la diagonale de $\boldsymbol{\Omega}_0$ et toutes les composantes des matrices $\mathbf{A}_0, \mathbf{B}_0$ et \mathbf{C}_0 . Ce vecteur appartient à l'espace des paramètres $\Theta^m = \Theta^{(1)} \times \dots \times \Theta^{(m)}$, dont un élément général est noté par $\boldsymbol{\theta} = (\boldsymbol{\theta}^{(1)'}, \dots, \boldsymbol{\theta}^{(m)'})'$.

Un estimateur de $\boldsymbol{\theta}_0$ est défini par $\hat{\boldsymbol{\theta}}_n = (\hat{\boldsymbol{\theta}}_n^{(1)'}, \dots, \hat{\boldsymbol{\theta}}_n^{(m)'})'$ qui est une collection des EbE estimateurs.

La matrice $\boldsymbol{\Omega}_0$ peut être entièrement estimée comme suit

$$\text{vech}^0(\hat{\boldsymbol{\Omega}}_n) = \text{vech}^0 \left(\hat{\Sigma}_{\varepsilon n} - \hat{\mathbf{A}}_n \hat{\Sigma}_{\varepsilon n} \hat{\mathbf{A}}_n' - \hat{\mathbf{B}}_n \hat{\Sigma}_{\varepsilon n} \hat{\mathbf{B}}_n' - \hat{\mathbf{C}}_n \hat{\Sigma}_{\mathbf{x}n} \hat{\mathbf{C}}_n' \right), \quad (1.47)$$

où $\hat{\mathbf{A}}_n, \hat{\mathbf{B}}_n$ et $\hat{\mathbf{C}}_n$ sont respectivement les estimateurs des matrices $\mathbf{A}_0, \mathbf{B}_0$ et \mathbf{C}_0 et $\hat{\Sigma}_{\varepsilon n} = \frac{1}{n} \sum_{t=1}^n \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'$ et $\hat{\Sigma}_{\mathbf{x}n} = \frac{1}{n} \sum_{t=1}^n \mathbf{x}_t \mathbf{x}_t'$ sont respectivement les estimateurs empiriques de $\Sigma_{\varepsilon} = E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t')$ et $\Sigma_{\mathbf{x}} = E(\mathbf{x}_t \mathbf{x}_t')$.

L'estimation du modèle (1.42) est donc l'estimation du vecteur des paramètres

$$\boldsymbol{\vartheta}_0 = (\boldsymbol{\theta}'_0, \boldsymbol{\gamma}'_{\varepsilon 0}, \boldsymbol{\gamma}'_{\mathbf{x}0})', \quad \boldsymbol{\gamma}_{\varepsilon 0} = \text{vech}(\Sigma_{\varepsilon}), \quad \boldsymbol{\gamma}_{\mathbf{x}0} = \text{vech}(\Sigma_{\mathbf{x}}).$$

Les estimateurs sont donnés par $\hat{\boldsymbol{\vartheta}}_n = (\hat{\boldsymbol{\theta}}_n', \hat{\boldsymbol{\gamma}}_{\varepsilon n}', \hat{\boldsymbol{\gamma}}_{\mathbf{x}n}')'$, où $\hat{\boldsymbol{\gamma}}_{\varepsilon n} = \text{vech}(\hat{\Sigma}_{\varepsilon n})$ et $\hat{\boldsymbol{\gamma}}_{\mathbf{x}n} = \text{vech}(\hat{\Sigma}_{\mathbf{x}n})$.

1.4.2 L'inférence de l'estimation EbE

Nous ferons les hypothèses suivantes pour montrer la convergence forte de l'estimateur

C4 : $\theta_0^{(k)} \in \Theta^{(k)}$, $\Theta^{(k)}$ est compact, pour $k = 1, \dots, m$.

C5 : $\rho(\mathbf{A}_0 \otimes \mathbf{A}_0 + \mathbf{B}_0 \otimes \mathbf{B}_0) < 1$ et $\sum_{k=1}^m b_k^2 < 1$, $\forall \theta^{(k)} \in \Theta^{(k)}$.

C6 : Il existe $s > 0$ tel que $E|\varepsilon_{kt}|^s < \infty$ et $E|x_{kt}|^s < \infty$.

C7 : Pour tout $\ell^* = 1, \dots, m$, $\varepsilon_{\ell^*t}^2$ n'appartient pas à l'espace de Hilbert engendré par les combinaisons linéaires des $\varepsilon_{\ell u} \varepsilon_{\ell' u}$, des $x_{sv} x_{s'v}$ avec $u < t$, $v \leq t$, $\ell, \ell' = 1, \dots, m$, $s, s' = 1, \dots, r$ et des $\varepsilon_{\ell t} \varepsilon_{\ell' t}$ avec $(\ell, \ell') \neq (\ell^*, \ell^*)$.

C8 : Pour tout $s^* = 1, \dots, r$, $x_{s^*t}^2$ n'appartient pas à l'espace de Hilbert engendré par les combinaisons linéaires des $x_{sv} x_{s'v}$ avec $v < t$, $s, s' = 1, \dots, r$ et des $x_{st} x_{s't}$ avec $(s, s') \neq (s^*, s^*)$.

Remarque 1.11. Les hypothèses **C7** et **C8** sont les conditions d'identifiabilité. Prenons un exemple. Pour la simplicité, nous considérons (1.42) dans le cas où $m = 2$, $r = 2$ et la matrice de covariance conditionnelle

$$\mathbf{H}_t = \mathbf{\Omega} + \mathbf{A} \boldsymbol{\varepsilon}_{t-1} \boldsymbol{\varepsilon}'_{t-1} \mathbf{A}' + \mathbf{C} \mathbf{x}_{t-1} \mathbf{x}'_{t-1} \mathbf{C}', \quad (1.48)$$

où $\mathbf{\Omega} = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{pmatrix}$ est symétrique définie positive, $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ et $\mathbf{C} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}$, avec $a_{11} > 0$, $a_{21} > 0$, $c_{11} > 0$ et $c_{21} > 0$. La volatilité de la première composante, ε_{1t} , de $\boldsymbol{\varepsilon}_t$ est donc donnée par

$$\sigma_{1t}^2 = \omega_{11} + a_{11} \varepsilon_{1,t-1}^2 + 2a_{11} a_{12} \varepsilon_{1,t-1} \varepsilon_{2,t-1} + a_{12} \varepsilon_{2,t-1}^2 + c_{11} x_{1,t-1}^2 + 2c_{11} c_{12} x_{1,t-1} x_{2,t-1} + c_{12} x_{2,t-1}^2.$$

L'hypothèse **C7** exclut le cas où, par exemple $x_{1,t-1} = \varepsilon_{1,t-1}$. Ceci est évidemment nécessaire, sinon le modèle n'est pas identifiable.

De même, l'hypothèse **C8** élimine l'existence de combinaison linéaire d'un nombre fini des $x_{s,t-i} x_{s',t-i}$ qui est clairement une condition d'identifiabilité.

Nous pouvons maintenant énoncer le théorème de convergence suivant

Théorème 1.6. Sous les hypothèses **C1** - **C8**, l'estimateur de $\theta_0^{(k)}$ est consistant fort

$$\widehat{\theta}_n^{(k)} \rightarrow \theta_0^{(k)}, \text{ p.s. lorsque } n \rightarrow \infty.$$

Le résultat ci-dessous est une conséquence du Théorème 1.6.

Corollaire 1.4. *Sous les hypothèses du Théorème 1.6, l'estimateur $\widehat{\boldsymbol{\vartheta}}_n$ de $\boldsymbol{\vartheta}_0$ est consistant fort.*

Pour montrer la normalité asymptotique, nous avons besoin des hypothèses supplémentaires suivantes

C9 : $\boldsymbol{\theta}_0^{(k)}$ appartient à l'intérieur de l'espace des paramètres $\Theta^{(k)}$, pour $k = 1, \dots, m$.

C10 : $E\|\boldsymbol{\eta}_t\|^{4(1+\delta)} < \infty$, $E\|\boldsymbol{\varepsilon}_t\|^{4(1+\delta)} < \infty$ et $E\|\mathbf{x}_t\|^{4(1+\delta)} < \infty$ pour quelque $\delta > 0$.

C11 : Le processus $\mathbf{z}_t = (\mathbf{x}'_t, \boldsymbol{\varepsilon}'_t, \boldsymbol{\eta}'_t)'$ vérifie

$$E\|\boldsymbol{\varepsilon}_t\|^{(4+2\nu)(1+\delta)} < \infty, E\|\mathbf{x}_t\|^{(4+2\nu)(1+\delta)} < \infty \text{ and } E\|\boldsymbol{\eta}_t\|^{(4+2\nu)(1+\delta)} < \infty,$$

en plus les coefficients de mélange, $\alpha_{\mathbf{z}}(h)$, du processus (\mathbf{z}_t) sont tels que

$$\sum_{h=0}^{\infty} \{\alpha_{\mathbf{z}}(h)\}^{\nu/(2+\nu)} < \infty \text{ avec } \nu > 0 \text{ et } \delta > 0.$$

Soit $\underline{\mathbf{H}}_{t,s}(\boldsymbol{\vartheta})$ tel que, avec $s > 0$,

$$\text{vec}(\underline{\mathbf{H}}_{t,s}(\boldsymbol{\vartheta})) = \sum_{k=0}^s (\mathbf{B}^{\otimes 2})^k (\text{vec}(\boldsymbol{\Omega}) + \mathbf{A}^{\otimes 2} \text{vec}(\boldsymbol{\varepsilon}_{t-k-1} \boldsymbol{\varepsilon}'_{t-k-1}) + \mathbf{C}^{\otimes 2} \text{vec}(\mathbf{x}_{t-k-1} \mathbf{x}'_{t-k-1})),$$

où $A^{\otimes 2}$ dénote le produit de Kronecker de la matrice A et elle-même. Notons \mathcal{S} un sous-espace tel que pour tout $\boldsymbol{\vartheta} \in \Theta$, $\mathbf{H}_t(\boldsymbol{\vartheta}) \in \mathcal{S}$ et pour tout $s > 0$, $\underline{\mathbf{H}}_{t,s}(\boldsymbol{\vartheta}) \in \mathcal{S}$.

C12 : Il existe $K > 0$ tel que

$$\left\| \mathbf{H}_t^{1/2}(\boldsymbol{\vartheta}) - \mathbf{H}_t^{*1/2}(\boldsymbol{\vartheta}) \right\| \leq K \|\mathbf{H}_t(\boldsymbol{\vartheta}) - \mathbf{H}_t^*(\boldsymbol{\vartheta})\| \quad \text{pour tout } \mathbf{H}_t(\boldsymbol{\vartheta}), \mathbf{H}_t^*(\boldsymbol{\vartheta}) \in \mathcal{S}.$$

Remarque 1.12. *Les hypothèses C11 and C12 sont également requises pour le modèle BEKK-X estimé par le ciblage de variance dans le chapitre 3.*

Afin d'énoncer le théorème de la normalité asymptotique, nous introduisons les notations suivantes

Soit les matrices $\mathbf{J}_{ks} = E(\Delta_{kt} \Delta'_{st})$, $k, s = 1, \dots, m$, où $\Delta_{kt} = \frac{1}{\sigma_{kt}^2} \frac{\partial \sigma_{kt}^2(\boldsymbol{\theta}_0^{(k)})}{\partial \boldsymbol{\theta}^{(k)}}$ et soit la matrice par blocs $\mathbf{J} = \text{diag}\{\mathbf{J}_{11}, \dots, \mathbf{J}_{mm}\}$. Notons $\boldsymbol{\Delta}_t = \text{diag}(\Delta_{1t}, \dots, \Delta_{mt})$, $\mathbf{N}_1 = L_m(I_{m^2} -$

$\mathbf{A}_0^{\otimes 2} - \mathbf{B}_0^{\otimes 2})^{-1}(I_{m^2} - \mathbf{B}_0^{\otimes 2})$ et $\mathbf{N}_2 = L_m(I_{m^2} - \mathbf{A}_0^{\otimes 2} - \mathbf{B}_0^{\otimes 2})^{-1}\mathbf{C}_0^{\otimes 2}D_r$. Définissons également les matrices suivantes

$$\boldsymbol{\Sigma}_{11} = \sum_{h=-\infty}^{\infty} \text{cov}(\boldsymbol{\Upsilon}_{0t} \text{vec}(\boldsymbol{\eta}_t \boldsymbol{\eta}'_t), \boldsymbol{\Upsilon}_{0,t-h} \text{vec}(\boldsymbol{\eta}_{t-h} \boldsymbol{\eta}'_{t-h})), \quad (1.49)$$

$$\boldsymbol{\Sigma}_{22} = \sum_{h=-\infty}^{\infty} \text{cov}(\text{vech}(\mathbf{x}_t \mathbf{x}'_t), \text{vech}(\mathbf{x}_{t-h} \mathbf{x}'_{t-h})), \quad (1.50)$$

$$\boldsymbol{\Sigma}_{12} = \sum_{h=-\infty}^{\infty} \text{cov}(\text{vech}(\mathbf{x}_t \mathbf{x}'_t), \boldsymbol{\Upsilon}_{0,t-h} \text{vec}(\boldsymbol{\eta}_{t-h} \boldsymbol{\eta}'_{t-h})), \quad (1.51)$$

où

$$\boldsymbol{\Upsilon}_{0t} = \begin{pmatrix} \Delta_t T_m \left(\mathbf{D}_{0t}^{-1} \mathbf{H}_{0t}^{1/2} \right) \otimes \left(\mathbf{D}_{0t}^{-1} \mathbf{H}_{0t}^{1/2} \right) \\ \mathbf{H}_{0t}^{1/2} \otimes \mathbf{H}_{0t}^{1/2} \end{pmatrix}. \quad (1.52)$$

Le théorème suivant montre la normalité asymptotique du vecteur des paramètres estimés $\widehat{\boldsymbol{\theta}}_n$

Théorème 1.7. *Sous les hypothèses C1 - C12, lorsque $n \rightarrow \infty$,*

$$\sqrt{n} \begin{pmatrix} \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \\ \widehat{\boldsymbol{\gamma}}_{\varepsilon n} - \boldsymbol{\gamma}_{\varepsilon 0} \\ \widehat{\boldsymbol{\gamma}}_{\mathbf{x}n} - \boldsymbol{\gamma}_{\mathbf{x}0} \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Gamma} \boldsymbol{\Sigma} \boldsymbol{\Gamma}'), \quad (1.53)$$

où

$$\boldsymbol{\Gamma} = \begin{pmatrix} -\mathbf{J}^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{N}_1 & \mathbf{N}_2 \\ \mathbf{0} & \mathbf{0} & I_{r(r+1)/2} \end{pmatrix} \text{ et } \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}'_{12} & \boldsymbol{\Sigma}_{22} \end{pmatrix}. \quad (1.54)$$

Le vecteur des paramètres BEKK-X est noté

$$\boldsymbol{\xi}_0 = ((\text{vech}^0(\boldsymbol{\Omega}_0))', \boldsymbol{\theta}'_0)' \in \mathbb{R}^{m(m-1)/2+md}. \quad (1.55)$$

Un estimateur de $\boldsymbol{\xi}_0$ est défini par $\widehat{\boldsymbol{\xi}}_n = (\widehat{\boldsymbol{\omega}}'_n, \widehat{\boldsymbol{\theta}}'_n)'$, où $\widehat{\boldsymbol{\omega}}_n = \text{vech}^0(\widehat{\boldsymbol{\Omega}}_n)$ est un estimateur de $\boldsymbol{\omega}_0 = \text{vech}^0(\boldsymbol{\Omega}_0)$.

Le théorème suivant nous donne la convergence forte et la normalité asymptotique de l'estimateur $\widehat{\boldsymbol{\xi}}_n$ de $\boldsymbol{\xi}_0$.

Théorème 1.8. *Si les hypothèses C1 - C8 sont satisfaites, l'estimateur $\widehat{\boldsymbol{\xi}}_n$ de $\boldsymbol{\xi}_0$ est consis-*

tant fort

$$\widehat{\boldsymbol{\xi}}_n \rightarrow \boldsymbol{\xi}_0 \text{ a.s. as } n \rightarrow \infty.$$

Si, en plus, les hypothèses **C9** - **C12** sont satisfaites,

$$\sqrt{n} \begin{pmatrix} \widehat{\boldsymbol{\omega}}_n - \boldsymbol{\omega}_0 \\ \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Omega} \boldsymbol{\Sigma} \boldsymbol{\Omega}'), \quad (1.56)$$

où

$$\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_1 \\ \boldsymbol{\Omega}_2 \end{pmatrix}, \quad \boldsymbol{\Omega}_1 = \begin{pmatrix} \mathbf{A}^* & \mathbf{B}^* & \mathbf{C}^* & \mathbf{E}^* & \mathbf{X}^* \end{pmatrix} \boldsymbol{\Psi}, \quad (1.57)$$

$$\boldsymbol{\Omega}_2 = \begin{pmatrix} I_{md} & \mathbf{0}_{m(m+1)/2} & \mathbf{0}_{r(r+1)/2} \end{pmatrix},$$

avec

$$\begin{aligned} \mathbf{A}^* &= -P_m \{ I_m \otimes (\mathbf{A}_0 \boldsymbol{\Sigma}_\varepsilon) + ((\mathbf{A}_0 \boldsymbol{\Sigma}_\varepsilon) \otimes I_m) M_{mm} \}, \\ \mathbf{B}^* &= -P_m \{ I_m \otimes (\mathbf{B}_0 \boldsymbol{\Sigma}_\varepsilon) + (\mathbf{B}_0 \boldsymbol{\Sigma}_\varepsilon) \otimes I_m \}, \\ \mathbf{C}^* &= -P_m \{ I_m \otimes (\mathbf{C}_0 \boldsymbol{\Sigma}_x) + ((\mathbf{C}_0 \boldsymbol{\Sigma}_x) \otimes I_m) M_{mr} \}, \\ \mathbf{E}^* &= P_m (I_{m^2} - \mathbf{B}_0 \otimes \mathbf{B}_0 - \mathbf{A}_0 \otimes \mathbf{A}_0) D_m, \quad \mathbf{X}^* = -P_m (\mathbf{C}_0 \otimes \mathbf{C}_0) D_r. \end{aligned}$$

Des expériences de Monte Carlo réalisées au chapitre 4 montrent que les résultats théoriques sont valides. Les résultats numériques montrent également que la différence des estimateurs du VT et des estimateurs EbE est minuscule alors que l'approche EbE est bien meilleure que la méthode du VT en terme de temps de calcul.

Références

- Andrews, D.W.** (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.
- Andrews, D.W.** (1999) Estimation when a parameter is on a boundary. *Econometrica* 67, 1341–1383.
- Andrews, D.W.** (2001) Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica* 69, 683–734.
- Billingsley, P.** (1961) The Lindeberg-Levy theorem for martingales. *Proceedings of the American Mathematical Society* 12, 788–792.
- Billingsley, P.** (1995) *Probability and measure*. John Wiley.
- Bollerslev, T.** (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.
- Bollerslev, T., R.F. Engle and J.M. Wooldridge** (1988) A capital asset pricing model with time-varying covariances. *The Journal of Political Economy*, 116–131.
- Bollerslev, T.** (1990) Modelling the coherence in short-run nominal exchange rates : a multivariate generalized ARCH model. *The review of economics and statistics* 495–505.
- Bollerslev, T. and J.M. Wooldridge** (1992) Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric reviews* 11, 143–172.
- Bollerslev, T.** (2008) Glossary to ARCH (GARCH). *Volatility and Time Series Econometrics : Essays in Honor of Robert F. Engle* (eds. T.Bollerslev, J. R. Russell and M. Watson), Oxford University Press, Oxford, UK.
- Bouissou, M.B., J-J. Laffont and Q.H. Vuong** (1986) Tests of noncausality under Markov assumptions for qualitative panel data. *Econometrica : Journal of the Econometric Society*, 395–414.
- Bougerol, P. and N. Picard** (1992a) Stationarity of GARCH processes and of some non-negative time series. *Journal of Econometrics* 52, 115–127.
- Bougerol, P. and N. Picard** (1992b) Strict stationarity of generalized autoregressive processes. *Annals of Probability* 20, 1714–1729.
- Boussama, F., F. Fuchs and R. Stelzer** (2011) Stationarity and geometric ergodicity of BEKK multivariate GARCH models. *Stochastic Processes and their Applications* 121, 2331–2360.

- Brandt, A.** (1986) The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Advance in Applied Probability* 18, 221–254.
- Cakmakli, C. and D.J.C Van Dijk** (2010) Getting the most out of macroeconomic information for predicting stock returns and volatility.
- Christiansen, C., M. Schmeling and A. Schrimpf** (2012) A comprehensive look at financial volatility prediction by economic variables. *Journal of Applied Econometrics* 27, 956–977.
- Comte, F. and O. Lieberman** (2003) Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis* 84, 61–84.
- Ding, Z., C. Granger and R.F. Engle** (1993) A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1, 83–106.
- Engle, R.F., Hendry, D.F. and J.F. Richard** (1983) Exogeneity. *Econometrica* 51, 277–304.
- Engle, R.F., V.K. Ng and K.F. Kroner** (1990) Asset pricing with a factor-ARCH covariance structure : Empirical estimates for treasury bills. *Journal of Econometrics* 45, 213–237.
- Engle, R.F. and K.F. Kroner** (1995) Multivariate simultaneous generalized ARCH. *Econometric theory* 11, 122–150.
- Engle, R.F. and A.J. Patton** (2001) What good is a volatility model. *Quantitative finance* 1, 237–245.
- Engle, R.F. and K. Sheppar** (2001) Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. *National Bureau of Economic Research*.
- Engle, R.F.** (2002) Dynamic conditional correlation : A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics* 20, 339–350.
- Engle, R.F.** (2009) *Anticipating correlations : a new paradigm for risk management*. Princeton University Press.
- Escanciano, J.C.** (2009) Quasi-maximum likelihood estimation of semi-strong GARCH models. *Econometric Theory* 25, 561–570.
- Florens, J-P. and M. Mouchart** (1982) A note on noncausality. *Econometrica : Journal of the Econometric Society*, 583–591.

- Francq, C. and J-M. Zakoïan** (1998) Estimating linear representations of nonlinear processes. *Journal of Statistical Planning and Inference* 68, 145–165.
- Francq, C. and J-M. Zakoïan** (2000) Covariance matrix estimation for estimators of mixing weak ARMA models. *Journal of Statistical Planning and Inference* 83, 369–394.
- Francq, C. and J-M. Zakoïan** (2004) Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* 10, 605–637.
- Francq, C., R. Roy and J-M. Zakoïan** (2005) Diagnostic checking in ARMA models with uncorrelated errors. *Journal of the American Statistical Association* 100, 532–544.
- Francq, C. and J-M. Zakoïan** (2007) Quasi-Maximum Likelihood Estimation in GARCH Processes when some coefficients are equal to zero. *Stochastic Processes and their Applications* 117, 1265–1284.
- Francq, C. and J-M. Zakoïan** (2009) Testing the nullity of GARCH coefficients : correction of the standard tests and relative efficiency comparisons. *Journal of the American Statistical Association* 104, 313–324.
- Francq, C. and J-M. Zakoïan** (2010) *GARCH Models : Structure, Statistical Inference and Financial Applications*. John Wiley.
- Francq, C., L. Horvath and J-M. Zakoïan** (2011) Merits and drawbacks of variance targeting in GARCH models. *Journal of Financial Econometrics* 9, 619–656.
- Francq, C., L. Horváth and J-M. Zakoïan** (2014) Variance targeting estimation of multivariate GARCH models. *Journal of Financial Econometrics*, https://mpra.ub.uni-muenchen.de/57794/1/MPRA_paper_57794.pdf
- Francq, C. and Thieu, L.Q.** (2015) Qml inference for volatility models with covariates. *University Library of Munich, Germany*, No. 63198.
- Francq, C. and G. Sucarrat** (2015) Equation-by-Equation Estimation of a Multivariate Log-GARCH-X Model of Financial Returns.
- Fuertes, A.M., M. Izzeldin and E. Kalotychou** (2009) On forecasting daily stock volatility : the role of intraday information and market conditions. *International Journal of Forecasting* 25, 259–281.
- Glosten, L.R., R. Jaganathan and D. Runkle** (1993) On the relation between the expected values and the volatility of the nominal excess return on stocks. *Journal of Finance* 48, 1779–1801.

- Hafner, C.M. and A. Preminger** (2009) On asymptotic theory for multivariate GARCH models. *Journal of Multivariate Analysis* 100, 2044–2054.
- Hall, P. and Q. Yao** (2003) Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica* 71, 285–317.
- Hamadeh, T. and J-M. Zakoïan** (2011) Asymptotic properties of LS and QML estimators for a class of nonlinear GARCH processes. *Journal of Statistical Planning and Inference* 141, 488–507.
- Han, H. and D. Kristensen** (2014) Asymptotic Theory for the QMLE in GARCH-X Models With Stationary and Nonstationary Covariates. *Journal of Business & Economic Statistics* 32, 416–429.
- Herrndorf, N.** (1984) A functional central limit theorem for weakly dependent sequences of random variables. *The Annals of Probability*, 141–153.
- Kenneth S.M.** (1981) On the Inverse of the Sum of Matrices. *Mathematics Magazine*. <http://www.jstor.org/stable/2690437>, 54, 67–72
- Koopmans, T.C. and O. Reiersol** (1950) The identification of structural characteristics. *The Annals of Mathematical Statistics* 21, 165–181.
- Laurent, S., C. Lecourt and F.C. Palm** (2014) Testing for jumps in conditionally Gaussian ARMA-GARCH models, a robust approach. *Computational Statistics & Data Analysis*. <http://dx.doi.org/10.1016/j.csda.2014.05.015>
- Mezrich, J. and R. F. Engle** (1996) GARCH for groups. *Risk* 9, 36–40.
- Nelson, D. B.** (1991) Conditional heteroskedasticity in asset returns : A new approach. *Econometrica* 59, 347–370.
- Newey, W.K. and K.D. West** (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix. *Econometrica* 55, 703–708.
- Nijman, T. and E. Sentana** (1996) Marginalization and contemporaneous aggregation in multivariate GARCH processes. *Journal of Econometrics* 71, 71–87.
- Pan, J., Wang, H. and H. Tong** (2008) Estimation and tests for power-transformed and threshold GARCH models. *Journal of Econometrics* 142, 352–378.
- Pedersen, R.S. and A. Rahbek** (2014) Multivariate variance targeting in the BEKK-GARCH model. *The Econometrics Journal* 17, 24–55.
- Pedersen, R. S.** (2015) Inference and Testing on the Boundary in Extended Constant Conditional Correlation GARCH Models. Available at SSRN : <http://dx.doi.org/10.2139/ssrn.2656058>

- Phillips, P.C.B., Y. Sun and S. Jin** (2003) Consistent HAC estimation and robust regression testing using sharp origin kernels with no truncation. *Discussion paper, Yale University*.
- Silvennoinen, A. and T. Teräsvirta** (2009) Modeling multivariate autoregressive conditional heteroskedasticity with the double smooth transition conditional correlation GARCH model. *Journal of Financial Econometrics* 7, 373–411.
- Sucarrat, G. and A. Escribano** (2010) The Power Log-GARCH Model. *Working document*, Economic Series 10-13, University Carlos III, Madrid.
- Taylor, S.J.** (1986) *Modelling Financial Time Series*, New-York : Wiley.
- Wintenberger, O.** (2013) Continuous invertibility and stable QML estimation of the EGARCH(1,1) model. *Scandinavian Journal of Statistics* 40, 846–867.
- Zakoïan, J-M.** (1994) Threshold heteroskedastic models. *Journal of Economic Dynamics and Control* 18, 931–955.

Chapitre 2

QML inference for volatility models with covariates

Abstract. The asymptotic distribution of the Gaussian quasi-maximum likelihood estimator (QMLE) is obtained for a wide class of asymmetric GARCH models with exogenous covariates. The true value of the parameter is not restricted to belong to the interior of the parameter space, which allows us to derive tests for the significance of the parameters. In particular, the relevance of the exogenous variables can be assessed. The results are obtained without assuming that the innovations are independent, which allows conditioning on different information sets. Monte Carlo experiments and applications to financial series illustrate the asymptotic results. In particular, an empirical study demonstrates that the realized volatility can be an helpful covariate for predicting squared returns.

Keywords : APARCH model augmented with explanatory variables, Boundary of the parameter space, Consistency and asymptotic distribution of the Gaussian quasi-maximum likelihood estimator, GARCH-X models, Power-transformed and Threshold GARCH with exogenous covariates.

2.1 Introduction

The GARCH-type models are of the form

$$\varepsilon_t = \sigma_t \eta_t, \tag{2.1}$$

where the squared volatility σ_t^2 is the best predictor of ε_t^2 given a certain information set \mathcal{F}_{t-1} available at time t . More precisely, it is assumed that $E(\varepsilon_t^2 | \mathcal{F}_{t-1}) = \sigma_t^2 > 0$, or equivalently that $\sigma_t > 0$, $\sigma_t \in \mathcal{F}_{t-1}$ and $E(\eta_t^2 | \mathcal{F}_{t-1}) = 1$. For the usual GARCH models, \mathcal{F}_{t-1} is simply the sigma-field generated by the past returns $\{\varepsilon_u, u < t\}$, and the volatility has a parametric form $\sigma_t = \sigma(\varepsilon_u, u < t; \boldsymbol{\theta}_0)$, where $\boldsymbol{\theta}_0$ is a vector of parameters. It is however often the case that some extra information is available, under the form of a vector \mathbf{x}_{t-1} of exogenous covariates, such as the daily volume of transactions, or high frequency intraday data, or even series of other returns. It is natural to try to take advantage of the extra information, in order to improve the prediction of the squares. To incorporate the information conveyed by $\{\mathbf{x}_u, u < t\}$ into \mathcal{F}_{t-1} , researchers have considered GARCH models augmented with additional explanatory variables, the so-called GARCH-X models, which are of the form $\sigma_t = \sigma(\varepsilon_u, \mathbf{x}_u, u < t; \boldsymbol{\vartheta}_0)$, where $\boldsymbol{\vartheta}_0$ is a vector of parameters including a parameter $\boldsymbol{\theta}_0$ specific to the past returns and a parameter $\boldsymbol{\pi}_0$ related to the exogenous covariates (see *e.g.* Engle and Patton (2001) and the references therein).

In practice, the difficulties are the choice of the parametric form (as illustrated by Bollerslev (2008), there exists a plethora of GARCH formulations) and the estimation of the parameter $\boldsymbol{\vartheta}_0$. The two problems are closely related. For GARCH, as well as for GARCH-X models, the coefficients are generally positively constrained, and tests of nullity of some components of $\boldsymbol{\vartheta}_0$ help to find a parsimonious GARCH-X formulation. The usual estimator of the GARCH models is the quasi-maximum likelihood estimator (QMLE), which does not require to specify a particular distribution for the error term η_t . The consistency of the QMLE does not even require that (η_t) be iid, which is particularly relevant for GARCH-X models (see Remarks 3.2 and 2.4 below). The asymptotic normality however requires that the true value of the parameter belongs to the interior of the parameter space, which is generally not the case when components of $\boldsymbol{\vartheta}_0$ are equal to zero.

Questions that seem particularly relevant in the GARCH-X framework are : is it really useful to introduce covariates in the volatility ? which covariates should we add to \mathcal{F}_{t-1} ? how many lagged values should we consider in the GARCH formulation ?

Each of these questions can be discussed by testing the nullity of certain components of $\boldsymbol{\vartheta}_0$. It is thus of interest to study the behaviour of the estimator $\hat{\boldsymbol{\vartheta}}_n$ of $\boldsymbol{\vartheta}_0$ when this parameter may stand at the boundary of the parameter space. To our knowledge, this problem has not yet been explicitly considered for GARCH-X models. This will be the focus of this paper. We now present the class of GARCH-X that we will consider, and then we detail the main objectives of the paper.

2.1.1 The model

Let $x^+ = \max(x, 0)$ and $x^- = \max(-x, 0)$. We consider the model defined by

$$\begin{cases} \varepsilon_t = h_t^{1/\delta} \eta_t \\ h_t = \omega_0 + \sum_{i=1}^q \{ \alpha_{0i+} (\varepsilon_{t-i}^+)^{\delta} + \alpha_{0i-} (\varepsilon_{t-i}^-)^{\delta} \} + \sum_{j=1}^p \beta_{0j} h_{t-j} + \boldsymbol{\pi}'_0 \mathbf{x}_{t-1}, \end{cases} \quad (2.2)$$

where $\mathbf{x}_t = (x_{1,t}, \dots, x_{r,t})'$ is a vector of r exogenous covariates. To ensure that $h_t > 0$ with probability one, assume that the covariates are almost surely positive and that the coefficients satisfy $\alpha_{0i+} \geq 0$, $\alpha_{0i-} \geq 0$, $\beta_{0j} \geq 0$, $\omega_0 > 0$, $\delta > 0$ and $\boldsymbol{\pi}_0 = (\pi_{01}, \dots, \pi_{0r})' \geq 0$ componentwise.

In absence of covariates, *i.e.* when $\boldsymbol{\pi}_0 = 0$, this equation corresponds to the Asymmetric Power GARCH (APARCH) model introduced by [Ding et al. \(1993\)](#). Model (2.2) can thus be called APARCH-X. The APARCH is rather general, since it nests numerous ARCH-type parameterizations used by the practitioners. The standard GARCH is obtained with $\delta = 2$ and $\alpha_{0i-} = \alpha_{0i+}$. Motivated by the fact that the autocorrelations are often larger for $|\varepsilon_t|$ than for ε_t^2 , [Taylor \(1986\)](#) proposed the model with $\delta = 1$ and $\alpha_{0i-} = \alpha_{0i+}$. When $\alpha_{0i-} > \alpha_{0i+}$, a negative return has a higher impact on the future volatility than a positive return of the same magnitude. This is a well-documented stylized fact that is called "leverage effect". Two widely used models that allow for the leverage effect are the TARARCH of [Zakoïan \(1994\)](#), obtained with $\delta = 1$, and the GJR of ([Glosten et al., 1993](#)), obtained with $\delta = 2$. One popular ARCH formulation that is not nested by the APARCH is the EGARCH model of [Nelson \(1991\)](#). The inference of the EGARCH is however quite difficult, and the behaviour of the QMLE is still partially unknown for this model (see [Wintenberger \(2013\)](#)). Another exponential formulation that is not encompassed by (2.2) is the log-GARCH model (see [Sucarrat and Escribano \(2010\)](#)).

2.1.2 The objectives

The most comprehensible results concerning the inference of the APARCH model can be found in [Pan et al. \(2008\)](#) and in [Hamadeh and Zakoïan \(2011\)](#) (HZ hereafter).¹ To our knowledge, there exists no general result concerning the estimation of the APARCH-X model. Actually, even if practitioners often add exogenous variables to volatility models, the probabilistic properties and the statistical inference of ARCH models with exogenous variables have not been yet extensively studied. In particular, [Han and Kristensen \(2014\)](#) studied the inference of the standard GARCH(1,1) model augmented by an additional covariate. With the aim to detect simultaneously causality-in-mean and causality-in-variance, [Guo et al. \(2014\)](#) studied the factor double autoregressive model, whose a particular case is an ARCH-X model in which the volatility depends on lagged values of a covariate. [Kristensen and Rahbek \(2005\)](#) derived the asymptotic properties of the QMLE for a wide class of

1. Note that the APARCH model is called Power-Transformed Threshold GARCH in these two papers

ARCH models in which the volatility of the observed process y_t depends on a observed covariate x_{t-1} . Since this covariate is assumed to be measurable with respect to the past observations (see their assumption C3 (ii)), their model does not fall exactly into the class of the ARCH-X.

To summarize, the few existing references dealing with the estimation of volatility models with exogenous variables concern specific models. In particular, these models are unable to take into account the leverage effect that is usually observed in financial series. More importantly, a common assumption to the above-mentioned papers is that the true value of the parameter belongs to the interior of the parameter space. Under this assumption, and other regularity conditions, the QMLE is asymptotically normally distributed. When the parameter belongs to the boundary of the parameter space, the asymptotic distribution of the QMLE may be non standard (see [Andrews \(2001\)](#) for a general reference, [Francq and Zakoïan \(2007\)](#) for an application to univariate GARCH, and [Pedersen \(2015\)](#) for a recent application to multivariate GARCH models). An important consequence of the non-normality of the QMLE is that the usual t -ratio or the Wald tests employed to identify the order p and q are also non standard (see *e.g.* [Francq and Zakoïan \(2009b\)](#) and the reference therein).

Our first objective is thus to study the asymptotic distribution of the QMLE of the APARCH-X model when the parameter is not restricted to belong to the interior of the parameter space. For the applications we have in mind, the covariates can be for instance lagged values of other squared returns, or realized volatilities, or positive and negative parts of relative volume increments. The covariates will be supposed to be positive and stationary, but they are allowed to be strongly correlated, and also correlated with η_t . Therefore, the covariates will not be weakly or strongly exogenous in the sense of [Engle et al. \(1983\)](#), but we can say that the $x_{i,t}$'s are exogenous variables in the sense that their dynamics is not specified by the APARCH-X model.

Our second objective is to propose tests of nullity for one or several components of ϑ_0 . This is closely related to the first objective because, due to the positivity constraints on the components of ϑ_0 , under the null hypothesis, the true parameter stands at the boundary of the parameter space.

The remainder of the paper is organized as follows. In [Section 2.2](#), we first discuss the strict stationarity. We then introduce the Gaussian quasi-maximum likelihood estimator for APARCH-X model ([2.2](#)) and derive conditions for its consistency. The asymptotic distribution of the QMLE is studied conditioning on different information sets. We also consider the problem of testing the nullity of certain coefficients. Our asymptotic results are obtained under regularity conditions that are close to be optimal. They even allow to improve the existing results for GARCH models without exogenous variables (see [Remark 2.8](#) below). The simulation results and two real data applications are presented in [Section 3.4](#). [Section 3.5](#) concludes the paper. All the proofs are collected in [Section 3.6](#).

2.2 Main results

We first discuss the strict stationarity, which will be the main condition for the consistency of the QMLE.

2.2.1 Strict stationarity

Without loss of generality, assume that $p \geq 2$ and $q \geq 2$. Let the vector of dimension $d_0 = 2q + p - 2$

$$\mathbf{Y}_t = \left(h_{t+1}, \dots, h_{t-p+2}, (\varepsilon_t^+)^{\delta}, (\varepsilon_t^-)^{\delta}, \dots, (\varepsilon_{t-q+2}^+)^{\delta}, (\varepsilon_{t-q+2}^-)^{\delta} \right)'$$

It is easy to see that (ε_t) satisfies (2.2) if and only if

$$\mathbf{Y}_t = \mathbf{C}_{0t} \mathbf{Y}_{t-1} + \mathbf{B}_{0t}, \quad (2.3)$$

where $\mathbf{B}_{0t} = (\omega_0 + \boldsymbol{\pi}'_0 \mathbf{x}_t, 0, \dots, 0)'$ is a vector of dimension d_0 and \mathbf{C}_{0t} is a $d_0 \times d_0$ matrix. Denoting by $\mathbf{C}_{0t}(i, j)$ the element of the row i and column j of \mathbf{C}_{0t} , we have

$$\begin{aligned} \mathbf{C}_{0t}(1, 1) &= \alpha_{01+} (\eta_t^+)^{\delta} + \alpha_{01-} (\eta_t^-)^{\delta} + \beta_{01}, & \mathbf{C}_{0t}(1, j) &= \beta_{0j} \quad \text{for } j = 2, \dots, p, \\ \mathbf{C}_{0t}(1, p+2i-3) &= \alpha_{0i+} \text{ and } \mathbf{C}_{0t}(1, p+2i-2) &= \alpha_{0i-} \quad \text{for } i = 2, \dots, q, \\ \mathbf{C}_{0t}(i, i-1) &= 1 \quad \text{for } i = 2, \dots, p, \\ \mathbf{C}_{0t}(p+1, 1) &= (\eta_t^+)^{\delta}, & \mathbf{C}_{0t}(p+2, 1) &= (\eta_t^-)^{\delta} \\ \mathbf{C}_{0t}(p+2i-1, p+2i-3) &= \mathbf{C}_{0t}(p+2i, p+2i-2) &= 1 \quad \text{for } i = 2, \dots, q-1, \end{aligned}$$

and all the other elements of \mathbf{C}_{0t} are equal to zero. Now assume that

A1 : (η_t, \mathbf{x}'_t) is a strictly stationary and ergodic process, and there exists $s > 0$ such that $E|\eta_1|^s < \infty$ and $E\|\mathbf{x}_1\|^s < \infty$.

Note that, for GARCH-type models of the form (2.2), the sequence (η_t) is usually assumed to be a white noise, but this assumption is not necessary. Following Brandt (1986) and Bougerol and Picard (1992a), the stationarity relies on the top Lyapunov exponent

$$\gamma := \lim_{t \rightarrow \infty} \frac{1}{t} \log \|\mathbf{C}_{0t} \mathbf{C}_{0,t-1} \cdots \mathbf{C}_{01}\| \quad \text{a.s.},$$

which is well defined in $[-\infty, +\infty)$ because $E \log^+ \|\mathbf{C}_{01}\| < \infty$ under the condition $E|\eta_1|^s < \infty$ (see (A.5) in Pan et al. (2008)). It is shown in the previous reference that when (η_t) is iid and satisfies some regularity conditions, there exists a unique strictly stationary solution to the APARCH model if and only if $\gamma < 0$. The following lemma shows that the condition is the same for the APARCH-X.

Lemma 2.1. *Suppose that **A1** is satisfied. If $\gamma < 0$, the APARCH-X equation (2.2) (or equivalently (2.3)) admits a unique strictly stationary, non anticipative and ergodic solution. The solution of (2.3) is given by*

$$\mathbf{Y}_t = \mathbf{B}_{0t} + \sum_{k=1}^{\infty} \left(\prod_{i=1}^k \mathbf{C}_{0,t-i+1} \right) \mathbf{B}_{0,t-k}.$$

When $\gamma \geq 0$, there exists no stationary solution to (2.2) and to (2.3).

The proof is omitted since it is quite similar to Bougerol and Picard (1992b). Pan et al. (2008) also obtained the same result for the APARCH without exogenous variables. They used different assumptions, but a careful examination of their proof and of Lemma 3.4 in Bougerol and Picard (1992b) reveals that the assumption that (η_t) is iid, as well as their regularity condition **A1**, are useless in our framework. Moreover, the introduction of exogenous variables does not complicate anything since, under our assumption **A1**, the proof is unchanged when the intercept ω_0 is replaced by $\varpi_t = \omega_0 + \boldsymbol{\pi}'_0 \mathbf{x}_t$.

Remark 2.1. *Note that Lemma 2.1 can also be used when $p < 2$ or $q < 2$. Consider for instance the APARCH-X(1,1) model, i.e. (2.2) with $p = q = 1$. This is equivalent to considering a model with $p = q = 2$ and $\alpha_{02+} = \alpha_{02-} = \beta_{02} = 0$. In that case we have*

$$\mathbf{C}_{0t} = \begin{pmatrix} a(\eta_t) & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ (\eta_t^+)^{\delta} & 1 & 0 & 0 \\ (\eta_t^-)^{\delta} & 0 & 1 & 0 \end{pmatrix}, \quad a(z) = \alpha_{0+}(z^+)^{\delta} + \alpha_{0-}(z^-)^{\delta} + \beta_0,$$

with the simplified notations $\alpha_{0+} = \alpha_{01+}$, $\alpha_{0-} = \alpha_{01-}$ and $\beta_0 = \beta_{01}$. In view of the shape of \mathbf{C}_{0t} , the top Lyapunov then takes the explicit form

$$\gamma = E \log a(\eta_1). \quad (2.4)$$

The condition $\alpha_{0+}E(\eta_1^+)^{\delta} + \alpha_{0-}E(\eta_1^-)^{\delta} + \beta_0 < 1$ then implies the strict stationarity condition $\gamma < 0$. It can be checked directly that, under **A1** and $\gamma < 0$, the APARCH-X(1,1) model admits a strictly stationary solution, whose volatility is given by

$$h_t = \sum_{k=0}^{\infty} \prod_{i=1}^k a(\eta_{t-i}) \varpi_{t-k-1}, \quad (2.5)$$

with the convention $\prod_{i=1}^k a(\eta_{t-i}) = 1$ when $k = 0$, and $\varpi_t = \omega_0 + \boldsymbol{\pi}'_0 \mathbf{x}_t$.

Remark 2.2. *It has to be noted that the top Lyapunov coefficient γ does not depend on the exogenous variables \mathbf{x}_t . Therefore, as long as **A1** holds, the presence of exogenous variables does not modify the*

stationarity constraints. This is not surprising because, for the APARCH model without exogenous variables, it is known that the stationarity condition does not depend on the intercept ω_0 and, as illustrated by (2.5) in the APARCH(1,1) case, adding exogenous variables essentially amounts to replace ω_0 by $\varpi_t = \omega_0 + \boldsymbol{\pi}'_0 \mathbf{x}_t$. It is not forbidden to take $\mathbf{x}_t = \varepsilon_t$, but of course Assumption **A1** entails that (\mathbf{x}_t) is stationary, and in this case, the lemma becomes trivial.

2.2.2 Strong consistency of the QMLE

Hamadeh and Zakoian (2011) showed that, for APARCH models, the power parameter δ is difficult to be estimated in practice. The quasi-likelihood being very flat in the direction of δ , estimating this parameter leads to imprecise results and considerably slows down the optimization routine. We therefore consider that δ is fixed. In many applications, $\delta = 1$ (as in the TARCH) or $\delta = 2$ (as in the GJR model). Let $d = 2q + p + r + 1$ be the remaining number of unknown parameters. A generic element of the parameter space $\Theta \subseteq (0, +\infty) \times [0, +\infty)^{d-1}$ is denoted by

$$\boldsymbol{\vartheta} = (\omega, \alpha_{1+}, \alpha_{1-}, \dots, \alpha_{q+}, \alpha_{q-}, \beta_1, \dots, \beta_p, \boldsymbol{\pi}')'.$$

Denote by $\overset{\circ}{\Theta}$ the interior of Θ . Let $(\varepsilon_1, \dots, \varepsilon_n)$ be a realization of length n of the stationary solution (ε_t) of the APARCH-X model (2.2), and let $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ be the corresponding observations of the exogenous variables. Given initial values $\varepsilon_{1-q}, \dots, \varepsilon_0, \tilde{\sigma}_{1-p} \geq 0, \dots, \tilde{\sigma}_0 \geq 0, \mathbf{x}_0 \geq 0$, the Gaussian quasi-likelihood is given by

$$L_n(\boldsymbol{\vartheta}) = L_n(\boldsymbol{\vartheta}, \varepsilon_1, \dots, \varepsilon_n, \mathbf{x}_1, \dots, \mathbf{x}_n) = \prod_{t=1}^n \frac{1}{\sqrt{2\pi\tilde{\sigma}_t^2}} \exp\left\{\frac{-\varepsilon_t^2}{2\tilde{\sigma}_t^2}\right\},$$

where the $\tilde{\sigma}_t$'s are defined recursively, for $t \geq 1$, by

$$\tilde{\sigma}_t^\delta = \tilde{\sigma}_t^\delta(\boldsymbol{\vartheta}) = \omega + \sum_{i=1}^q \left\{ \alpha_{i+} (\varepsilon_{t-i}^+)^{\delta} + \alpha_{i-} (\varepsilon_{t-i}^-)^{\delta} \right\} + \sum_{j=1}^p \beta_j \tilde{\sigma}_{t-j}^\delta + \boldsymbol{\pi}' \mathbf{x}_{t-1}.$$

The QMLE of $\boldsymbol{\vartheta}_0$ is defined as any measurable solution $\hat{\boldsymbol{\vartheta}}_n$ of

$$\hat{\boldsymbol{\vartheta}}_n = \arg \max_{\boldsymbol{\vartheta} \in \Theta} L_n(\boldsymbol{\vartheta}) = \arg \min_{\boldsymbol{\vartheta} \in \Theta} \tilde{Q}_n(\boldsymbol{\vartheta}), \quad (2.6)$$

where

$$\tilde{Q}_n(\boldsymbol{\vartheta}) = \frac{1}{n} \sum_{t=1}^n \tilde{\ell}_t, \quad \tilde{\ell}_t = \tilde{\ell}_t(\boldsymbol{\vartheta}) = \frac{\varepsilon_t^2}{\tilde{\sigma}_t^2} + \ln \tilde{\sigma}_t^2. \quad (2.7)$$

Let $\mathcal{A}_{\boldsymbol{\vartheta}+}(z) = \sum_{i=1}^q \alpha_{i+} z^i$, $\mathcal{A}_{\boldsymbol{\vartheta}-}(z) = \sum_{i=1}^q \alpha_{i-} z^i$ and $\mathcal{B}_{\boldsymbol{\vartheta}}(z) = 1 - \sum_{j=1}^p \beta_j z^j$. Denote by \mathcal{F}_{t-1} the σ -field generated by $\{\varepsilon_u, \mathbf{x}_u, u < t\}$ and denote by $\mathcal{F}_{t,i}$ the σ -field generated by $\{\eta_{t-j}, j >$

$i, \mathbf{x}_{t-k}, k > 0\}$ (so that $\mathcal{F}_{t-1} = \mathcal{F}_{t,0}$). To show the strong consistency of the QMLE, we need the following assumptions.

A2 : $E(\eta_t | \mathcal{F}_{t-1}) = 0$ and $E(\eta_t^2 | \mathcal{F}_{t-1}) = 1$.

A3 : $\vartheta_0 \in \Theta$, Θ is compact.

A4 : for all $i \geq 1$ (and all $t \in \mathbb{Z}$), the support of the distribution of η_{t-i} given $\mathcal{F}_{t,i}$ is not included in $[0, \infty)$ or in $(-\infty, 0]$ and contains at least three points.

Assumption **A4** is an identifiability condition which prevents taking redundant explanatory variables in the volatility, for instance $\mathbf{x}_{t-1} = (\varepsilon_{t-i}^+)^{\delta}$ (see Remarks 2.5 and 2.6 below). If Assumption **A4** does not hold, then there exists $i > 0$ such that the law of η_{t-i} given the information $\mathcal{F}_{t,i}$ contained in $\{\eta_{t-i-1}, \eta_{t-i-1}, \dots; x_{t-1}, x_{t-2}, \dots\}$ is "degenerated". In such a case, either ε_{t-i}^+ or ε_{t-i}^- can be "predicted" from $\mathcal{F}_{t,i}$, and the volatility can be written in several ways.

A5 : $\gamma < 0$ and $\sum_{j=1}^p \beta_j < 1$ for all $\vartheta \in \Theta$.

A6 : there exists $s > 0$, such that $Eh_t^s < \infty$ and $E|\varepsilon_t|^s < \infty$.

A7 : if $p > 0$, $\mathcal{B}_{\vartheta_0}(z)$ has no common root with $\mathcal{A}_{\vartheta_0+}(z)$ and $\mathcal{A}_{\vartheta_0-}(z)$; $\mathcal{A}_{\vartheta_0+}(1) + \mathcal{A}_{\vartheta_0-}(1) \neq 0$ and $\alpha_{0q+} + \alpha_{0q-} + \beta_{0p} \neq 0$ (with the notation $\alpha_{00+} = \alpha_{00-} = \beta_{00} = 1$).

A8 : if \mathbf{c} is a non- zero vector of \mathbb{R}^r then $\mathbf{c}'\mathbf{x}_1$ is not degenerated.

Assumptions **A3**, **A5** and **A7** have already been used to show the consistency of the QMLE for GARCH models. Assumption **A8** is an identifiability condition which is obviously necessary to avoid multicollinearity of the explanatory variables. The following remarks concern respectively **A2** and **A4**.

Remark 2.3. *Assumptions **A1** and **A2** entail that (η_t, \mathcal{F}_t) is a conditionally homoscedastic martingale difference. For the GARCH-type models, it is usual to assume the stronger assumption that (η_t) is iid $(0, 1)$. Note, however, that Escanciano (2009) and Han and Kristensen (2014) employed **A2**. The advantage of using **A2** is that (2.2) becomes a semi-strong model, that can be satisfied for different σ -fields \mathcal{F}_t , corresponding for example to different sequences of exogenous variables (\mathbf{x}_t) . Indeed, **A2** is satisfied for a model of the form (2.1) whenever $E(\varepsilon_t | \mathcal{F}_{t-1}) = 0$ and $E(\varepsilon_t^2 | \mathcal{F}_{t-1}) = \sigma_t^2 > 0$. With APARCH-X models, for which several information sets \mathcal{F}_t can be naturally investigated, Assumption **A2** seems thus more flexible than the iid assumption.*

Remark 2.4. *Let us give an example of a data generated process for which several GARCH-X models of the form (2.2) coexist under the semi-strong noise Assumption **A2**. Assume that $\mathbf{X}_t = (\varepsilon_t, y_t)'$ follows the bivariate GARCH model $\mathbf{X}_t = \Sigma_t^{1/2} \boldsymbol{\eta}_t$, where $\Sigma_t = \text{diag}(\sigma_{1,t}^2, \sigma_{2,t}^2)$ with $\boldsymbol{\eta}_t$ iid $\mathcal{N}(0, \mathbf{I}_2)$, and $\sigma_{i,t}^2 = \omega_i + \alpha_i \varepsilon_{t-1}^2 + \beta_i \sigma_{i,t-1}^2 + \pi_i y_{t-1}^2$ for $i = 1, 2$. The process (ε_t) thus follows a (strong) GARCH-X(1,1) model with exogenous variable $x_t = y_t^2$. Nijman and Sentana (1996) showed that (ε_t) also follows a GARCH(2,2) model, without exogenous variable, but with a semi-strong noise satisfying **A2**, which is not independent in general.*

Remark 2.5. Note that when there is no covariate and when (η_t) is iid, Assumption **A4** reduces to

$$P[\eta_1 > 0] \in (0, 1) \text{ and the support of the distribution of } \eta_1 \text{ contains at least 3 points,}$$

which is exactly the identifiability condition **A2** of HZ. When there exist covariates, **A4** rules out the existence of collinearities between the exogenous variables and the functions of the past returns involved in the volatility. For example, the assumption precludes that $\mathbf{c}'\mathbf{x}_{t-1} = (\varepsilon_{t-i}^+)^{\delta}$ with $\mathbf{c} \in \mathbb{R}^r$ (otherwise the variable $(\eta_{t-i}^+)^{\delta}$ given $\mathcal{F}_{t,i}$ would be degenerated, which is impossible under **A4**).

The following lemma, which was obtained by Berkes et al. (2003) for standard GARCH and by Ling (2007) for ARMA-GARCH models, shows that **A6** can be suppressed when (η_t) is iid.

Lemma 2.2. If $\gamma < 0$ and Assumptions **A1-A2** hold with (η_t) iid $(0,1)$, then **A6** is satisfied.

We are now in a position to show the consistency of the QMLE.

Theorem 2.1. Let $\hat{\boldsymbol{\vartheta}}_n$ be a sequence of QMLE satisfying (2.6). Then, under **A1-A8**,

$$\hat{\boldsymbol{\vartheta}}_n \rightarrow \boldsymbol{\vartheta}_0 \text{ a.s. as } n \rightarrow \infty.$$

Remark 2.6. The case of the standard GARCH(1,1) with one covariate ($\delta = 2$, $p = q = r = 1$ and $\alpha_{i+} = \alpha_{i-}$) has been considered by Han and Kristensen (2014), with possibly non-stationary (x_t) . However, our regularity conditions do not coincide with those of this paper. In particular, the identifiability condition given in Han and Kristensen (2014) is, with the notation of the present paper, for any $(a, b) \in \mathbb{R}^2$

$$(a, b) \neq (0, 0) \implies a\eta_t^2 + bx_t^2 \mid \mathcal{F}_{t-1} \text{ has a non degenerated distribution.} \quad (2.8)$$

One has to preclude the covariate $x_t^2 = \varepsilon_t^2$, since the parameters α and π of the volatility $\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \pi x_{t-1}^2$ are clearly not identifiable in that case. Note that we then have $\mathcal{F}_{t,1} = \sigma(\eta_{t-1}^2, \eta_{t-2}, \eta_{t-3}, \dots)$, and the distribution of $\eta_{t-1} \mid \mathcal{F}_{t,1}$ is concentrated on two points. The condition **A4** thus excludes the covariate $x_t^2 = \varepsilon_t^2$. This is not the case of (2.8) because $a\eta_t^2 + bx_t^2 \mid \mathcal{F}_{t-1} = (a + b\sigma_t^2)\eta_t^2 \mid \mathcal{F}_{t-1}$ is not degenerated when $(a, b) \neq (0, 0)$ (and η_t^2 is not degenerated). Thus (2.8) is not sufficient to ensure the identifiability.

2.2.3 Asymptotic distribution of the QMLE

For the computation of (2.7), it is necessary to have $\tilde{\sigma}_t(\boldsymbol{\vartheta}) > 0$ almost surely, for any $\boldsymbol{\vartheta} \in \Theta$. This is why the components of $\boldsymbol{\vartheta} \in \Theta$ are constrained to be non negative. More precisely, it can be assumed that, for $i = 2, \dots, d$, the i -th section of Θ is $[0, K_i]$ with $K_i > 0$ (the first section being $[\underline{\omega}, \bar{\omega}]$ with $0 < \underline{\omega} < \bar{\omega}$). If Θ is of this form and is large enough (to avoid, for instance, that the i -th component $\boldsymbol{\vartheta}_{0i}$ of $\boldsymbol{\vartheta}_0$ be greater than K_i), the following assumption is satisfied.

A9 : $\mathcal{C} := \cup_{n=1}^{\infty} \{\sqrt{n}(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0) : \boldsymbol{\vartheta} \in \Theta\} = \prod_{i=1}^d \mathcal{C}_i$, where $\mathcal{C}_i = [0, +\infty)$ when $\boldsymbol{\vartheta}_{0i} = 0$ and $\mathcal{C}_i = \mathbb{R}$ otherwise.

The set $\mathcal{C} = \mathcal{C}_{\boldsymbol{\vartheta}_0}$ will be called the local parameter space. Take the simple example of Model (2.2) with $q = 1, p = 0$ and without exogenous variables. Assume that the parameter space is of the form $\Theta = [\underline{\omega}, \bar{\omega}] \times [0, \bar{\alpha}_+] \times [0, \bar{\alpha}_-]$, with obvious notations. Assumption **A9** is satisfied if $\underline{\omega} < \omega_0 < \bar{\omega}$, $\bar{\alpha}_+ > \alpha_{01+}$, and $\bar{\alpha}_- > \alpha_{01-}$ (*i.e.* if Θ is large enough). We then have, for instance, $\mathcal{C} = \mathbb{R} \times [0, \infty) \times \mathbb{R}$ when $\alpha_{01+} = 0$ and $\alpha_{01-} > 0$. More generally, \mathcal{C} is a convex cone, which is equal to \mathbb{R}^d if and only if $\boldsymbol{\vartheta}_0$ belongs to the interior $\overset{\circ}{\Theta}$ of Θ , *i.e.* if all the components of $\boldsymbol{\vartheta}_0$ are non-zero, under **A9**.

For standard GARCH models, without covariates and with (η_t) iid, note that η_t is independent of \mathcal{F}_{t-1} . In that situation, it is known that no moment condition on ε_t is needed for the consistency and asymptotic normality (CAN) of the QMLE when the GARCH parameter belongs to $\overset{\circ}{\Theta}$, whereas moments conditions are required when the parameter stands at the boundary of the parameter space (see the example given in Section 3.1 of Francq and Zakoïan (2007)). When the model is semi-strong, *i.e.* when η_t is not independent of \mathcal{F}_{t-1} , stronger moment conditions will be required. We thus distinguish four cases :

- Case A : η_t is independent of \mathcal{F}_{t-1} and all the components of $\boldsymbol{\vartheta}_0$ are strictly positive ;
- Case B : η_t is independent of \mathcal{F}_{t-1} and at least one component of $\boldsymbol{\vartheta}_0$ is equal to zero ;
- Case C : η_t is not independent of \mathcal{F}_{t-1} and all the components of $\boldsymbol{\vartheta}_0$ are strictly positive ;
- Case D : η_t is not independent of \mathcal{F}_{t-1} and at least one component of $\boldsymbol{\vartheta}_0$ is equal to zero.

It is however worth considering Cases B and D, in particular, because this gives the asymptotic distribution of the QMLE under the null that $\boldsymbol{\pi}_0 = 0$. For simplicity, these four cases are referred to respectively as strong in the interior, strong at the boundary, semi-strong in the interior and semi-strong at the boundary. We assume that

A10 : $E\eta_t^4 < \infty$ in Cases A and B, and $E|\eta_t|^{4+\nu} < \infty$ for some $\nu > 0$ in Cases C and D.

A11 : For the power δ of the APARCH model (2.2) : in Case B we have $E|\varepsilon_t|^{2\delta+\nu_1} < \infty$ and $E\|\mathbf{x}_t\|^{2+\nu_1} < \infty$ for some $\nu_1 > 0$; in Case D we have $E|\varepsilon_t|^{2\delta+8\delta/\nu} < \infty$ and $E\|\mathbf{x}_t\|^{2+8/\nu} < \infty$ with ν satisfying **A10**.

It will be convenient to approximate the sequence $(\tilde{\ell}_t(\boldsymbol{\vartheta}))$ by an ergodic stationary sequence. Therefore, denote by $(\sigma_t^\delta)_t = \{\sigma_t^\delta(\boldsymbol{\vartheta})\}_t$ the strictly stationary, ergodic and non-anticipative solution of

$$\sigma_t^\delta = \omega + \sum_{i=1}^q \left[\alpha_{i+} (\varepsilon_{t-i}^+)^{\delta} + \alpha_{i-} (\varepsilon_{t-i}^-)^{\delta} \right] + \sum_{j=1}^p \beta_j \sigma_{t-j}^\delta + \boldsymbol{\pi}' \mathbf{x}_{t-1}. \quad (2.9)$$

Note that $\sigma_t^\delta(\boldsymbol{\vartheta}_0) = h_t$. Let $Q_n(\boldsymbol{\vartheta})$ and ℓ_t be obtained by replacing $\tilde{\sigma}_t^\delta$ with σ_t^δ in $\tilde{Q}_n(\boldsymbol{\vartheta})$ and $\tilde{\ell}_t$.

Remark 2.7. In order to explain the need of the moment condition **A11**, let us consider the simple example of an ARCH(1) with one covariate : $\sigma_t^2 = \omega_0 + \alpha_0 \varepsilon_{t-1}^2 + \pi_0 x_{t-1}$. The score is then equal to

$$\frac{\partial \ell_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}} = (1 - \eta_t^2) \mathbf{D}_t, \quad \mathbf{D}_t = \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}} = \frac{1}{\omega_0 + \alpha_0 \varepsilon_{t-1}^2 + \pi_0 x_{t-1}} \begin{pmatrix} 1 \\ \varepsilon_{t-1}^2 \\ x_{t-1} \end{pmatrix}.$$

In Cases A and C, the components of \mathbf{D}_t are bounded respectively by $1/\omega_0$, $1/\alpha_0$ and $1/\pi_0$. For more general GARCH models, the components of \mathbf{D}_t are not bounded, but they admit moments of any order. This explains why, when $\boldsymbol{\vartheta}_0 \in \overset{\circ}{\Theta}$, no moment assumption is required, except those of **A10**, for the existence of the variance of the score (which is a condition for having a CLT for the sequence of the standardized sum of the scores). On the contrary, when $\alpha_0 = \pi_0 = 0$, the existence of the variance of the score requires $E\eta_t^4 \varepsilon_{t-1}^4 < \infty$ and $E\eta_t^4 x_{t-1}^2 < \infty$, which gives $E\eta_t^4 < \infty$, $E\varepsilon_t^4 < \infty$ and $Ex_t^2 < \infty$ in Case B, and a stronger condition in Case D.

Hall and Yao (2003) showed that the asymptotic distribution of the QMLE is in general non standard when the score does not admit a second order moment. Even in the case where the score belongs to the domain of attraction of the Gaussian law, they found a rate of convergence different from \sqrt{n} .

Under the previous assumptions, Lemma 2.4 in Section 3.6 below shows that the matrix

$$\mathbf{J} := E \left(\frac{\partial^2 \ell_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \right) = \frac{4}{\delta^2} E \left(\frac{1}{\sigma_t^{2\delta}(\boldsymbol{\vartheta}_0)} \frac{\partial \sigma_t^\delta(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}} \frac{\partial \sigma_t^\delta(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}'} \right) \quad (2.10)$$

is positive definite. Let us thus consider the norm $\|\mathbf{x}\|_{\mathbf{J}}^2 = \mathbf{x}' \mathbf{J} \mathbf{x}$ and the scalar product $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{J}} = \mathbf{x}' \mathbf{J} \mathbf{y}$ for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$. In the sense of this scalar product, the orthogonal projection of a vector $\mathbf{Z} \in \mathbb{R}^d$ on \mathcal{C} is defined by

$$\mathbf{Z}^{\mathcal{C}} = \arg \inf_{\mathbf{C} \in \mathcal{C}} \|\mathbf{C} - \mathbf{Z}\|_{\mathbf{J}} \quad (2.11)$$

or equivalently by

$$\mathbf{Z}^{\mathcal{C}} \in \mathcal{C} \quad \text{and} \quad \langle \mathbf{Z} - \mathbf{Z}^{\mathcal{C}}, \mathbf{C} - \mathbf{Z}^{\mathcal{C}} \rangle_{\mathbf{J}} \leq 0, \quad \forall \mathbf{C} \in \mathcal{C}. \quad (2.12)$$

Theorem 2.2. Under the assumptions of Theorem 2.1 and **A9–A11**, as $n \rightarrow \infty$,

$$\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) \xrightarrow{d} \mathbf{Z}^{\mathcal{C}}, \quad \text{where } \mathbf{Z} \sim \mathcal{N}\{0, \boldsymbol{\Sigma} := \mathbf{J}^{-1} \mathbf{I} \mathbf{J}^{-1}\}, \quad (2.13)$$

\mathbf{J} is defined by (2.10) and

$$\mathbf{I} = \frac{4}{\delta^2} E \left[\left\{ E(\eta_t^4 | \mathcal{F}_{t-1}) - 1 \right\} \frac{1}{\sigma_t^{2\delta}(\boldsymbol{\vartheta}_0)} \frac{\partial \sigma_t^\delta(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}} \frac{\partial \sigma_t^\delta(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}'} \right].$$

Remark 2.8. *The previous theorem provides the asymptotic distribution of the QMLE in each of the Cases A-D. In all cases, Assumptions A1-A9 are required. Note that, in Cases A and B, we have $\mathbf{I} = (E\eta_1^4 - 1)\mathbf{J}$. In Cases A and C, the local parameter space is $\mathcal{C} = \mathbb{R}^d$, and the asymptotic distribution of the QMLE is thus normal :*

$$\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) \xrightarrow{d} \mathcal{N}\{0, (E\eta_1^4 - 1)\mathbf{J}^{-1}\} \quad \text{in Case A} \quad (2.14)$$

and

$$\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) \xrightarrow{d} \mathcal{N}\{0, \mathbf{J}^{-1}\mathbf{I}\mathbf{J}^{-1}\} \quad \text{in Case C.} \quad (2.15)$$

This result is obtained under the assumption that $E\eta_t^4 < \infty$ in Case A and a slightly stronger condition in Case C (see A10), but without moment condition on the observed process ε_t . When there is no covariate ($r = 0$), we retrieve the results obtained by Francq (2004) in the GARCH case ($\delta = 2$ and $\alpha_{0i+} = \alpha_{0i-}$) and when (η_t) is iid, by Escanciano (2009) in the GARCH case when (η_t) is a conditionally homoscedastic martingale difference, and by HZ in the general APARCH case.

Under the moment condition $E|\varepsilon_t|^6 < \infty$, and other regularity conditions similar to ours, Francq and Zakoian (2007) obtained (2.13) in the particular case of GARCH models with (η_t) iid and $r = 0$ (see also Andrews (1999) and the references therein for the boundary problem in a more general estimation framework). Note that A11 in Case B is a milder condition than $E|\varepsilon_t|^6 < \infty$, therefore Theorem 2.2 slightly improves the results available for the standard GARCH models. Even when $r = 0$, the authors are not aware of the existence of conditions entailing (2.13) for the general APARCH model, or even for the subclass of the GARCH model in Case D.

To the best of our knowledge, when $\boldsymbol{\vartheta}_0$ stands at the boundary of the parameter space (Cases B and D), a result similar to (2.13) for GARCH models with covariates is not available in the literature.

The next proposition provides estimations for the matrices \mathbf{I} and \mathbf{J} required to apply Theorem 2.2.

Proposition 2.1. *Under the assumptions of Theorem 2.2, strongly consistent estimators of \mathbf{J} and \mathbf{I} are given by*

$$\widehat{\mathbf{J}}_n = \frac{4}{\delta^2} \frac{1}{n} \sum_{t=1}^n \frac{1}{\widetilde{\sigma}_t^{2\delta}(\widehat{\boldsymbol{\vartheta}}_n)} \frac{\partial \widetilde{\sigma}_t^\delta(\widehat{\boldsymbol{\vartheta}}_n)}{\partial \boldsymbol{\vartheta}} \frac{\partial \widetilde{\sigma}_t^\delta(\widehat{\boldsymbol{\vartheta}}_n)}{\partial \boldsymbol{\vartheta}'}, \quad (2.16)$$

and

$$\widehat{\mathbf{I}}_n = \frac{4}{\delta^2} \frac{1}{n} \sum_{t=1}^n (\widehat{\eta}_t^4 - 1) \frac{1}{\widetilde{\sigma}_t^{2\delta}(\widehat{\boldsymbol{\vartheta}}_n)} \frac{\partial \widetilde{\sigma}_t^\delta(\widehat{\boldsymbol{\vartheta}}_n)}{\partial \boldsymbol{\vartheta}} \frac{\partial \widetilde{\sigma}_t^\delta(\widehat{\boldsymbol{\vartheta}}_n)}{\partial \boldsymbol{\vartheta}'}, \quad (2.17)$$

with $\widehat{\eta}_t = \varepsilon_t / \widetilde{\sigma}_t(\widehat{\boldsymbol{\vartheta}}_n)$.

Remark 2.9. Note that, for Theorem 2.2 and Proposition 2.1, a moment assumption is required on the observed process in Cases B and D. To see what may happen when this moment condition is not met, consider again the example of the ARCH(1)-X of Remark 2.7 with $\alpha_0 = \pi_0 = 0$ (so that the data generating process is a noise). In Case B, the law of large numbers for non integrable positive random variables (Billingsley, 1995, p. 284 and 495) and the arguments of the proof of Proposition 2.1 entail that

$$\widehat{\mathbf{I}}_n(2,2) = \frac{1}{n} \sum_{t=1}^n (\widehat{\eta}_t^4 - 1) \frac{1}{\widehat{\sigma}_t^4(\widehat{\boldsymbol{\vartheta}}_n)} \varepsilon_{t-1}^4 \rightarrow +\infty \quad a.s.$$

when $E\varepsilon_t^4 = +\infty$. This divergence was expected because, in the same framework, the information matrix is such that

$$\mathbf{I}(2,2) = (E\eta_t^4 - 1) \frac{1}{\omega_0^2} E\varepsilon_t^4 = +\infty.$$

When the information is not finite, the asymptotic distribution of the (Q)MLE may be non standard (Hall and Yao, 2003), although the estimator remains consistent (Theorem 2.1 does not require moment conditions). We made simulation experiments showing that the inference based on $\widehat{\boldsymbol{\vartheta}}_n$ can however work nicely even if the moment conditions of Assumption A11 do not hold true.

Remark 2.10. In Cases A and B, in view of Remark 2.8, the estimator defined by (2.17) can be replaced by

$$\widehat{\mathbf{I}}_n = \left(\frac{1}{n} \sum_{t=1}^n \widehat{\eta}_t^4 - 1 \right) \widehat{\mathbf{J}}_n. \quad (2.18)$$

Theorem 2.2 and Proposition 2.1 allow to test if one or several GARCH coefficients are equal to zero, which is important for identifying the orders of the model and the relevant covariates. The Wald, Rao score (or Lagrange multiplier) and likelihood ratio tests are the most widely used parametric tests in econometrics. The Rao score test presents the advantage that its asymptotic distribution remains generally unchanged when, under the null, the parameter is at the boundary of the parameter space or not (Guo et al., 2014). In the sequel, we however concentrate on the Wald test because it has generally a higher local asymptotic power than the score test (see Francq and Zakoïan (2009b) for a comparison of the Bahadur slopes and local asymptotic powers of different tests on standard GARCH), and reduces to the popular t -test when testing the nullity of one coefficient.

We first concentrate on this one-dimensional testing problem. Let \mathbf{e}_k be the k -th element of the canonical basis of \mathbb{R}^d . We will test the hypothesis that the k -th element of $\boldsymbol{\vartheta}_0$ is equal to zero, assuming the other elements are positive :

$$H_0 : \mathbf{e}'_k \boldsymbol{\vartheta}_0 = 0 \text{ and } \mathbf{e}'_\ell \boldsymbol{\vartheta}_0 > 0 \quad \forall \ell \neq k \quad \text{against} \quad H_1 : \mathbf{e}'_k \boldsymbol{\vartheta}_0 > 0. \quad (2.19)$$

For this testing problem, the Student t -test statistic is defined by

$$t_n(k) = \sqrt{n} \frac{\mathbf{e}'_k \widehat{\boldsymbol{\vartheta}}_n}{\sqrt{\mathbf{e}'_k \widehat{\boldsymbol{\Sigma}} \mathbf{e}_k}}, \quad \widehat{\boldsymbol{\Sigma}} = \widehat{\mathbf{J}}_n^{-1} \widehat{\mathbf{I}}_n \widehat{\mathbf{J}}_n^{-1}.$$

Denote, respectively, by $\Phi^{-1}(\alpha)$ and $\chi^2_\ell(\alpha)$ the α -quantiles of the standard normal and the chi-squared distribution with ℓ degrees of freedom. As a corollary of Theorem 2.2 and Proposition 2.1, we obtain the following result.

Corollary 2.1. *Under the assumptions of Proposition 2.1, the test of rejection region*

$$\{t_n(k) > \Phi^{-1}(1 - \alpha)\} \Leftrightarrow \{t_n^2(k) > \chi_1^2(1 - 2\alpha)\} \quad (2.20)$$

has the asymptotic level α under H_0 and is consistent under H_1 defined in (2.19).

Remark 2.11. *Under H_0 , it can be shown that the QMLE saturates the positivity constraint (i.e. that $\mathbf{e}'_k \widehat{\boldsymbol{\vartheta}}_n = 0$) with probability 1/2. The Wald statistic $t_n^2(k)$ then follows the chi-bar-square distribution $\delta_0/2 + \chi_1^2/2$, which is the mixture of the Dirac at 0 and of the χ_1^2 with equal weights. The p -value of the test of rejection region (2.20) is then equal to*

$$P\left(\frac{\delta_0}{2} + \frac{\chi_1^2}{2} > t_n^2(k)\right) = \frac{1}{2}P(\chi_1^2 > t_n^2(k)).$$

When $t_n^2(k) = 0$, the p -value is then equal to 0.5. The standard two-sided t -ratio test of rejection region $\{|t_n^2(k)| > \Phi^{-1}(1 - \alpha/2)\}$ (i.e. $\{t_n^2(k) > \chi_1^2(1 - \alpha)\}$) would have the wrong asymptotic level $\alpha/2$ instead of α . By contrast, the standard one-sided test has a correct asymptotic level.

Now consider a more general Wald test of the form

$$H_0^W : \mathbf{R}\boldsymbol{\vartheta}_0 = \mathbf{r} \quad \text{against} \quad H_1^W : \mathbf{R}\boldsymbol{\vartheta}_0 \neq \mathbf{r}, \quad (2.21)$$

where \mathbf{R} is a $r_0 \times d$ full-rank matrix and \mathbf{r} is a vector of dimension r_0 . Noting that, under H_0^W , we have $\mathbf{R}\widehat{\boldsymbol{\vartheta}}_n - \mathbf{r} = \mathbf{R}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0)$, Theorem 2.2 and Proposition 2.1 immediately show that the standard Wald test statistic

$$W_n = n \left(\mathbf{R}\widehat{\boldsymbol{\vartheta}}_n - \mathbf{r} \right)' \left(\mathbf{R}\widehat{\boldsymbol{\Sigma}}\mathbf{R}' \right)^{-1} \left(\mathbf{R}\widehat{\boldsymbol{\vartheta}}_n - \mathbf{r} \right)$$

has the following asymptotic behavior.

Corollary 2.2. *Under the assumptions of Proposition 2.1 and H_0^W ,*

$$W_n \xrightarrow{d} W = \left\| \mathbf{R}\mathbf{Z}^c \right\|_{(\mathbf{R}\boldsymbol{\Sigma}\mathbf{R}')^{-1}}^2, \quad \mathbf{Z} \sim \mathcal{N}(0, \boldsymbol{\Sigma}).$$

The previous result is essentially useful to test that a given subset of the components of $\boldsymbol{\vartheta}_0$ are zero and that the other components are non-zero, because in this case the local parameter space \mathcal{C} is determined by the null hypothesis. At the asymptotic level $\alpha \in (0, 1)$, the critical value q_α of the Wald test $\{W_n \geq q_\alpha\}$ can be obtained by taking the empirical $(1 - \alpha)$ -quantile of the set

$$\left\{ \left\| \mathbf{R} \widehat{\mathbf{Z}}_i^{\mathcal{C}} \right\|_{(\mathbf{R} \widehat{\boldsymbol{\Sigma}} \mathbf{R}')^{-1}}^2, i = 1, \dots, N \right\},$$

where the $\widehat{\mathbf{Z}}_i$'s are independent and $\mathcal{N}(0, \widehat{\boldsymbol{\Sigma}})$ -distributed random vectors, and N is a large number, say $N = 10,000$ or more (see Algorithm 1 of Pedersen, 2015). A R code of the algorithm can be found on the web pages of the authors.

2.2.4 Testing the relevance of the covariates

For GARCH-X models, the most interesting tests concern the relevance of the covariates. When the GARCH orders are small, for instance in the case of a GARCH-X(1,1) model, the GARCH coefficients are typically far from zero. In that case, it is reasonable to assume that the GARCH parameter $\boldsymbol{\theta}_0$ is strictly positive component-wise. Under this assumption, no moment condition on the observed process ε_t is needed for the QMLE to have the asymptotic distribution given by (2.13). We only assume

A12 : in Case B we have $E \|\mathbf{x}_t\|^{2+\nu_1} < \infty$ for some $\nu_1 > 0$, and in Case D we have $E \|\mathbf{x}_t\|^{2+8/\nu} < \infty$ with ν as in **A10**.

Theorem 2.3. *Under **A1**–**A10** and **A12**, when $\boldsymbol{\theta}_0 > 0$, the conclusion of Theorem 2.2 holds true.*

Without loss of generality, one can assume that we test that the last $r_0 \geq 1$ exogenous variables are irrelevant and the others are relevant. Let

$$\boldsymbol{\pi}_0 = (\boldsymbol{\pi}'_{01}, \boldsymbol{\pi}'_{02})', \quad \boldsymbol{\pi}_{01} = (\pi_{01}, \dots, \pi_{0,r-r_0})', \quad \boldsymbol{\pi}_{02} = (\pi_{0,r-r_0+1}, \dots, \pi_{0r})'. \quad (2.22)$$

Under the maintained assumption $\boldsymbol{\theta}_0 > 0$ and $\boldsymbol{\pi}_{01} > 0$ element-wise, we are interested in testing

$$H_0^\pi : \boldsymbol{\pi}_{02} = \mathbf{0}_{r_0} \quad \text{against} \quad H_1^\pi : \boldsymbol{\pi}_{02} \neq \mathbf{0}_{r_0}. \quad (2.23)$$

Let $\mathbf{K} = (\mathbf{0}_{r_0 \times (d-r_0)} : I_{r_0})$ where I_k denotes the identity matrix of size k and $\mathbf{0}_{k_1 \times k_2}$ denotes the null matrix of size $k_1 \times k_2$, with $\mathbf{0}_{k_1 \times k_2} = \mathbf{0}_{k_1}$ when $k_2 = 1$. Notice that under H_0 ,

$$\sqrt{n} \widehat{\boldsymbol{\pi}}_{2n} \xrightarrow{d} \mathbf{K} \mathbf{Z}^{\mathcal{C}}, \quad \text{as } n \rightarrow \infty$$

where $\hat{\boldsymbol{\pi}}_{2n} = \mathbf{K}\hat{\boldsymbol{\vartheta}}_n$. The usual Wald test statistic is defined by

$$W_n = n \|\hat{\boldsymbol{\pi}}_{2n}\|_{(\mathbf{K}\hat{\boldsymbol{\Sigma}}\mathbf{K}')^{-1}}^2. \quad (2.24)$$

Corollary 2.3. *Under the assumptions of Theorem 2.3 and H_0^π ,*

$$W_n \xrightarrow{d} W = \|\mathbf{K}\mathbf{Z}^{\mathcal{C}}\|_{(\mathbf{K}\boldsymbol{\Sigma}\mathbf{K}')^{-1}}^2, \quad (2.25)$$

where $\boldsymbol{\Sigma} = \mathbf{J}^{-1}\mathbf{I}\mathbf{J}^{-1}$ and $\mathcal{C} = \mathbb{R}^{d-r_0} \times [0, \infty)^{r_0}$.

2.3 Numerical illustrations

We now illustrate our asymptotic results on Monte Carlo simulations and on financial series of daily returns and volumes, as well as high frequency intraday data.

2.3.1 Simulation experiments

The aim of this section is to study the finite sample behavior of the QMLE and of tests of significance of the form (2.19), in the different frameworks corresponding to Cases A-D. We thus simulated the following TAR-CH-X(1,1) model with 2 lagged values of an exogenous variable

$$\begin{cases} \varepsilon_t = \sigma_t \eta_t \\ \sigma_t = \omega_0 + \alpha_{0+}(\varepsilon_{t-1}^+) + \alpha_{0-}(\varepsilon_{t-1}^-) + \beta_0 \sigma_{t-1} + \pi_{01} x_{t-1} + \pi_{02} x_{t-2}. \end{cases} \quad (2.26)$$

The exogenous variable x_t is assumed to be the stochastic volatility defined by

$$x_t = e^{y_t}, \quad y_t = a y_{t-1} + e_t \quad (2.27)$$

where $a = 0.8$ and (e_t) is independently and $\mathcal{N}(0, 1)$ distributed. In Cases A and B, we also assume that η_t is iid $\mathcal{N}(0, 1)$ and independent of (e_t) . In Cases C and D, we assume that, given \mathcal{F}_{t-1} , the variable $\sqrt{\nu_t/(\nu_t - 2)}\eta_t$ is generated according to the Student distribution with ν_t degrees of freedom, where $\nu_t = 5 + x_{t-1}$. This specification of ν_t , in particular the existence of the constant 5, guarantees that **A10** is satisfied. Indeed, the fourth order moment exists because

$$E\eta_t^4 = E(\eta_t^4 | \mathcal{F}_{t-1}) = E\left(\frac{3(\nu_t - 2)}{\nu_t - 4}\right) \leq E(3(\nu_t - 2)) < \infty$$

and, by a similar argument, it can be shown that a moment of order larger than 4 also exists. We took the parameter $\boldsymbol{\vartheta}_0 = (0.046, 0.027, 0.092, 0.843, 0.089, \pi_{02})$ where $\pi_{02} = 0$ in Cases B and D (this value of parameter corresponds to the estimated value of the parameter for the series BA

TABLE 2.1 – Sampling distribution of the QMLE of $\boldsymbol{\vartheta}_0$ over 1000 replications for the TARCH-X(1,1) model in Case A

parameter	true val.	bias	RMSE	min	Q_1	Q_2	Q_3	max
$n = 1,000$								
ω	0.046	0.0007	0.0157	0.0005	0.0352	0.0459	0.0571	0.0994
α_+	0.027	-0.0056	0.0182	0.0000	0.0059	0.0180	0.0331	0.0938
α_-	0.092	-0.0063	0.0232	0.0119	0.0708	0.0853	0.1004	0.1536
β	0.843	0.0042	0.0158	0.7922	0.8372	0.8485	0.8576	0.8917
π_1	0.084	0.0000	0.0271	0.0000	0.0668	0.0830	0.1025	0.1664
π_2	0.084	0.0001	0.0307	0.0000	0.0631	0.0847	0.1043	0.1831
$n = 2,000$								
ω	0.046	0.0000	0.0108	0.0096	0.0387	0.0456	0.0529	0.0876
α_+	0.027	-0.0022	0.0147	0.0000	0.0139	0.0242	0.0344	0.0796
α_-	0.092	-0.0030	0.0162	0.0378	0.0778	0.0893	0.1002	0.1455
β	0.843	0.0018	0.0117	0.8001	0.8370	0.8452	0.8532	0.8738
π_1	0.084	-0.0005	0.0195	0.0194	0.0705	0.0826	0.0971	0.1505
π_2	0.084	0.0010	0.0225	0.0136	0.0688	0.0855	0.1007	0.1544

RMSE is the Root Mean Square Error, Q_i , $i = 1, 3$, denote the quartiles.

studied in Section 2.3.3 below) and $\pi_{01} = \pi_{02} = 0.089$ in Cases A and C. In each of the four cases, we simulated 500 independent replications of model (4.18) for the two sample sizes $n = 1,000$ and $n = 2,000$. To attenuate the effect of the initial values, the first 200 values of each simulation have been eliminated. Tables 2.1–2.4 summarize the distribution of the estimation errors in the four cases. As expected, the accuracy of the estimators always increases with n . It can be noted that the estimators are more accurate when the model is strong (Cases A and B) than when it is semi-strong (Cases C and D). One can note the presence of more frequent outliers in the semi-strong case. Also, in accordance with the asymptotic theory, the distribution of the estimation errors is clearly non Gaussian, especially for the estimation of π_{02} , when the true value of the parameter stands at the boundary of the parameter space (Cases B and D). In particular, Tables 2.2 and 2.4 show that the distribution of $\hat{\pi}_2$ has a mass at zero (since the first or second quantiles of $\hat{\pi}_2$ are equal to zero), which is incompatible with a Gaussian distribution.

Table 2.5 gives the empirical frequencies of rejection of the hypotheses $\pi_{01} = 0$ and $\pi_{02} = 0$. The test of the null hypothesis $\pi_{01} = 0$ (which is false in the four cases) is more powerful in Cases A

TABLE 2.2 – Sampling distribution of the QMLE of ϑ_0 over 1000 replications for the TARX(1,1) model in Case B

parameter	true val.	bias	RMSE	min	Q_1	Q_2	Q_3	max
$n = 1,000$								
ω	0.046	0.0012	0.0119	0.0000	0.0394	0.0468	0.0547	0.0951
α_+	0.027	-0.0055	0.0185	0.0000	0.0055	0.0181	0.0335	0.0925
α_-	0.092	-0.0056	0.0236	0.0113	0.0712	0.0856	0.1013	0.1576
β	0.843	0.0020	0.0167	0.7892	0.8349	0.8464	0.8558	0.8933
π_1	0.084	-0.0053	0.0180	0.0106	0.0716	0.0800	0.0856	0.2615
π_2	0.000	0.0078	0.0118	0.0000	0.0000	0.0002	0.0129	0.1012
$n = 2,000$								
ω	0.046	0.0009	0.0076	0.0107	0.0418	0.0466	0.0516	0.0073
α_+	0.027	-0.0021	0.0150	0.0000	0.0136	0.0240	0.0347	0.0146
α_-	0.092	-0.0025	0.0166	0.0378	0.0781	0.0898	0.1006	0.0157
β	0.843	0.0004	0.0121	0.7963	0.8354	0.8438	0.8521	0.0120
π_1	0.084	-0.0048	0.0087	0.0390	0.0749	0.0802	0.0849	0.0127
π_2	0.000	0.0060	0.0082	0.0000	0.0000	0.0014	0.0105	0.0137

RMSE is the Root Mean Square Error, Q_i , $i = 1, 3$, denote the quartiles.

TABLE 2.3 – Sampling distribution of the QMLE of ϑ_0 over 1000 replications for the TARX(1,1) model in Case C

parameter	true val.	bias	RMSE	min	Q_1	Q_2	Q_3	max
$n = 1,000$								
ω	0.046	-0.0011	0.0227	0.0000	0.0290	0.0438	0.0586	0.1286
α_+	0.027	-0.0027	0.0240	0.0000	0.0000	0.0193	0.0390	0.1451
α_-	0.092	-0.0065	0.0326	0.0000	0.0640	0.0848	0.1051	0.2547
β	0.843	0.0022	0.0215	0.7414	0.8327	0.8467	0.8591	0.9015
π_1	0.084	0.0021	0.0328	0.0000	0.0648	0.0867	0.1090	0.1893
π_2	0.084	0.0029	0.0683	0.0000	0.0559	0.0832	0.1090	1.1950
$n = 2,000$								
ω	0.046	-0.0009	0.0154	0.0027	0.0346	0.0440	0.0541	0.1139
α_+	0.027	-0.0028	0.0187	0.0000	0.0089	0.0228	0.0370	0.0912
α_-	0.092	-0.0026	0.0229	0.0111	0.0744	0.0886	0.1023	0.1946
β	0.843	0.0019	0.0144	0.7848	0.8360	0.8451	0.8541	0.8908
π_1	0.084	0.0008	0.0244	0.0082	0.0682	0.0849	0.1011	0.1576
π_2	0.084	-0.0005	0.0296	0.0000	0.0650	0.0820	0.1026	0.1945

RMSE is the Root Mean Square Error, Q_i , $i = 1, 3$, denote the quartiles.

TABLE 2.4 – Sampling distribution of the QMLE of ϑ_0 over 1000 replications for the TARCH-X(1,1) model in Case D

parameter	true val.	bias	RMSE	min	Q_1	Q_2	Q_3	max
$n = 1,000$								
ω	0.046	-0.0002	0.0180	0.0000	0.0346	0.0451	0.0565	0.1137
α_+	0.027	-0.0028	0.0243	0.0000	0.0000	0.0183	0.0388	0.1429
α_-	0.092	-0.0056	0.0336	0.0000	0.0640	0.0847	0.1068	0.2508
β	0.843	-0.0005	0.0230	0.7141	0.8293	0.8440	0.8569	0.9041
π_1	0.084	-0.0051	0.0229	0.0000	0.0697	0.0797	0.0869	0.2429
π_2	0.000	0.0112	0.0250	0.0000	0.0000	0.0000	0.0155	0.4328
$n = 2,000$								
ω	0.046	0.0005	0.0111	0.0000	0.0391	0.0457	0.0531	0.0957
α_+	0.027	-0.0027	0.0191	0.0000	0.0083	0.0226	0.0374	0.0944
α_-	0.092	-0.0020	0.0234	0.0107	0.0748	0.0891	0.1032	0.1965
β	0.843	0.0001	0.0153	0.7783	0.8337	0.8433	0.8529	0.8919
π_1	0.084	-0.0057	0.0120	0.0188	0.0724	0.0801	0.0853	0.2045
π_2	0.000	0.0073	0.0115	0.0000	0.0000	0.0000	0.0124	0.1112

RMSE is the Root Mean Square Error, Q_i , $i = 1, 3$, denote the quartiles.

and B (corresponding to a strong model) than C and D (corresponding to a semi-strong model). This is not surprising since the semi-strong model is less accurately estimated than the strong one. Less obviously, the test is slightly more powerful in Cases B and D (when $\pi_{02} = 0$) than in Cases A and C (when $\pi_{02} = 0.089$). Turning to the test of the null hypothesis $\pi_{02} = 0$ (which is true in Cases B and D), one can see that the type 1 errors are well controlled when $n = 2,000$. Indeed, when the nominal level is 5%, the empirical relative frequency of rejection over the 500 independent replications should vary between 2.2% and 6.6% with probability of approximately 95%. When the nominal level is 1%, it varies from 0.4% to 2.0% with the same probability. All the relative rejection frequencies displayed in Table 2.5 are within these 95% limits when $n = 2,000$.

2.3.2 SP500 with realized range, volume and other indices

In this section, we built a model which aims to explain the volatility of the daily returns of the SP500 index by its past values, the realized range, the volume and other stock returns. The data set has been downloaded from <http://finance.yahoo.com/> and covers the period from January 3, 1985 to June 29, 2007. We considered the series of the relative range $rr_t = (high_t - low_t)/low_t$, where $high_t$ and low_t denote respectively the highest and lowest prices of the day. We also measured

TABLE 2.5 – Relative frequencies (in %) of rejection of the assumptions that the first and second lagged values of the exogenous variable do not appear in the conditional variance

	$H_0^{\pi_{01}} : \pi_{01} = 0$				$H_0^{\pi_{02}} : \pi_{02} = 0$			
	$\alpha = 1\%$		$\alpha = 5\%$		$\alpha = 1\%$		$\alpha = 5\%$	
	$n = 1,000$	$n = 2,000$	$n = 1,000$	$n = 2,000$	$n = 1,000$	$n = 2,000$	$n = 1,000$	$n = 2,000$
A	83.00	99.40	96.00	99.80	66.80	91.00	85.80	98.40
B	99.40	100.00	100.00	100.00	3.60	1.40	7.60	5.20
C	72.80	92.00	88.00	98.00	50.80	77.20	70.60	92.60
D	96.40	98.80	99.20	98.80	2.20	2.00	6.40	5.80

the relative volume by the formula $v_t = \left| \frac{vol_t}{\frac{1}{20} \sum_{i=1}^{20} vol_{t-i}} - 1 \right|$, where vol_t denotes the daily number of shares traded. We did not consider directly (vol_t) as covariate because this series is non stationary (the Augmented Dickey-Fuller (ADF) test does not reject the unit root hypothesis, whereas the KPSS test clearly rejects the stationarity). The indicator v_t compares the present volume with the averaged volume over the past 20 days, which is a technique used by some traders. Figure 2.3.2 displays the series of the returns ε_t , the ranges rr_t and the relative volumes v_t , for which the output of the ADF and KPSS tests is compatible with stationarity. We also added the returns of the Nikkei, Nik_t , and of the FTSE, Ft_t , as potential explanatory variables for the SP500 volatility. We fitted APARCH-X(1,1) models with $\delta \in \{0.5, 1, 1.5, 2\}$. The model with the largest likelihood is obtained for $\delta = 1$, and is given by $\varepsilon_t = h_t \eta_t$ with

$$h_t = 0.004 + \frac{0.000}{(0.022) 0.500} \varepsilon_{t-1}^+ + \frac{0.113}{(0.043) 0.004} \varepsilon_{t-1}^- + \frac{0.880}{(0.027) 0.000} h_{t-1} + \frac{2.964}{(1.642) 0.036} rr_{t-1} \\ + \frac{0.053}{(0.031) 0.044} v_{t-1} + \frac{0.000}{(0.008) 0.500} Nik_{t-1}^2 + \frac{0.037}{(0.018) 0.018} Ft_{t-1}^2.$$

Under the estimated value of each coefficient, the estimated standard deviation is given into brackets, followed by the p -value of the test that the coefficient is equal to zero. The estimated standard deviation of the i -th parameter, defined by $\sqrt{\widehat{\Sigma}(i, i)/n}$ where $\widehat{\Sigma} = \widehat{\mathbf{J}}_n^{-1} \widehat{\mathbf{I}}_n \widehat{\mathbf{J}}_n^{-1}$ with (2.17)-(2.16), is consistent in Cases A and C. Each p -value is obtained from an application of Corollary 2.1 and corresponds to a test of the form (2.19). Since two estimated parameters are equal to zero, a test of nullity of a single coefficient of the form (2.19) does not seem appropriate. Denoting by π_{04} the theoretical value of the coefficient in front of the variable Nik_{t-1}^2 , we thus performed the test

$$\alpha_{0+} = \pi_{04} = 0 \quad \text{against} \quad \alpha_{0+} \neq 0 \text{ or } \pi_{04} \neq 0,$$

under the maintained assumption that the other coefficients are non-zero. This test is of the form (2.21). The estimated p -value of this test being 0.7758, we re-estimated the model under the null,

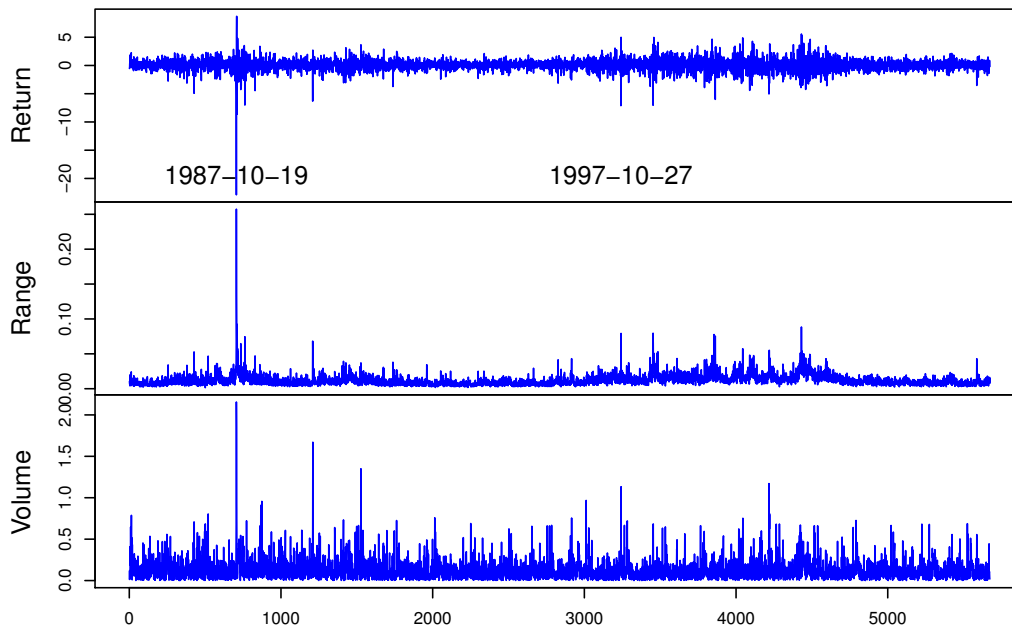


FIGURE 2.1 – Return, range and volume of the SP500 index from January 4, 1985 to August 26, 2011 (October 19, 1987 corresponds to the black Monday, and October 27, 1997 is the date of a mini-crash that was caused by an economic crisis in Asia).

and we found

$$\begin{aligned}
 h_t = & 0.004 + \frac{0.113}{(0.034)} \varepsilon_{t-1}^- + \frac{0.880}{(0.026)} h_{t-1} + \frac{2.964}{(1.146)} rr_{t-1} \\
 & + \frac{0.053}{(0.033)} v_{t-1} + \frac{0.037}{(0.017)} Ft_{t-1}^2.
 \end{aligned} \tag{2.28}$$

One can see that the range rr_{t-1} and the return series Ft_{t-1}^2 are significant covariates, whereas there exists a doubt concerning the significance of the volume v_{t-1} . This is in accordance with several empirical studies showing that the realized range, and to a lesser extent, the volume can help to predict the volatility (see *e.g.* Fuertes et al., 2009).

To assess the adequacy of the model, we also plotted in Figure 2.3.2 the correlogram of the series ε_t of the SP500 returns. Many sample autocorrelations lie outside the significance bounds $\pm 1.96/\sqrt{n}$, but these bounds, derived from the standard Bartlett's formula for linear processes, are only valid when the observations are iid. The significance bounds delimited by the full lines

are based on a generalized Bartlett's formula (see [Francq and Zakoïan, 2009a](#)) which is valid for GARCH and more general non linear processes. In view of these significance bounds, there is no evidence against the weak white noise assumption. Figure 2.3.2 also displays the autocorrelations of the squared residuals of the APARCH-X model, which are all very close to zero. The portmanteau test [Li and Mak \(1994\)](#) and the weighted version of [Fisher and Gallagher \(2012\)](#) give respectively p -values of 0.1182 and 0.1096, and thus do not reject the adequacy of the model.

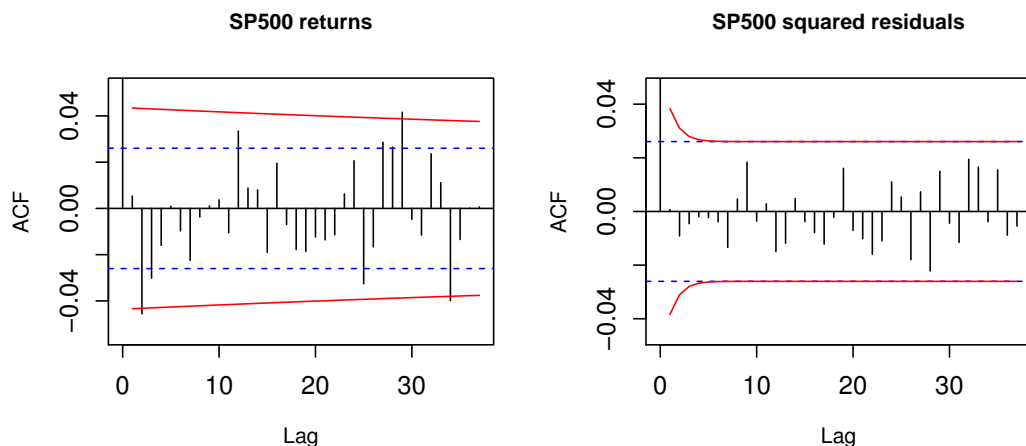


FIGURE 2.2 – Autocorrelations of the SP500 returns and of the squared residuals of model (2.28), with the significance bands obtained from the standard Bartlett's formula (in dotted lines) and the generalized Bartlett's formula (in full lines).

2.3.3 US stocks with realized volatility

The data used in this section come from Section 4.2 of [Laurent et al. \(2014\)](#)² and concern 49 large capitalization stocks of american stock exchanges, covering the period from January 4, 1999 to December 31, 2008 (2,489 trading days). At the end of each trading day t , the log-return in percentage ε_t and the realized volatility rv_t (computed as the sum of intraday squared 5-minute log-returns) are available.

The question that we are interested in is whether the realized volatility is useful to predict the squared returns or not. More precisely, we would like to know how many lagged values of the realized volatility have to be considered in the volatility equation.

2. The authors are grateful to Sébastien Laurent who has kindly provided them with the data set, and who has provided very useful comments on a preliminary version of this paper.

In order to answer this question, we estimated APARCH-X(1,1) models of the form

$$\begin{cases} \varepsilon_t = h_t^{1/\delta} \eta_t \\ h_t = \omega + \alpha_+(\varepsilon_{t-1}^+)^{\delta} + \alpha_-(\varepsilon_{t-1}^-)^{\delta} + \beta h_{t-1} + \pi_1 r v_{t-1}^{\delta/2} + \pi_2 r v_{t-2}^{\delta/2}, \end{cases} \quad (2.29)$$

with $\delta \in \{0.5, 1, 1.5, 2\}$. The variables $r v_{t-1}$ and $r v_{t-2}$ are raised to the power $\delta/2$ in order to have the same unit of measure for ε_t^2 , the squared volatility $h_t^{2/\delta}$ and the realized volatility $r v_t$, regardless of δ . The selected value of δ is that which leads to the maximum value of the quasi-likelihood. Table 2.6 displays the fitted model on each of the 49 stocks. For all the estimated models, except 3 over the 49, one observes that $\alpha_- > \alpha_+$, which is in accordance with the leverage effect (*i.e.* the fact that the volatility tends to increase more after a negative return than after a positive return of the same magnitude). We mostly find π_1 significantly non-zero and π_2 close to zero. From this table, it is clear that yesterday's realized volatility often helps in predicting today's squared return.

2.4 Conclusion

In this paper, we studied the asymptotic behavior of the QMLE for the versatile class of the semi-strong PGARCH models augmented with exogenous variables. The main assumptions on the exogenous variables are the stationarity and non-colinearity with the other explanatory variables of the volatility. This allows to incorporate some additional covariates for predicting the volatility of the financial returns, such as the volumes, the realized ranges, other past squared returns or intraday realized volatilities. Since the true value of the parameter is not constrained to belong to the interior of the parameter space, we were able to derive tests for the significance of the exogenous variables. For the asymptotic distribution of the QMLE, we investigated four different situations corresponding to strong or semi-strong models, and to parameters inside or at the boundary of the parameter space. When the GARCH-X parameter belongs to the interior of the parameter space, the asymptotic distribution of the QMLE is normal, whereas it is the projection of a normal distribution on a convex cone when one or several coefficients are equal to zero. For models with positive GARCH coefficients, the asymptotic distribution is obtained under very mild conditions, in particular, without any moment condition on the observed process. When the GARCH parameter stands at the boundary, fourth-order moment conditions are required for the information matrix to be finite. Our asymptotic results are obtained under conditions that are only marginally stronger than these optimal moment conditions, which extends and improves the results that existed for GARCH models without covariates.

The asymptotic theory developed in the paper has been applied to simulations and real series. Our empirical results are in accordance with numerous applied studies, and complement them by

TABLE 2.6 – APARCH-X(1,1) models (2.29) fitted by QMLE on daily returns of US stock with two lagged values of realized volatilities as covariates. The estimated standard deviations are displayed into parentheses. For the estimated values of π_1 and π_2 , one star (*) means a p -value $p \in [0.01, 0.05)$ for testing the nullity of the coefficient, two stars (**) means $p \in [0.001, 0.01)$, and three stars (***) means $p < 0.001$. The last column gives the selected value of the power δ .

	ω	α_+	α_-	β	π_1	π_2	δ
AAPL	0.080 (0.029)	0.042 (0.015)	0.055 (0.013)	0.796 (0.056)	0.120* (0.072)	0.000 (0.079)	0.5
ABT	0.046 (0.027)	0.023 (0.024)	0.019 (0.027)	0.661 (0.100)	0.285*** (0.071)	0.000 (0.102)	0.5
AXP	0.027 (0.010)	0.000 (0.019)	0.074 (0.019)	0.809 (0.038)	0.155** (0.061)	0.000 (0.069)	1
BA	0.046 (0.022)	0.027 (0.020)	0.092 (0.023)	0.843 (0.037)	0.084 (0.065)	0.000 (0.071)	2
BAC	0.007 (0.010)	0.009 (0.026)	0.090 (0.029)	0.813 (0.045)	0.151** (0.063)	0.000 (0.083)	1
BMJ	0.000 (0.020)	0.051 (0.018)	0.077 (0.025)	0.880 (0.031)	0.072 (0.106)	0.000 (0.108)	1
BP	0.017 (0.015)	0.010 (0.020)	0.043 (0.017)	0.682 (0.068)	0.191* (0.086)	0.106 (0.092)	0.5
C	0.013 (0.011)	0.019 (0.025)	0.123 (0.028)	0.744 (0.056)	0.178** (0.059)	0.007 (0.077)	1
CAT	0.045 (0.022)	0.000 (0.016)	0.011 (0.017)	0.780 (0.066)	0.183** (0.074)	0.000 (0.087)	0.5
CL	0.127 (0.054)	0.032 (0.021)	0.212 (0.065)	0.424 (0.105)	0.156** (0.057)	0.193** (0.073)	2
CSCO	0.013 (0.015)	0.000 (0.020)	0.054 (0.020)	0.848 (0.034)	0.131 (0.081)	0.000 (0.084)	1
CVX	0.082 (0.030)	0.014 (0.022)	0.068 (0.025)	0.716 (0.071)	0.102* (0.062)	0.088 (0.088)	2
DELL	0.000 (0.006)	0.030 (0.012)	0.055 (0.012)	0.874 (0.034)	0.093 (0.080)	0.000 (0.085)	0.5
DIS	0.042 (0.021)	0.000 (0.016)	0.070 (0.022)	0.807 (0.056)	0.132* (0.063)	0.002 (0.069)	2
EK	0.182 (0.091)	0.071 (0.029)	0.111 (0.036)	0.583 (0.161)	0.224* (0.103)	0.000 (0.155)	0.5
EXC	0.092 (0.033)	0.056 (0.032)	0.157 (0.039)	0.677 (0.060)	0.196** (0.072)	0.000 (0.077)	1.5
F	0.060 (0.046)	0.068 (0.037)	0.075 (0.025)	0.740 (0.050)	0.091 (0.088)	0.103 (0.097)	1
FDX	0.033 (0.018)	0.012 (0.021)	0.025 (0.021)	0.803 (0.070)	0.162* (0.074)	0.000 (0.081)	0.5
GE	0.005 (0.010)	0.000 (0.019)	0.052 (0.022)	0.802 (0.050)	0.180* (0.085)	0.000 (0.076)	1
GM	0.026 (0.023)	0.018 (0.015)	0.047 (0.023)	0.881 (0.031)	0.094 (0.093)	0.000 (0.099)	2
HD	0.007 (0.010)	0.000 (0.015)	0.030 (0.013)	0.850 (0.037)	0.135* (0.070)	0.000 (0.071)	0.5
HNZ	0.007 (0.009)	0.050 (0.018)	0.084 (0.022)	0.840 (0.038)	0.105* (0.059)	0.000 (0.064)	1
HON	0.015 (0.012)	0.000 (0.023)	0.108 (0.019)	0.860 (0.030)	0.092 (0.068)	0.000 (0.065)	1
IBM	0.011 (0.007)	0.000 (0.015)	0.044 (0.016)	0.858 (0.033)	0.116* (0.063)	0.000 (0.070)	0.5
INTC	0.013 (0.009)	0.000 (0.012)	0.031 (0.013)	0.862 (0.029)	0.119* (0.061)	0.000 (0.063)	0.5
JNJ	0.022 (0.011)	0.004 (0.023)	0.176 (0.033)	0.757 (0.049)	0.144** (0.061)	0.000 (0.070)	1.5
KO	0.009 (0.016)	0.010 (0.023)	0.084 (0.028)	0.702 (0.057)	0.248** (0.082)	0.000 (0.093)	1
LLY	0.021 (0.089)	0.071 (0.029)	0.065 (0.036)	0.245 (0.146)	0.108 (0.097)	0.599* (0.316)	0.5
MCD	0.015 (0.010)	0.031 (0.015)	0.043 (0.016)	0.863 (0.029)	0.096 (0.065)	0.000 (0.070)	0.5

providing a formal test for the significance of the exogenous variables. In particular, we generally find useful the volume, the realized range and the intraday realized volatility for predicting the squares of the financial returns.

2.5 Proofs and technical lemmas

Proof of Lemma 2.2. The result is known when $\pi_0 = 0$ (see Proposition A.1 in HZ). For notational simplicity, we give the proof for general π_0 when $p = q = 1$. If $\gamma < 0$, h_t is given by (2.5). Using the elementary inequality $(\sum_i u_i)^s \leq \sum_i u_i^s$ for any sequence of positive numbers u_i and any $s \in (0, 1]$,

TABLE 2.6 – (continued)

	ω	α_+	α_-	β	π_1	π_2	δ
MMM	0.035 (0.020)	0.011 (0.024)	0.015 (0.024)	0.777 (0.057)	0.181* (0.104)	0.000 (0.104)	0.5
MOT	0.011 (0.010)	0.004 (0.014)	0.066 (0.015)	0.888 (0.025)	0.081 (0.067)	0.000 (0.070)	1
MRK	0.022 (0.013)	0.017 (0.017)	0.085 (0.024)	0.904 (0.026)	0.046 (0.073)	0.000 (0.065)	1
MS	0.015 (0.016)	0.011 (0.019)	0.058 (0.022)	0.720 (0.080)	0.251*** (0.078)	0.000 (0.102)	0.5
MSFT	0.000 (0.011)	0.046 (0.019)	0.038 (0.015)	0.731 (0.066)	0.237*** (0.073)	0.000 (0.100)	0.5
ORCL	0.000 (0.010)	0.001 (0.014)	0.050 (0.016)	0.888 (0.024)	0.095 (0.063)	0.000 (0.065)	1
PEP	0.011 (0.010)	0.042 (0.017)	0.070 (0.021)	0.842 (0.035)	0.084 (0.058)	0.000 (0.066)	2
PFE	0.005 (0.006)	0.014 (0.010)	0.041 (0.010)	0.956 (0.010)	0.014 (0.031)	0.000 (0.029)	2
PG	0.032 (0.021)	0.000 (0.027)	0.134 (0.035)	0.649 (0.074)	0.269*** (0.080)	0.000 (0.100)	1
QCOM	0.051 (0.027)	0.029 (0.019)	0.110 (0.024)	0.819 (0.038)	0.116* (0.065)	0.000 (0.070)	1.5
SLB	0.116 (0.049)	0.003 (0.017)	0.015 (0.019)	0.827 (0.045)	0.121* (0.067)	0.000 (0.073)	2
T	0.008 (0.009)	0.003 (0.013)	0.050 (0.018)	0.881 (0.023)	0.087 (0.055)	0.000 (0.058)	2
TWX	0.000 (0.030)	0.041 (0.028)	0.150 (0.033)	0.564 (0.063)	0.211** (0.071)	0.166* (0.081)	1.5
UN	0.020 (0.010)	0.039 (0.021)	0.108 (0.039)	0.705 (0.064)	0.189** (0.072)	0.000 (0.078)	2
VZ	0.012 (0.011)	0.050 (0.019)	0.054 (0.020)	0.787 (0.051)	0.162** (0.063)	0.000 (0.074)	0.5
WFC	0.000 (0.011)	0.020 (0.024)	0.091 (0.029)	0.734 (0.056)	0.120* (0.069)	0.100 (0.073)	1
WMT	0.002 (0.006)	0.010 (0.012)	0.047 (0.013)	0.916 (0.016)	0.050 (0.061)	0.000 (0.061)	2
WYE	0.000 (0.008)	0.012 (0.013)	0.042 (0.013)	0.877 (0.029)	0.099 (0.063)	0.005 (0.069)	0.5
XOM	0.073 (0.030)	0.021 (0.021)	0.066 (0.023)	0.742 (0.073)	0.118* (0.060)	0.048 (0.085)	2
XRX	0.000 (0.017)	0.010 (0.019)	0.012 (0.020)	0.828 (0.049)	0.170** (0.056)	0.000 (0.084)	0.5

and the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}
Eh_t^s &\leq \sum_{k=0}^{\infty} E \left(\prod_{i=1}^k a(\eta_{t-i}) \varpi_{t-k-1} \right)^s = \sum_{k=0}^{\infty} E \left(\prod_{i=1}^k a^s(\eta_{t-i}) \varpi_{t-k-1}^s \right) \\
&\leq \sum_{k=0}^{\infty} \left(\prod_{i=1}^k E a^{2s}(\eta_{t-i}) E \varpi_{t-k-1}^{2s} \right)^{1/2}.
\end{aligned}$$

By Assumption **A1**, there exists $s > 0$ such that $E\varpi_t^{2s} < \infty$. Moreover, the fact that $\gamma = E \log a(\eta_t) < 0$ and that $Ea^r(\eta_t) < \infty$ for some $r > 0$ entails that there exists $s > 0$ such that $Ea^{2s}(\eta_t) < 1$ (see *e.g.* Lemma 2.2 in [Francq and Zakoïan, 2010](#)). It follows that $Eh_t^s < \infty$, and thus $Eh_t^{s/\delta} < \infty$ for some $s > 0$. By the Cauchy-Schwarz inequality and **A1**, we deduce that $E|\varepsilon_t|^{s/2} < \infty$. \square

The following lemma is useful to show the identifiability of the parameters under **A4**.

Lemma 2.3. *Let X be a random variable which takes at least three values and $P(X > 0) \in (0, 1)$. If $a(X^+)^{\delta} + b(X^-)^{\delta} = c$ a.s., with $a, b, c \in \mathbb{R}$, then $a = b = 0$.*

Proof. If $a = 0$ and $b \neq 0$ then $b(X^-)^{\delta} = c$ with probability one. Since $P(X > 0) > 0$, we must have $c = 0$. It follows that $X^- = 0$ a.s. which is in contradiction with the assumption that

$P(X > 0) < 1$.

If $a \neq 0$ and $b \neq 0$, we obtain $X = \left(\frac{c}{a}\right)^{1/\delta}$ when $X > 0$ and $X = -\left(\frac{c}{b}\right)^{1/\delta}$ when $X < 0$ which is in contradiction with assumption that X has at least three values. \square

Proof of Theorem 2.1. In the sequel, a neighborhood of ϑ is denoted by $V(\vartheta)$. The consistency can be shown by establishing the following intermediate results :

- i) $\lim_{n \rightarrow \infty} \sup_{\vartheta \in \Theta} |Q_n(\vartheta) - \tilde{Q}_n(\vartheta)| = 0$ a.s.
- ii) If $\sigma_t(\vartheta) = \sigma_t(\vartheta_0)$ a.s. then $\vartheta = \vartheta_0$.
- iii) $E|\ell_t(\vartheta_0)| < \infty$ and if $\vartheta \neq \vartheta_0$, $E|\ell_t(\vartheta)| > E|\ell_t(\vartheta_0)|$.
- iv) For any $\vartheta \neq \vartheta_0$, there exists $V(\vartheta)$ such that

$$\liminf_{n \rightarrow \infty} \inf_{\vartheta^* \in V(\vartheta)} Q_n(\vartheta^*) > E\ell_1(\vartheta_0) \text{ a.s.}$$

The proofs of i), iii) and iv) being essentially the same as when the parameter π is not present, we only give the proof of ii).

Assume that $\sigma_t(\vartheta) = \sigma_t(\vartheta_0)$ a.s. By the second part of **A5**, the polynomials $\mathcal{B}_\vartheta(B)$ and $\mathcal{B}_{\vartheta_0}(B)$ are invertible. We thus have

$$\begin{aligned} & \left\{ \frac{\mathcal{A}_{\vartheta_+}(B)}{\mathcal{B}_\vartheta(B)} - \frac{\mathcal{A}_{\vartheta_0+}(B)}{\mathcal{B}_{\vartheta_0}(B)} \right\} (\varepsilon_t^+)^{\delta} + \left\{ \frac{\mathcal{A}_{\vartheta_-}(B)}{\mathcal{B}_\vartheta(B)} - \frac{\mathcal{A}_{\vartheta_0-}(B)}{\mathcal{B}_{\vartheta_0}(B)} \right\} (\varepsilon_t^-)^{\delta} \\ & + \left\{ \frac{\pi' B}{\mathcal{B}_\vartheta(B)} - \frac{\pi'_0 B}{\mathcal{B}_{\vartheta_0}(B)} \right\} \mathbf{x}_t = \frac{\omega_0}{\mathcal{B}_\vartheta(1)} - \frac{\omega}{\mathcal{B}_{\vartheta_0}(1)} \text{ a.s.} \end{aligned} \quad (2.30)$$

If $\frac{\mathcal{A}_{\vartheta_+}(B)}{\mathcal{B}_\vartheta(B)} \neq \frac{\mathcal{A}_{\vartheta_0+}(B)}{\mathcal{B}_{\vartheta_0}(B)}$ or $\frac{\mathcal{A}_{\vartheta_-}(B)}{\mathcal{B}_\vartheta(B)} \neq \frac{\mathcal{A}_{\vartheta_0-}(B)}{\mathcal{B}_{\vartheta_0}(B)}$, then there exist $c_{i_0}^+ \neq 0$ or $c_{i_0}^- \neq 0$, a constant e and a sequence of vectors (\mathbf{d}_i) such that

$$\sum_{i=i_0}^{\infty} c_i^+ B^i (\varepsilon_t^+)^{\delta} + \sum_{i=i_0}^{\infty} c_i^- B^i (\varepsilon_t^-)^{\delta} + \sum_{i=1}^{\infty} \mathbf{d}'_i \mathbf{x}_{t-i} = e.$$

Since $(\varepsilon_t^+)^{\delta} = \sigma_t^{\delta} (\eta_t^+)^{\delta}$ and $(\varepsilon_t^-)^{\delta} = \sigma_t^{\delta} (\eta_t^-)^{\delta}$, there exists $(a, b)' \in \mathbb{R}^2 \setminus (0, 0)$ such that

$$a(\eta_{t-i_0}^+)^{\delta} + b(\eta_{t-i_0}^-)^{\delta} = c_{t,i_0}$$

where $c_{t,i}$ is a measurable function of $\{\eta_{t-j}, j > i, \mathbf{x}_{t-k}, k > 0\}$. It follows that

$$\mathcal{L} \left(a(\eta_{t-i_0}^+)^{\delta} + b(\eta_{t-i_0}^-)^{\delta} \middle| \mathcal{F}_{t,i_0} \right) = \mathcal{L}(c_{t,i_0} | \mathcal{F}_{t,i_0})$$

where $\mathcal{L}(X|Y)$ denotes the distribution of X given Y . By lemma 2.3 and **A4**, we obtain $a = b = 0$.

Therefore, we have

$$\frac{\mathcal{A}_{\boldsymbol{\vartheta}+}(B)}{\mathcal{B}_{\boldsymbol{\vartheta}}(B)} = \frac{\mathcal{A}_{\boldsymbol{\vartheta}_0+}(B)}{\mathcal{B}_{\boldsymbol{\vartheta}_0}(B)} \text{ and } \frac{\mathcal{A}_{\boldsymbol{\vartheta}-}(B)}{\mathcal{B}_{\boldsymbol{\vartheta}}(B)} = \frac{\mathcal{A}_{\boldsymbol{\vartheta}_0-}(B)}{\mathcal{B}_{\boldsymbol{\vartheta}_0}(B)} \quad (2.31)$$

By **A7**, we obtain $\mathcal{A}_{\boldsymbol{\vartheta}+}(B) = \mathcal{A}_{\boldsymbol{\vartheta}_0+}(B)$, $\mathcal{A}_{\boldsymbol{\vartheta}-}(B) = \mathcal{A}_{\boldsymbol{\vartheta}_0-}(B)$ and $\mathcal{B}_{\boldsymbol{\vartheta}}(B) = \mathcal{B}_{\boldsymbol{\vartheta}_0}(B)$. Then (2.30) becomes

$$(\boldsymbol{\pi} - \boldsymbol{\pi}_0)' \mathbf{x}_{t-1} = \omega_0 - \omega,$$

which entails $\boldsymbol{\pi} = \boldsymbol{\pi}_0$ and $\omega = \omega_0$ under **A8**. Hence, (ii) is proved. \square

The proof of the asymptotic distribution of the QMLE is split into several technical lemmas.

Lemma 2.4. *Under the assumptions of Theorem 2.2,*

$$(i) \quad E \left\| \frac{\partial \ell_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}} \frac{\partial \ell_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}'} \right\| < \infty, \quad E \left\| \frac{\partial^2 \ell_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \right\| < \infty.$$

$$(ii) \quad \mathbf{J} \text{ is non-singular and } \text{var} \left\{ \frac{\partial \ell_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}} \right\} = \mathbf{I}.$$

Proof of Lemma 2.4. First note that the derivatives of $\ell_t(\boldsymbol{\vartheta}) = \frac{\varepsilon_t^2}{\sigma_t^2} + \ln \sigma_t^2$ are

$$\frac{\partial \ell_t(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} = \frac{\partial}{\partial \boldsymbol{\vartheta}} \left\{ \frac{\varepsilon_t^2}{(\sigma_t^\delta)^{2/\delta}} + \frac{2}{\delta} \ln \sigma_t^\delta \right\} = \frac{2}{\delta} \left\{ 1 - \frac{\varepsilon_t^2}{\sigma_t^2} \right\} \left\{ \frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \boldsymbol{\vartheta}} \right\} \quad (2.32)$$

and

$$\frac{\partial^2 \ell_t(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} = \frac{2}{\delta} \left\{ 1 - \frac{\varepsilon_t^2}{\sigma_t^2} \right\} \left\{ \frac{1}{\sigma_t^\delta} \frac{\partial^2 \sigma_t^\delta}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \right\} + \frac{2}{\delta} \left\{ \frac{\delta + 2}{\delta} \frac{\varepsilon_t^2}{\sigma_t^2} - 1 \right\} \left\{ \frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \boldsymbol{\vartheta}} \right\} \left\{ \frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \boldsymbol{\vartheta}'} \right\}. \quad (2.33)$$

In Cases A and B (strong model), one can thus prove (i) by showing that

$$E \left\| \frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}} \right\|^2 < \infty \quad (2.34)$$

and

$$E \left\| \frac{1}{\sigma_t^\delta} \frac{\partial^2 \sigma_t^\delta(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \right\| < \infty. \quad (2.35)$$

In Cases C and D (semi-strong model), using the Hölder inequality and $E|\eta_t|^{4+\nu} < \infty$, the existence of the first expectation in (i) can be proven by showing that

$$E \left\| \frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}} \right\|^{2+8/\nu} < \infty. \quad (2.36)$$

Thanks to $E(\eta_t^2 | \mathcal{F}_{t-1}) = 1$ in **A2**, the second expectation of (i) is still obtained by showing (2.34) and (2.35).

The existence of the moments in (2.34)–(2.36) is already known when $\boldsymbol{\pi}_0$ is absent, (η_t) is iid and $\boldsymbol{\vartheta}_0 \in \overset{\circ}{\Theta}$ (see the proof of (i) and (ii) of Theorem 2.2 in HZ). We now explain the changes in the proof induced by our particular framework in the case $p = q = 1$. The proof can be easily extended to the general case.

Since $\sigma_t^\delta = \sum_{j=0}^{\infty} \beta^j c_{t,j}$ with $c_{t,j} = \omega + \alpha_+ (\varepsilon_{t-j-1}^+)^{\delta} + \alpha_- (\varepsilon_{t-j-1}^-)^{\delta} + \boldsymbol{\pi}' \mathbf{x}_{t-j-1}$, we have

$$\begin{aligned} \frac{\partial \sigma_t^\delta}{\partial \omega} &= \sum_{j=0}^{\infty} \beta^j = \frac{1}{1-\beta}, \\ \frac{\partial \sigma_t^\delta}{\partial \alpha_+} &= \sum_{j=0}^{\infty} \beta^j (\varepsilon_{t-j-1}^+)^{\delta}, \quad \frac{\partial \sigma_t^\delta}{\partial \alpha_-} = \sum_{j=0}^{\infty} \beta^j (\varepsilon_{t-j-1}^-)^{\delta}, \\ \frac{\partial \sigma_t^\delta}{\partial \beta} &= \sum_{j=1}^{\infty} j \beta^{j-1} c_{t,j}, \\ \frac{\partial \sigma_t^\delta}{\partial \boldsymbol{\pi}} &= \sum_{j=0}^{\infty} \beta^j \mathbf{x}_{t-j-1}. \end{aligned} \tag{2.37}$$

Similar expressions hold for the second order derivatives. Noting also that

$$\underline{\omega} := \inf_{\boldsymbol{\vartheta} \in \overset{\circ}{\Theta}} \sigma_t^\delta > 0, \tag{2.38}$$

under the moment conditions in **A11**, we have (2.34) and (2.35) in Case B, and (2.35) and (2.36) in Case D. In Cases A and C, we have

$$\frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \omega} \leq \frac{1}{\omega}, \quad \frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \alpha_+} \leq \frac{1}{\alpha_+}, \quad \frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \alpha_-} \leq \frac{1}{\alpha_-}, \quad \frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \pi_i} \leq \frac{1}{\pi_i},$$

with the notation $\boldsymbol{\pi} = (\pi_1, \dots, \pi_r)'$. Using the inequalities $\frac{x}{1+x} \leq x^s$ and $(x+y)^s \leq x^s + y^s$, for $x, y \geq 0$ and $s \in (0, 1]$, we also have

$$\frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \beta} \leq \frac{1}{\beta \omega^s} \sum_{i=1}^{\infty} i \beta^{is} \left\{ \omega^s + \alpha_+^s (\varepsilon_{t-i-1}^+)^{\delta s} + \alpha_-^s (\varepsilon_{t-i-1}^-)^{\delta s} + (\boldsymbol{\pi}' \mathbf{x}_{t-i-1})^s \right\}.$$

Note that, for any $q \geq 1$, under the moments assumptions in **A1** and **A6**, we have

$$\left\| (\varepsilon_{t-i-1}^+)^{\delta s} \right\|_q < \infty, \quad \left\| (\varepsilon_{t-i-1}^-)^{\delta s} \right\|_q < \infty \text{ and } \left\| (\boldsymbol{\pi}' \mathbf{x}_{t-i-1})^s \right\|_q < \infty$$

for sufficiently small $s > 0$. It follows that (2.34) and (2.36), for any $\nu > 0$, hold true when all the components of $\boldsymbol{\vartheta}_0$ are non-zero. By the same arguments, we can even show the stronger result that

for any $s_0 > 0$ there exists $V(\boldsymbol{\vartheta}_0)$ included in Θ such that

$$E \sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0)} \left\| \frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} \right\|^{s_0} < \infty \quad \text{in Cases A and C.} \quad (2.39)$$

We obtain (2.35) by similar arguments, which completes the proof of (i) in the two remaining cases A and C.

We now turn to the proof of (ii). Note that the second equality in (2.10) comes from (2.33) and the fact that $E(\eta_t^2 | \mathcal{F}_{t-1}) = 1$. Using (2.32), we also have $\text{var}(\partial \ell_t(\boldsymbol{\vartheta}_0)/\partial \boldsymbol{\vartheta}) = \mathbf{I}$. It remains to show that \mathbf{J} is invertible. If it is not the case, then there exists $\mathbf{c} \in \mathbb{R}^d$ such that

$$\mathbf{c}' \mathbf{J} \mathbf{c} = \frac{4}{\delta^2} E \left\{ \frac{1}{\sigma_t^{2\delta}(\boldsymbol{\vartheta}_0)} \left(\mathbf{c}' \frac{\partial \sigma_t^\delta(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}} \right)^2 \right\} = 0.$$

Then a.s., $\mathbf{c}' \frac{\partial \sigma_t^\delta(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}} = 0$. In view of (2.9), this implies

$$\mathbf{c}' \left(1, (\varepsilon_{t-1}^+)^{\delta}, (\varepsilon_{t-1}^-)^{\delta}, \dots, (\varepsilon_{t-q}^+)^{\delta}, (\varepsilon_{t-q}^-)^{\delta}, \sigma_{t-1}^{\delta}(\boldsymbol{\vartheta}_0), \dots, \sigma_{t-p}^{\delta}(\boldsymbol{\vartheta}_0), \mathbf{x}_{t-1} \right) = 0.$$

By the arguments used to prove (ii) of Theorem 2.1, this is impossible with $\mathbf{c} \neq 0$, which completes the proof of (ii). \square

Lemma 2.5. *Under the assumptions of Theorem 2.2, as $n \rightarrow \infty$, we have*

$$\sqrt{n} \left\| \frac{\partial}{\partial \boldsymbol{\vartheta}} \tilde{Q}_n(\boldsymbol{\vartheta}_0) - \frac{\partial}{\partial \boldsymbol{\vartheta}} Q_n(\boldsymbol{\vartheta}_0) \right\| = o(1) \quad a.s., \quad (2.40)$$

$$\sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0) \cap \Theta} \left\| \frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \tilde{Q}_n(\boldsymbol{\vartheta}) - \frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} Q_n(\boldsymbol{\vartheta}) \right\| = o(1) \quad a.s. \quad (2.41)$$

for some neighborhood $V(\boldsymbol{\vartheta}_0)$ of $\boldsymbol{\vartheta}_0$,

$$\frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \tilde{Q}_n(\boldsymbol{\vartheta}_n) \rightarrow \mathbf{J} \text{ in probability when } \boldsymbol{\vartheta}_n \rightarrow \boldsymbol{\vartheta}_0 \text{ in probability,} \quad (2.42)$$

$$\sqrt{n} \frac{\partial Q_n(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}} \xrightarrow{d} \mathcal{N}\{0, \mathbf{I}\}. \quad (2.43)$$

Proof of Lemma 2.5. In this proof, K and ρ denote generic constants whose values can be modified and such that $K > 0$ and $\rho \in (0, 1)$.

By the definition of $Q_n(\boldsymbol{\vartheta})$ and $\tilde{Q}_n(\boldsymbol{\vartheta})$, (2.40) and (2.41) are entailed by

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \left\| \frac{\partial \ell_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}} - \frac{\partial \tilde{\ell}_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}} \right\| \rightarrow 0 \quad \text{a.s.}, \quad (2.44)$$

$$\sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0) \cap \Theta} \left\| \frac{\partial^2 \ell_t(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} - \frac{\partial^2 \tilde{\ell}_t(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \right\| \rightarrow 0 \quad \text{a.s.} \quad (2.45)$$

By the arguments used to show (7.60) in ?, we have

$$\left| \frac{\partial \ell_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}_i} - \frac{\partial \tilde{\ell}_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}_i} \right| \leq K \rho^t (1 + K \eta_t^2) \left| 1 + \frac{1}{\sigma_t^\delta(\boldsymbol{\vartheta}_0)} \frac{\partial \sigma_t^\delta(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}_i} \right|.$$

Under **A10-A11**, we have $E \eta_t^4 < \infty$ and (2.34), and thus the expectation of the right-hand side of the inequality is bounded by $K \rho^t$. It follows that

$$\sum_{t=1}^{\infty} \left| \frac{\partial \ell_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}_i} - \frac{\partial \tilde{\ell}_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}_i} \right|$$

has a finite expectation, and thus is finite almost surely, which entails (2.44). The convergence (2.45) is shown by arguments which follow the scheme of the proof of the last part of (d) on Page 167 in [Francq and Zakoian \(2010\)](#).

To establish (2.42), first note that

$$P \left(\left\| \frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \tilde{Q}_n(\boldsymbol{\vartheta}_n) - \mathbf{J} \right\| \geq \epsilon \right) \leq a_1 + a_2 + a_3 + a_4,$$

where

$$\begin{aligned} a_1 &= P \left(\sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0) \cap \Theta} \left\| \frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \tilde{Q}_n(\boldsymbol{\vartheta}) - \frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} Q_n(\boldsymbol{\vartheta}) \right\| \geq \frac{\epsilon}{3} \right), \\ a_2 &= P \left(\sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0) \cap \Theta} \left\| \frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} Q_n(\boldsymbol{\vartheta}) - \frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} Q_n(\boldsymbol{\vartheta}_0) \right\| \geq \frac{\epsilon}{3} \right), \\ a_3 &= P \left(\left\| \frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} Q_n(\boldsymbol{\vartheta}_0) - \mathbf{J} \right\| \geq \frac{\epsilon}{3} \right), \quad a_4 = P \{ \boldsymbol{\vartheta}_n \notin V(\boldsymbol{\vartheta}_0) \} \end{aligned}$$

for any $\epsilon > 0$ and any neighborhood $V(\boldsymbol{\vartheta}_0)$. By the assumption that $\boldsymbol{\vartheta}_n \rightarrow \boldsymbol{\vartheta}_0$ in probability, we have $a_4 \rightarrow 0$ as $n \rightarrow \infty$. By (2.41), for any $\epsilon > 0$ and when $V(\boldsymbol{\vartheta}_0)$ is sufficiently small, $a_1 \rightarrow 0$. The ergodic theorem and Lemma 2.4 imply that $a_3 \rightarrow 0$ for any $\epsilon > 0$. To prove that $a_2 \rightarrow 0$, it suffices

to show that, for all $\epsilon > 0$, there exists $V(\boldsymbol{\vartheta}_0)$ satisfying

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0) \cap \Theta} \left\| \frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \ell_t(\boldsymbol{\vartheta}) - \frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \ell_t(\boldsymbol{\vartheta}_0) \right\| \leq \epsilon \quad \text{a.s.}$$

The result follows from the ergodic theorem, the dominated convergence theorem, the uniform continuity of the second order derivatives of $\ell_t(\boldsymbol{\vartheta})$, and by showing that

$$E \sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0) \cap \Theta} \left\| \frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \ell_t(\boldsymbol{\vartheta}) \right\| < \infty \quad (2.46)$$

for some neighborhood $V(\boldsymbol{\vartheta}_0)$.

We begin by showing that, for any $s_0 > 0$, there exists $V(\boldsymbol{\vartheta}_0)$ such that $V(\boldsymbol{\vartheta}_0) \subset \Theta$ and

$$E \sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0)} \left| \frac{\sigma_t^2(\boldsymbol{\vartheta}_0)}{\sigma_t^2(\boldsymbol{\vartheta})} \right|^{s_0} < \infty. \quad (2.47)$$

Let us first consider the case $p = q = 1$. When $\boldsymbol{\vartheta}_0 \in \overset{\circ}{\Theta}$, there exists $V(\boldsymbol{\vartheta}_0)$ such that for all $\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0)$

$$\frac{\sigma_t^\delta(\boldsymbol{\vartheta}_0)}{\sigma_t^\delta(\boldsymbol{\vartheta})} \leq K + K \sum_{j=0}^{\infty} \frac{\beta_0^j}{\beta^j} \left\{ \frac{\frac{\alpha_+}{\omega} \beta^j \varepsilon_{t-j-1}^{+\delta}}{1 + \frac{\alpha_+}{\omega} \beta^j \varepsilon_{t-j-1}^{+\delta}} + \frac{\frac{\alpha_-}{\omega} \beta^j \varepsilon_{t-j-1}^{+\delta}}{1 + \frac{\alpha_-}{\omega} \beta^j \varepsilon_{t-j-1}^{-\delta}} + \frac{\frac{1}{\omega} \beta^j \boldsymbol{\pi}' \mathbf{x}_{t-j-1}}{1 + \frac{1}{\omega} \beta^j \boldsymbol{\pi}' \mathbf{x}_{t-j-1}} \right\}.$$

Now, consider the case where $\boldsymbol{\vartheta}_0$ belongs to the boundary of Θ . Assume for instance that $\alpha_{0+} = 0$, $\alpha_{0-} > 0$, $\beta_0 > 0$ and $\boldsymbol{\pi}_0 > 0$. Then there exists $V(\boldsymbol{\vartheta}_0)$ such that for all $\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0)$

$$\frac{\sigma_t^\delta(\boldsymbol{\vartheta}_0)}{\sigma_t^\delta(\boldsymbol{\vartheta})} \leq K + K \sum_{j=0}^{\infty} \frac{\beta_0^j}{\beta^j} \left\{ \frac{\frac{\alpha_-}{\omega} \beta^j \varepsilon_{t-j-1}^{+\delta}}{1 + \frac{\alpha_-}{\omega} \beta^j \varepsilon_{t-j-1}^{-\delta}} + \frac{\frac{1}{\omega} \beta^j \boldsymbol{\pi}' \mathbf{x}_{t-j-1}}{1 + \frac{1}{\omega} \beta^j \boldsymbol{\pi}' \mathbf{x}_{t-j-1}} \right\}.$$

Similar inequality hold true in the other cases. For all $\xi > 0$, there exists $V(\boldsymbol{\vartheta}_0)$ such that $\beta_0 \leq (1 + \xi)\beta$ for all $\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0)$. It follows that for all $\xi > 0$ and $s \in (0, 1)$, using the inequality $x/(1+x) \leq x^s$ for all $x \geq 0$, there exists $V(\boldsymbol{\vartheta}_0)$ such that

$$\sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0)} \frac{\sigma_t^\delta(\boldsymbol{\vartheta}_0)}{\sigma_t^\delta(\boldsymbol{\vartheta})} \leq K + K \sum_{j=0}^{\infty} (1 + \xi)^j \rho^{js} |\varepsilon_{t-j-1}|^{\delta s} + K \sum_{j=0}^{\infty} (1 + \xi)^j \rho^{js} \|\mathbf{x}_{t-j-1}\|^s.$$

The same inequality can be obtained for general orders p and q . Without loss of generality, it can be assumed that $2s_0/\delta \geq 1$. By the Minskowski inequality, choosing s such that $E|\varepsilon_1|^{2ss_0} < \infty$ and

$E\|\mathbf{x}_1\|^{2ss_0/\delta} < \infty$ and choosing for instance $\xi = \frac{1 - \rho^s}{2\rho^s}$, we have

$$\begin{aligned} & \left\| \sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0)} \frac{\sigma_t^2(\boldsymbol{\vartheta}_0)}{\sigma_t^2(\boldsymbol{\vartheta})} \right\|_{s_0}^{\delta/2} = \left\| \sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0)} \frac{\sigma_t^\delta(\boldsymbol{\vartheta}_0)}{\sigma_t^\delta(\boldsymbol{\vartheta})} \right\|_{2s_0/\delta} \\ & \leq K + K \sum_{j=0}^{\infty} (1 + \xi)^j \rho^{js} \left\| |\varepsilon_1|^{\delta s} \right\|_{2s_0/\delta} + K \sum_{j=0}^{\infty} (1 + \xi)^j \rho^{js} \|\mathbf{x}_1\|^s \left\| \right\|_{2s_0/\delta} < \infty. \end{aligned}$$

We thus have shown (2.47).

Now, let us prove (2.46) in Cases A and C. In view of (2.33), **A2** and the Cauchy-Schwarz inequality entail

$$\begin{aligned} E \sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0)} \left\| \frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \ell_t(\boldsymbol{\vartheta}) \right\| & \leq K \left\| 1 + \sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0)} \left\| \frac{\sigma_t^2(\boldsymbol{\vartheta}_0)}{\sigma_t^2(\boldsymbol{\vartheta})} \right\|_2 \left\{ \left\| \sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0)} \left\| \frac{1}{\sigma_t^\delta} \frac{\partial^2 \sigma_t^\delta(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \right\|_2 \right. \right. \\ & \left. \left. + \left\| \sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0)} \left\| \frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}} \right\|_4 \right\|^2 \right\}. \end{aligned}$$

Hence, in Cases A and C, (2.46) comes from (2.39) and (2.47), and the analog (2.39) for second order derivatives.

It remains to prove (2.46) in Cases B and D. Using (2.33), **A2**, (2.47) and the Hölder inequality it suffices to show that

$$E \sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0)} \left\| \frac{1}{\sigma_t^\delta} \frac{\partial^2 \sigma_t^\delta(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \right\|^{1+s_1} < \infty, \quad E \sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0)} \left\| \frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} \right\|^{2+s_1} < \infty \quad (2.48)$$

for some $s_1 > 0$. Since σ_t is bounded away from zero, to show (3.6), it suffices to establish that

$$E \sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0) \cap \Theta} \left\| \frac{\partial^2 \sigma_t^\delta(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \right\|^{1+s_1} < \infty, \quad E \sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0) \cap \Theta} \left\| \frac{\partial \sigma_t^\delta(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} \right\|^{2+s_1} < \infty. \quad (2.49)$$

Now, (2.9), the second part of **A5** and the compactness of Θ entail that

$$\frac{\partial \sigma_t^\delta}{\partial \boldsymbol{\vartheta}_i} = d_{i,0}(\boldsymbol{\vartheta}) + \sum_{k=1}^{\infty} d_{i,k}^+(\boldsymbol{\vartheta}) (\varepsilon_{t-k}^+)^{\delta} + d_{i,k}^-(\boldsymbol{\vartheta}) (\varepsilon_{t-k}^-)^{\delta} + \boldsymbol{\pi}'_{i,k}(\boldsymbol{\vartheta}) \mathbf{x}_{t-k},$$

with

$$\sup_{\boldsymbol{\vartheta} \in \Theta} |d_{i,0}(\boldsymbol{\vartheta})| \leq K, \quad \sup_{\boldsymbol{\vartheta} \in \Theta} \max \left\{ |d_{i,k}^+(\boldsymbol{\vartheta})|, |d_{i,k}^-(\boldsymbol{\vartheta})|, \|\boldsymbol{\pi}_{i,k}(\boldsymbol{\vartheta})\| \right\} \leq K \rho^k.$$

The second moment condition in (2.49) thus follows from **A11**. The first moment condition is obtained by doing similar developments for the second order derivatives. We thus have shown (2.46)

in all cases, which completes the proof of (2.42).

By (2.32), Lemma 2.4 and **A2**, the last result, (2.43), is a consequence of the central limit theorem for square integrable martingale difference of Billingsley (1961). \square

Proof of Theorem 2.2. A Taylor expansion of $\tilde{Q}_n(\boldsymbol{\vartheta})$ around $\boldsymbol{\vartheta}_0$ gives

$$\tilde{Q}_n(\boldsymbol{\vartheta}) - \tilde{Q}_n(\boldsymbol{\vartheta}_0) = \frac{\partial Q_n(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}'} (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0) + \frac{1}{2} (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0)' \mathbf{J} (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0) + R_n(\boldsymbol{\vartheta}),$$

where

$$R_n(\boldsymbol{\vartheta}) = \left\{ \frac{\partial \tilde{Q}_n(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}'} - \frac{\partial Q_n(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}'} \right\} (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0) + \frac{1}{2} (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0)' \left(\frac{\partial^2 \tilde{Q}_n(\boldsymbol{\vartheta}^*)}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} - \mathbf{J} \right) (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0),$$

and $\boldsymbol{\vartheta}^*$ is between $\boldsymbol{\vartheta}$ and $\boldsymbol{\vartheta}_0$. In view of (2.40), (2.41) and (2.42), as $n \rightarrow \infty$, we have

$$nR_n(\boldsymbol{\vartheta}_n) = o_P \{ \sqrt{n}(\boldsymbol{\vartheta}_n - \boldsymbol{\vartheta}_0) \} + o_P \{ n \|\boldsymbol{\vartheta}_n - \boldsymbol{\vartheta}_0\|^2 \} \quad (2.50)$$

when $\boldsymbol{\vartheta}_n - \boldsymbol{\vartheta}_0 = o_P(1)$. Therefore

$$nR_n(\boldsymbol{\vartheta}_n) = o_P(1) \quad \text{when} \quad \sqrt{n}(\boldsymbol{\vartheta}_n - \boldsymbol{\vartheta}_0) = O_P(1). \quad (2.51)$$

Letting

$$\mathbf{Z}_n = -\mathbf{J}^{-1} \sqrt{n} \frac{\partial Q_n(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}},$$

we obtain the following quadratic approximation of the objective function

$$\tilde{Q}_n(\boldsymbol{\vartheta}) - \tilde{Q}_n(\boldsymbol{\vartheta}_0) = \frac{1}{2n} \|\sqrt{n}(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0) - \mathbf{Z}_n\|_{\mathbf{J}}^2 - \frac{1}{2n} \|\mathbf{Z}_n\|_{\mathbf{J}}^2 + R_n(\boldsymbol{\vartheta}). \quad (2.52)$$

Let

$$\boldsymbol{\vartheta}_{\mathbf{Z}_n} = \arg \inf_{\boldsymbol{\vartheta} \in \Theta} \|\sqrt{n}(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0) - \mathbf{Z}_n\|_{\mathbf{J}}.$$

Note that

$$\sqrt{n}(\boldsymbol{\vartheta}_{\mathbf{Z}_n} - \boldsymbol{\vartheta}_0) = \mathbf{Z}_n^{\mathcal{C}} \text{ for } n \text{ large enough,} \quad (2.53)$$

where $\mathbf{Z}_n^{\mathcal{C}}$ denotes the projection of \mathbf{Z}_n on \mathcal{C} . We have

$$\begin{aligned} 0 &\leq \|\sqrt{n}(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) - \mathbf{Z}_n\|_{\mathbf{J}}^2 - \|\sqrt{n}(\boldsymbol{\vartheta}_{\mathbf{Z}_n} - \boldsymbol{\vartheta}_0) - \mathbf{Z}_n\|_{\mathbf{J}}^2 \\ &= 2n \left\{ \tilde{Q}_n(\hat{\boldsymbol{\vartheta}}_n) - \tilde{Q}_n(\boldsymbol{\vartheta}_{\mathbf{Z}_n}) \right\} + 2n \left\{ R_n(\boldsymbol{\vartheta}_{\mathbf{Z}_n}) - R_n(\hat{\boldsymbol{\vartheta}}_n) \right\} \\ &\leq 2n \left\{ R_n(\hat{\boldsymbol{\vartheta}}_n) - R_n(\boldsymbol{\vartheta}_{\mathbf{Z}_n}) \right\}, \end{aligned}$$

where the first inequality comes from the definition of $\boldsymbol{\vartheta}_{\mathbf{Z}_n}$, the equality from (2.52), and the second

inequality from the definition of $\widehat{\boldsymbol{\vartheta}}_n$. By (2.53), it follows that

$$\left| \|\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) - \mathbf{Z}_n\|_{\mathbf{J}}^2 - \|\mathbf{Z}_n^{\mathbf{C}} - \mathbf{Z}_n\|_{\mathbf{J}}^2 \right| \leq 2n \left\{ R_n(\widehat{\boldsymbol{\vartheta}}_n) - R_n(\boldsymbol{\vartheta}_{\mathbf{Z}_n}) \right\}. \quad (2.54)$$

We now show that $R_n(\widehat{\boldsymbol{\vartheta}}_n) - R_n(\boldsymbol{\vartheta}_{\mathbf{Z}_n}) = o_P(1)$. By definition of $\boldsymbol{\vartheta}_{\mathbf{Z}_n}$ and since $\boldsymbol{\vartheta}_0 \in \Theta$, we also have

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

The Minskowski inequality then entails that

$$\|\sqrt{n}(\boldsymbol{\vartheta}_{\mathbf{Z}_n} - \boldsymbol{\vartheta}_0)\|_{\mathbf{J}} \leq \|\sqrt{n}(\boldsymbol{\vartheta}_{\mathbf{Z}_n} - \boldsymbol{\vartheta}_0) - \mathbf{Z}_n\|_{\mathbf{J}} + \|\mathbf{Z}_n\|_{\mathbf{J}} \leq 2\|\mathbf{Z}_n\|_{\mathbf{J}}.$$

By (2.43), we have $\|\mathbf{Z}_n\|_{\mathbf{J}} = O_P(1)$, and thus $\sqrt{n}(\boldsymbol{\vartheta}_{\mathbf{Z}_n} - \boldsymbol{\vartheta}_0) = O_P(1)$. In view of (2.51), this entails $nR_n(\boldsymbol{\vartheta}_{\mathbf{Z}_n}) = o_P(1)$. By definition of $\widehat{\boldsymbol{\vartheta}}_n$ and (2.52), we have

$$0 \leq 2n\widetilde{Q}_n(\boldsymbol{\vartheta}_0) - 2n\widetilde{Q}_n(\widehat{\boldsymbol{\vartheta}}_n) = \|\mathbf{Z}_n\|_{\mathbf{J}}^2 - \left\| \sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) - \mathbf{Z}_n \right\|_{\mathbf{J}}^2 - 2nR_n(\widehat{\boldsymbol{\vartheta}}_n).$$

It follows that, by the c_r -inequality,

$$\begin{aligned} \left\| \sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) \right\|_{\mathbf{J}}^2 &\leq 2 \left(\left\| \sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) - \mathbf{Z}_n \right\|_{\mathbf{J}}^2 + \|\mathbf{Z}_n\|_{\mathbf{J}}^2 \right) \\ &\leq 4\|\mathbf{Z}_n\|_{\mathbf{J}}^2 - 4nR_n(\widehat{\boldsymbol{\vartheta}}_n). \end{aligned}$$

We now show that $\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) = O_P(1)$, arguing by contradiction. The consistency of $\widehat{\boldsymbol{\vartheta}}_n$ and (2.50) then entail that $nR_n(\widehat{\boldsymbol{\vartheta}}_n) = o_P\left(\left\| \sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) \right\|_{\mathbf{J}}^2\right)$. It follows that $\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) = O_P(1)$. From (2.51), we deduce that $nR_n(\widehat{\boldsymbol{\vartheta}}_n) = o_P(1)$. Thus, (2.54) can be rewritten as

$$\|\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) - \mathbf{Z}_n\|_{\mathbf{J}}^2 - \|\mathbf{Z}_n^{\mathbf{C}} - \mathbf{Z}_n\|_{\mathbf{J}}^2 = o_P(1). \quad (2.55)$$

Using the characterization (2.12) of the projection on a convex cone, we have

$$\begin{aligned} \|\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) - \mathbf{Z}_n\|_{\mathbf{J}}^2 &= \|\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) - \mathbf{Z}_n^{\mathbf{C}}\|_{\mathbf{J}}^2 + \|\mathbf{Z}_n^{\mathbf{C}} - \mathbf{Z}_n\|_{\mathbf{J}}^2 \\ &\quad + 2 \left\langle \sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) - \mathbf{Z}_n^{\mathbf{C}}, \mathbf{Z}_n^{\mathbf{C}} - \mathbf{Z}_n \right\rangle_{\mathbf{J}} \\ &\geq \|\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) - \mathbf{Z}_n^{\mathbf{C}}\|_{\mathbf{J}}^2 + \|\mathbf{Z}_n^{\mathbf{C}} - \mathbf{Z}_n\|_{\mathbf{J}}^2. \end{aligned}$$

Using (2.55), we thus obtain

$$\|\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) - \mathbf{Z}_n^{\mathbf{C}}\|_{\mathbf{J}}^2 \leq \|\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) - \mathbf{Z}_n\|_{\mathbf{J}}^2 - \|\mathbf{Z}_n^{\mathbf{C}} - \mathbf{Z}_n\|_{\mathbf{J}}^2 = o_P(1).$$

The continuous mapping theorem and (2.43) entail $\mathbf{Z}_n^c \xrightarrow{d} \mathbf{Z}^c$, and the conclusion follows. \square

Proof of Proposition 2.1. We only show (2.17) because (2.16) and (2.18) are obtained directly from arguments used in the proof of Theorem 2.1. Let

$$\mathbf{I}_n(\boldsymbol{\vartheta}) = \frac{4}{\delta^2} \frac{1}{n} \sum_{t=1}^n S_t(\boldsymbol{\vartheta}), \quad S_t(\boldsymbol{\vartheta}) = \left(\frac{\varepsilon_t^4}{\sigma_t^4(\boldsymbol{\vartheta})} - 1 \right) \frac{1}{\sigma_t^{2\delta}(\boldsymbol{\vartheta})} \frac{\partial \sigma_t^\delta(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} \frac{\partial \sigma_t^\delta(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}'}. \quad (2.56)$$

The difference between $\mathbf{I}_n(\widehat{\boldsymbol{\vartheta}}_n)$ and $\widehat{\mathbf{I}}_n$ is due to the initial values used to compute $\widetilde{\sigma}_t^\delta(\boldsymbol{\vartheta})$. Because the difference $\sigma_t^\delta(\boldsymbol{\vartheta}) - \widetilde{\sigma}_t^\delta(\boldsymbol{\vartheta})$ does not depend on the covariates (\mathbf{x}_t) , one can use standard arguments to show that

$$\lim_{n \rightarrow \infty} \left\| \mathbf{I}_n(\widehat{\boldsymbol{\vartheta}}_n) - \widehat{\mathbf{I}}_n \right\| = 0 \quad \text{a.s.}$$

The ergodic theorem shows that

$$\lim_{n \rightarrow \infty} \left\| \mathbf{I}_n(\boldsymbol{\vartheta}_0) - \mathbf{I} \right\| = 0 \quad \text{a.s.}$$

Using the strong convergence of $\widehat{\boldsymbol{\vartheta}}_n$ to $\boldsymbol{\vartheta}_0$, it remains to show that, for any $\epsilon > 0$, there exists $V(\boldsymbol{\vartheta}_0)$ such that

$$\lim_{n \rightarrow \infty} \sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0)} \left\| \mathbf{I}_n(\boldsymbol{\vartheta}) - \mathbf{I}_n(\boldsymbol{\vartheta}_0) \right\| \leq \epsilon.$$

It suffices to show that there exists $V(\boldsymbol{\vartheta}_0)$ such that

$$E \sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0) \cap \Theta} \|S_t(\boldsymbol{\vartheta})\| < \infty, \quad (2.57)$$

and the result will follow from the ergodic theorem applied to the sequence $\left\{ \sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0)} \|S_t(\boldsymbol{\vartheta}) - S_t(\boldsymbol{\vartheta}_0)\| \right\}_t$, the dominated convergence theorem, and the uniform continuity of $\boldsymbol{\vartheta} \mapsto S_t(\boldsymbol{\vartheta})$ in the neighborhood of $\boldsymbol{\vartheta}_0$.

In Cases A and C, using the Hölder inequality, (2.57) follows from **A10**, (2.39) and (2.47). In Cases B, (2.57) follows from **A10**, (2.47) and the second part of (3.6). In Cases D, (2.57) follows from **A10**, **A11**, (2.47) and the second part of (3.6) with $s_1 = 8/\nu$. \square

Proof of Theorem 2.3. The proof is the same as that of Theorem 2.2, except that the proofs of Lemmas 2.4 and 2.5 have to be modified in Cases B and D. First note that $\boldsymbol{\theta}_0 > 0$ entails that for any $s_0 > 0$ there exists $V(\boldsymbol{\vartheta}_0)$ included in Θ such that

$$E \sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0)} \left\| \frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\|^{s_0} < \infty \quad \text{and} \quad E \sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0)} \left\| \frac{1}{\sigma_t^\delta} \frac{\partial^2 \sigma_t^\delta(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right\|^{s_0} < \infty. \quad (2.58)$$

To show (2.34), it remains to show that

$$E \left\| \frac{\partial \sigma_t^\delta(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\pi}} \right\|^2 < \infty,$$

which follows from (2.37) and the existence of $E\|\boldsymbol{x}_t\|^2$ implied by **A12**. Similarly the condition $E\|\boldsymbol{x}_t\|^{2+8/\nu} < \infty$ in **A12** entails (2.36) in Case D, and **A12** entails (2.35) in the two cases B and D. The results of Lemma 2.4 thus continue to hold. Turning to the proof of Lemma 2.5, in view of (2.33), one can prove (2.46) by showing that

$$E \sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0) \cap \Theta} \left\| \eta_t^2 \frac{\sigma_t^2(\boldsymbol{\vartheta}_0)}{\sigma_t^2(\boldsymbol{\vartheta})} \frac{1}{\sigma_t^\delta(\boldsymbol{\vartheta})} \frac{\partial \sigma_t^\delta}{\partial \boldsymbol{\theta}} \frac{1}{\sigma_t^\delta(\boldsymbol{\vartheta})} \frac{\partial \sigma_t^\delta}{\partial \boldsymbol{\theta}'}(\boldsymbol{\vartheta}) \right\| < \infty, \quad (2.59)$$

$$E \sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0) \cap \Theta} \left\| \eta_t^2 \frac{\sigma_t^2(\boldsymbol{\vartheta}_0)}{\sigma_t^2(\boldsymbol{\vartheta})} \frac{1}{\sigma_t^\delta(\boldsymbol{\vartheta})} \frac{\partial^2 \sigma_t^\delta}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}(\boldsymbol{\vartheta}) \right\| < \infty, \quad (2.60)$$

$$E \sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0) \cap \Theta} \left\| \eta_t^2 \frac{\sigma_t^2(\boldsymbol{\vartheta}_0)}{\sigma_t^2(\boldsymbol{\vartheta})} \frac{\partial \sigma_t^\delta}{\partial \boldsymbol{\pi}} \frac{\partial \sigma_t^\delta}{\partial \boldsymbol{\pi}'}(\boldsymbol{\vartheta}) \right\| < \infty, \quad E \sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0) \cap \Theta} \left\| \eta_t^2 \frac{\sigma_t^2(\boldsymbol{\vartheta}_0)}{\sigma_t^2(\boldsymbol{\vartheta})} \frac{\partial^2 \sigma_t^\delta}{\partial \boldsymbol{\pi} \partial \boldsymbol{\pi}'}(\boldsymbol{\vartheta}) \right\| < \infty \quad (2.61)$$

and similar moments for cross derivatives with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\pi}$. Using Hölder's inequality, the moment conditions (2.59)–(2.60) come from (2.47), (2.58) and $E(\eta_t^2 \mid \mathcal{F}_{t-1}) = 1$, and the moment conditions (2.61) come from the same arguments and **A12**. \square

Proof of Corollaries 2.1, 2.2 and 2.3. As in Francq and Zakoïan (2009b). \square

References

- Andrews, D.W.** (1999) Estimation when a parameter is on a boundary. *Econometrica* 67, 1341–1383.
- Andrews, D.W.** (2001) Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica* 69, 683–734.
- Berkes, I., Horváth, L. and P. Kokoszka** (2003) GARCH processes : structure and estimation. *Bernoulli* 9, 201–227.
- Billingsley, P.** (1961) The Lindeberg-Levy theorem for martingales. *Proceedings of the American Mathematical Society* 12, 788–792.
- Billingsley, P.** (1995) *Probability and measure*. John Wiley.
- Bollerslev, T.** (2008) Glossary to ARCH (GARCH). *Volatility and Time Series Econometrics : Essays in Honor of Robert F. Engle* (eds. T. Bollerslev, J. R. Russell and M. Watson), Oxford University Press, Oxford, UK.
- Bougerol, P. and N. Picard** (1992a) Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics* 52, 115–127.

- Bougerol, P. and N. Picard** (1992b) Strict stationarity of generalized autoregressive processes. *Annals of Probability* 20, 1714–1729.
- Brandt, A.** (1986) The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Advance in Applied Probability* 18, 221–254.
- Ding, Z., Granger, C. and R.F. Engle** (1993) A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1, 83–106.
- Engle, R.F., Hendry, D.F. and J.F. Richard** (1983) Exogeneity. *Econometrica* 51, 277–304.
- Engle, R.F. and A.J. Patton** (2001) What good is a volatility model. *Quantitative finance* 1, 237–245.
- Escanciano, J.C.** (2009) Quasi-maximum likelihood estimation of semi-strong GARCH models. *Econometric Theory* 25, 561–570.
- Fisher, T.J. and C.M. Gallagher** (2012) New weighted portmanteau statistics for time series goodness of fit testing. *Journal of the American Statistical Association* 107, 777–787.
- Francq, C. and J-M. Zakoïan** (2004) Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* 10, 605–637.
- Francq, C. and J-M. Zakoïan** (2007) Quasi-Maximum Likelihood Estimation in GARCH Processes when some coefficients are equal to zero. *Stochastic Processes and their Applications* 117, 1265–1284.
- Francq, C. and J-M. Zakoïan** (2009a) Bartlett’s formula for a general class of nonlinear processes. *Journal of Time Series Analysis* 30, 449–465.
- Francq, C. and J-M. Zakoïan** (2009b) Testing the nullity of GARCH coefficients : correction of the standard tests and relative efficiency comparisons. *Journal of the American Statistical Association* 104, 313–324.
- Francq, C. and J-M. Zakoïan** (2010) *GARCH Models : Structure, Statistical Inference and Financial Applications*. John Wiley.
- Fuertes, A.M., Izzeldin, M. and E. Kalotychou** (2009) On forecasting daily stock volatility : the role of intraday information and market conditions. *International Journal of Forecasting* 25, 259–281.
- Glosten, L.R., Jaganathan, R. and D. Runkle** (1993) On the relation between the expected values and the volatility of the nominal excess return on stocks. *Journal of Finance* 48, 1779–1801.
- Guo, S., Ling, S., and Zhu, K.** (2014) Factor double autoregressive models with application to simultaneous causality testing. *Journal of Statistical Planning and Inference* 148, 82–94.

- Hall, P. and Q. Yao** (2003) Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica* 71, 285–317.
- Hamadeh, T. and J-M. Zakoïan** (2011) Asymptotic properties of LS and QML estimators for a class of nonlinear GARCH processes. *Journal of Statistical Planning and Inference* 141, 488–507.
- Han, H. and D. Kristensen** (2014) Asymptotic Theory for the QMLE in GARCH-X Models With Stationary and Nonstationary Covariates. *Journal of Business & Economic Statistics* 32, 416–429.
- Kristensen, D. and A. Rahbek** (2005) Asymptotics of the QMLE for a class of ARCH(q) models. *Econometric Theory* 21, 946–961.
- Laurent, S., Lecourt, C. and F.C. Palm** (2016) Testing for jumps in conditionally Gaussian ARMA-GARCH models, a robust approach. *Computational Statistics & Data Analysis* 100, 383–400.
- Li, W.K. and T.K. Mak** (1994) On the Squared Residual Autocorrelations in Non-Linear Time Series With Conditional Heteroskedasticity. *Journal of Time Series Analysis* 15, 627–636.
- Ling, S.** (2007) Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models. *Journal of Econometrics* 140, 849–873.
- Nelson, D. B.** (1991) Conditional heteroskedasticity in asset returns : A new approach. *Econometrica* 59, 347–370.
- Nijman, T. and E. Sentana** (1996) Marginalization and contemporaneous aggregation in multivariate GARCH processes. *Journal of Econometrics* 71, 71–87.
- Pan, J., Wang, H. and H. Tong** (2008) Estimation and tests for power-transformed and threshold GARCH models. *Journal of Econometrics* 142, 352–378.
- Pedersen, R. S.** (2017) Inference and Testing on the Boundary in Extended Constant Conditional Correlation GARCH Models. *Journal of Econometrics* 196, 23–36.
- Sucarrat, G. and A. Escribano** (2010) The Power Log-GARCH Model. *Working document*, Economic Series 10-13, University Carlos III, Madrid.
- Taylor, S.J.** (1986) *Modelling Financial Time Series*, New-York : Wiley.
- Wintenberger, O.** (2013) Continuous invertibility and stable QML estimation of the EGARCH(1,1) model. *Scandinavian Journal of Statistics* 40, 846–867.
- Zakoïan, J-M.** (1994) Threshold heteroskedastic models. *Journal of Economic Dynamics and Control* 18, 931–955.

Chapitre 3

Variance targeting estimation of the BEKK-X model

Abstract. This paper studies the BEKK model with exogenous variables (BEKK-X), which intends to take into account the influence of explanatory variables on the conditional covariance of the asset returns. Strong consistency and asymptotic normality of a variance targeting estimator (VTE) is proved. Monte Carlo experiments and an application to financial series illustrate the asymptotic results.

Keywords : BEKK model augmented with exogenous variables, BEKK-X model, Variance targeting estimation (VTE).

3.1 Introduction

Analysing asset return covariances is important since it is a crucial input, in particular, for portfolio selection, asset management and risk assessment. Forecasting sequences of covariance matrices can be done by using multivariate conditional heteroskedastic (GARCH) models, see [Bauwens and Rombouts \(2006\)](#) and [Silvennoinen and Terasvirta \(2009\)](#) for extensive surveys. The first generation of models, for example the VEC model of [Bollerslev and Wooldridge \(1988\)](#) and the BEKK model of [Engle and Kroner \(1995\)](#), are direct extensions of the univariate GARCH model of [Bollerslev \(1986\)](#). These models take into account the information contained in the past of the individual asset returns, but can not exploit external information. In practice, the relations between the volatilities and co-volatilities of several markets can be affected by external influences that we call exogenous variables. In this paper, we employ this term in a wide sens for any external explanatory variable (see [Koopmans and Reiersol \(1950\)](#), [Engle et al. \(1983\)](#), [Florens and Mouchart \(1982\)](#) and [Bouissou](#)

and Vuong (1986) for other concepts of exogeneity). Engle (2009) provide evidence that economic fundamentals such as inflation and industrial production growth drive stock market volatility. Cakmakli and Dijk (2010) demonstrate that a number of macroeconomic variables can help predicting US stocks volatility. Christiansen and Schrimpf (2012) get similar results for the foreign exchange, the commodities, and the bond market.

Despite the fact that such additional information in financial and macroeconomic variables are widely used to explain and forecast volatility in financial markets, there are, however, relatively few results on the asymptotic behavior of the estimation in presence of exogenous variables. In the univariate case, Han and Kristensen (2014) give conditions for the Consistency and Asymptotic Normality (CAN) of the Gaussian QMLE for the standard GARCH(1,1) augmented by a single covariate. Francq and Thieu (2015) study the asymptotic distribution of the QMLE for a versatile class of model : The Asymmetric Power ARCH(p, q)-X model with an unrestrictive number of the exogenous variables. In the multivariate case, Francq and Sucarrat (2015) provide the proof of the CAN of an estimator of the volatilities for the components of a vectorial log-GARCH model with covariates. Their framework does not directly specify the conditional covariance. Engle and Kroner (1995) suggest the BEKK model augmented by exogenous variables (BEKK-X). In their model, the covariates can affect all the volatilities and co-volatilities of the returns. However, they only provide the estimation of the model without exogenous influences. Moreover, the CAN of their estimator is not proved. In this paper, the estimation of the BEKK-X model will be presented and its CAN will be established. There are several advantages with the BEKK model. First, although the asymptotic theory for multivariate GARCH has been less investigated than the asymptotic theory for univariate models, several papers have established asymptotic results for different methods of estimation of the BEKK model without covariates. Comte and Lieberman (2003) show the CAN of the QMLE. However, the BEKK model contains a large number of parameters, even for moderate dimensions. This implies that it is difficult to estimate the model by the classical QMLE. Pedersen and Rahbek (2014) consider a simplified estimation method, the variance targeting estimation (VTE), and provide its CAN. The VTE method has been proposed by Engle and Mezrich (1996) to alleviate the numerical difficulties encountered in the maximization of the quasi likelihood. The VTE is numerically more efficient than the QMLE, in particular, in the presence of exogenous variables, because it requires an optimization of lower dimension. This estimator has also the advantage of being relatively robust for long term predictions (see Francq and Zakoïan (2011)). In the present paper, we establish the CAN of the VTE for the BEKK-X model.

The paper is organized as follows. Section 3.2 introduces the BEKK model augmented with explanatory variables and presents the VTE method. The consistency and asymptotic distribution of the VTE are investigated in Section 3.3. Numerical illustrations are presented in Section 3.4. Section 3.5 concludes the paper. All the proofs are collected in Section 3.6.

3.2 The model and variance targeting estimation

3.2.1 The model

Let $\{\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \dots, \varepsilon_{mt})'\}$ be a m -dimensional process and $\mathbf{x}_t = (x_{1t}, \dots, x_{rt})' \in \mathbb{R}^r$ be a vector of r exogenous variables. Denote \mathcal{F}_{t-1} the σ -field generated by the past of $\boldsymbol{\varepsilon}_t$ and \mathbf{x}_t ; *i.e.* $\mathcal{F}_{t-1} = \sigma\{\boldsymbol{\varepsilon}_u, \mathbf{x}_v; u < t, v < t\}$. Assume that

$$E(\boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1}) = \mathbf{0}, \quad \text{Var}(\boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1}) = \mathbf{H}_t \text{ exists and is positive definite.} \quad (3.1)$$

The $m \times m$ matrix \mathbf{H}_t is specified as a function of the past values of $\boldsymbol{\varepsilon}_t$ and \mathbf{x}_t .

We consider the following BEKK(1,1)-X model

$$\begin{cases} \boldsymbol{\varepsilon}_t = \mathbf{H}_t^{1/2} \boldsymbol{\eta}_t \\ \mathbf{H}_t = \boldsymbol{\Omega} + \mathbf{A}\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}'_{t-1}\mathbf{A}' + \mathbf{B}\mathbf{H}_{t-1}\mathbf{B}' + \mathbf{C}\mathbf{x}_{t-1}\mathbf{x}'_{t-1}\mathbf{C}', \end{cases} \quad (3.2)$$

where $\boldsymbol{\Omega}, \mathbf{A}, \mathbf{B}$ are $m \times m$ parameter matrices and \mathbf{C} is $m \times r$ parameter matrix. To ensure the positivity of the conditional covariance \mathbf{H}_t , we assume that the coefficient matrix $\boldsymbol{\Omega} > \mathbf{0}$, where the symbol $>$ denotes the positive definiteness of a matrix. Note that in the present paper, we focus on the BEKK(1,1,1) model, the BEKK(1,1) model hereafter, as it is the most widely used in applications (see [Silvennoinen and Terasvirta \(2009\)](#))

The attractive property of the BEKK model is that the conditional covariance matrices are positive definite by construction. Introducing the explanatory variables under the form $\mathbf{C}\mathbf{x}_{t-1}\mathbf{x}'_{t-1}\mathbf{C}'$ still guarantees the positive definiteness of \mathbf{H}_t . Furthermore, they are not restricted to a single variable.

Let $\|A\| = \sqrt{\text{Tr}(A'A)}$ be the Euclidean norm of a vector or a matrix A , where $\text{Tr}(\cdot)$ is the trace of a square matrix. The following assumptions are made throughout the paper.

B1 : $E(\boldsymbol{\eta}_t | \mathcal{F}_{t-1}) = \mathbf{0}$ and $E(\boldsymbol{\eta}_t \boldsymbol{\eta}'_t | \mathcal{F}_{t-1}) = \mathbf{I}_m$.

B2 : $(\boldsymbol{\varepsilon}_t, \mathbf{x}_t)$ is a strictly stationary and ergodic process.

B3 : $E\|\mathbf{x}_t\|^2 < \infty$ and $E\|\boldsymbol{\varepsilon}_t\|^2 < \infty$.

Remark 3.1. [Boussama and Stelzer \(2011\)](#) provide sufficient conditions for the existence of a unique stationary and ergodic solution to BEKK multivariate GARCH models. For the model (3.2) without covariate, for example, this solution exists if the following assumptions are satisfied

- i) The innovation $\boldsymbol{\eta}_t$ admits a density absolutely continuous with respect to the Lebesgue measure on \mathbb{R}^m and positive in a neighborhood of the point zero
- ii) $\rho(\mathbf{A}_0 \otimes \mathbf{A}_0 + \mathbf{B}_0 \otimes \mathbf{B}_0) < 1$

where the symbol \otimes stands for the Kronecker product and $\rho(\cdot)$ denotes the spectral radius of a matrix which is the maximum among the absolute values of the eigenvalues of a matrix. Furthermore, under these conditions, the existence $E\|\varepsilon_t\|^2 < \infty$ in **B3** is satisfied.

Remark 3.2. Under Assumptions **B2**, **B3**, the intercept matrix in the volatility of (3.2) can be represented as a function of the unconditional covariance of the observations and of the unconditional second moment of covariates. That allows us to apply the variance targeting estimation method.

3.2.2 Variance targeting estimation

The VTE is a two-step estimation technique whose advantages are to reduce the computational complexity of the optimization produce (see Pedersen and Rahbek (2014), Francq and Zakoian (2011) and Francq and Zakoian (2014)) and to guarantee that the implied variance is equal to the sample variance. This method is based on a reparameterization of the volatility equation, in which the intercept is replaced by the returns unconditional variance in case that there is no covariates.

Denote by $\Sigma_\varepsilon := Var(\varepsilon_t) = E(\varepsilon_t \varepsilon_t') = E(\mathbf{H}_t)$ the variance matrix of the observations and $\Sigma_x := E(\mathbf{x}_t \mathbf{x}_t')$ the second-order moment of the vector of exogenous variables. These matrices are well defined under Assumption **B3**. By taking the expectation of the two hand sides of (3.2), we get

$$\Sigma_\varepsilon = \Omega + \mathbf{A}\Sigma_\varepsilon\mathbf{A}' + \mathbf{B}\Sigma_\varepsilon\mathbf{B}' + \mathbf{C}\Sigma_x\mathbf{C}'. \quad (3.3)$$

Then (3.2) can be reparameterized by

$$\begin{cases} \varepsilon_t = \mathbf{H}_t^{1/2} \eta_t, \\ \mathbf{H}_t = (\Sigma_\varepsilon - \mathbf{A}\Sigma_\varepsilon\mathbf{A}' - \mathbf{B}\Sigma_\varepsilon\mathbf{B}' - \mathbf{C}\Sigma_x\mathbf{C}') + \mathbf{A}\varepsilon_{t-1}\varepsilon_{t-1}'\mathbf{A}' + \mathbf{B}\mathbf{H}_{t-1}\mathbf{B}' \\ \quad + \mathbf{C}\mathbf{x}_{t-1}\mathbf{x}_{t-1}'\mathbf{C}'. \end{cases} \quad (3.4)$$

Note that in this reparametrization, the constraint of the positive definiteness of the intercept $\Omega > 0$ in (3.2) becomes

$$\Sigma_\varepsilon - \mathbf{A}\Sigma_\varepsilon\mathbf{A}' - \mathbf{B}\Sigma_\varepsilon\mathbf{B}' - \mathbf{C}\Sigma_x\mathbf{C}' > 0. \quad (3.5)$$

The generic parameter of the model (3.4) consists of the elements of the matrices Σ_ε , Σ_x and the ones of the matrices \mathbf{A} , \mathbf{B} and \mathbf{C} . As mentioned, the parameters of the model will be estimated in the two steps VTE. In the first step, the matrices Σ_ε and Σ_x will be empirically estimated. In the second step, the other parameters will be estimated by QML optimization. The vector of unknown parameters is thus decomposed by $\vartheta_0 = (\gamma_0', \theta_0')' \in \mathbb{R}^d$ with

$$\begin{aligned} \gamma_0 &= (\gamma_{x0}', \gamma_{\varepsilon0}') = (\text{vech}'(\Sigma_x), \text{vech}'(\Sigma_\varepsilon))' \in \mathbb{R}^{d_1}, \quad d_1 = \frac{r(r+1)}{2} + \frac{m(m+1)}{2}, \\ \theta_0 &= (\text{vec}'(\mathbf{A}_0), \text{vec}'(\mathbf{B}_0), \text{vec}'(\mathbf{C}_0))' \in \mathbb{R}^{d_2}, \quad d_2 = 2m^2 + mr \end{aligned}$$

and $d = d_1 + d_2$ is the total number of unknown parameter of (3.4), where vec denotes the operator that stacks all columns of a matrix into a column vector, and $vech$ denotes the one that stacks only the lower triangular part including the diagonal of a symmetric matrix into a vector. Likewise, we define the parameter space

$$\Theta := \Theta_\gamma \times \Theta_\theta \subset \mathbb{R}^{d_1} \times \mathbb{R}^{d_2},$$

whose a generic parameter vector is denoted by

$$\vartheta = (\gamma', \theta')' = (\gamma'_x, \gamma'_\varepsilon, \theta')' = (vech'(\Sigma_x), vech'(\Sigma_\varepsilon), vec'(\mathbf{A}), vec'(\mathbf{B}), vec'(\mathbf{C}))'.$$

To emphasize that the conditional covariance matrix in (3.4) depends on the parameters γ and θ and that they are independently estimated, we write $\mathbf{H}_t(\gamma, \theta)$.

Let $(\varepsilon_1, \dots, \varepsilon_n)$ be a realization of length n of the stationary ergodic process (ε_t) and $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ be n observations of the exogenous variables (\mathbf{x}_t) . Conditionally on the initial values $\varepsilon_0, \mathbf{x}_0$ and $\widetilde{\mathbf{H}}_0$, the conditional covariance matrix can be recursively defined, for $t \geq 1$, as follows

$$\begin{aligned} \widetilde{\mathbf{H}}_t(\gamma, \theta) &= (\Sigma_\varepsilon - \mathbf{A}\Sigma_\varepsilon\mathbf{A}' - \mathbf{B}\Sigma_\varepsilon\mathbf{B}' - \mathbf{C}\Sigma_x\mathbf{C}') + \mathbf{A}\varepsilon_{t-1}\varepsilon'_{t-1}\mathbf{A}' + \mathbf{B}\widetilde{\mathbf{H}}_{t-1}(\gamma, \theta)\mathbf{B}' \\ &\quad + \mathbf{C}\mathbf{x}_{t-1}\mathbf{x}'_{t-1}\mathbf{C}'. \end{aligned} \quad (3.6)$$

Let us define the functions

$$\widetilde{Q}_n(\gamma, \theta) = \frac{1}{n} \sum_{t=1}^n \widetilde{\ell}_t(\gamma, \theta), \quad \widetilde{\ell}_t(\gamma, \theta) = \varepsilon'_t \widetilde{\mathbf{H}}_t^{-1}(\gamma, \theta) \varepsilon_t + \log \det \left(\widetilde{\mathbf{H}}_t(\gamma, \theta) \right). \quad (3.7)$$

As mentioned, in the first stage of VT estimation method, the parameter $\gamma = (\gamma'_x, \gamma'_\varepsilon)'$ is pre-estimated directly from the sample by the method of moments :

$$\widehat{\gamma}_n = (\widehat{\gamma}'_{x_n}, \widehat{\gamma}'_{\varepsilon_n})' = \left(vech'(\widehat{\Sigma}_{x_n}), vech'(\widehat{\Sigma}_{\varepsilon_n}) \right)', \quad (3.8)$$

where $\widehat{\Sigma}_{\varepsilon_n} = \frac{1}{n} \sum_{t=1}^n \varepsilon_t \varepsilon'_t$ and $\widehat{\Sigma}_{x_n} = \frac{1}{n} \sum_{t=1}^n \mathbf{x}_t \mathbf{x}'_t$ are the empirical estimators of the covariance matrix of ε_t and the second-order moment of \mathbf{x}_t , respectively. Once $\widehat{\gamma}_n$ is obtained, the parameter θ is estimated by using the quasi likelihood, conditioning on the parameters estimated in the first stage

$$\widetilde{Q}_n(\widehat{\gamma}_n, \theta) = \frac{1}{n} \sum_{t=1}^n \widetilde{\ell}_t(\widehat{\gamma}_n, \theta) = \frac{1}{n} \sum_{t=1}^n \varepsilon'_t \widetilde{\mathbf{H}}_t^{-1}(\widehat{\gamma}_n, \theta) \varepsilon_t + \log \det \left(\widetilde{\mathbf{H}}_t(\widehat{\gamma}_n, \theta) \right), \quad (3.9)$$

where the covariance process $\widetilde{\mathbf{H}}_t(\widehat{\gamma}_n, \theta)$ can be recursively calculated by replacing Σ_ε and Σ_x in (3.6) by $\widehat{\Sigma}_{\varepsilon_n}$ and $\widehat{\Sigma}_{x_n}$ respectively. The estimator of the parameter θ is thus defined as any

measurable solution $\widehat{\boldsymbol{\theta}}_n$ of

$$\widehat{\boldsymbol{\theta}}_n = \arg \min_{\boldsymbol{\theta} \in \Theta_{\boldsymbol{\theta}}} \widetilde{Q}_n(\widehat{\boldsymbol{\gamma}}_n, \boldsymbol{\theta}). \quad (3.10)$$

The VTE of $\boldsymbol{\vartheta}_0$ is then given by $\widehat{\boldsymbol{\vartheta}}_n = (\widehat{\boldsymbol{\gamma}}_n', \widehat{\boldsymbol{\theta}}_n')'$.

The estimation of $\boldsymbol{\Omega}$ in the original BEKK-X model (3.2) can be obtained by

$$\widehat{\boldsymbol{\Omega}}_n = \widehat{\boldsymbol{\Sigma}}_{\varepsilon n} - \widehat{\mathbf{A}}_n \widehat{\boldsymbol{\Sigma}}_{\varepsilon n} \widehat{\mathbf{A}}_n' - \widehat{\mathbf{B}}_n \widehat{\boldsymbol{\Sigma}}_{\varepsilon n} \widehat{\mathbf{B}}_n' - \widehat{\mathbf{C}}_n \widehat{\boldsymbol{\Sigma}}_{\mathbf{x}n} \widehat{\mathbf{C}}_n', \quad (3.11)$$

where $\widehat{\mathbf{A}}_n$, $\widehat{\mathbf{B}}_n$ and $\widehat{\mathbf{C}}_n$ are the QML estimators of \mathbf{A} , \mathbf{B} and \mathbf{C} respectively. Then the estimator of original parameter vector, denoted by $\boldsymbol{\xi}_0 = (\text{vech}(\boldsymbol{\Omega}_0)', \boldsymbol{\theta}_0')'$, of (3.2) can be given by $\widehat{\boldsymbol{\xi}}_n = (\text{vech}'(\widehat{\boldsymbol{\Omega}}_n), \widehat{\boldsymbol{\theta}}_n')'$.

3.3 VTE inference

In this section, the asymptotic properties of the VTE will be established. The computation of the asymptotic covariance matrix will be also given.

3.3.1 Consistency and asymptotic normality

For the strong consistency of the VTE, the following assumptions are required

B4 : The true parameter $\boldsymbol{\vartheta}_0 \in \Theta$ and Θ is compact.

B5 : $\rho(\mathbf{B}) < 1$ and $\rho(\mathbf{A} \otimes \mathbf{A} + \mathbf{B} \otimes \mathbf{B}) < 1$ for all $\boldsymbol{\vartheta} \in \Theta$.

B6 : If for any $\boldsymbol{\theta} \in \Theta_{\boldsymbol{\theta}}$, $\mathbf{H}_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta}) = \mathbf{H}_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)$ a.s., then $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

B7 : If $\boldsymbol{\pi}$ is a non zero vector of \mathbb{R}^r then $\boldsymbol{\pi}' \mathbf{x}_1$ is non-degenerate.

Remark 3.3. *Assumption B6 is a condition for the identifiability of the model. Note that, Comte and Lieberman (2003), Hafner and Preminger (2009) and Pedersen and Rahbek (2014) give an identification condition equivalent to Assumption B6 for BEKK models without covariates.*

Remark 3.4. *Assumption B7 is an identifiability condition which is obviously necessary to avoid multicollinearity of the exogenous variables.*

Theorem 3.1. *Under Assumptions B1 - B7,*

$$\widehat{\boldsymbol{\vartheta}}_n \rightarrow \boldsymbol{\vartheta}_0, \text{ a.s. as } n \rightarrow \infty. \quad (3.12)$$

To establish the asymptotic normality of VTE the following additional assumptions are needed.

B8 : The true parameter $\boldsymbol{\theta}_0$ belongs to the interior of $\Theta_{\boldsymbol{\theta}}$.

B9 : $E \|\varepsilon_t\|^6 < \infty$ and $E \|\mathbf{x}_t\|^6 < \infty$.

We denote by $\alpha_X(h)$ the strong mixing coefficient of a stationary process $X = (X_t)$

$$\alpha_X(h) = \sup_{A \in \sigma(X_u, u \leq t), B \in \sigma(X_u, u \geq t+h)} |P(A \cap B) - P(A)P(B)|.$$

B10 : $\mathbf{z}_t = (\mathbf{x}'_t, \varepsilon'_t, \boldsymbol{\eta}'_t)'$ is a α -mixing process such that, for some $\nu > 0$ and $\delta > 0$,

$$E \|\varepsilon_t\|^{(4+2\nu)(1+1/\delta)} < \infty, \quad E \|\mathbf{x}_t\|^{(4+2\nu)(1+1/\delta)} < \infty, \quad E \|\boldsymbol{\eta}_t\|^{(4+2\nu)(1+\delta)} < \infty$$

$$\text{and } \sum_{h=0}^{\infty} \{\alpha_{\mathbf{z}}(h)\}^{\nu/(2+\nu)} < \infty.$$

Let $\underline{\mathbf{H}}_{t,s}(\boldsymbol{\vartheta})$ be such that, for $s > 0$,

$$\text{vec}(\underline{\mathbf{H}}_{t,s}(\boldsymbol{\vartheta})) = \sum_{k=0}^s (\mathbf{B}^{\otimes 2})^k (\text{vec}(\boldsymbol{\Omega}) + \mathbf{A}^{\otimes 2} \text{vec}(\varepsilon_{t-k-1} \varepsilon'_{t-k-1}) + \mathbf{C}^{\otimes 2} \text{vec}(\mathbf{x}_{t-k-1} \mathbf{x}'_{t-k-1})),$$

where $A^{\otimes 2}$ denotes the Kronecker product of a matrix A and itself. Let also \mathcal{S} be a subspace such that for all $\boldsymbol{\vartheta} \in \Theta$, $\mathbf{H}_t(\boldsymbol{\vartheta}) \in \mathcal{S}$ and for all $s > 0$, $\underline{\mathbf{H}}_{t,s}(\boldsymbol{\vartheta}) \in \mathcal{S}$.

B11 : There exists $K > 0$ such that

$$\left\| \mathbf{H}_t^{1/2}(\boldsymbol{\vartheta}) - \mathbf{H}_t^{*1/2}(\boldsymbol{\vartheta}) \right\| \leq K \|\mathbf{H}_t(\boldsymbol{\vartheta}) - \mathbf{H}_t^*(\boldsymbol{\vartheta})\| \quad \text{for all } \mathbf{H}_t(\boldsymbol{\vartheta}), \mathbf{H}_t^*(\boldsymbol{\vartheta}) \in \mathcal{S}.$$

Remark 3.5. *The condition that the observations ε_t admit finite moment of order 6 is also found in the existing body of literature on asymptotic normality of the QMLE (see [Hafner and Preminger \(2009\)](#)) or the one of the VTE (see [Pedersen and Rahbek \(2014\)](#)) of the models without covariates. Assumption **B9** is needed to show the existence of moment of second-order derivatives of the log-likelihood function and its uniform convergence on the parameter space.*

Remark 3.6. *For the BEKK model without covariates, [Comte and Lieberman \(2003\)](#) and [Pedersen and Rahbek \(2014\)](#) establish the asymptotic normality of QMLE and VTE respectively by applying the CLT for an ergodic martingale difference sequence under iid assumptions. In the present paper, under Assumption **B1**, $(\boldsymbol{\eta}_t, \mathcal{F}_t)$ is a conditionally homoscedastic martingale difference and (3.2) becomes a semi-strong model. The exogenous variables need not to be independent on the innovations $\boldsymbol{\eta}_t$. The mixing assumption in **B10** is used to apply the central limit theorem (CLT) of [Herrndorf \(1984\)](#) to specify the limiting distribution in Theorem 3.2. When (3.2) is a strong model, i.e. when $\boldsymbol{\eta}_t$ is iid, the moment conditions in **B10** can be weakened as follows*

B10* : $\mathbf{z}_t = (\mathbf{x}'_t, \varepsilon'_t, \boldsymbol{\eta}'_t)'$ is a α -mixing process such that $E \|\mathbf{z}_t\|^{4+2\nu} < \infty$ and $\sum_{h=0}^{\infty} \{\alpha_{\mathbf{z}}(h)\}^{\nu/(2+\nu)} < \infty$, for some $\nu > 0$.

Remark 3.7. In univariate case, Assumption **B11** is always satisfied. Indeed, for simplicity, we consider the univariate GARCH(1, 1) model $\varepsilon_t = \sigma_t^2(\theta)\eta_t$, where $\sigma_t^2(\theta) = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2(\theta)$, with $\omega \geq \underline{\omega} > 0$. For any $\sigma_t^2(\theta)$ and $\sigma_t^{*2}(\theta)$, we have

$$|\sigma_t(\theta) - \sigma_t^*(\theta)| = \left| \sqrt{\sigma_t^{*2}(\theta) + (\sigma_t^2(\theta) - \sigma_t^{*2}(\theta)) \frac{1}{2\sqrt{\sigma_t^{*2}(\theta)}}} - \sqrt{\sigma_t^{*2}(\theta)} \right| \leq K |\sigma_t^2(\theta) - \sigma_t^{*2}(\theta)|,$$

where $\bar{\sigma}_t^2(\theta)$ is between $\sigma_t^2(\theta)$ and $\sigma_t^{*2}(\theta)$ and the inequality follows from $\bar{\sigma}_t^2(\theta) \geq \underline{\omega}$ for all θ .

Let $Q_n(\boldsymbol{\gamma}, \boldsymbol{\theta})$ and $\ell_t(\boldsymbol{\gamma}, \boldsymbol{\theta})$ be obtained by replacing $\widetilde{\mathbf{H}}_t(\boldsymbol{\gamma}, \boldsymbol{\theta})$ by $\mathbf{H}_t(\boldsymbol{\gamma}, \boldsymbol{\theta})$ in $\widetilde{Q}_n(\boldsymbol{\gamma}, \boldsymbol{\theta})$ and $\widetilde{\ell}_t(\boldsymbol{\gamma}, \boldsymbol{\theta})$. We define the following matrices.

$$\mathbf{J} = E \left(\frac{\partial^2 \ell_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right), \quad \mathbf{K}_\varepsilon = E \left(\frac{\partial^2 \ell_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\gamma}'_\varepsilon} \right), \quad \mathbf{K}_x = E \left(\frac{\partial^2 \ell_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\gamma}'_x} \right) \quad (3.13)$$

and

$$\boldsymbol{\Sigma}_{11} = \sum_{h=-\infty}^{\infty} \text{cov}(\text{vech}(\mathbf{x}_t \mathbf{x}'_t), \text{vech}(\mathbf{x}_{t-h} \mathbf{x}'_{t-h})), \quad (3.14)$$

$$\boldsymbol{\Sigma}_{22} = \sum_{h=-\infty}^{\infty} \text{cov}(\boldsymbol{\Upsilon}_{0t} \text{vec}(\boldsymbol{\eta}_t \boldsymbol{\eta}'_t), \boldsymbol{\Upsilon}_{0,t-h} \text{vec}(\boldsymbol{\eta}_{t-h} \boldsymbol{\eta}'_{t-h})), \quad (3.15)$$

$$\boldsymbol{\Sigma}_{12} = \sum_{h=-\infty}^{\infty} \text{cov}(\text{vech}(\mathbf{x}_t \mathbf{x}'_t), \boldsymbol{\Upsilon}_{0,t-h} \text{vec}(\boldsymbol{\eta}_{t-h} \boldsymbol{\eta}'_{t-h})), \quad (3.16)$$

where

$$\boldsymbol{\Upsilon}_{0t} = \begin{pmatrix} \mathbf{H}_{0t}^{1/2} \otimes \mathbf{H}_{0t}^{1/2} \\ -\frac{\partial \text{vec}'(\mathbf{H}_{0t})}{\partial \boldsymbol{\theta}} (\mathbf{H}_{0t}^{-1/2} \otimes \mathbf{H}_{0t}^{-1/2})' \end{pmatrix}. \quad (3.17)$$

Denote by D_m and L_m the duplication matrix and elimination matrix defined such that, for any symmetric $(m \times m)$ matrix A , $\text{vec}(A) = D_m \text{vech}(A)$ and $\text{vech}(A) = L_m \text{vec}(A)$.

The following theorem gives the asymptotic distribution of VTE estimators.

Theorem 3.2. Under Assumptions **B1** - **B11**, as $n \rightarrow \infty$,

$$\sqrt{n} \begin{pmatrix} \widehat{\boldsymbol{\gamma}}_{xn} - \boldsymbol{\gamma}_{x0} \\ \widehat{\boldsymbol{\gamma}}_{\varepsilon n} - \boldsymbol{\gamma}_{\varepsilon 0} \\ \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Gamma} \boldsymbol{\Phi} \boldsymbol{\Sigma} \boldsymbol{\Phi}' \boldsymbol{\Gamma}'), \quad (3.18)$$

where

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}'_{12} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \quad \boldsymbol{\Gamma} = \begin{pmatrix} I_{r(r+1)/2} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I_{m(m+1)/2} & \mathbf{0} \\ -\mathbf{J}^{-1} \mathbf{K}_x & -\mathbf{J}^{-1} \mathbf{K}_\varepsilon & -\mathbf{J}^{-1} \end{pmatrix} \quad (3.19)$$

and

$$\Phi = \begin{pmatrix} I_{r(r+1)/2} & \mathbf{0} & \mathbf{0} \\ L_m(I_{m^2} - \mathbf{A}_0^{\otimes 2} - \mathbf{B}_0^{\otimes 2})^{-1} \mathbf{C}_0^{\otimes 2} D_r & L_m(I_{m^2} - \mathbf{A}_0^{\otimes 2} - \mathbf{B}_0^{\otimes 2})^{-1} (I_{m^2} - \mathbf{B}_0^{\otimes 2}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -I_{2m^2+mr} \end{pmatrix}. \quad (3.20)$$

The asymptotic normality of the estimation of the original parameters is given in the following corollary.

Corollary 3.1. *Under the assumptions of Theorem 3.2, the VTE of ξ_0 satisfies*

$$\sqrt{n} (\hat{\xi}_n - \xi_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Delta \Gamma \Phi \Sigma \Phi' \Gamma' \Delta'), \quad (3.21)$$

where

$$\Delta = \begin{pmatrix} \Delta_1 & \Delta_2 \\ \mathbf{0}_{(2m^2+mr) \times (m^2+r^2)} & I_{(2m^2+mr) \times (2m^2+mr)} \end{pmatrix} \quad (3.22)$$

with

$$\begin{aligned} \Delta_1 &= \begin{pmatrix} -L_m(\mathbf{C}_0 \otimes \mathbf{C}_0) D_d & L_m(I_{m^2} - \mathbf{A}_0 \otimes \mathbf{A}_0 - \mathbf{B}_0 \otimes \mathbf{B}_0) D_m \end{pmatrix}, \\ \Delta_2 &= -L_m(I_{m^2} + M_{mm}) \begin{pmatrix} (\mathbf{A}_0 \Sigma_{\varepsilon 0} \otimes I_m) & (\mathbf{B}_0 \Sigma_{\varepsilon 0} \otimes I_m) & (\mathbf{C}_0 \Sigma_{\varepsilon 0} \otimes I_m) \end{pmatrix} \end{aligned}$$

and M_{pq} denotes the commutation matrix such that, for any $(p \times q)$ matrix A , $M_{pq} \text{vec}(A) = \text{vec}(A')$.

3.3.2 Estimating the asymptotic covariance matrix

In the econometric literature the nonparametric kernel estimator, also called heterosceastic autocorrelation consistent (HAC) estimator (see Newey and West (1987), Andrews (1991) and Phillips and Jin (2003)) is widely used to estimate covariance matrix of the form Σ_{11} . The consistent estimators $\hat{\Sigma}_{11n}$, $\hat{\Sigma}_{22n}$ and $\hat{\Sigma}_{12n}$ of Σ_{11} , Σ_{22} and Σ_{12} , respectively, can thus be given by

$$\begin{aligned} \hat{\Sigma}_{11n} &= \frac{1}{n} \sum_{t,s=1}^n w_{|t-s|} \text{vech}(\mathbf{x}_t \mathbf{x}_t') \text{vech}'(\mathbf{x}_s \mathbf{x}_s'), \\ \hat{\Sigma}_{22n} &= \frac{1}{n} \sum_{t,s=1}^n w_{|t-s|} \hat{\mathbf{Y}}_t \text{vec}(\boldsymbol{\eta}_t \boldsymbol{\eta}_t') \text{vec}'(\boldsymbol{\eta}_s \boldsymbol{\eta}_s') \hat{\mathbf{Y}}_t', \\ \hat{\Sigma}_{12n} &= \frac{1}{n} \sum_{t,s=1}^n w_{|t-s|} \text{vech}(\mathbf{x}_t \mathbf{x}_t') \text{vec}'(\boldsymbol{\eta}_s \boldsymbol{\eta}_s') \hat{\mathbf{Y}}_s', \end{aligned}$$

where w_0, \dots, w_{n-1} is a sequence of weights (see [Newey and West \(1987\)](#), [Andrews \(1991\)](#) and [Phillips and Jin \(2003\)](#) for the problem of the choice of weights) and

$$\widehat{\Upsilon}_t = \begin{pmatrix} \widetilde{\mathbf{H}}_t^{1/2}(\widehat{\vartheta}_n) \otimes \widetilde{\mathbf{H}}_t^{1/2}(\widehat{\vartheta}_n) \\ -\frac{\partial \text{vec}'(\widetilde{\mathbf{H}}_t(\widehat{\vartheta}_n))}{\partial \boldsymbol{\theta}} \left(\widetilde{\mathbf{H}}_t^{-1/2}(\widehat{\vartheta}_n) \otimes \widetilde{\mathbf{H}}_t^{-1/2}(\widehat{\vartheta}_n) \right)' \end{pmatrix}.$$

Let

$$\widehat{\mathbf{J}}_n = \frac{1}{n} \sum_{t=1}^n \frac{\partial^2 \widetilde{\ell}_t(\widehat{\vartheta}_n)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}, \quad \widehat{\mathbf{K}}_{xn} = \frac{1}{n} \sum_{t=1}^n \frac{\partial^2 \widetilde{\ell}_t(\widehat{\vartheta}_n)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\gamma}'_x}, \quad \widehat{\mathbf{K}}_{\varepsilon n} = \frac{1}{n} \sum_{t=1}^n \frac{\partial^2 \widetilde{\ell}_t(\widehat{\vartheta}_n)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\gamma}'_\varepsilon}. \quad (3.23)$$

Then under the assumptions of [Theorem 3.2](#), strongly consistent estimators of $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ are given by

$$\widehat{\boldsymbol{\Gamma}}_n = \begin{pmatrix} I_{r(r+1)/2} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I_{m(m+1)/2} & \mathbf{0} \\ -\widehat{\mathbf{J}}_n^{-1} \widehat{\mathbf{K}}_{xn} & -\widehat{\mathbf{J}}_n^{-1} \widehat{\mathbf{K}}_{\varepsilon n} & -\widehat{\mathbf{J}}_n^{-1} \end{pmatrix} \quad \text{and} \quad \widehat{\boldsymbol{\Sigma}}_n = \begin{pmatrix} \widehat{\boldsymbol{\Sigma}}_{11n} & \widehat{\boldsymbol{\Sigma}}_{12n} \\ \widehat{\boldsymbol{\Sigma}}'_{12n} & \widehat{\boldsymbol{\Sigma}}_{22n} \end{pmatrix} \quad (3.24)$$

Note that, the computation of the matrices $\widehat{\boldsymbol{\Gamma}}_n$ and $\widehat{\boldsymbol{\Sigma}}_n$ requires the evaluation of complicated first and second-order derivatives. More precisely, for $\widehat{\boldsymbol{\Sigma}}_{22n}$ and $\widehat{\boldsymbol{\Sigma}}_{12n}$ one needs to compute $\partial \text{vec}'(\widetilde{\mathbf{H}}_t(\widehat{\vartheta}_n))/\partial \boldsymbol{\theta}$. [Francq and Zakoïan \(2014\)](#) show that these n vectors of derivatives cannot be numerically calculated within a reasonable amount of time. They provide thus recursive formulas for a rapid computation of these derivatives.

3.4 Numerical illustration

In this section, we illustrate our asymptotic results of [Section 3.3](#) on Monte Carlo simulations and on US stock series.

3.4.1 A Monte Carlo experiment

This subsection presents the results from a series of Monte Carlo experiments that allow us to evaluate the performance of the BEKK framework when the exogenous variables are introduced.

In order to reduce the computation burden of the simulations, we consider a simplified version of the bivariate BEKK(1,1)-X model [\(3.2\)](#) with \mathbf{B} a diagonal matrix. The vector of the exogenous variables is $\mathbf{x}_t = (\mathbf{x}_{1t}, \mathbf{x}_{2t})'$ where \mathbf{x}_{1t} and \mathbf{x}_{2t} are two lagged values of an APARCH(1,1)

$$\begin{cases} \mathbf{z}_t = \sigma_t e_t, \\ \sigma_t = 0.046 + 0.027 \mathbf{z}_{t-1}^+ + 0.092 \mathbf{z}_{t-1}^- + 0.843 \sigma_{t-1}, \end{cases} \quad (3.25)$$

where $\sqrt{2}e_t$ is i.i.d and follows a Student distribution with 4 degrees of freedom. Two components of $\boldsymbol{\eta}_t$ are independent and normally distributed $\mathcal{N}(0, 1)$. The true the parameter matrix are taken as follows

$$\boldsymbol{\Omega}_0 = \begin{pmatrix} 0.3 & 0.2 \\ 0.2 & 0.4 \end{pmatrix}, \boldsymbol{A}_0 = \begin{pmatrix} 0.15 & 0.1 \\ 0.1 & 0.2 \end{pmatrix}, \boldsymbol{B}_0 = \begin{pmatrix} 0.8 & 0.0 \\ 0.0 & 0.9 \end{pmatrix}, \boldsymbol{C}_0 = \begin{pmatrix} 0.15 & 0.05 \\ 0.1 & 0.2 \end{pmatrix}. \quad (3.26)$$

We investigate samples with $n = 1000$ and $n = 5000$ observations. All simulations are repeated 500 times. For each data series, we simulated $(n + 500)$ observations of $\boldsymbol{\varepsilon}_t$ and then the first 500 observations are discarded in each simulation to minimize the effect of the initial values. In order to assess the statistical properties of the estimates we have computed the bias, the root mean squared error (RMSE) and the quartiles of the estimated parameters $\hat{\boldsymbol{\xi}}_n$

$$bias(\hat{\boldsymbol{\xi}}_n) = \frac{1}{500} \sum_{i=1}^{500} (\hat{\boldsymbol{\xi}}_n^{(i)} - \boldsymbol{\xi}_0)$$

$$RMSE(\hat{\boldsymbol{\xi}}_n) = \left(\frac{1}{500} \sum_{i=1}^{500} (\hat{\boldsymbol{\xi}}_n^{(i)} - \bar{\boldsymbol{\xi}})^2 \right)^{1/2}$$

where $\hat{\boldsymbol{\xi}}_n^{(i)}$ is the estimator at the i^{th} replication and $\bar{\boldsymbol{\xi}}$ is their empirical mean. The results of the simulation study are presented in Table 4.1. They are in accordance with the consistency of the VTE, in particular the medians of the estimated parameters are close to the true values. As expected, the accuracy of the estimation increases as the sample size increases from $n = 1000$ to $n = 5000$.

3.4.2 An application to stocks US

Is the intraday realized volatility useful for predicting the the volatility of the financial returns? In the univariate case, Francq and Thieu (2015) demonstrate that, for the capitalization stocks of American stock exchanges, yesterday's realized volatility often helps in predicting today's squared returns. Another question that we would like to investigate is whether the realized volatilities of some series returns affect their co-volatilities. The aim of this subsection is to apply the BEKK-X model in order to answer this question. For illustration purposes we restrict our attention to only 3 indices, the MSFT (Microsoft Corporation), the AAPL (Apple) and the DELL, and initially we only include one exogenous variable that is the yesterday's realized volatility of the MSFT.

The data come from Section 4.2 of Laurent et al. (2014), covering the period from January 4, 1999 to December 31, 2008 (2,489 trading days). In the end of each trading day t , the log-return in percentage ε_{kt} and the realized volatility rv_{kt} (computed as the sum of intraday squared 5-minute log-returns) are available.

TABLE 3.1 – Sampling distribution of the VTE of ϑ_0 over 500 replications for the BEKK(1,1)-X model

parameter	true val.	bias	RMSE	min	Q_1	Q_2	Q_3	max
$n = 1,000$								
$vec(\mathbf{\Omega})$	0.30	-0.0024	0.0942	0.0376	0.2411	0.2952	0.3544	0.6045
	0.20	0.0047	0.1053	-0.0810	0.1409	0.1998	0.2692	0.5817
	0.40	0.0101	0.2037	0.0005	0.2842	0.4037	0.5079	2.2127
\mathbf{A}	0.15	-0.0101	0.0790	0.0000	0.0841	0.1420	0.1960	0.3428
	0.10	-0.0124	0.1404	-0.3823	0.0185	0.1094	0.2027	0.5215
	0.10	-0.0004	0.0404	-0.0447	0.0734	0.1000	0.1261	0.2087
	0.20	-0.0138	0.0770	0.0000	0.1400	0.1908	0.2394	0.4012
$diag(\mathbf{B})$	0.80	-0.0011	0.0353	0.6854	0.7801	0.7984	0.8224	0.9135
	0.90	-0.0030	0.0227	0.6665	0.8857	0.8998	0.9112	0.9519
\mathbf{C}	0.15	-0.0010	0.0154	0.1033	0.1390	0.1492	0.1600	0.2123
	0.10	-0.0001	0.0265	0.0079	0.0813	0.1005	0.1174	0.1814
	0.05	0.0007	0.0164	0.0014	0.0405	0.0510	0.0616	0.1082
	0.20	-0.0006	0.0251	0.1275	0.1834	0.1990	0.2147	0.2854
$n = 5,000$								
$vec(\mathbf{\Omega})$	0.30	0.0015	0.0385	0.0777	0.2752	0.3008	0.3270	0.4342
	0.20	0.0028	0.0393	-0.0660	0.1770	0.2012	0.2284	0.3717
	0.40	0.0046	0.0687	0.0560	0.3600	0.4010	0.4469	0.6909
\mathbf{A}	0.15	-0.0027	0.0324	0.0392	0.1295	0.1481	0.1683	0.2403
	0.10	0.0013	0.0543	-0.0714	0.0651	0.0999	0.1352	0.2885
	0.10	0.0011	0.0168	0.0377	0.0905	0.1011	0.1118	0.1557
	0.20	-0.0019	0.0307	0.0871	0.1799	0.1992	0.2172	0.4138
$diag(\mathbf{B})$	0.80	-0.0013	0.0146	0.7575	0.7898	0.8000	0.8084	0.8487
	0.90	-0.0008	0.0075	0.8785	0.8943	0.8996	0.9044	0.9309
\mathbf{C}	0.15	0.0001	0.0071	0.1305	0.1456	0.1498	0.1547	0.1752
	0.10	0.0003	0.0118	0.0684	0.0927	0.0998	0.1086	0.1404
	0.05	0.0002	0.0072	0.0254	0.0457	0.0504	0.0547	0.0721
	0.20	0.0002	0.0112	0.1561	0.1928	0.2004	0.2075	0.2563

RMSE is the Root Mean Square Error, Q_i , $i = 1, 3$, denote the quartiles.

With obvious notations (in particular the estimated standard deviations, obtained from the empirical estimator (3.24) in Section 3.3, are into brackets), the estimated parameters can be written

as

$$\hat{\Omega}_n^{VTE} = \begin{pmatrix} 0.0218 & & \\ (0.0418) & & \\ 0.0118 & 0.0257 & \\ (0.0186) & (0.0218) & \\ 0.0041 & -0.0070 & 0.0052 \\ (0.0065) & (0.0047) & (0.0070) \end{pmatrix}, \quad \hat{A}_n^{VTE} = \begin{pmatrix} 0.1888 & -0.0032 & -0.0015 \\ (0.0320) & (0.0113) & (0.0184) \\ 0.0062 & 0.1378 & -0.0157 \\ (0.0673) & (0.0457) & (0.0143) \\ -0.0049 & 0.0455 & 0.2014 \\ (0.0330) & (0.0138) & (0.0204) \end{pmatrix}$$

$$\hat{B}_n^{VTE} = \begin{pmatrix} 0.9721 & 0 & 0 \\ (0.0142) & & \\ 0 & 0.9888 & 0 \\ & (0.0006) & \\ 0 & 0 & 0.9731 \\ & & (0.0056) \end{pmatrix}, \quad \hat{C}_n^{VTE} = \begin{pmatrix} 0.0390 \\ (0.0523) \\ 0.0181 \\ (0.0126) \\ 0.0167 \\ (0.0264) \end{pmatrix}$$

3.5 Conclusion

In this paper we establish the asymptotic behavior of the variance-targeting estimator of the parameters for the multivariate BEKK augmented by exogenous variables. We do not restrict the number of covariates that we want to investigate. They are introduced in the conditional covariance equation such that the positivity of the conditional covariance matrix is still assured. The model BEKK-X is reparameterized such that the unconditional covariance matrix of the observed process and the second moment matrix of the explanatory variables appear explicitly in the model equation. We demonstrate the strong consistency of the VTE under the existence of the second-order moments of the observations and the covariates. We also establish the asymptotic normality under the conditions that the process and the exogenous variables have finite sixth-order moments and that the exogenous variables follow an α -mixing process. We also provide the asymptotic distribution of the original parameters. One Monte-Carlo simulation and one empirical application illustrate the usefulness of the results.

Acknowledgements

The author is very grateful to Professor Christian Francq for valuable comments and encouragement.

3.6 Proofs

Proof of Theorem 3.1. The strong convergence of $\widehat{\gamma}_n$ to γ_0 is a direct consequence of the ergodic theorem and Assumption **B3**. To show the strong consistency of $\widehat{\boldsymbol{\theta}}_n$, it suffices to establish the following results :

- i) $\lim_{n \rightarrow \infty} \sup_{\boldsymbol{\theta} \in \Theta_{\boldsymbol{\theta}}} \left| Q_n(\gamma_0, \boldsymbol{\theta}) - \widetilde{Q}_n(\widehat{\gamma}_n, \boldsymbol{\theta}) \right| = 0$ a.s.
- ii) $E \left(\sup_{\boldsymbol{\theta} \in \Theta} |\ell_t(\gamma, \boldsymbol{\theta})| \right) < \infty$ and if $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, $E(\ell_t(\gamma_0, \boldsymbol{\theta})) > E(\ell_t(\gamma_0, \boldsymbol{\theta}_0))$.
- iii) For any $\bar{\boldsymbol{\theta}} \neq \boldsymbol{\theta}_0$, there exists a neighborhood $V(\bar{\boldsymbol{\theta}})$ such that

$$\liminf_{n \rightarrow \infty} \inf_{\boldsymbol{\theta} \in V(\bar{\boldsymbol{\theta}})} \widetilde{Q}_n(\widehat{\gamma}_n, \boldsymbol{\theta}) > E\ell_1(\gamma_0, \boldsymbol{\theta}_0) \text{ a.s.}$$

For notation simplicity, we denote $f_t := f_t(\gamma, \boldsymbol{\theta})$ for any function f_t depending on parameters $(\gamma, \boldsymbol{\theta})$ and denote f_{0t} when $(\gamma, \boldsymbol{\theta}) = (\gamma_0, \boldsymbol{\theta}_0)$. In the sequel, K and ϱ denote generic constants such that $K > 0$ and $\varrho \in (0, 1)$ whose exact values are unimportant.

Assumption **B5** implies that $\rho(\mathbf{B}^{\otimes 2}) < 1$ for all $\boldsymbol{\vartheta} \in \Theta$. By the compactness of Θ , we even have

$$\sup_{\boldsymbol{\vartheta} \in \Theta} \rho(\mathbf{B}^{\otimes 2}) < 1. \quad (3.27)$$

Using the relation $\text{vec}(ABC) = (C' \otimes A)\text{vec}(B)$, the vec representation of (3.4) is given by

$$\begin{aligned} \text{vec}(\mathbf{H}_t) = & (I_{m^2} - \mathbf{A}^{\otimes 2} - \mathbf{B}^{\otimes 2})\text{vec}(\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}) - \mathbf{C}^{\otimes 2}\text{vec}(\boldsymbol{\Sigma}_{\mathbf{x}}) + \mathbf{A}^{\otimes 2}\text{vec}(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}'_{t-1}) \\ & + \mathbf{B}^{\otimes 2}\text{vec}(\mathbf{H}_{t-1}) + \mathbf{C}^{\otimes 2}\text{vec}(\mathbf{x}_{t-1}\mathbf{x}'_{t-1}). \end{aligned} \quad (3.28)$$

Iteratively using equation (3.28), we deduce that almost surely

$$\sup_{\boldsymbol{\vartheta} \in \Theta} \left\| \widetilde{\mathbf{H}}_t - \mathbf{H}_t \right\| \leq K \varrho^t, \quad \forall t. \quad (3.29)$$

Observe that K is a random variable that depends on the past values $\{\boldsymbol{\varepsilon}_s, \mathbf{x}_s; s \leq 0\}$ but does not depend on t . It can thus be considered as a constant, such as ρ . Applying the inequality $\det(A + B) \geq \det(A)$, for $A > 0$ and $B \geq 0$, where \geq denotes that the matrix is positive semi-definite, we have $\det(\mathbf{H}_t) > 0$, for all t and for all $\boldsymbol{\vartheta} \in \Theta$. It implies that \mathbf{H}_t is invertible. Moreover, by using the equality $0 < \text{tr}((A + B)^{-1}) \leq \text{tr}(B^{-1})$, we have

$$\|\mathbf{H}_t^{-1}\| \leq \|\mathbf{H}_t^{-1/2}\|^2 = \text{tr}(\mathbf{H}_t^{-1}) \leq \text{tr}(\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} - \mathbf{A}\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}\mathbf{A}' - \mathbf{B}\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}\mathbf{B}' - \mathbf{C}\boldsymbol{\Sigma}_{\mathbf{x}}\mathbf{C}')^{-1} \leq K.$$

Hence, it yields

$$\sup_{\boldsymbol{\vartheta} \in \Theta} \|\mathbf{H}_t^{-1}\| < K. \quad (3.30)$$

By the same arguments, $\widetilde{\mathbf{H}}_t$ is also invertible and, for some constant K

$$\sup_{\vartheta \in \Theta} \|\widetilde{\mathbf{H}}_t^{-1}\| < K. \quad (3.31)$$

We thus have almost surely,

$$\sup_{\vartheta \in \Theta} \|\mathbf{H}_t^{-1} - \widetilde{\mathbf{H}}_t^{-1}\| \leq \sup_{\vartheta \in \Theta} \|\widetilde{\mathbf{H}}_t^{-1}\| \|\widetilde{\mathbf{H}}_t - \mathbf{H}_t\| \|\mathbf{H}_t^{-1}\| \leq K \varrho^t. \quad (3.32)$$

Now

$$\begin{aligned} & \sup_{\vartheta \in \Theta} \left| Q_n(\gamma, \boldsymbol{\theta}) - \widetilde{Q}_n(\gamma, \boldsymbol{\theta}) \right| \\ & \leq \frac{1}{n} \sum_{t=1}^n \sup_{\vartheta \in \Theta} \left| \text{tr} \left(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t \left(\mathbf{H}_t^{-1} - \widetilde{\mathbf{H}}_t^{-1} \right) \right) \right| + \frac{1}{n} \sum_{t=1}^n \sup_{\vartheta \in \Theta} \left| \log \frac{\det(\mathbf{H}_t)}{\det(\widetilde{\mathbf{H}}_t)} \right| \\ & \leq \frac{1}{n} \sum_{t=1}^n \sup_{\vartheta \in \Theta} \left(\|\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t\| \|\mathbf{H}_t^{-1} - \widetilde{\mathbf{H}}_t^{-1}\| \right) + \frac{1}{n} \sum_{t=1}^n \sup_{\vartheta \in \Theta} \left| \log \frac{\det(\mathbf{H}_t)}{\det(\widetilde{\mathbf{H}}_t)} \right|. \end{aligned} \quad (3.33)$$

The first sum converges to zero almost surely by using Assumption **B3**, (3.32) and the same arguments to show Theorem 11.7(a) in [Francq and Zakoian \(2010\)](#). The convergence to zero of the second sum is also showed as on page 297 – 298 of the previous reference. Therefore we obtain

$$\lim_{n \rightarrow \infty} \sup_{\vartheta \in \Theta} \left| Q_n(\gamma, \boldsymbol{\theta}) - \widetilde{Q}_n(\gamma, \boldsymbol{\theta}) \right| = 0 \text{ a.s.} \quad (3.34)$$

Now we have

$$\begin{aligned} & \sup_{\boldsymbol{\theta} \in \Theta_{\boldsymbol{\theta}}} \left| Q_n(\gamma_0, \boldsymbol{\theta}) - \widetilde{Q}_n(\widehat{\gamma}_n, \boldsymbol{\theta}) \right| \\ & \leq \sup_{\boldsymbol{\theta} \in \Theta_{\boldsymbol{\theta}}} \left| Q_n(\gamma_0, \boldsymbol{\theta}) - Q_n(\widehat{\gamma}_n, \boldsymbol{\theta}) \right| + \sup_{\boldsymbol{\theta} \in \Theta_{\boldsymbol{\theta}}} \left| Q_n(\widehat{\gamma}_n, \boldsymbol{\theta}) - \widetilde{Q}_n(\widehat{\gamma}_n, \boldsymbol{\theta}) \right| \\ & \leq \sup_{\boldsymbol{\theta} \in \Theta_{\boldsymbol{\theta}}} \left| Q_n(\gamma_0, \boldsymbol{\theta}) - Q_n(\widehat{\gamma}_n, \boldsymbol{\theta}) \right| + \sup_{\vartheta \in \Theta} \left| Q_n(\gamma, \boldsymbol{\theta}) - \widetilde{Q}_n(\gamma, \boldsymbol{\theta}) \right|. \end{aligned} \quad (3.35)$$

To show point *i*), it thus remains to show that the first term in (3.35) also almost surely converges to zero. For m large enough, let $V_m(\gamma_0)$ be the open ball of center γ_0 and radius $1/m$. Because of the consistency of $\widehat{\gamma}_n$, for n large enough, we have

$$\sup_{\boldsymbol{\theta} \in \Theta_{\boldsymbol{\theta}}} \left| Q_n(\widehat{\gamma}_n, \boldsymbol{\theta}) - Q_n(\gamma_0, \boldsymbol{\theta}) \right| \leq \sup_{\boldsymbol{\theta} \in \Theta_{\boldsymbol{\theta}}} \frac{1}{n} \sum_{t=1}^n \sup_{\gamma \in V_m(\gamma_0)} \left| \ell_t(\gamma, \boldsymbol{\theta}) - \ell_t(\gamma_0, \boldsymbol{\theta}) \right|.$$

Then

$$\begin{aligned} \lim_{n \rightarrow \infty} \sup_{\boldsymbol{\theta} \in \Theta_{\boldsymbol{\theta}}} |Q_n(\hat{\boldsymbol{\gamma}}_n, \boldsymbol{\theta}) - Q_n(\boldsymbol{\gamma}_0, \boldsymbol{\theta})| &\leq \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \sup_{\boldsymbol{\theta} \in \Theta_{\boldsymbol{\theta}}} \sup_{\boldsymbol{\gamma} \in V_m(\boldsymbol{\gamma}_0)} |\ell_t(\boldsymbol{\gamma}, \boldsymbol{\theta}) - \ell_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta})| \\ &= E \sup_{\boldsymbol{\theta} \in \Theta_{\boldsymbol{\theta}}} \sup_{\boldsymbol{\gamma} \in V_m(\boldsymbol{\gamma}_0)} |\ell_t(\boldsymbol{\gamma}, \boldsymbol{\theta}) - \ell_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta})| \end{aligned}$$

where the last equality follows by the ergodicity and the existence of the expectation of the term under the summation symbol. By Lebesgue's dominated convergence theorem, the latter expectation tends to zero when the neighborhood $V_m(\boldsymbol{\gamma}_0)$ shrinks to the singleton $\boldsymbol{\gamma}_0$. The point *i*) is proved.

We turn now to prove ii). Iteratively using equation (3.28) and then using (3.27) and Assumption **B3**, we easily get

$$E \left(\sup_{\boldsymbol{\vartheta} \in \Theta} \|\text{vec}(\mathbf{H}_t)\| \right) \leq \sum_{k=0}^{\infty} K \varrho^k (1 + E\|\boldsymbol{\varepsilon}_{t-k-1}\|^2 + E\|\mathbf{x}_{t-k-1}\|^2) < \infty. \quad (3.36)$$

Then using the inequalities $|\text{tr}(AB)| \leq \|A\| \|B\|$ and $\log|A| \leq \text{Tr}(A)$, for matrix $A > 0$, we have

$$\begin{aligned} E \left(\sup_{\boldsymbol{\vartheta} \in \Theta} |\ell_t(\boldsymbol{\gamma}, \boldsymbol{\theta})| \right) &\leq E \left(\sup_{\boldsymbol{\vartheta} \in \Theta} \|\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t\| \|\mathbf{H}_t^{-1}\| \right) + E \left(\sup_{\boldsymbol{\vartheta} \in \Theta} |\text{tr}(\mathbf{H}_t)| \right) \\ &\leq E \left(\sup_{\boldsymbol{\vartheta} \in \Theta} \|\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t\| \|\mathbf{H}_t^{-1}\| \right) + \sqrt{m} E \left(\sup_{\boldsymbol{\vartheta} \in \Theta} \|\mathbf{H}_t\| \right) \\ &< \infty. \end{aligned}$$

Let λ_{kt} , $k = 1, \dots, m$ be the eigenvalues of matrix $\mathbf{H}_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0) \mathbf{H}_t^{-1}(\boldsymbol{\gamma}_0, \boldsymbol{\theta})$. Using the same arguments used to show the point (c) in the proof of Theorem 11.7 in Francq and Zakoian (2010), we can obtain

$$E(\ell_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta})) - E(\ell_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)) = \sum_{k=0}^m E(\lambda_{kt} - 1 - \log(\lambda_{kt})) \geq 0.$$

The inequality is strict unless if, for all k , $\lambda_{kt} = 1$ a.s., that is, if $\mathbf{H}_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0) = \mathbf{H}_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta})$ a.s. which implies that $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0$. The second inequality of the point ii) is thus obtained.

It now remains to show the point iii). For any $\bar{\boldsymbol{\theta}} \neq \boldsymbol{\theta}_0$, let $V_k(\bar{\boldsymbol{\theta}})$ be the open ball with center $\bar{\boldsymbol{\theta}}$ and radius $1/k$. By properties of the supremum and infimum of a function and using successively

i), the ergodic theorem, the monotone convergence theorem and *ii*), we obtain almost surely

$$\begin{aligned}
& \liminf_{n \rightarrow \infty} \inf_{\boldsymbol{\theta} \in V_k(\hat{\boldsymbol{\theta}}) \cap \Theta_{\boldsymbol{\theta}}} \tilde{Q}_n(\hat{\boldsymbol{\gamma}}_n, \boldsymbol{\theta}) \\
& \geq \liminf_{n \rightarrow \infty} \inf_{\boldsymbol{\theta} \in V_k(\hat{\boldsymbol{\theta}}) \cap \Theta_{\boldsymbol{\theta}}} Q_n(\boldsymbol{\gamma}_0, \boldsymbol{\theta}) - \limsup_{n \rightarrow \infty} \sup_{\boldsymbol{\theta} \in \Theta_{\boldsymbol{\theta}}} \left| \tilde{Q}_n(\hat{\boldsymbol{\gamma}}_n, \boldsymbol{\theta}) - Q_n(\boldsymbol{\gamma}_0, \boldsymbol{\theta}) \right| \\
& \geq \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \inf_{\boldsymbol{\theta} \in V_k(\hat{\boldsymbol{\theta}}) \cap \Theta_{\boldsymbol{\theta}}} \ell_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta}) \\
& = E \inf_{\boldsymbol{\theta} \in V_k(\hat{\boldsymbol{\theta}}) \cap \Theta_{\boldsymbol{\theta}}} \ell_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta}) \\
& > E \ell_1(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)
\end{aligned}$$

for k large enough. □

Proof of the asymptotic normality in Theorem 3.2

For the proof of the asymptotic distribution we need a few elementary results on the differentiation of expressions involving matrices.

If $X \in \mathcal{M}_n(\mathbb{R})$ is a symmetric matrix then

$$\frac{\partial \text{vec}(AXB)}{\partial \text{vech}(X)'} = (B' \otimes A) D_m. \quad (3.37)$$

If $X \in \mathcal{M}_{m \times n}(\mathbb{R})$ and $A \in \mathcal{M}_n(\mathbb{R})$ is a symmetric matrix then

$$\frac{\partial \text{vec}(XAX')}{\partial \text{vec}(X)'} = (I_{m^2} + M_{mm})(XA \otimes I_m). \quad (3.38)$$

Let x be a vector

$$\frac{\partial \text{vec}(Y^k)}{\partial x'} = \sum_{i=0}^{k-1} \left((Y')^{k-i-1} \otimes Y^i \right) \frac{\partial \text{vec}(Y)}{\partial x'}. \quad (3.39)$$

The proof is based on several technical lemmas.

Lemma 3.1. *Under Assumptions B1-B10,*

$$E \left(\sup_{\boldsymbol{\vartheta} \in \Theta} \left\| \frac{\partial^2 \ell_t(\boldsymbol{\gamma}, \boldsymbol{\theta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \right\| \right) < \infty \quad (3.40)$$

for all $i, j = 1, \dots, d$.

Proof of Lemma 3.1. We have

$$\frac{\partial \ell_t}{\partial \boldsymbol{\vartheta}_i} = \text{Tr} \left((\mathbf{H}_t^{-1} - \mathbf{H}_t^{-1} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' \mathbf{H}_t^{-1}) \frac{\partial \mathbf{H}_t}{\partial \boldsymbol{\vartheta}_i} \right), \quad (3.41)$$

$$\begin{aligned} \frac{\partial^2 \ell_t}{\partial \boldsymbol{\vartheta}_i \partial \boldsymbol{\vartheta}_j} &= -\text{Tr} \left(\mathbf{H}_t^{-1} \frac{\partial \mathbf{H}_t}{\partial \boldsymbol{\vartheta}_j} \mathbf{H}_t^{-1} \frac{\partial \mathbf{H}_t}{\partial \boldsymbol{\vartheta}_i} \right) + \text{Tr} \left(\mathbf{H}_t^{-1} \frac{\partial^2 \mathbf{H}_t}{\partial \boldsymbol{\vartheta}_i \partial \boldsymbol{\vartheta}_j} \right) \\ &\quad + 2\text{Tr} \left(\mathbf{H}_t^{-1} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' \mathbf{H}_t^{-1} \frac{\partial \mathbf{H}_t}{\partial \boldsymbol{\vartheta}_j} \mathbf{H}_t^{-1} \frac{\partial \mathbf{H}_t}{\partial \boldsymbol{\vartheta}_i} \right) - \text{Tr} \left(\mathbf{H}_t^{-1} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' \mathbf{H}_t^{-1} \frac{\partial^2 \mathbf{H}_t}{\partial \boldsymbol{\vartheta}_i \partial \boldsymbol{\vartheta}_j} \right). \end{aligned} \quad (3.42)$$

The triangle inequality and $|\text{Tr}(AB)| \leq \|A\| \|B\|$ give

$$\begin{aligned} \sup_{\boldsymbol{\vartheta} \in \Theta} \left| \frac{\partial^2 \ell_t}{\partial \boldsymbol{\vartheta}_i \partial \boldsymbol{\vartheta}_j} \right| &\leq \sup_{\boldsymbol{\vartheta} \in \Theta} \|\mathbf{H}_t^{-1}\|^2 \left\| \frac{\partial \mathbf{H}_t}{\partial \boldsymbol{\vartheta}_i} \right\| \left\| \frac{\partial \mathbf{H}_t}{\partial \boldsymbol{\vartheta}_j} \right\| + \sup_{\boldsymbol{\vartheta} \in \Theta} \|\mathbf{H}_t^{-1}\| \left\| \frac{\partial^2 \mathbf{H}_t}{\partial \boldsymbol{\vartheta}_i \partial \boldsymbol{\vartheta}_j} \right\| \\ &\quad + \sup_{\boldsymbol{\vartheta} \in \Theta} 2 \|\mathbf{H}_t^{-1}\|^3 \|\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'\| \left\| \frac{\partial \mathbf{H}_t}{\partial \boldsymbol{\vartheta}_i} \right\| \left\| \frac{\partial \mathbf{H}_t}{\partial \boldsymbol{\vartheta}_j} \right\| + \sup_{\boldsymbol{\vartheta} \in \Theta} \|\mathbf{H}_t^{-1}\|^2 \|\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'\| \left\| \frac{\partial^2 \mathbf{H}_t}{\partial \boldsymbol{\vartheta}_i \partial \boldsymbol{\vartheta}_j} \right\|. \end{aligned}$$

Note that $\sup_{\boldsymbol{\vartheta} \in \Theta} \|\mathbf{H}_t^{-1}\|^3 < K$ follows from (3.30). Then by Hölder's inequality and assumption **B9**, the existence of the second-order moment of the score will be proved by showing

$$E \left(\sup_{\boldsymbol{\vartheta} \in \Theta} \left\| \frac{\partial \mathbf{H}_t}{\partial \boldsymbol{\vartheta}_i} \right\|^3 \right) = E \left(\sup_{\boldsymbol{\vartheta} \in \Theta} \left\| \frac{\partial \text{vec}(\mathbf{H}_t)}{\partial \boldsymbol{\vartheta}_i} \right\|^3 \right) < \infty \quad (3.43)$$

and

$$E \left(\sup_{\boldsymbol{\vartheta} \in \Theta} \left\| \frac{\partial^2 \mathbf{H}_t}{\partial \boldsymbol{\vartheta}_i \partial \boldsymbol{\vartheta}_j} \right\|^2 \right) = E \left(\sup_{\boldsymbol{\vartheta} \in \Theta} \left\| \frac{\partial^2 \text{vec}(\mathbf{H}_t)}{\partial \boldsymbol{\vartheta}_i \partial \boldsymbol{\vartheta}_j} \right\|^2 \right) < \infty, \quad (3.44)$$

for any $i, j = 1, \dots, d$.

Denote $\mathbf{a} = \text{vec}(\mathbf{A})$, $\mathbf{b} = \text{vec}(\mathbf{B})$ and $\mathbf{c} = \text{vec}(\mathbf{C})$. Using (3.37), we can calculate the components of the first derivation of $\text{vec}(\mathbf{H}_t)$ as the following

$$\frac{\partial \text{vec}(\mathbf{H}_t)}{\partial \boldsymbol{\gamma}'_{\boldsymbol{\varepsilon}}} = \sum_{k=0}^{\infty} (\mathbf{B}^{\otimes 2})^k (I_{m^2} - \mathbf{A}^{\otimes 2} - \mathbf{B}^{\otimes 2}) D_m, \quad (3.45)$$

$$\frac{\partial \text{vec}(\mathbf{H}_t)}{\partial \boldsymbol{\gamma}'_{\mathbf{x}}} = - \sum_{k=0}^{\infty} (\mathbf{B}^{\otimes 2})^k \mathbf{C}^{\otimes 2} D_m, \quad (3.46)$$

$$\begin{aligned} \frac{\partial \text{vec}(\mathbf{H}_t)}{\partial \mathbf{a}'} &= (I_{m^2} + M_{mm}) (\mathbf{A}(\boldsymbol{\varepsilon}_{t-1} \boldsymbol{\varepsilon}'_{t-1} - \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}) \otimes I_m) + \mathbf{B}^{\otimes 2} \frac{\partial \text{vec}(\mathbf{H}_{t-1})}{\partial \mathbf{a}'}, \\ &= \sum_{k=0}^{\infty} (\mathbf{B}^{\otimes 2})^k (I_{m^2} + M_{mm}) (\mathbf{A}(\boldsymbol{\varepsilon}_{t-k-1} \boldsymbol{\varepsilon}'_{t-k-1} - \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}) \otimes I_m) \end{aligned} \quad (3.47)$$

$$\begin{aligned} \frac{\partial \text{vec}(\mathbf{H}_t)}{\partial \mathbf{c}'} &= (I_{m^2} + M_{mm}) (\mathbf{C}(\mathbf{x}_{t-1} \mathbf{x}'_{t-1} - \boldsymbol{\Sigma}_{\mathbf{x}}) \otimes I_m) + \mathbf{B}^{\otimes 2} \frac{\partial \text{vec}(\mathbf{H}_{t-1})}{\partial \mathbf{c}'}, \\ &= \sum_{k=0}^{\infty} (\mathbf{B}^{\otimes 2})^k (I_{m^2} + M_{mm}) (\mathbf{C}(\mathbf{x}_{t-k-1} \mathbf{x}'_{t-k-1} - \boldsymbol{\Sigma}_{\mathbf{x}}) \otimes I_m), \end{aligned} \quad (3.48)$$

and

$$\begin{aligned} \frac{\partial \text{vec}(\mathbf{H}_t)}{\partial \mathbf{b}'} &= \sum_{k=1}^{\infty} \left((\mathbf{A}^{\otimes 2} \text{vec}(\boldsymbol{\varepsilon}_{t-k-1} \boldsymbol{\varepsilon}'_{t-k-1} - \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}) \right. \\ &\quad \left. + \mathbf{C}^{\otimes 2} \text{vec}(\mathbf{x}_{t-k-1} \mathbf{x}'_{t-k-1} - \boldsymbol{\Sigma}_{\mathbf{x}}))' \otimes I_{m^2} \right) \frac{\partial \text{vec}((\mathbf{B}^{\otimes 2})^k)}{\partial \mathbf{b}'}, \end{aligned} \quad (3.49)$$

where, using (3.39),

$$\frac{\partial \text{vec}((\mathbf{B}^{\otimes 2})^k)}{\partial \mathbf{b}'} = \sum_{i=0}^k \left((\mathbf{B}^{\otimes 2})^{k-i-1} \otimes (\mathbf{B}^{\otimes 2})^i \right) \frac{\partial \text{vec}(\mathbf{B}^{\otimes 2})}{\partial \mathbf{b}'} \quad (3.50)$$

To show (3.43), it suffices to prove that $E \left(\sup_{\boldsymbol{\vartheta} \in \Theta} \left\| \frac{\partial \mathbf{H}_t}{\partial \mathbf{d}'} \right\|^3 \right) = E \left(\sup_{\boldsymbol{\vartheta} \in \Theta} \left\| \frac{\partial \text{vec}(\mathbf{H}_t)}{\partial \mathbf{d}'} \right\|^3 \right) < \infty$, where $\mathbf{d} = \boldsymbol{\gamma}_{\boldsymbol{\varepsilon}}, \boldsymbol{\gamma}_{\mathbf{x}}, \mathbf{a}, \mathbf{b}, \mathbf{c}$. We immediately see from (3.45) and (3.46) that the derivatives with respect to the elements of $\boldsymbol{\gamma}$ are obviously bounded. Using the inequalities $\|A \otimes B\|_{sp} = \|A\|_{sp} \|B\|_{sp}$, $\|A\|_{sp} \leq \|A\| \leq \sqrt{m} \|A\|_{sp}$, $E(\sum a_i)^3 \leq \left(\sum (E a_i^3)^{1/3} \right)^3$, (3.27) and assumption **B9**, we get

$$\begin{aligned} E \left(\sup_{\boldsymbol{\vartheta} \in \Theta} \left\| \frac{\partial \text{vec}(\mathbf{H}_t)}{\partial \mathbf{a}'} \right\|^3 \right) &\leq K \left(\sum_{k=0}^{\infty} \left(E \sup_{\boldsymbol{\vartheta} \in \Theta} \|\mathbf{B}^{\otimes 2}\|^{3k} \|\text{vec}(\boldsymbol{\varepsilon}_{t-k-1} \boldsymbol{\varepsilon}'_{t-k-1} - \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})\|^3 \right)^{1/3} \right)^3 \\ &\leq K \left(\sum_{k=0}^{\infty} \varrho^k (E \|\boldsymbol{\varepsilon}_{t-k-1} \boldsymbol{\varepsilon}'_{t-k-1} - \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}\|^3)^{1/3} \right)^3 < \infty. \end{aligned} \quad (3.51)$$

Similary we also obtain

$$E \left(\sup_{\boldsymbol{\vartheta} \in \Theta} \left\| \frac{\partial \text{vec}(\mathbf{H}_t)}{\partial \mathbf{c}'} \right\|^3 \right) < \infty. \quad (3.52)$$

From (3.27)

$$\sup_{\boldsymbol{\vartheta} \in \Theta} \sum_{i=0}^k \left\| \left(\mathbf{B}^{\otimes 2'} \right)^{k-i-1} \otimes \left(\mathbf{B}^{\otimes 2} \right)^i \right\| \leq \sum_{i=0}^k \sup_{\boldsymbol{\vartheta} \in \Theta} \left\| \mathbf{B}^{\otimes 2'} \right\|^{k-i-1} \sup_{\boldsymbol{\vartheta} \in \Theta} \left\| \left(\mathbf{B}^{\otimes 2} \right)^i \right\| < Kk\varrho^k.$$

Using the same arguments to show (3.51) and (3.52), the following result is also obtained

$$E \left(\sup_{\boldsymbol{\vartheta} \in \Theta} \left\| \frac{\partial \text{vec}(\mathbf{H}_t)}{\partial \mathbf{b}'} \right\|^3 \right) < \infty.$$

(3.43) is thus shown. The second moment condition (3.44) is obtained by doing similar developments for the second order derivatives. \square

Lemma 3.2. *Under Assumptions B1-B10, \mathbf{J} is non-singular.*

Proof of Lemma 3.2. We apply the approach of Comte and Lieberman (2003) to prove the invertibility of the matrix \mathbf{J} . Starting by writing \mathbf{J} as a function of \mathbf{H}_t and of its derivatives. From (3.42), we have

$$\begin{aligned} E \left(\frac{\partial^2 \ell_t(\gamma_0, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_j} \mid \mathcal{F}_{t-1} \right) &= \text{Tr} \left(\mathbf{H}_{0t}^{-1} \frac{\partial \mathbf{H}_{0t}}{\partial \boldsymbol{\theta}_i} \mathbf{H}_{0t}^{-1} \frac{\partial \mathbf{H}_{0t}}{\partial \boldsymbol{\theta}_j} \right) \\ &= \left(\left(\mathbf{H}_{0t}^{-1/2} \right)^{\otimes 2} \text{vec} \left(\frac{\partial \mathbf{H}_{0t}}{\partial \boldsymbol{\theta}_i} \right) \right)' \left(\left(\mathbf{H}_{0t}^{-1/2} \right)^{\otimes 2} \text{vec} \left(\frac{\partial \mathbf{H}_{0t}}{\partial \boldsymbol{\theta}_j} \right) \right). \end{aligned}$$

Let $\mathbf{u}_{ti} = \left(\mathbf{H}_{0t}^{-1/2} \right)^{\otimes 2} \mathbf{v}_{ti}$, $\mathbf{v}_{ti} = \text{vec} \left(\frac{\partial \mathbf{H}_{0t}}{\partial \boldsymbol{\theta}_i} \right)$ and the matrices $\mathbf{u}_t = (\mathbf{u}_{t1} \mid \cdots \mid \mathbf{u}_{td_2})$ and $\mathbf{v}_t = (\mathbf{v}_{t1} \mid \cdots \mid \mathbf{v}_{td_2})$. Then $\mathbf{u}_t = \left(\mathbf{H}_{0t}^{-1/2} \right)^{\otimes 2} \mathbf{v}_t$ and $\mathbf{J} = E(\mathbf{u}_t' \mathbf{u}_t)$. If \mathbf{J} is singular, there exists $\mathbf{c} = (c_1, \dots, c_{d_2})' \in \mathbb{R}^{d_2}$ such that $\mathbf{c} \neq 0$ and

$$\mathbf{c}' \mathbf{J} \mathbf{c} = \mathbf{c}' E(\mathbf{u}_t' \mathbf{u}_t) \mathbf{c} = E((\mathbf{u}_t \mathbf{c})' (\mathbf{u}_t \mathbf{c})) = 0 \text{ a.s.}$$

Due to the positivity of $(\mathbf{u}_t \mathbf{c})' (\mathbf{u}_t \mathbf{c})$, then almost surely

$$(\mathbf{u}_t \mathbf{c})' (\mathbf{u}_t \mathbf{c}) = \mathbf{c}' \mathbf{u}_t' \mathbf{u}_t \mathbf{c} = \mathbf{c}' \mathbf{v}_t' \left(\left(\mathbf{H}_{0t}^{-1/2} \right)^{\otimes 2} \right)' \left(\mathbf{H}_{0t}^{-1/2} \right)^{\otimes 2} \mathbf{v}_t \mathbf{c} = 0 \text{ a.s.}$$

Because $\left(\left(\mathbf{H}_{0t}^{-1/2}\right)^{\otimes 2}\right)' \left(\mathbf{H}_{0t}^{-1/2}\right)^{\otimes 2}$ is strictly positive definite with probability one, it follows that

$$\mathbf{v}_t \mathbf{c} = \sum_{i=1}^{d_2} c_i \text{vec} \left(\frac{\partial \mathbf{H}_{0t}}{\partial \boldsymbol{\theta}_i} \right) = 0 \text{ a.s.}$$

Let $\boldsymbol{\Omega}_0^* = (I_{m^2} - \mathbf{A}_0^{\otimes 2} - \mathbf{B}_0^{\otimes 2}) \text{vec}(\boldsymbol{\Sigma}_\varepsilon) - \mathbf{C}_0^{\otimes 2} \text{vec}(\boldsymbol{\Sigma}_x)$. Then we have

$$0 = \bar{\boldsymbol{\Omega}}_0 + \bar{\mathbf{A}}_0 \text{vec}(\boldsymbol{\varepsilon}_{t-1} \boldsymbol{\varepsilon}'_{t-1}) + \bar{\mathbf{B}}_0 \text{vec}(\mathbf{H}_{0t-1}) + \bar{\mathbf{C}}_0 \text{vec}(\mathbf{x}_{t-1} \mathbf{x}'_{t-1}),$$

where $\bar{\boldsymbol{\Omega}}_0 = \sum_{i=1}^{d_2} c_i \frac{\partial \boldsymbol{\Omega}_0^*}{\partial \boldsymbol{\theta}_i}$, $\bar{\mathbf{A}}_0 = \sum_{i=1}^{d_2} c_i \frac{\partial \mathbf{A}_0^{\otimes 2}}{\partial \boldsymbol{\theta}_i}$, $\bar{\mathbf{B}}_0 = \sum_{i=1}^{d_2} c_i \frac{\partial \mathbf{B}_0^{\otimes 2}}{\partial \boldsymbol{\theta}_i}$, $\bar{\mathbf{C}}_0 = \sum_{i=1}^{d_2} c_i \frac{\partial \mathbf{C}_0^{\otimes 2}}{\partial \boldsymbol{\theta}_i}$. Then $\text{vec}(\mathbf{H}_{0t})$ can be represented by

$$\begin{aligned} \text{vec}(\mathbf{H}_{0t}) &= (\boldsymbol{\Omega}_0^* - \bar{\boldsymbol{\Omega}}_0) + (\mathbf{A}_0^{\otimes 2} - \bar{\mathbf{A}}_0) \text{vec}(\boldsymbol{\varepsilon}_{t-1} \boldsymbol{\varepsilon}'_{t-1}) + (\mathbf{B}_0^{\otimes 2} - \bar{\mathbf{B}}_0) \text{vec}(\mathbf{H}_{0t-1}) \\ &\quad + (\mathbf{C}_0^{\otimes 2} - \bar{\mathbf{C}}_0) \text{vec}(\mathbf{x}_{t-1} \mathbf{x}'_{t-1}). \end{aligned}$$

Because $\mathbf{c} \neq 0$, we have found another representation of $\text{vec}(\mathbf{H}_{0t})$, which contradicts Assumption **B6**. Hence, \mathbf{J} must be non-singular. \square

Lemma 3.3. *Under Assumptions B1-B11,*

$$\sqrt{n} \left\| \frac{\partial \tilde{Q}_n(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} - \frac{\partial Q_n(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right\| \rightarrow 0 \text{ in probability when } n \rightarrow \infty. \quad (3.53)$$

$$\sqrt{n} \sup_{\boldsymbol{\vartheta} \in \Theta} \left\| \frac{\partial^2 \tilde{Q}_n(\boldsymbol{\gamma}, \boldsymbol{\theta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} - \frac{\partial^2 Q_n(\boldsymbol{\gamma}, \boldsymbol{\theta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \right\| \rightarrow 0 \text{ in probability when } n \rightarrow \infty. \quad (3.54)$$

Proof of Lemma 3.3. This lemma means that the effect of the initial values on the derivatives of the criterion vanishes asymptotically. By the definition of $Q_n(\boldsymbol{\gamma}, \boldsymbol{\theta})$ and $\tilde{Q}_n(\boldsymbol{\gamma}, \boldsymbol{\theta})$, (3.53) and (3.54) are entailed by showing that

$$\frac{1}{\sqrt{n}} \sum_{t=0}^n \left\| \frac{\partial \tilde{\ell}_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} - \frac{\partial \ell_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right\| \rightarrow 0 \text{ in probability when } n \rightarrow \infty, \quad (3.55)$$

$$\frac{1}{\sqrt{n}} \sum_{t=0}^n \sup_{\boldsymbol{\vartheta} \in \Theta} \left\| \frac{\partial^2 \tilde{\ell}_t(\boldsymbol{\gamma}, \boldsymbol{\theta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} - \frac{\partial^2 \ell_t(\boldsymbol{\gamma}, \boldsymbol{\theta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \right\| \rightarrow 0 \text{ in probability when } n \rightarrow \infty. \quad (3.56)$$

For any $i = 1, \dots, 3m^2$, we have

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{t=0}^n \left| \frac{\partial \tilde{\ell}_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_i} - \frac{\partial \ell_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_i} \right| \\ & \leq \frac{1}{\sqrt{n}} \sum_{t=0}^n \left| \text{Tr} \left(\tilde{\mathbf{H}}_{0t}^{-1} \frac{\partial \tilde{\mathbf{H}}_{0t}}{\partial \boldsymbol{\theta}_i} - \mathbf{H}_{0t}^{-1} \frac{\partial \mathbf{H}_{0t}}{\partial \boldsymbol{\theta}_i} \right) \right| \\ & \quad + \frac{1}{\sqrt{n}} \sum_{t=0}^n \left| \text{Tr} \left(\tilde{\mathbf{H}}_{0t}^{-1} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' \tilde{\mathbf{H}}_{0t}^{-1} \frac{\partial \tilde{\mathbf{H}}_{0t}}{\partial \boldsymbol{\theta}_i} - \mathbf{H}_{0t}^{-1} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' \mathbf{H}_{0t}^{-1} \frac{\partial \mathbf{H}_{0t}}{\partial \boldsymbol{\theta}_i} \right) \right|. \end{aligned} \quad (3.57)$$

We will show that two terms in the right-hand side of the last inequality (3.57) tend to zero as $n \rightarrow \infty$. For the first term, we have

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{t=0}^n \left| \text{Tr} \left(\tilde{\mathbf{H}}_{0t}^{-1} \frac{\partial \tilde{\mathbf{H}}_{0t}}{\partial \boldsymbol{\theta}_i} - \mathbf{H}_{0t}^{-1} \frac{\partial \mathbf{H}_{0t}}{\partial \boldsymbol{\theta}_i} \right) \right| \\ & \leq \frac{1}{\sqrt{n}} \sum_{t=0}^n \left\| \tilde{\mathbf{H}}_{0t}^{-1} - \mathbf{H}_{0t}^{-1} \right\| \left\| \frac{\partial \tilde{\mathbf{H}}_{0t}}{\partial \boldsymbol{\theta}_i} \right\| + \frac{1}{\sqrt{n}} \sum_{t=0}^n \left\| \mathbf{H}_{0t}^{-1} \right\| \left\| \frac{\partial \tilde{\mathbf{H}}_{0t}}{\partial \boldsymbol{\theta}_i} - \frac{\partial \mathbf{H}_{0t}}{\partial \boldsymbol{\theta}_i} \right\| \\ & \leq \frac{1}{\sqrt{n}} \sum_{t=0}^n \sup_{\boldsymbol{\vartheta} \in \Theta} \left\| \tilde{\mathbf{H}}_t^{-1} - \mathbf{H}_t^{-1} \right\| \sup_{\boldsymbol{\vartheta} \in \Theta} \left\| \frac{\partial \tilde{\mathbf{H}}_t}{\partial \boldsymbol{\vartheta}_i} \right\| + \frac{1}{\sqrt{n}} \sum_{t=0}^n \sup_{\boldsymbol{\vartheta} \in \Theta} \left\| \mathbf{H}_t^{-1} \right\| \sup_{\boldsymbol{\vartheta} \in \Theta} \left\| \frac{\partial \tilde{\mathbf{H}}_t}{\partial \boldsymbol{\vartheta}_i} - \frac{\partial \mathbf{H}_t}{\partial \boldsymbol{\vartheta}_i} \right\|. \end{aligned}$$

The *vec* representation of $\tilde{\mathbf{H}}_t$ is obtained by replacing \mathbf{H}_t in (3.28) with $\tilde{\mathbf{H}}_t$. By simple differentiation and using Assumption **B3** and (3.27), we can obtain

$$E \left(\sup_{\boldsymbol{\vartheta} \in \Theta} \left\| \frac{\partial \tilde{\mathbf{H}}_t}{\partial \boldsymbol{\vartheta}_i} \right\| \right) < \infty \text{ and } E \left(\sup_{\boldsymbol{\vartheta} \in \Theta} \left\| \text{vec} \left(\frac{\partial \tilde{\mathbf{H}}_t}{\partial \boldsymbol{\vartheta}_i} - \frac{\partial \mathbf{H}_t}{\partial \boldsymbol{\vartheta}_i} \right) \right\| \right) = O(t \varrho^t). \quad (3.58)$$

Using Markov's inequality, (3.32) and (3.58), we can show that, for any $\epsilon > 0$,

$$P \left(\frac{1}{\sqrt{n}} \sum_{t=0}^n \sup_{\boldsymbol{\vartheta} \in \Theta} \left\| \tilde{\mathbf{H}}_t^{-1} - \mathbf{H}_t^{-1} \right\| \sup_{\boldsymbol{\vartheta} \in \Theta} \left\| \frac{\partial \tilde{\mathbf{H}}_t}{\partial \boldsymbol{\vartheta}_i} \right\| > \epsilon \right) \rightarrow 0$$

and

$$P \left(\frac{1}{\sqrt{n}} \sum_{t=0}^n \sup_{\boldsymbol{\vartheta} \in \Theta} \left\| \mathbf{H}_t^{-1} \right\| \sup_{\boldsymbol{\vartheta} \in \Theta} \left\| \frac{\partial \tilde{\mathbf{H}}_t}{\partial \boldsymbol{\vartheta}_i} - \frac{\partial \mathbf{H}_t}{\partial \boldsymbol{\vartheta}_i} \right\| > \epsilon \right) \rightarrow 0.$$

For the second term, we have

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{t=0}^n \left| \text{Tr} \left(\widetilde{\mathbf{H}}_{0t}^{-1} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t \widetilde{\mathbf{H}}_{0t}^{-1} \frac{\partial \widetilde{\mathbf{H}}_{0t}}{\partial \boldsymbol{\theta}_i} - \mathbf{H}_{0t}^{-1} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t \mathbf{H}_{0t}^{-1} \frac{\partial \mathbf{H}_{0t}}{\partial \boldsymbol{\theta}_i} \right) \right| \\ & \leq \frac{1}{\sqrt{n}} \sum_{t=0}^n \left| \text{Tr} \left(\widetilde{\mathbf{H}}_{0t}^{-1} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t \widetilde{\mathbf{H}}_{0t}^{-1} \left(\frac{\partial \widetilde{\mathbf{H}}_{0t}}{\partial \boldsymbol{\theta}_i} - \frac{\partial \mathbf{H}_{0t}}{\partial \boldsymbol{\theta}_i} \right) \right) \right| \\ & \quad + \frac{1}{\sqrt{n}} \sum_{t=0}^n \left| \text{Tr} \left(\left(\widetilde{\mathbf{H}}_{0t}^{-1} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t \widetilde{\mathbf{H}}_{0t}^{-1} - \mathbf{H}_{0t}^{-1} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t \mathbf{H}_{0t}^{-1} \right) \frac{\partial \mathbf{H}_{0t}}{\partial \boldsymbol{\theta}_i} \right) \right|. \end{aligned}$$

Using the inequality $|\text{Tr}(AB)| \leq \|A\| \|B\|$ and $\text{Tr}(AB) = \text{Tr}(BA)$ for matrices of appropriate sizes, we have

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{t=0}^n \left| \text{Tr} \left(\widetilde{\mathbf{H}}_{0t}^{-1} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t \widetilde{\mathbf{H}}_{0t}^{-1} \left(\frac{\partial \widetilde{\mathbf{H}}_{0t}}{\partial \boldsymbol{\theta}_i} - \frac{\partial \mathbf{H}_{0t}}{\partial \boldsymbol{\theta}_i} \right) \right) \right| \\ & \leq \frac{1}{\sqrt{n}} \sum_{t=0}^n \left\| \widetilde{\mathbf{H}}_{0t}^{-1} \right\|^2 \|\mathbf{H}_{0t}\| \|\boldsymbol{\eta}_t \boldsymbol{\eta}'_t\| \left\| \frac{\partial \widetilde{\mathbf{H}}_{0t}}{\partial \boldsymbol{\theta}_i} - \frac{\partial \mathbf{H}_{0t}}{\partial \boldsymbol{\theta}_i} \right\| \\ & \leq K \frac{1}{\sqrt{n}} \sum_{t=0}^n \sup_{\boldsymbol{\vartheta} \in \Theta} \|\mathbf{H}_t\| \|\boldsymbol{\eta}_t \boldsymbol{\eta}'_t\| \sup_{\boldsymbol{\vartheta} \in \Theta} \left\| \frac{\partial \widetilde{\mathbf{H}}_t}{\partial \boldsymbol{\theta}_i} - \frac{\partial \mathbf{H}_t}{\partial \boldsymbol{\theta}_i} \right\|. \end{aligned}$$

Then by Hölder's inequality, we can get

$$P \left(K \frac{1}{\sqrt{n}} \sum_{t=0}^n \sup_{\boldsymbol{\vartheta} \in \Theta} \|\mathbf{H}_t\| \|\boldsymbol{\eta}_t \boldsymbol{\eta}'_t\| \sup_{\boldsymbol{\vartheta} \in \Theta} \left\| \frac{\partial \widetilde{\mathbf{H}}_t}{\partial \boldsymbol{\theta}_i} - \frac{\partial \mathbf{H}_t}{\partial \boldsymbol{\theta}_i} \right\| > \epsilon \right) \rightarrow 0.$$

It implies that almost surely, as $n \rightarrow \infty$

$$\frac{1}{\sqrt{n}} \sum_{t=0}^n \left| \text{Tr} \left(\widetilde{\mathbf{H}}_{0t}^{-1} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t \widetilde{\mathbf{H}}_{0t}^{-1} \left(\frac{\partial \widetilde{\mathbf{H}}_{0t}}{\partial \boldsymbol{\theta}_i} - \frac{\partial \mathbf{H}_{0t}}{\partial \boldsymbol{\theta}_i} \right) \right) \right| \rightarrow 0. \quad (3.59)$$

Applying the same arguments to show (3.59), we also get

$$\frac{1}{\sqrt{n}} \sum_{t=0}^n \left| \text{Tr} \left(\left(\widetilde{\mathbf{H}}_{0t}^{-1} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t \widetilde{\mathbf{H}}_{0t}^{-1} - \mathbf{H}_{0t}^{-1} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t \mathbf{H}_{0t}^{-1} \right) \frac{\partial \mathbf{H}_{0t}}{\partial \boldsymbol{\theta}_i} \right) \right| \rightarrow 0 \text{ a.s.}$$

It follows that the second term in the right-hand of (3.57) almost surely tends to zero. The proof of *i*) is thus obtained. By the same arguments, (3.56) can be also showed. \square

Lemma 3.4. *Under Assumptions B1-B10,*

$$\frac{\partial^2 \widetilde{Q}_n(\boldsymbol{\vartheta}_n)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\vartheta}'} \rightarrow E \left(\frac{\partial^2 \ell_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\vartheta}'} \right) \text{ in probability when } \boldsymbol{\vartheta}_n \rightarrow \boldsymbol{\vartheta}_0 \text{ in probability.} \quad (3.60)$$

Proof of Lemma 3.4. First note that

$$P \left(\left\| \frac{\partial^2 \tilde{Q}_n(\boldsymbol{\vartheta}_n)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\vartheta}'} - E \left(\frac{\partial^2 \ell_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\vartheta}'} \right) \right\| \geq \epsilon \right) \leq p_1 + p_2 + p_3 + p_4,$$

where

$$\begin{aligned} p_1 &= P \left(\sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0) \cap \Theta} \left\| \frac{\partial^2 \tilde{Q}_n(\boldsymbol{\vartheta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\vartheta}'} - \frac{\partial^2 Q_n(\boldsymbol{\vartheta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\vartheta}'} \right\| \geq \frac{\epsilon}{3} \right), \\ p_2 &= P \left(\sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0) \cap \Theta} \left\| \frac{\partial^2 Q_n(\boldsymbol{\vartheta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\vartheta}'} - \frac{\partial^2 Q_n(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\vartheta}'} \right\| \geq \frac{\epsilon}{3} \right), \\ p_3 &= P \left(\left\| \frac{\partial^2 Q_n(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\vartheta}'} - E \left(\frac{\partial^2 \ell_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\vartheta}'} \right) \right\| \geq \frac{\epsilon}{3} \right), \quad p_4 = P \{ \boldsymbol{\vartheta}_n \notin V(\boldsymbol{\vartheta}_0) \} \end{aligned}$$

for any $\epsilon > 0$ and any neighborhood $V(\boldsymbol{\vartheta}_0)$. By the assumption that $\boldsymbol{\vartheta}_n \rightarrow \boldsymbol{\vartheta}_0$ in probability, we have $p_4 \rightarrow 0$ as $n \rightarrow \infty$. By (3.54), for any $\epsilon > 0$ and when $V(\boldsymbol{\vartheta}_0)$ is sufficiently small, $p_1 \rightarrow 0$. Because $\frac{\partial^2 \ell_t(\boldsymbol{\gamma}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\vartheta}'}$ is a function of $(\boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t-1}, \dots)$ and $(\mathbf{x}_t, \mathbf{x}_{t-1}, \dots)$, under Assumption **B2** it is strictly stationary and ergodic. The uniform law of large numbers for stationary ergodic processes and Lemma 3.1 thus imply that $p_3 \rightarrow 0$ for any $\epsilon > 0$. To prove that $p_2 \rightarrow 0$, it suffices to show that, for all $\epsilon > 0$, there exists $V(\boldsymbol{\vartheta}_0)$ satisfying

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0) \cap \Theta} \left\| \frac{\partial^2 \ell_t(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} - \frac{\partial^2 \ell_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \right\| \leq \epsilon \quad \text{a.s.}$$

The result follows from the ergodic theorem, the dominated convergence theorem, the uniform continuity of the second order derivatives of $\ell_t(\boldsymbol{\vartheta})$, and by Lemma 3.1. \square

Lemma 3.5. *Under Assumptions B1 - B10,*

$$\begin{pmatrix} \sqrt{n} (\hat{\boldsymbol{\gamma}}_{\mathbf{x},n} - \boldsymbol{\gamma}_{\mathbf{x},0}) \\ \sqrt{n} (\hat{\boldsymbol{\gamma}}_{\boldsymbol{\varepsilon},n} - \boldsymbol{\gamma}_{\boldsymbol{\varepsilon},0}) \\ \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\partial \ell_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \end{pmatrix} = \boldsymbol{\Phi} \frac{1}{\sqrt{n}} \begin{pmatrix} \sum_{t=1}^n \text{vech}(\mathbf{x}_t \mathbf{x}_t' - E \mathbf{x}_1 \mathbf{x}_1') \\ \sum_{t=1}^n \boldsymbol{\Upsilon}_{0t} \text{vec}(\boldsymbol{\eta}_t \boldsymbol{\eta}_t' - I_m) \end{pmatrix} + o_p(1), \quad (3.61)$$

where $\boldsymbol{\Upsilon}_{0t}$ and $\boldsymbol{\Phi}$ are given in (4.12) and (3.20), respectively.

Proof of Lemma 3.5. Introduce the martingale difference

$$\boldsymbol{\nu}_t = \text{vec}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') - \text{vec}(\mathbf{H}_{0t}) = \left(\mathbf{H}_{0t}^{1/2} \right)^{\otimes 2} \text{vec}(\boldsymbol{\eta}_t \boldsymbol{\eta}_t' - I_m).$$

In the representation of $\text{vec}(\mathbf{H}_{0t})$ obtained from (3.2), we replace $\text{vec}(\mathbf{H}_{0t})$ by $\text{vec}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t) - \boldsymbol{\nu}_t$. Then, we get

$$\begin{aligned} \text{vec}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t - E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t)) &= (\mathbf{A}_0^{\otimes 2} + \mathbf{B}_0^{\otimes 2}) \text{vec}(\boldsymbol{\varepsilon}_{t-1} \boldsymbol{\varepsilon}'_{t-1} - E(\boldsymbol{\varepsilon}_{t-1} \boldsymbol{\varepsilon}'_{t-1})) \\ &\quad + \mathbf{C}_0^{\otimes 2} \text{vec}(\mathbf{x}_{t-1} \mathbf{x}'_{t-1} - E(\mathbf{x}_{t-1} \mathbf{x}'_{t-1})) + (\boldsymbol{\nu}_t - \mathbf{B}_0^{\otimes 2} \boldsymbol{\nu}_{t-1}). \end{aligned}$$

Note that under assumption **B5**, the matrix $I_{m^2} - \mathbf{A}_0^{\otimes 2} - \mathbf{B}_0^{\otimes 2}$ is invertible. Taking the average of the two side of the equality for $t = 1, \dots, n$ gives

$$\begin{aligned} \widehat{\boldsymbol{\gamma}}_{\boldsymbol{\varepsilon}, n} - \boldsymbol{\gamma}_{\boldsymbol{\varepsilon}, 0} &= L_m (I_{m^2} - \mathbf{A}_0^{\otimes 2} - \mathbf{B}_0^{\otimes 2})^{-1} (I_{m^2} - \mathbf{B}_0^{\otimes 2}) \frac{1}{n} \sum_{t=1}^n \boldsymbol{\nu}_t \\ &\quad + L_m (I_{m^2} - \mathbf{A}_0^{\otimes 2} - \mathbf{B}_0^{\otimes 2})^{-1} \mathbf{C}_0^{\otimes 2} D_r (\widehat{\boldsymbol{\gamma}}_{\mathbf{x}, n} - \boldsymbol{\gamma}_{\mathbf{x}, 0}) + o_p(1), \quad a.s. \end{aligned}$$

Using the relation $\text{Tr}(A'B) = \text{vec}'(A)\text{vec}(B)$, we have, for $i = 1, \dots, d_2$

$$\begin{aligned} \frac{\partial \ell_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_i} &= \text{Tr} \left(\left(\mathbf{H}_{0t}^{-1} - \mathbf{H}_{0t}^{-1} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t \mathbf{H}_{0t}^{-1} \right) \frac{\partial \mathbf{H}_{0t}}{\partial \boldsymbol{\theta}_i} \right) \\ &= \text{vec}' \left(\frac{\partial \mathbf{H}_{0t}}{\partial \boldsymbol{\theta}_i} \right) \text{vec} \left(\left(\mathbf{H}_{0t}^{-1/2} \right)' (I_m - \boldsymbol{\eta}_t \boldsymbol{\eta}'_t) \left(\mathbf{H}_{0t}^{-1/2} \right) \right) \\ &= -\text{vec}' \left(\frac{\partial \mathbf{H}_{0t}}{\partial \boldsymbol{\theta}_i} \right) \left(\left(\mathbf{H}_{0t}^{-1/2} \right)^{\otimes 2} \right)' \text{vec}(\boldsymbol{\eta}_t \boldsymbol{\eta}'_t - I_m). \end{aligned}$$

It follows that

$$\frac{\partial \ell_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} = -\frac{\partial \text{vec}'(\mathbf{H}_{0t})}{\partial \boldsymbol{\theta}} \left(\left(\mathbf{H}_{0t}^{-1/2} \right)^{\otimes 2} \right)' \text{vec}(\boldsymbol{\eta}_t \boldsymbol{\eta}'_t - I_m) \quad (3.62)$$

and (3.61) is thus obtained. \square

Lemma 3.6. *Under Assumptions B1-B10,*

$$\frac{1}{\sqrt{n}} \left(\begin{array}{c} \sum_{t=1}^n \text{vech}(\mathbf{x}_t \mathbf{x}'_t - E\mathbf{x}_1 \mathbf{x}'_1) \\ \sum_{t=1}^n \boldsymbol{\Upsilon}_{0t} \text{vec}(\boldsymbol{\eta}_t \boldsymbol{\eta}'_t - I_m) \end{array} \right) \xrightarrow{d} \mathcal{N} \left(0, \boldsymbol{\Sigma} = \left(\begin{array}{cc} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}'_{12} & \boldsymbol{\Sigma}_{22} \end{array} \right) \right)$$

where $\boldsymbol{\Sigma}$ is given in Theorem 3.2.

Proof of Lemma 3.6. From (3.2), we write, for $s > 0$, $\mathbf{H}_{0t} = \underline{\mathbf{H}}_{0t,s} + \overline{\mathbf{H}}_{0t,s}$ such that

$$\text{vec}(\overline{\mathbf{H}}_{0t,s}) = \sum_{k=s+1}^{\infty} (\mathbf{B}_0^{\otimes 2})^k (\text{vec}(\boldsymbol{\Omega}_0) + \mathbf{A}_0^{\otimes 2} \text{vec}(\boldsymbol{\varepsilon}_{t-k-1} \boldsymbol{\varepsilon}'_{t-k-1}) + \mathbf{C}_0^{\otimes 2} \text{vec}(\mathbf{x}_{t-k-1} \mathbf{x}'_{t-k-1})).$$

Let $\mathbf{H}_{0t}^{1/2} = \underline{\mathbf{H}}_{0t,s}^{1/2} + R_{t,s}$. Note that $\underline{\mathbf{H}}_{0t,s}^{1/2}$ is invertible. Then $\mathbf{H}_{0t}^{-1/2} = \underline{\mathbf{H}}_{0t,s}^{-1/2} + R_{t,s}^*$, where $R_{t,s}^* = -\underline{\mathbf{H}}_{0t,s}^{-1/2} R_{t,s} \left(I + \underline{\mathbf{H}}_{0t,s}^{-1/2} R_{t,s} \right)^{-1} \underline{\mathbf{H}}_{0t,s}^{-1/2}$ (see Miller (1981) for the inverse of the sum of two

matrices). Using the elementary relation $(A + B) \otimes (C + D) = A \otimes C + A \otimes D + B \otimes C + B \otimes D$, we can write $\Upsilon_{0t} \text{vec}(\boldsymbol{\eta}_t \boldsymbol{\eta}_t' - I_m) = \underline{\mathbf{Y}}_{t,s} + \mathbf{R}_{t,s}$, where

$$\underline{\mathbf{Y}}_{t,s} = \left(\begin{array}{c} \underline{\mathbf{H}}_{0t,s}^{1/2} \otimes \underline{\mathbf{H}}_{0t,s}^{1/2} \\ -\frac{\partial \text{vec}'(\underline{\mathbf{H}}_{0t,s})}{\partial \boldsymbol{\theta}} \left(\underline{\mathbf{H}}_{0t,s}^{-1/2} \otimes \underline{\mathbf{H}}_{0t,s}^{-1/2} \right)' \end{array} \right) \text{vec}(\boldsymbol{\eta}_t \boldsymbol{\eta}_t' - I_m)$$

and $\mathbf{R}_{t,s}$ is the rest of the development. Note that the processes $(\underline{\mathbf{Y}}_{t,s})_t$ and $(\mathbf{R}_{t,s})_t$ are stationary and centered. Using the relations $(A \otimes B)(C \otimes D) = (AC \otimes BD)$, $(A \otimes B)' = A' \otimes B'$ and $\text{Tr}(A \otimes B) = \text{Tr}(A)\text{Tr}(B)$, we have

$$E \left\| \underline{\mathbf{H}}_{0t,s}^{1/2} \otimes \underline{\mathbf{H}}_{0t,s}^{1/2} \right\|^2 = E \left(\text{Tr} \left(\underline{\mathbf{H}}_{0t,s} \otimes \underline{\mathbf{H}}_{0t,s} \right) \right) = E \left(\text{Tr}(\underline{\mathbf{H}}_{0t,s}) \right)^2 \leq KE \|\underline{\mathbf{H}}_{0t,s}\|^2.$$

Then using the Holder inequality and Assumption **B10**, we have, for some ν and $\delta > 0$

$$\begin{aligned} \left\| \left(\underline{\mathbf{H}}_{0t,s}^{1/2} \otimes \underline{\mathbf{H}}_{0t,s}^{1/2} \right) \text{vec}(\boldsymbol{\eta}_t \boldsymbol{\eta}_t') \right\|_{2+\nu} &\leq \left\| \underline{\mathbf{H}}_{0t,s}^{1/2} \otimes \underline{\mathbf{H}}_{0t,s}^{1/2} \right\|_{(2+\nu)(1+1/\delta)} \left\| \text{vec}(\boldsymbol{\eta}_t \boldsymbol{\eta}_t') \right\|_{(2+\nu)(1+\delta)} \\ &\leq \left\| \underline{\mathbf{H}}_{0t,s} \right\|_{(2+\nu)(1+1/\delta)} \left\| \text{vec}(\boldsymbol{\eta}_t \boldsymbol{\eta}_t') \right\|_{(2+\nu)(1+\delta)} < \infty. \end{aligned}$$

Similarly,

$$\left\| \frac{\partial \text{vec}'(\underline{\mathbf{H}}_{0t,s})}{\partial \boldsymbol{\theta}} \left(\underline{\mathbf{H}}_{0t,s}^{-1/2} \otimes \underline{\mathbf{H}}_{0t,s}^{-1/2} \right)' \text{vec}(\boldsymbol{\eta}_t \boldsymbol{\eta}_t') \right\|_{2+\nu} \leq K \left\| \frac{\partial \text{vec}'(\underline{\mathbf{H}}_{0t,s})}{\partial \boldsymbol{\theta}} \text{vec}(\boldsymbol{\eta}_t \boldsymbol{\eta}_t') \right\|_{2+\nu} < \infty.$$

It entails that $\|\underline{\mathbf{Y}}_{t,s}\|_{2+\nu} < \infty$. Therefore, under Assumption **B10** and s fixed, the process $(\underline{\mathbf{Y}}_{t,s})_t$ is strongly mixing, with mixing coefficients $\alpha_{\underline{\mathbf{Y}}}(h) \leq \alpha_{\mathbf{z}}(\max\{0, h - s\})$. Applying the CLT of Herrndorf (1984) for mixing processes, we directly obtain

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \underline{\mathbf{Y}}_{t,s} \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma}_{22,s}), \quad \boldsymbol{\Sigma}_{22,s} = \sum_{h=-\infty}^{\infty} \text{cov}(\underline{\mathbf{Y}}_{t,s}, \underline{\mathbf{Y}}_{t-h,s}).$$

Let $\mathbf{x}_t = \text{vech}(\mathbf{x}_t \mathbf{x}_t' - E \mathbf{x}_1 \mathbf{x}_1')$ and $\boldsymbol{\Sigma}_{12,s} = \sum_{h=-\infty}^{\infty} \text{cov}(\mathbf{x}_t, \underline{\mathbf{Y}}_{t-h,s})$. As in Francq and Zakoian (1998) Lemma 3, we can show that the matrices $\boldsymbol{\Sigma}_{22} = \lim_{s \rightarrow \infty} \boldsymbol{\Sigma}_{22,s}$ and $\boldsymbol{\Sigma}_{12} = \lim_{s \rightarrow \infty} \boldsymbol{\Sigma}_{12,s}$ exist. Using Assumption **B11** and the arguments given in the proof of Lemma 4 in the precedent reference, one can show that

$$\lim_{s \rightarrow \infty} \limsup_{n \rightarrow \infty} P \left(\left\| n^{-1/2} \sum_{t=1}^n \mathbf{R}_{t,s} \right\| > \epsilon \right) = 0$$

for any $\epsilon > 0$. Then using Assumption **B10** and the CLT of Herrndorf (1984), we get

$$\frac{1}{\sqrt{n}} \begin{pmatrix} \sum_{t=1}^n \text{vech}(\mathbf{x}_t \mathbf{x}_t' - E \mathbf{x}_1 \mathbf{x}_1') \\ \sum_{t=1}^n \mathbf{\Upsilon}_{0t} \text{vec}(\boldsymbol{\eta}_t \boldsymbol{\eta}_t' - I_m) \end{pmatrix} = \frac{1}{\sqrt{n}} \begin{pmatrix} \sum_{t=1}^n \mathbf{x}_t \\ \sum_{t=1}^n \mathbf{Y}_{t,s} \end{pmatrix} + \frac{1}{\sqrt{n}} \begin{pmatrix} \mathbf{0} \\ \sum_{t=1}^n \mathbf{R}_{t,s} \end{pmatrix} \\ \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma}).$$

□

Proof of Theorem 3.2.

By the strong consistency, assumption **B4** and the definition of $\widehat{\boldsymbol{\theta}}_n$, for n large enough, $\widehat{\boldsymbol{\theta}}_n$ is contained *a.s.* in an arbitrary small neighborhood of $\boldsymbol{\theta}_0$ that belongs to interior of the parameter set $\Theta_{\boldsymbol{\theta}}$. The first-order condition

$$0 = \frac{1}{n} \sum_{t=1}^n \frac{\partial \tilde{\ell}_t(\widehat{\boldsymbol{\gamma}}_n, \widehat{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}} \quad (3.63)$$

is thus satisfied.

Let us define the following matrices

$$\mathbf{K}_{\epsilon n}(\boldsymbol{\vartheta}) = \frac{\partial^2 \tilde{Q}_n(\boldsymbol{\vartheta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\gamma}'_{\epsilon}} = \frac{1}{n} \sum_{t=1}^n \frac{\partial^2 \tilde{\ell}_t(\boldsymbol{\vartheta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\gamma}'_{\epsilon}}. \quad (3.64)$$

$$\mathbf{K}_{x n}(\boldsymbol{\vartheta}) = \frac{\partial^2 \tilde{Q}_n(\boldsymbol{\vartheta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\gamma}'_x} = \frac{1}{n} \sum_{t=1}^n \frac{\partial^2 \tilde{\ell}_t(\boldsymbol{\vartheta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\gamma}'_x}. \quad (3.65)$$

$$\mathbf{J}_n(\boldsymbol{\vartheta}) = \frac{\partial^2 \tilde{Q}_n(\boldsymbol{\vartheta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \frac{1}{n} \sum_{t=1}^n \frac{\partial^2 \tilde{\ell}_t(\boldsymbol{\vartheta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}. \quad (3.66)$$

The mean-value theorem applied to each element of the right-hand side of the first-order condition gives

$$0 = \frac{1}{n} \sum_{t=1}^n \frac{\partial \tilde{\ell}_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} + \frac{1}{n} \sum_{t=1}^n \frac{\partial^2 \tilde{\ell}_t(\boldsymbol{\vartheta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\gamma}'_{\epsilon}} (\widehat{\boldsymbol{\gamma}}_{\epsilon n} - \boldsymbol{\gamma}_{\epsilon 0}) + \frac{1}{n} \sum_{t=1}^n \frac{\partial^2 \tilde{\ell}_t(\boldsymbol{\vartheta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\gamma}'_x} (\widehat{\boldsymbol{\gamma}}_{x n} - \boldsymbol{\gamma}_{x 0}) \\ + \frac{1}{n} \sum_{t=1}^n \frac{\partial^2 \tilde{\ell}_t(\boldsymbol{\vartheta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \\ = \frac{1}{n} \sum_{t=1}^n \frac{\partial \tilde{\ell}_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} + \mathbf{K}_{\epsilon n}(\boldsymbol{\vartheta}^*) (\widehat{\boldsymbol{\gamma}}_{\epsilon n} - \boldsymbol{\gamma}_{\epsilon 0}) + \mathbf{K}_{x n}(\boldsymbol{\vartheta}^*) (\widehat{\boldsymbol{\gamma}}_{x n} - \boldsymbol{\gamma}_{x 0}) + \mathbf{J}_n(\boldsymbol{\vartheta}^*) (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$$

where $\boldsymbol{\vartheta}^*$ is between $\widehat{\boldsymbol{\vartheta}}_n$ and $\boldsymbol{\vartheta}_0$. By Lemma 3.2, Lemma 3.3, Lemma 3.4 and the consistency of $\widehat{\boldsymbol{\vartheta}}_n$, the matrix $\mathbf{J}_n(\boldsymbol{\vartheta}^*)$ is *a.s.* invertible for sufficiently large n . Hence multiplying by \sqrt{n} and solving

for $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ gives

$$\begin{aligned} \sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) &= -\mathbf{J}_n^{-1}(\boldsymbol{\vartheta}^*) \left(\frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\partial \tilde{\ell}_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right) - \mathbf{J}_n^{-1}(\boldsymbol{\vartheta}^*) \mathbf{K}_{\varepsilon n}(\boldsymbol{\vartheta}^*) \sqrt{n}(\widehat{\boldsymbol{\gamma}}_{\varepsilon n} - \boldsymbol{\gamma}_{\varepsilon 0}) \\ &\quad - \mathbf{J}_n^{-1}(\boldsymbol{\vartheta}^*) \mathbf{K}_{x n}(\boldsymbol{\vartheta}^*) \sqrt{n}(\widehat{\boldsymbol{\gamma}}_{x n} - \boldsymbol{\gamma}_{x 0}). \end{aligned}$$

Hence, notice that $\frac{\partial \tilde{Q}_n(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} = \frac{1}{n} \sum_{t=1}^n \frac{\partial \tilde{\ell}_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}$, we have the following representation

$$\sqrt{n} \begin{pmatrix} \widehat{\boldsymbol{\gamma}}_{x n} - \boldsymbol{\gamma}_{x 0} \\ \widehat{\boldsymbol{\gamma}}_{\varepsilon n} - \boldsymbol{\gamma}_{\varepsilon 0} \\ \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \end{pmatrix} = \boldsymbol{\Gamma}_n \begin{pmatrix} \sqrt{n}(\widehat{\boldsymbol{\gamma}}_{x n} - \boldsymbol{\gamma}_{x 0}) \\ \sqrt{n}(\widehat{\boldsymbol{\gamma}}_{\varepsilon n} - \boldsymbol{\gamma}_{\varepsilon 0}) \\ \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\partial \tilde{\ell}_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \end{pmatrix}$$

where

$$\boldsymbol{\Gamma}_n = \begin{pmatrix} I_{r(r+1)/2} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I_{m(m+1)/2} & \mathbf{0} \\ -\mathbf{J}_n^{-1}(\boldsymbol{\vartheta}^*) \mathbf{K}_{x n}(\boldsymbol{\vartheta}^*) & -\mathbf{J}_n^{-1}(\boldsymbol{\vartheta}^*) \mathbf{K}_{\varepsilon n}(\boldsymbol{\vartheta}^*) & -\mathbf{J}_n^{-1}(\boldsymbol{\vartheta}^*) \end{pmatrix}$$

By Lemma 3.4 and the consistency of $\widehat{\boldsymbol{\vartheta}}_n$, we get

$$\boldsymbol{\Gamma}_n \rightarrow \boldsymbol{\Gamma} \text{ in probability as } n \rightarrow \infty$$

and by Lemma 3.3, Lemma 3.5 and Lemma 3.6,

$$\begin{pmatrix} \sqrt{n}(\widehat{\boldsymbol{\gamma}}_{x n} - \boldsymbol{\gamma}_{x 0}) \\ \sqrt{n}(\widehat{\boldsymbol{\gamma}}_{\varepsilon n} - \boldsymbol{\gamma}_{\varepsilon 0}) \\ \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\partial \tilde{\ell}_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Phi} \boldsymbol{\Sigma} \boldsymbol{\Phi}')$$

The asymptotic normality of the VTE now follows the Slutsky theorem. \square

Proof of Corollary 3.1. The proof of this corollary can be obtained by applying directly the delta method (see Theorem 3.1 in van der Vaart, 1998). Indeed, let φ be the map which transforms $\boldsymbol{\vartheta}_0$

into ξ_0 . This linear map is differentiable at ϑ_0 , and is described by the Jacobian matrix

$$\frac{\partial \varphi}{\partial \vartheta'_0} = \begin{pmatrix} \frac{\partial \text{vec}(\Omega_0)}{\partial \gamma'_{x0}} & \frac{\partial \text{vec}(\Omega_0)}{\partial \gamma'_{\varepsilon 0}} & \frac{\partial \text{vec}(\Omega_0)}{\partial \mathbf{a}'_0} & \frac{\partial \text{vec}(\Omega_0)}{\partial \mathbf{b}'_0} & \frac{\partial \text{vec}(\Omega_0)}{\partial \mathbf{c}'_0} \\ \frac{\partial \gamma'_{x0}}{\partial \mathbf{a}_0} & \frac{\partial \gamma'_{\varepsilon 0}}{\partial \mathbf{a}_0} & \frac{\partial \mathbf{a}'_0}{\partial \mathbf{a}_0} & \frac{\partial \mathbf{b}'_0}{\partial \mathbf{a}_0} & \frac{\partial \mathbf{c}'_0}{\partial \mathbf{a}_0} \\ \frac{\partial \gamma'_{x0}}{\partial \mathbf{b}_0} & \frac{\partial \gamma'_{\varepsilon 0}}{\partial \mathbf{b}_0} & \frac{\partial \mathbf{a}'_0}{\partial \mathbf{b}_0} & \frac{\partial \mathbf{b}'_0}{\partial \mathbf{b}_0} & \frac{\partial \mathbf{c}'_0}{\partial \mathbf{b}_0} \\ \frac{\partial \gamma'_{x0}}{\partial \mathbf{c}_0} & \frac{\partial \gamma'_{\varepsilon 0}}{\partial \mathbf{c}_0} & \frac{\partial \mathbf{a}'_0}{\partial \mathbf{c}_0} & \frac{\partial \mathbf{b}'_0}{\partial \mathbf{c}_0} & \frac{\partial \mathbf{c}'_0}{\partial \mathbf{c}_0} \\ \frac{\partial \gamma'_{x0}}{\partial \gamma'_{\varepsilon 0}} & \frac{\partial \gamma'_{\varepsilon 0}}{\partial \gamma'_{\varepsilon 0}} & \frac{\partial \mathbf{a}'_0}{\partial \gamma'_{\varepsilon 0}} & \frac{\partial \mathbf{b}'_0}{\partial \gamma'_{\varepsilon 0}} & \frac{\partial \mathbf{c}'_0}{\partial \gamma'_{\varepsilon 0}} \end{pmatrix} \quad (3.67)$$

where $\mathbf{a}_0 = \text{vec}(\mathbf{A}_0)$, $\mathbf{b}_0 = \text{vec}(\mathbf{B}_0)$ and $\mathbf{c}_0 = \text{vec}(\mathbf{C}_0)$. Using (3.37) and (3.38), the components of the Jacobian matrix can be easily calculated and we get $\frac{\partial \varphi}{\partial \vartheta'_0} = \Delta$.

□

References

- Andrews, D.W.K.** (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.
- Bollerslev, T.** (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.
- Bollerslev, T., R.F. Engle and J.M. Wooldridge** (1988) A capital asset pricing model with time-varying covariances. *The Journal of Political Economy*, 116–131.
- Bouissou, M.B., J-J. Laffont and Q.H. Vuong** (1986) Tests of noncausality under Markov assumptions for qualitative panel data. *Econometrica : Journal of the Econometric Society*, 395–414.
- Boussama, F., F. Fuchs and R. Stelzer** (2011) Stationarity and geometric ergodicity of BEKK multivariate GARCH models. *Stochastic Processes and their Applications* 121, 2331–2360.
- Cakmakli, C. and D.J.C. Van Dijk** (2010) Getting the most out of macroeconomic information for predicting stock returns and volatility.
- Christiansen, C., M. Schmeling and A. Schrimpf** (2012) A comprehensive look at financial volatility prediction by economic variables. *Journal of Applied Econometrics* 27, 956–977.
- Comte, F. and O. Lieberman** (2003) Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis* 84, 61–84.
- Engle, R.F., Hendry, D.F. and J.F. Richard** (1983) Exogeneity. *Econometrica* 51, 277–304.
- Engle, R.F., D.F. Hendry and J-F. Richard** (1983) Exogeneity. *Econometrica : Journal of the Econometric Society*, 277–304.

- Engle, R.F. and K.F. Kroner (1995) Multivariate simultaneous generalized ARCH. *Econometric theory* 11, 122–150.
- Engle, R. (2009) *Anticipating correlations : a new paradigm for risk management*. Princeton University Press.
- Florens, J-P. and M. Mouchart (1982) A note on noncausality. *Econometrica : Journal of the Econometric Society*, 583–591.
- Francq, C. and J-M. Zakoïan (1998) Estimating linear representations of nonlinear processes. *Journal of Statistical Planning and Inference* 68, 145–165.
- Francq, C. and J-M. Zakoïan (2000) Covariance matrix estimation for estimators of mixing weak ARMA models. *Journal of Statistical Planning and Inference* 83, 369–394.
- Francq, C., R. Roy and J-M. Zakoïan (2005) Diagnostic checking in ARMA models with uncorrelated errors. *Journal of the American Statistical Association* 100, 532–544.
- Francq, C. and J-M. Zakoïan (2010) *GARCH Models : Structure, Statistical Inference and Financial Applications*. John Wiley.
- Francq, C., L. Horvath and J-M Zakoïan (2011) Merits and drawbacks of variance targeting in GARCH models. *Journal of Financial Econometrics* 9, 619–656.
- Francq, C., L. Horváth and J-M Zakoïan (2014) Variance targeting estimation of multivariate GARCH models. *Journal of Financial Econometrics*, https://mpra.ub.uni-muenchen.de/57794/1/MPRA_paper_57794.pdf
- Francq, C. and Thieu, L.Q. (2015) Qml inference for volatility models with covariates. *University Library of Munich, Germany*, No. 63198.
- Francq, C. and G. Sucarrat (2015) Equation-by-Equation Estimation of a Multivariate Log-GARCH-X Model of Financial Returns.
- Hafner, C.M. and A. Preminger (2009) On asymptotic theory for multivariate GARCH models. *Journal of Multivariate Analysis* 100, 2044–2054.
- Han, H. and D. Kristensen (2014) Asymptotic Theory for the QMLE in GARCH-X Models With Stationary and Nonstationary Covariates. *Journal of Business & Economic Statistics* 32, 416–429.
- Herrndorf, N. (1984) A functional central limit theorem for weakly dependent sequences of random variables. *The Annals of Probability*, 141–153.
- Kenneth S.M. (1981) On the Inverse of the Sum of Matrices. *Mathematics Magazine*. <http://www.jstor.org/stable/2690437>, 54, 67–72
- Koopmans, T.C. and O. Reiersol (1950) The identification of structural characteristics. *The Annals of Mathematical Statistics* 21, 165–181.

- Mezrich, J. and R. F. Engle** (1996) GARCH for groups. *Risk* 9, 36–40.
- Newey, W.K. and K.D. West** (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix. *Econometrica* 55, 703–708.
- Pedersen, R.S. and A. Rahbek** (2014) Multivariate variance targeting in the BEKK-GARCH model. *The Econometrics Journal* 17, 24–55.
- Phillips, P.C.B, Y. Sun and S. Jin** (2003) Consistent HAC estimation and robust regression testing using sharp origin kernels with no truncation. *Discussion paper, Yale University*.
- Silvennoinen, A. and T. Teräsvirta** (2009) Modeling multivariate autoregressive conditional heteroskedasticity with the double smooth transition conditional correlation GARCH model. *Journal of Financial Econometrics* 7, 373–411.

Chapitre 4

Equation by equation estimation of the semi-diagonal BEKK model with covariates

Abstract. This paper provide the asymptotic normality of the Equation by Equation estimator for the semi-diagonal BEKK models augmented by the exogenous variables. The results are obtained without assuming that the innovations are independent, which allows investigate different additional explanatory variables into the information set.

Keywords : BEKK-X, Equation by equation estimation, exogenous variables, covariates, semi-diagonal BEKK-X

4.1 Introduction

Volatility modeling plays a crucial role in financial mathematics, economics and statistics. Understanding volatilities of financial asset returns is important in hedging, risk management, and portfolio optimization. The family of GARCH has been widely used to model and forecast volatility (see [Bollerslev and Wooldridge \(1992\)](#) ; Engle, 2004, for a comprehensive review). In particular, multivariate GARCH models are popular for taking into account financial volatilities co-movements by estimating a conditional covariance matrix. Recent developments in the estimation of multivariate second moment specifications include, among others, the constant conditional correlation (CCC) model of [Bollerslev \(1990\)](#), the BEKK model of [Engle and Kroner \(1995\)](#), the Factor GARCH model of [Engle and Rothschild \(1990\)](#), and the Dynamic Conditional Correlation (DCC) model of [Engle](#)

(2002). One of the challenges in modeling second moments is to ensure the positive definiteness of the resultant covariance matrix without spuriously imposing a temporal pattern in conditional covariances. To cope with this issue, the BEKK model is one of the most flexible GARCH models that guarantees the time-varying covariance matrix to be positive definite by construction. However, the conditional covariance matrix is only explained by the past returns and volatilities while, in practice, many extra information, under the form of exogenous variables, can help explaining and forecasting financial volatility. For example, stock or portfolio covariances may be dependent on own financial characteristics such as liquidity measures, profitability measures, or valuation measures like price-to-book or cash-flow-to-book. In addition, overall market forces such as general credit market conditions might impact conditional covariances. It is, therefore, natural to ask for possible extensions of time series models to accommodate the wealth of information. Engle and Kroner (1995) suggest the BEKK model augmented by exogenous covariates (so-called BEKK-X). However, they only provide the estimation of the BEKK model without impact of the covariates. Thieu (2016) presents the variance targeting estimation (VTE) of the BEKK-X model and establishes the Consistency and Asymptotic Normality (CAN) of the VTE. The BEKK-X model is also related to recent literature on GARCH models extended by additional explanatory variables (GARCH-X) with the aim of explaining and forecasting the volatility. Examples of such GARCH-X models include, amongst other, the heavy model of Shephard and Sheppard (2010), the GARCH-X(1, 1) of Han and Kristensen (2014), the Power ARCH(p, q)-X model of Francq and Thieu (2015), the Multivariate Log-GARCH-X model of Francq and Sucarrat (2015).

Another challenge in modeling a covariance matrix is the "curse of dimensionality", in particular, in the presence of exogenous variables. Indeed, when the dimension of the time series is large, or when the number of covariates is large, the number of parameters can become very large in MGARCH models. The log-likelihood function may therefore contain a numerous number of local maxima, and different starting-values may thus lead to different outcomes. A way of reducing this dimensionality curse is via so-called targeting which imposes a structure on the model intercept based on sample information. Variance targeting estimation is originally proposed by Engle and Mezrich (1996). Pedersen and Rahbek (2014) and Thieu (2015) consider this estimation method for the BEKK model and the BEKK-X model, respectively. Despite the potential benefits of the VT estimation, it remains some difficulties. Enforcing positive definiteness of the conditional covariance matrices the model with targeting implies a set of model constraints that are nonlinear in parameters. This becomes very complicated, except for the scalar case. Moreover, an important question is how much we lose in terms of statistical fit in moving from a BEKK/BEKK-X model to a restricted model with fewer parameters to estimate. Finally, in presence of covariates, the curse of dimensionality is still problematic.

In the present paper, I consider an approach that is so-called equation by equation (EbE) estimation to the semi-diagonal BEKK model with presence of the exogenous variables. This statistical

method is initially proposed by Engle and Sheppard (2001) and Engle (2002) in the context of DCC models. The EbE method deals very effectively with high-dimensional problems and computationally costly log-likelihood. The asymptotic results for the EbE estimator (EbEE) of the volatility parameters, based on Quasi-Maximum Likelihood (QML) are developed in Francq and Zakoïan (2016). In their frame work, they only provide the asymptotic normality of the volatility parameter of the semi-diagonal without covariates. The asymptotic distribution of the estimator of the matrix intercept has not been given. The first goal of the present paper is to establish the strong consistency and the asymptotic distribution for the EbEE of the individual volatilities parameters of the semi-diagonal BEKK-X model. The second goal is to provide asymptotic results for the estimators of all matrix parameters.

The rest of paper is organized as follows. Section 4.2 introduces the BEKK model augmented with exogenous variables and presents the EbE method. The consistency and asymptotic behavior of the EbE estimators are investigated in Section 4.3. Section 4.4 contains the numerical illustrations. Section 4.5 concludes. All auxiliary lemmas and mathematical proofs are contained in Section 4.6.

Some notation and definition throughout the paper : For $m, n \in \mathbb{N}$, I_m denotes the $(n \times n)$ identity matrix and $O_{m \times n}$ denotes the $(m \times n)$ zero matrix. The trace of A is denoted by $tr(A)$, the determinant is denoted by $det(A)$ and the spectral radius of A is denote by $\rho(A)$, *i.e.*, $\rho(A)$ is the maximum among the absolute values of the eigenvalues of A . The operator vec stacks all columns of a matrix into a column vector, $vech$ denotes the operator that stacks only the lower triangular part including the diagonal of a symmetric matrix into a vector, and $vech^0$ is the operator which stacks the sub-diagonal elements (excluding the diagonal) of a matrix. The $(mn \times mn)$ commutation matrix M_{mn} is defined such that, for any $(m \times n)$ matrix A , $M_{mn}vec(A) = vec(A')$. D_m and L_m denote the duplication matrix and elimination matrix defined such that , for any symmetric $(m \times m)$ matrix A , $vec(A) = D_m vech(A)$ and $vech(A) = L_m vec(A)$. Denote T_m be a $m \times m(m+1)/2$ matrix such that $T_m vec(A) = diag(A)$ for any $m \times m$ matrix A . Let also P_m be a $(m(m-1)/2 \times m^2)$ matrix such that $vech^0(A) = P_m vec(A)$, for any symmetric $(m \times m)$ matrix A . The Kronecker product of A and B is defined by $A \otimes B = \{a_{ij}B\}$. The Euclidean norm of the matrix, or vector A , is defined as $\|A\| = \sqrt{tr(A'A)}$, and the spectral norm is defined as $\|A\|_{sp} = \sqrt{\rho(A'A)}$.

4.2 The model and EbE estimation

4.2.1 The model

Let $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \dots, \varepsilon_{mt})'$ denote a $(m \times 1)$ vector of random variables and let $\boldsymbol{x}_t = (x_{1t}, \dots, x_{rt})'$ be a r -dimensional vector of exogenous variables. Assume the existence of the $(m \times m)$ positive

definite matrix \mathbf{H}_t such that

$$E(\boldsymbol{\varepsilon}_t|\mathcal{F}_{t-1}) = \mathbf{0}, \quad E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'|\mathcal{F}_{t-1}) = \mathbf{H}_t, \quad (4.1)$$

where $\mathcal{F}_t = \sigma\{\boldsymbol{\varepsilon}_u, \mathbf{x}'_u; u, u' \leq t\}$ implies the information set at time t . Note that \mathbf{H}_t is the conditional covariance of $\boldsymbol{\varepsilon}_t$ given \mathcal{F}_{t-1} the information set until time $t - 1$.

We consider the following model

$$\begin{cases} \boldsymbol{\varepsilon}_t = \mathbf{H}_t^{1/2}\boldsymbol{\eta}_t \\ \mathbf{H}_t = \boldsymbol{\Omega} + \mathbf{A}\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}'_{t-1}\mathbf{A}' + \mathbf{B}\mathbf{H}_{t-1}\mathbf{B} + \mathbf{C}\mathbf{x}_{t-1}\mathbf{x}'_{t-1}\mathbf{C}' \end{cases} \quad (4.2)$$

where $\mathbf{A} = (a_{kl})_{1 \leq k, \ell \leq m}$, $\mathbf{B} = \text{diag}(b_1, \dots, b_m)$, $\mathbf{C} = (c_{kl})_{1 \leq k \leq m, 1 \leq \ell \leq r}$ and $\boldsymbol{\Omega} = (\omega_{kl})_{1 \leq k, \ell \leq m}$ is a positive definite $(m \times m)$ matrix.

When there is no covariate (4.2) is called the semi-diagonal BEKK (see [Francq and Zakoian \(2016\)](#)). Our model (4.2) can be so-called the semi-diagonal BEKK-X model.

Throughout of the paper, the following assumptions are made

C1 : $E(\boldsymbol{\eta}_t|\mathcal{F}_{t-1}) = \mathbf{0}$, $\text{Var}(\boldsymbol{\eta}_t|\mathcal{F}_{t-1}) = \mathbf{I}_m$.

C2 : $(\boldsymbol{\varepsilon}_t, \mathbf{x}_t)$ is strictly stationary and ergodic process.

C3 : $E(\|\boldsymbol{\varepsilon}_t\|^2) < \infty$ and $E(\|\mathbf{x}_t\|^2) < \infty$.

Remark 4.1. *In many multivariate GARCH models, it is usual to assume that $(\boldsymbol{\eta}_t)$ is an iid white noise vector with zero mean and identity variance matrix. [Comte and Lieberman \(2003\)](#) and [Pedersen and Rahbek \(2014\)](#) establish the CAN of QMLE and VTE, respectively, for the BEKK model under this assumption. [Francq and Zakoian \(2016\)](#) also provide the CAN of the EbEE for the semi-diagonal BEKK without covariate also under the assumption that the innovation process is iid. However, in presence of the exogenous variables, Assumption **C1** is weaker and seems more flexible than the iid assumption as several information sets \mathcal{F}_t with different explanatory variables can be investigated. In univariate case, [Francq and Thieu \(2015\)](#) study the APARCH-X model under an assumption like **C1**.*

Denoting by σ_{kt}^2 the k -th diagonal element of \mathbf{H}_t , that is the variance of the k -th component, ε_{kt} , of $\boldsymbol{\varepsilon}_t$ conditional on \mathcal{F}_{t-1}

$$\text{Var}(\varepsilon_{kt}|\mathcal{F}_{t-1}) = \sigma_{kt}^2.$$

Define \mathbf{D}_t as the diagonal matrix containing the conditional variances σ_{kt}^2 , i.e. $\mathbf{D}_t = \text{diag}(\sigma_{1t}, \dots, \sigma_{mt})$ and let $\boldsymbol{\eta}_t^* = \mathbf{D}_t^{-1}\boldsymbol{\varepsilon}_t$ be the standardized returns. From (4.1), we have $E(\boldsymbol{\eta}_t^*|\mathcal{F}_{t-1}) = \mathbf{0}$ and $\text{Var}(\boldsymbol{\eta}_t^*|\mathcal{F}_{t-1}) = \mathbf{D}_t^{-1}\mathbf{H}_t\mathbf{D}_t^{-1}$. It follows that the components η_{kt}^* of $\boldsymbol{\eta}_t^*$ satisfy, for $k = 1, \dots, m$,

$$E(\eta_{kt}^*|\mathcal{F}_{t-1}) = 0, \quad \text{Var}(\eta_{kt}^*|\mathcal{F}_{t-1}) = 1. \quad (4.3)$$

The individual volatilities are then parameterized as follows

$$\begin{cases} \varepsilon_{kt} = \sigma_{kt} \eta_{kt}^*, \\ \sigma_{kt}^2 = \omega_{kk} + \left(\sum_{\ell=1}^m a_{k\ell} \varepsilon_{\ell,t-1} \right)^2 + b_k^2 \sigma_{k,t-1}^2 + \left(\sum_{s=1}^r c_{ks} x_{s,t-1} \right)^2. \end{cases} \quad (4.4)$$

To ensure the positivity of the volatilities, we assume that $\omega_{kk} > 0$. In view of (4.3), the process (η_t^*) can be called the vector of equation by equation (EbE) innovations of (ε_t) .

Let $\underline{a}_k^0 = (a_{k1}^0, \dots, a_{km}^0)'$ and $\underline{c}_k^0 = (c_{k1}^0, \dots, c_{kr}^0)'$ be, respectively, the k -th row vectors of the matrices \mathbf{A}_0 and \mathbf{C}_0 . Then the vector of unknown parameters involved in the k -th equation (4.4) can be denoted by $\boldsymbol{\theta}_0^{(k)} = (\omega_{kk}^0, \underline{a}_k^{0'}, (b_k^0)^2, \underline{c}_k^{0'})' \in \mathbb{R}^d, d = m + r + 2$. It is clear that an identifiability condition must be required such that σ_{kt}^2 is invariant to a change of sign of the vectors $\underline{a}_k^0, \underline{c}_k^0$ and b_k^0 . Without loss of generality, we can assume that $a_{k1}^0 > 0, b_k^0 > 0$ and $c_{k1}^0 > 0$, for $k = 1, \dots, m$.

Let $\boldsymbol{\theta}^{(k)} = (\omega_{kk}, \underline{a}_k', b_k^2, \underline{c}_k')' = (\omega_{kk}, a_{k1}, \dots, a_{km}, b_k, c_{k1}, \dots, c_{kr})'$ be a generic parameter vector of the parameter space $\Theta^{(k)}$ which is an any compact subset of

$$(0, +\infty)^2 \times \mathbb{R}^{m-1} \times [0, 1) \times (0, +\infty) \times \mathbb{R}^{r-1}.$$

4.2.2 Equation-by-equation estimation of parameters

Let $\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_n$ be observations of a process satisfying the semi-diagonal BEKK-X representation (4.2) and $\mathbf{x}_1, \dots, \mathbf{x}_n$ be observations of a process of the explanatory variables. For all $\boldsymbol{\theta}^{(k)} \in \Theta^{(k)}$, we recursively define $\tilde{\sigma}_{kt}^2(\boldsymbol{\theta}^{(k)})$ for $t = 1, \dots, n$ by

$$\tilde{\sigma}_{kt}^2(\boldsymbol{\theta}^{(k)}) = \omega_{kk} + \left(\sum_{\ell=1}^m a_{k\ell} \varepsilon_{\ell,t-1} \right)^2 + b_k^2 \tilde{\sigma}_{k,t-1}^2(\boldsymbol{\theta}^{(k)}) + \left(\sum_{s=1}^r c_{ks} x_{s,t-1} \right)^2 \quad (4.5)$$

with the arbitrary initial values $\tilde{\varepsilon}_0, \tilde{\sigma}_0$ and \mathbf{x}_0 . Let

$$\tilde{Q}_n^{(k)}(\boldsymbol{\theta}^{(k)}) = \frac{1}{n} \sum_{t=1}^n \tilde{\ell}_{kt}(\boldsymbol{\theta}^{(k)}), \quad \tilde{\ell}_{kt}(\boldsymbol{\theta}^{(k)}) = \log \tilde{\sigma}_{kt}^2(\boldsymbol{\theta}^{(k)}) + \frac{\varepsilon_{kt}^2}{\tilde{\sigma}_{kt}^2(\boldsymbol{\theta}^{(k)})}.$$

The EbE estimator, denoted by $\hat{\boldsymbol{\theta}}_n^{(k)}$, of the true parameter vector $\boldsymbol{\theta}_0^{(k)}$ is defined as a measurable solution of the following equation

$$\hat{\boldsymbol{\theta}}_n^{(k)} = \arg \min_{\boldsymbol{\theta}^{(k)} \in \Theta^{(k)}} \tilde{Q}_n^{(k)}(\boldsymbol{\theta}^{(k)}). \quad (4.6)$$

Let $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}_0^{(1)'}, \dots, \boldsymbol{\theta}_0^{(m)'})'$. Note that $\boldsymbol{\theta}_0$ includes the diagonal elements of $\boldsymbol{\Omega}_0$ and all components of the matrices $\mathbf{A}_0, \mathbf{B}_0$ and \mathbf{C}_0 . This parameter vector belongs to the parameter space $\Theta^m =$

$\Theta^{(1)} \times \dots \times \Theta^{(m)}$, whose generic element is denoted by $\boldsymbol{\theta} = (\boldsymbol{\theta}^{(1)'}, \dots, \boldsymbol{\theta}^{(m)'})'$. The estimator of $\boldsymbol{\theta}_0$ is given by $\widehat{\boldsymbol{\theta}}_n = (\widehat{\boldsymbol{\theta}}_n^{(1)'}, \dots, \widehat{\boldsymbol{\theta}}_n^{(m)'})'$ which is the collection of the equation by equation estimators.

Once the EbEE estimators $\widehat{\mathbf{A}}_n$, $\widehat{\mathbf{B}}_n$ and $\widehat{\mathbf{C}}_n$ of the matrices \mathbf{A}_0 , \mathbf{B}_0 and \mathbf{C}_0 , respectively, are obtained, the matrix $\boldsymbol{\Omega}_0$ can be fully estimated as follows

$$vech^0(\widehat{\boldsymbol{\Omega}}_n) = vech^0\left(\widehat{\Sigma}_{\varepsilon n} - \widehat{\mathbf{A}}_n \widehat{\Sigma}_{\varepsilon n} \widehat{\mathbf{A}}_n' - \widehat{\mathbf{B}}_n \widehat{\Sigma}_{\varepsilon n} \widehat{\mathbf{B}}_n' - \widehat{\mathbf{C}}_n \widehat{\Sigma}_{\mathbf{x}n} \widehat{\mathbf{C}}_n'\right), \quad (4.7)$$

where $\widehat{\Sigma}_{\varepsilon n} = \frac{1}{n} \sum_{t=1}^n \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'$ and $\widehat{\Sigma}_{\mathbf{x}n} = \frac{1}{n} \sum_{t=1}^n \mathbf{x}_t \mathbf{x}_t'$ are the empirical estimators of the second order moment matrices $\Sigma_{\varepsilon} = E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t')$ and $\Sigma_{\mathbf{x}} = E(\mathbf{x}_t \mathbf{x}_t')$, respectively. The estimation of the model (4.2) is thus nothing else than the estimation of

$$\boldsymbol{\vartheta}_0 = (\boldsymbol{\theta}'_0, \boldsymbol{\gamma}'_{\varepsilon 0}, \boldsymbol{\gamma}'_{\mathbf{x}0})', \quad \boldsymbol{\gamma}_{\varepsilon 0} = vech(\Sigma_{\varepsilon}), \quad \boldsymbol{\gamma}_{\mathbf{x}0} = vech(\Sigma_{\mathbf{x}}).$$

Its estimator can be given by $\widehat{\boldsymbol{\vartheta}}_n = (\widehat{\boldsymbol{\theta}}_n', \widehat{\boldsymbol{\gamma}}_{\varepsilon n}', \widehat{\boldsymbol{\gamma}}_{\mathbf{x}n}')'$, where $\widehat{\boldsymbol{\gamma}}_{\varepsilon n} = vech(\widehat{\Sigma}_{\varepsilon n})$ and $\widehat{\boldsymbol{\gamma}}_{\mathbf{x}n} = vech(\widehat{\Sigma}_{\mathbf{x}n})$.

4.3 EbE estimation inference

For the consistency of the estimator, the assumptions following will be made

C4 : $\boldsymbol{\theta}_0^{(k)} \in \Theta^{(k)}$, $\Theta^{(k)}$ is compact, for $k = 1, \dots, m$.

C5 : $\rho(\mathbf{A}_0 \otimes \mathbf{A}_0 + \mathbf{B}_0 \otimes \mathbf{B}_0) < 1$ and $\sum_{k=1}^m b_k^2 < 1$, for all $\boldsymbol{\theta}^{(k)} \in \Theta^{(k)}$.

C6 : There exists $s > 0$ such that $E|\varepsilon_{kt}|^s < \infty$ and $E|x_{kt}|^s < \infty$.

C7 : For all $\ell^* = 1, \dots, m$, $\varepsilon_{\ell^* t}^2$ does not belong to the Hilbert space generated by the linear combinations of the $\varepsilon_{\ell u} \varepsilon_{\ell' u}$'s, the $x_{s v} x_{s' v}$'s for $u < t$, $v \leq t$, $\ell, \ell' = 1, \dots, m$, $s, s' = 1, \dots, r$ and the $\varepsilon_{\ell t} \varepsilon_{\ell' t}$ for $(\ell, \ell') \neq (\ell^*, \ell^*)$.

C8 : For all $s^* = 1, \dots, r$, $x_{s^* t}^2$ does not belong to the Hilbert space generated by the linear combinations of the the $x_{s v} x_{s' v}$'s for $v < t$, $s, s' = 1, \dots, r$ and the $x_{s t} x_{s' t}$ for $(s, s') \neq (s^*, s^*)$.

Remark 4.2. Assumptions **C7** and **C8** are identification conditions. For simplicity, let us consider (4.2) when $m = 2$, $r = 2$ and the conditional covariance matrix is given by

$$\mathbf{H}_t = \boldsymbol{\Omega} + \mathbf{A} \boldsymbol{\varepsilon}_{t-1} \boldsymbol{\varepsilon}_{t-1}' \mathbf{A}' + \mathbf{C} \mathbf{x}_{t-1} \mathbf{x}_{t-1}' \mathbf{C}', \quad (4.8)$$

where $\boldsymbol{\Omega} = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{pmatrix}$ is a symmetric positive definite matrix, $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ and $\mathbf{C} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}$, with $a_{11} > 0$, $a_{21} > 0$, $c_{11} > 0$ and $c_{21} > 0$. The volatility of the first component, ε_{1t} ,

of ε_t is thus given by

$$\sigma_{1t}^2 = \omega_{11} + a_{11}\varepsilon_{1,t-1}^2 + 2a_{11}a_{12}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + a_{12}\varepsilon_{2,t-1}^2 + c_{11}x_{1,t-1}^2 + 2c_{11}c_{12}x_{1,t-1}x_{2,t-1} + c_{12}x_{2,t-1}^2.$$

Assumption C7 precludes, for example, that $x_{1,t-1} = \varepsilon_{1,t-1}$ for which the model is not identifiable.

Similarly, Assumption C8 rules out the existence of linear combination of a finite number of the $x_{s,t-i}x_{s',t-i}$ that is obviously necessary for the identifiability.

Theorem 4.1. Under C1 - C8, the EbEE of $\theta_0^{(k)}$ is strongly consistent

$$\widehat{\theta}_n^{(k)} \rightarrow \theta_0^{(k)}, \text{ a.s. as } n \rightarrow \infty.$$

The following result is an immediate consequence of Theorem 4.1.

Corollary 4.1. Under the assumptions of Theorem 4.1, $\widehat{\vartheta}_n$ is a strongly consistent estimator of ϑ_0 .

Now we turn to the asymptotic distribution of the estimation. We need some additional assumptions.

C9 : $\theta_0^{(k)}$ belongs to the interior of the parameter space $\Theta^{(k)}$, for $k = 1, \dots, m$.

C10 : $E\|\eta_t\|^{4(1+\delta)} < \infty$, $E\|\varepsilon_t\|^{4(1+1/\delta)} < \infty$ and $E\|\mathbf{x}_t\|^{4(1+1/\delta)} < \infty$ for some $\delta > 0$.

C11 : The process $\mathbf{z}_t = (\mathbf{x}'_t, \varepsilon'_t, \boldsymbol{\eta}'_t)'$ satisfies

$$E\|\varepsilon_t\|^{(4+2\nu)(1+1/\delta)} < \infty, E\|\mathbf{x}_t\|^{(4+2\nu)(1+1/\delta)} < \infty \text{ and } E\|\boldsymbol{\eta}_t\|^{(4+2\nu)(1+\delta)} < \infty,$$

moreover the strong mixing coefficients, $\alpha_{\mathbf{z}}(h)$, of the process (\mathbf{z}_t) are such that

$$\sum_{h=0}^{\infty} \{\alpha_{\mathbf{z}}(h)\}^{\nu/(2+\nu)} < \infty \text{ for some } \nu > 0 \text{ and } \delta > 0.$$

Let $\underline{\mathbf{H}}_{t,s}(\boldsymbol{\vartheta})$ be such that, for $s > 0$,

$$\text{vec}(\underline{\mathbf{H}}_{t,s}(\boldsymbol{\vartheta})) = \sum_{k=0}^s (\mathbf{B}^{\otimes 2})^k (\text{vec}(\boldsymbol{\Omega}) + \mathbf{A}^{\otimes 2} \text{vec}(\varepsilon_{t-k-1} \varepsilon'_{t-k-1}) + \mathbf{C}^{\otimes 2} \text{vec}(\mathbf{x}_{t-k-1} \mathbf{x}'_{t-k-1})),$$

where $A^{\otimes 2}$ denotes the Kronecker product of a matrix A and itself. Let also \mathcal{S} be a subspace such that for all $\boldsymbol{\vartheta} \in \Theta$, $\mathbf{H}_t(\boldsymbol{\vartheta}) \in \mathcal{S}$ and for all $s > 0$, $\underline{\mathbf{H}}_{t,s}(\boldsymbol{\vartheta}) \in \mathcal{S}$.

C12 : There exists $K > 0$ such that

$$\left\| \mathbf{H}_t^{1/2}(\boldsymbol{\vartheta}) - \mathbf{H}_t^{*1/2}(\boldsymbol{\vartheta}) \right\| \leq K \|\mathbf{H}_t(\boldsymbol{\vartheta}) - \mathbf{H}_t^*(\boldsymbol{\vartheta})\| \quad \text{for all } \mathbf{H}_t(\boldsymbol{\vartheta}), \mathbf{H}_t^*(\boldsymbol{\vartheta}) \in \mathcal{S}.$$

Remark 4.3. Assumptions C11 and C12 are also required in Thieu (2016) for the BEKK-X model estimated by the variance targeting method.

In order to state the asymptotic normality, we have to introduce the following notations.

Let the $(d \times d)$ matrices $\mathbf{J}_{ks} = E(\Delta_{kt}\Delta'_{st})$, $k, s = 1, \dots, m$, where $\Delta_{kt} = \frac{1}{\sigma_{kt}^2} \frac{\partial \sigma_{kt}^2(\boldsymbol{\theta}_0^{(k)})}{\partial \boldsymbol{\theta}^{(k)}}$ and let $\mathbf{J} = \text{diag}\{\mathbf{J}_{11}, \dots, \mathbf{J}_{mm}\}$ in bloc-matrix notation. Let also $\boldsymbol{\Delta}_t = \text{diag}(\Delta_{1t}, \dots, \Delta_{mt})$, $\mathbf{N}_1 = L_m(I_{m^2} - \mathbf{A}_0^{\otimes 2} - \mathbf{B}_0^{\otimes 2})^{-1}(I_{m^2} - \mathbf{B}_0^{\otimes 2})$ and $\mathbf{N}_2 = L_m(I_{m^2} - \mathbf{A}_0^{\otimes 2} - \mathbf{B}_0^{\otimes 2})^{-1}\mathbf{C}_0^{\otimes 2}D_r$. We also define the following matrices

$$\boldsymbol{\Sigma}_{11} = \sum_{h=-\infty}^{\infty} \text{cov}(\boldsymbol{\Upsilon}_{0t} \text{vec}(\boldsymbol{\eta}_t \boldsymbol{\eta}'_t), \boldsymbol{\Upsilon}_{0,t-h} \text{vec}(\boldsymbol{\eta}_{t-h} \boldsymbol{\eta}'_{t-h})), \quad (4.9)$$

$$\boldsymbol{\Sigma}_{22} = \sum_{h=-\infty}^{\infty} \text{cov}(\text{vech}(\mathbf{x}_t \mathbf{x}'_t), \text{vech}(\mathbf{x}_{t-h} \mathbf{x}'_{t-h})), \quad (4.10)$$

$$\boldsymbol{\Sigma}_{12} = \sum_{h=-\infty}^{\infty} \text{cov}(\text{vech}(\mathbf{x}_t \mathbf{x}'_t), \boldsymbol{\Upsilon}_{0,t-h} \text{vec}(\boldsymbol{\eta}_{t-h} \boldsymbol{\eta}'_{t-h})), \quad (4.11)$$

where

$$\boldsymbol{\Upsilon}_{0t} = \begin{pmatrix} \boldsymbol{\Delta}_t I_m \left(\mathbf{D}_{0t}^{-1} \mathbf{H}_{0t}^{1/2} \right) \otimes \left(\mathbf{D}_{0t}^{-1} \mathbf{H}_{0t}^{1/2} \right) \\ \mathbf{H}_{0t}^{1/2} \otimes \mathbf{H}_{0t}^{1/2} \end{pmatrix}. \quad (4.12)$$

Theorem 4.2. Under C1 - C12, as $n \rightarrow \infty$,

$$\sqrt{n} \begin{pmatrix} \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \\ \widehat{\boldsymbol{\gamma}}_{\varepsilon n} - \boldsymbol{\gamma}_{\varepsilon 0} \\ \widehat{\boldsymbol{\gamma}}_{\mathbf{x}n} - \boldsymbol{\gamma}_{\mathbf{x}0} \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Gamma} \boldsymbol{\Sigma} \boldsymbol{\Gamma}'), \quad (4.13)$$

where

$$\boldsymbol{\Gamma} = \begin{pmatrix} -\mathbf{J}^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{N}_1 & \mathbf{N}_2 \\ \mathbf{0} & \mathbf{0} & I_{r(r+1)/2} \end{pmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}'_{12} & \boldsymbol{\Sigma}_{22} \end{pmatrix}. \quad (4.14)$$

The original parameter vector is denoted by

$$\boldsymbol{\xi}_0 = ((\text{vech}^0(\boldsymbol{\Omega}_0))', \boldsymbol{\theta}'_0)' \in \mathbb{R}^{m(m-1)/2+md}. \quad (4.15)$$

The estimator of $\boldsymbol{\xi}_0$ can be given by $\widehat{\boldsymbol{\xi}}_n = (\widehat{\boldsymbol{\omega}}'_n, \widehat{\boldsymbol{\theta}}'_n)'$, where $\widehat{\boldsymbol{\omega}}_n = \text{vech}^0(\widehat{\boldsymbol{\Omega}}_n)$ is the estimator of $\boldsymbol{\omega}_0 = \text{vech}^0(\boldsymbol{\Omega}_0)$.

The strong consistency and the asymptotic distribution of the estimation is given as follows

Theorem 4.3. *If C1 - C8 hold, the estimator $\widehat{\boldsymbol{\xi}}_n$ of $\boldsymbol{\xi}_0$ is strongly consistent :*

$$\widehat{\boldsymbol{\xi}}_n \rightarrow \boldsymbol{\xi}_0 \text{ a.s. as } n \rightarrow \infty.$$

If, in addition, C9 - C12 hold, then

$$\sqrt{n} \begin{pmatrix} \widehat{\boldsymbol{\omega}}_n - \boldsymbol{\omega}_0 \\ \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Omega} \boldsymbol{\Sigma} \boldsymbol{\Omega}'), \quad (4.16)$$

where

$$\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_1 \\ \boldsymbol{\Omega}_2 \end{pmatrix}, \quad \begin{aligned} \boldsymbol{\Omega}_1 &= \begin{pmatrix} \mathbf{A}^* & \mathbf{B}^* & \mathbf{C}^* & \mathbf{E}^* & \mathbf{X}^* \end{pmatrix} \boldsymbol{\Psi}, \\ \boldsymbol{\Omega}_2 &= \begin{pmatrix} I_{md} & \mathbf{0}_{m(m+1)/2} & \mathbf{0}_{r(r+1)/2} \end{pmatrix}, \end{aligned} \quad (4.17)$$

with

$$\begin{aligned} \mathbf{A}^* &= -P_m \{ I_m \otimes (\mathbf{A}_0 \boldsymbol{\Sigma}_\varepsilon) + ((\mathbf{A}_0 \boldsymbol{\Sigma}_\varepsilon) \otimes I_m) M_{mm} \}, \\ \mathbf{B}^* &= -P_m \{ I_m \otimes (\mathbf{B}_0 \boldsymbol{\Sigma}_\varepsilon) + (\mathbf{B}_0 \boldsymbol{\Sigma}_\varepsilon) \otimes I_m \}, \\ \mathbf{C}^* &= -P_m \{ I_m \otimes (\mathbf{C}_0 \boldsymbol{\Sigma}_x) + ((\mathbf{C}_0 \boldsymbol{\Sigma}_x) \otimes I_m) M_{mr} \}, \\ \mathbf{E}^* &= P_m (I_{m^2} - \mathbf{B}_0 \otimes \mathbf{B}_0 - \mathbf{A}_0 \otimes \mathbf{A}_0) D_m, \quad \mathbf{X}^* = -P_m (\mathbf{C}_0 \otimes \mathbf{C}_0) D_r. \end{aligned}$$

4.4 Numerical illustrations

This section presents the results of Monte Carlo simulations studies aimed at examining the performance of the EbEE of the semi-diagonal BEKK-X model and comparing the finite sample properties of the EbEE and VTE.

First, the quality of the EbEE will be evaluated through a Monte Carlo experiment. To keep the computation burden feasible, we focus on the bivariate case, $m = 2$, but the results should generalize in an obvious way to higher dimensions. The vector of the exogenous variables is $\mathbf{x}_t = (\mathbf{x}_{1t}, \mathbf{x}_{2t})' = (\mathbf{z}_{t-1}, \mathbf{z}_{t-2})'$, where \mathbf{z}_{t-1} and \mathbf{z}_{t-2} are two lagged values of an APARCH(1,1)

$$\mathbf{z}_t = \sigma_t e_t, \quad \sigma_t = 0.046 + 0.027 \mathbf{z}_{t-1}^+ + 0.092 \mathbf{z}_{t-1}^- + 0.843 \sigma_{t-1}, \quad (4.18)$$

where $\sqrt{2}e_t$ is i.i.d and follows a Student distribution with 4 degrees of freedom. As discussed in Remark 4.1, the $\boldsymbol{\eta}_t$'s do not need to be iid with zero mean and identity variance matrix. In this experiment, the $\boldsymbol{\eta}_t$'s are assumed to follow a bivariate Student distribution with $(4 + |\mathbf{x}_{1,t-1}|)$

degrees of freedom. The BEKK-X parameter are taken as follows

$$\mathbf{\Omega}_0 = \begin{pmatrix} 0.3 & 0.2 \\ 0.2 & 0.4 \end{pmatrix}, \mathbf{A}_0 = \begin{pmatrix} 0.15 & 0.1 \\ 0.1 & 0.2 \end{pmatrix}, \mathbf{B}_0 = \begin{pmatrix} 0.8 & 0.0 \\ 0.0 & 0.9 \end{pmatrix}, \mathbf{C}_0 = \begin{pmatrix} 0.15 & 0.05 \\ 0.1 & 0.2 \end{pmatrix}. \quad (4.19)$$

The data series are generated 500 times for $n = 1000$ and $n = 5000$ observations. For each data series, $(n + 500)$ observations of ε_t are simulated and then the first 500 observations are discarded in each simulation to minimize the effect of the initial values. The results of the simulation study are presented in Table 4.2. They are in accordance with the consistency of the EbEE, in particular the medians of the estimated parameters are close to the true values. As expected, the accuracy of the estimation increases as the sample size increases from $n = 1000$ to $n = 5000$.

Figures 4.1, 4.2, 4.3 and 4.4 show non parametric estimators of the density of the components of $\hat{\vartheta}_n - \vartheta_0$. As expected, the estimated densities of the estimators over the 500 simulations are close to a Gaussian density for n sufficiently large.

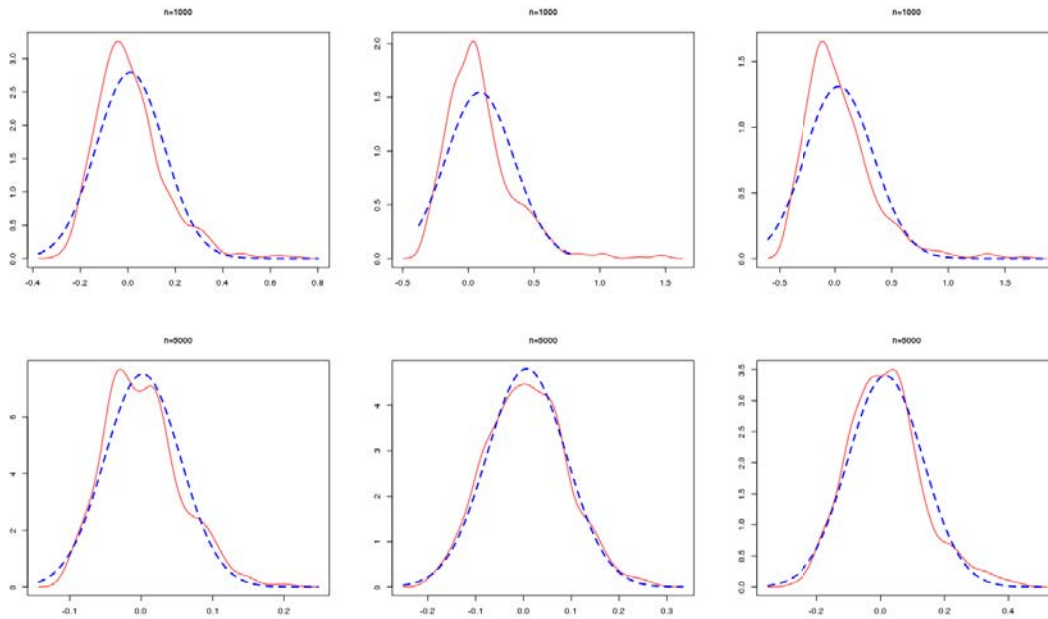


FIGURE 4.1 – Kernel density estimator (in full line) of the distribution of the EbEE errors for the estimation of the parameters involved $vech(\mathbf{\Omega})$.

Next, a Monte Carlo experiment is performed with the aim to compare the empirical accuracies of the EbEE and VTE. The true parameter matrices are taken as same as in the previous experiment. η_t are assumed to be independent and normally distributed $\mathcal{N}(0, 1)$. Table 4.2 summarizes the distributions of the two estimators over 500 independent simulations of the model, for the length

TABLE 4.1 – Sampling distribution of the EbEE of ϑ_0 over 500 replications for the semi-diagonal BEKK(1,1)-X model

parameter	true val.	bias	RMSE	min	Q_1	Q_2	Q_3	max
$n = 1,000$								
$vec(\mathbf{\Omega})$	0.30	0.0100	0.1427	0.0228	0.2136	0.2880	0.3826	1.0031
	0.20	0.0872	0.2576	-0.1338	0.1158	0.2442	0.3916	1.6704
	0.40	0.0328	0.3037	0.0000	0.2268	0.3775	0.5790	2.0654
\mathbf{A}	0.15	-0.0504	0.1244	-0.3250	0.0269	0.1158	0.1910	0.3889
	0.10	-0.0207	0.1871	-0.6000	-0.0121	0.1020	0.2039	0.6000
	0.10	-0.0210	0.0864	-0.2045	0.0552	0.0985	0.1294	0.2646
$diag(\mathbf{B})$	0.20	-0.0241	0.1197	-0.4465	0.1289	0.1890	0.2535	0.4824
	0.80	-0.0106	0.0617	0.5064	0.7577	0.7968	0.8287	0.9490
	0.90	-0.0083	0.0371	0.6658	0.8735	0.8954	0.9153	0.9765
\mathbf{C}	0.15	-0.0023	0.0289	-0.1420	0.1341	0.1494	0.1635	0.2115
	0.10	0.0081	0.0698	-0.2306	0.0618	0.1080	0.1576	0.3110
	0.05	-0.0008	0.0273	-0.0432	0.0316	0.0482	0.0656	0.1575
	0.20	-0.0162	0.0594	-0.2287	0.1533	0.1927	0.2201	0.3573
$n = 5,000$								
$vec(\mathbf{\Omega})$	0.30	0.0020	0.0530	0.1959	0.2631	0.2966	0.3308	0.5083
	0.20	0.0066	0.0829	0.0136	0.1479	0.2076	0.2622	0.4684
	0.40	0.0158	0.1171	0.1270	0.3354	0.4094	0.4790	0.8367
\mathbf{A}	0.15	-0.0007	0.0393	0.0004	0.1228	0.1515	0.1749	0.2602
	0.10	0.0049	0.0766	0.2135	0.0557	0.1036	0.1620	0.3341
	0.10	-0.0007	0.0209	0.0390	0.0856	0.0995	0.1128	0.1635
$diag(\mathbf{B})$	0.20	-0.0044	0.0392	0.0828	0.1711	0.1950	0.2244	0.3111
	0.80	-0.0025	0.0234	0.7033	0.7845	0.7990	0.8138	0.8536
	0.90	-0.0023	0.0133	0.8466	0.8898	0.8986	0.9063	0.9353
\mathbf{C}	0.15	0.0006	0.0099	0.1246	0.1444	0.1503	0.1569	0.1787
	0.10	0.0034	0.0310	0.0259	0.0820	0.1016	0.1217	0.2242
	0.05	-0.0005	0.0110	0.0166	0.0418	0.0491	0.0571	0.0803
	0.20	-0.0036	0.0234	0.0955	0.1844	0.1990	0.2113	0.2597

RMSE is the Root Mean Square Error, Q_i , $i = 1, 3$, denote the quartiles.

$n = 1000$, $n = 3000$ and $n = 5000$. From these results, the biases of the VTE are smaller than the one of the EbEE for $n = 1000$, $n = 3000$ and $n = 5000$. But the differences are tiny when n is sufficiently large, for example $n = 3000$ or $n = 5000$. Thus, the EbE approach does not seem to create a serious bias problem in the estimation of the dynamic parameters. Figure 4.5 displays

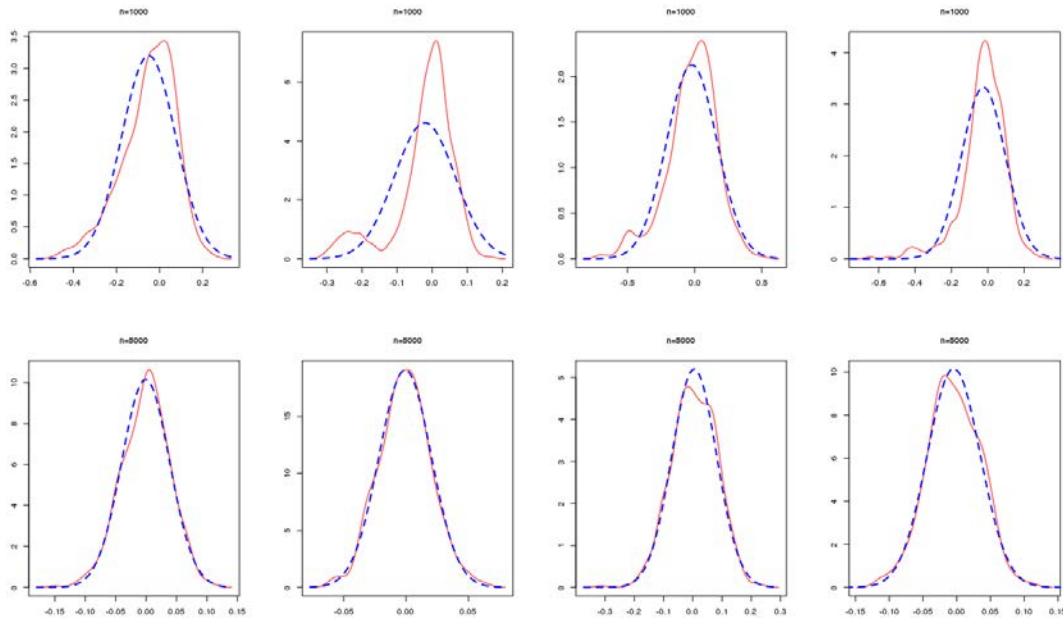


FIGURE 4.2 – Kernel density estimator (in full line) of the distribution of the EbEE errors for the estimation of the parameters involved \mathbf{A} .

the distribution of the estimation errors for simulations of length $n = 5000$. The upper-left, upper-right, bottom-left and bottom-right panels correspond respectively to the estimation errors for the parameters involved in $\text{vech}(\mathbf{\Omega})$, \mathbf{A} , $\text{diag}(\mathbf{B})$ and \mathbf{C} . The distributions of the EbEE and VTE are quite similar.

4.5 Conclusion

This paper suggest an effective approach, equation by equation estimation method, for estimation and inference on the semi-diagonal BEKK included the exogenous variables. This approach is recently widely used by researchers and practitioners, consisting in estimating separately the individual volatilities parameters in the first step, and estimating the remaining parameters of the intercept matrix in the second step. The strong consistency of the EbEE is showed under the assumptions that are weaker than the ones for the VTE of the same model. Under the mixing-conditions, the asymptotic distribution of the estimators of the parameters is normal. The main motivation for using the EbE method in application is the important gains in computational time and our experiments show that the reduction of computational time compared to the VT estimation can be effective.

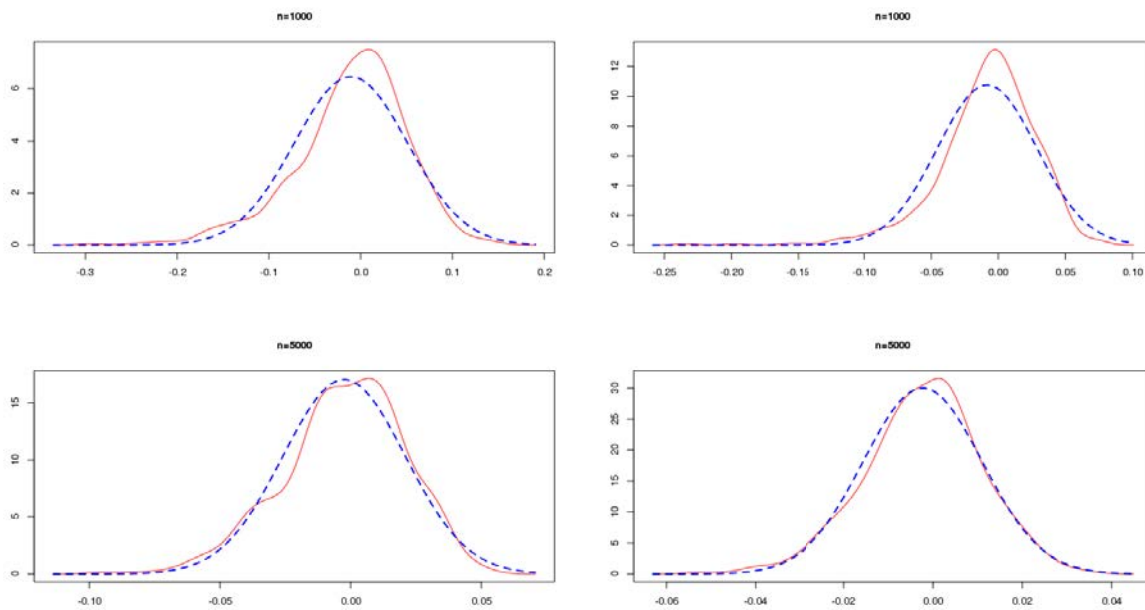


FIGURE 4.3 – Kernel density estimator (in full line) of the distribution of the EbEE errors for the estimation of the parameters involved $diag(\mathbf{B})$.

Acknowledgements

The author is very grateful to Professor Christian Francq for valuable comments and encouragement.

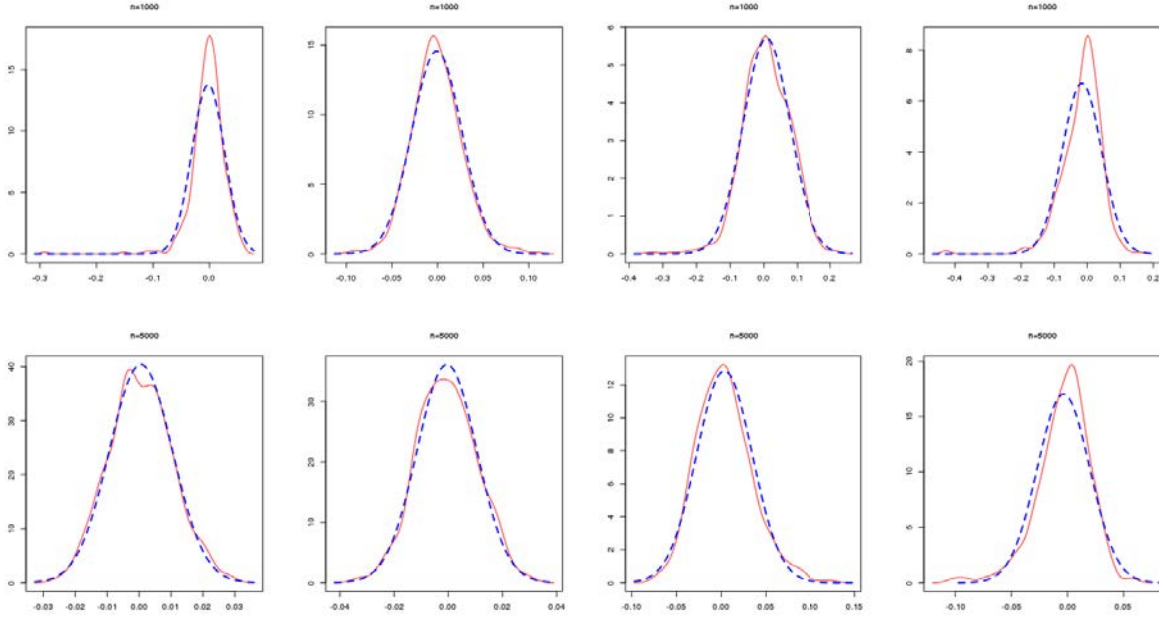


FIGURE 4.4 – Kernel density estimator (in full line) of the distribution of the EbEE errors for the estimation of the parameters involved \mathbf{C} .

4.6 Proofs

Proof of Theorem 4.1.

Let

$$Q_n^{(k)}(\boldsymbol{\theta}^{(k)}) = \frac{1}{n} \sum_{t=1}^n \ell_{kt}(\boldsymbol{\theta}^{(k)}), \quad \ell_{kt}(\boldsymbol{\theta}^{(k)}) = \log \sigma_{kt}^2(\boldsymbol{\theta}^{(k)}) + \frac{\varepsilon_{kt}^2}{\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})}.$$

To prove the theorem of the consistency of the EbEE, the following results have to be shown

- i) $\lim_{n \rightarrow \infty} \sup_{\boldsymbol{\theta}^{(k)} \in \Theta^{(k)}} \left| Q_n^{(k)}(\boldsymbol{\theta}^{(k)}) - \tilde{Q}_n^{(k)}(\boldsymbol{\theta}^{(k)}) \right| = 0$ a.s.
- ii) $\sigma_{kt}^2(\boldsymbol{\theta}_0^{(k)}) = \sigma_{kt}^2(\boldsymbol{\theta}^{(k)})$ a.s. iff $\boldsymbol{\theta}_0^{(k)} = \boldsymbol{\theta}^{(k)}$.
- iii) $E\ell_{kt}(\boldsymbol{\theta}_0^{(k)}) < \infty$ and if $\boldsymbol{\theta}^{(k)} \neq \boldsymbol{\theta}_0^{(k)}$ then $E\ell_{kt}(\boldsymbol{\theta}_0^{(k)}) < E\ell_{kt}(\boldsymbol{\theta}^{(k)})$.
- iv) There exists a neighborhood $\mathcal{V}(\boldsymbol{\theta}^{(k)})$ of any $\boldsymbol{\theta}^{(k)} \neq \boldsymbol{\theta}_0^{(k)}$ such that

$$\liminf_{n \rightarrow \infty} \inf_{\boldsymbol{\theta}^* \in \mathcal{V}(\boldsymbol{\theta}^{(k)})} \tilde{Q}_n^{(k)}(\boldsymbol{\theta}^*) > \limsup_{n \rightarrow \infty} \tilde{Q}_n^{(k)}(\boldsymbol{\theta}_0^{(k)}), \text{ a.s.}$$

The condition $\sum_{k=1}^m b_k^2 < 1$ of Assumption **C5** and the compactness of $\Theta^{(k)}$ imply that

$$\sup_{\boldsymbol{\theta}^{(k)} \in \Theta^{(k)}} b_k^2 < 1. \quad (4.20)$$

TABLE 4.2 – Sampling distribution of the EbEE and VTE of ϑ_0 over 500 replications for the BEKK-X(1,1) model

	true	estim.	Median	bias	RMSE	Median	bias	RMSE	Median	bias	RMSE
			$n = 1,000$			$n = 3,000$			$n = 5,000$		
Ω	0.30	EbEE	0.293	0.009	0.102	0.299	0.005	0.057	0.300	0.001	0.044
		VTE	0.295	-0.002	0.094	0.297	0.003	0.048	0.300	0.002	0.039
	0.20	EbEE	0.240	0.059	0.198	0.201	0.007	0.094	0.202	0.004	0.068
		VTE	0.199	0.005	0.105	0.202	0.005	0.051	0.201	0.003	0.039
	0.40	EbEE	0.401	0.031	0.224	0.403	0.013	0.107	0.399	0.008	0.085
		VTE	0.404	0.010	0.204	0.406	0.008	0.090	0.401	0.005	0.069
A	0.15	EbEE	0.140	-0.025	0.110	0.151	-0.001	0.047	0.149	-0.003	0.034
		VTE	0.142	-0.010	0.079	0.150	0.000	0.042	0.148	-0.003	0.032
	0.10	EbEE	0.094	-0.012	0.168	0.108	0.008	0.083	0.097	0.000	0.059
		VTE	0.109	0.012	0.140	0.109	0.008	0.074	0.100	0.001	0.054
	0.10	EbEE	0.097	-0.014	0.066	0.098	-0.003	0.029	0.100	0.001	0.018
		VTE	0.100	-0.001	0.040	0.098	-0.002	0.024	0.101	0.001	0.017
0.20	EbEE	0.194	-0.030	0.108	0.195	-0.008	0.050	0.199	-0.003	0.033	
	VTE	0.191	-0.014	0.077	0.195	-0.006	0.040	0.199	-0.002	0.031	
B	0.80	EbEE	0.799	-0.005	0.042	0.800	-0.002	0.023	0.799	-0.001	0.012
		VTE	0.798	-0.001	0.035	0.799	-0.001	0.018	0.800	-0.001	0.015
0.90	EbEE	0.899	-0.004	0.029	0.898	-0.002	0.013	0.899	-0.001	0.010	
	VTE	0.900	-0.003	0.023	0.900	-0.001	0.010	0.900	-0.001	0.008	
C	0.15	EbEE	0.150	-0.002	0.023	0.150	0.000	0.010	0.150	0.000	0.008
		VTE	0.149	-0.001	0.015	0.149	0.000	0.009	0.150	0.000	0.007
	0.10	EbEE	0.105	0.008	0.053	0.101	0.004	0.032	0.102	0.003	0.025
		VTE	0.101	0.000	0.027	0.100	0.000	0.016	0.100	0.000	0.012
	0.05	EbEE	0.050	0.000	0.025	0.050	0.000	0.023	0.051	0.000	0.010
		VTE	0.051	0.001	0.016	0.050	0.000	0.008	0.050	0.000	0.007
0.20	EbEE	0.195	-0.011	0.039	0.200	-0.003	0.023	0.199	-0.002	0.017	
	VTE	0.199	-0.001	0.025	0.200	0.001	0.013	0.200	0.000	0.011	

RMSE is the Root Mean Square Error.

By a simple recursion, we have

$$\sup_{\boldsymbol{\theta}^{(k)} \in \Theta^{(k)}} \left| \tilde{\sigma}_{kt}^2(\boldsymbol{\theta}^{(k)}) - \sigma_{kt}^2(\boldsymbol{\theta}^{(k)}) \right| = \sup_{\boldsymbol{\theta}^{(k)} \in \Theta^{(k)}} (b_k^2)^t \left| \tilde{\sigma}_{k0}^2(\boldsymbol{\theta}^{(k)}) - \sigma_{k0}^2(\boldsymbol{\theta}^{(k)}) \right| < K\rho^t \text{ a.s.}$$

where, here and the sequel of the paper, K and ρ denote generic constants whose the exact values are not important and $0 < \rho < 1$.

The proofs of i), iii) and i) are very similar to the ones given in [Francq and Zakoïan \(2010\)](#) for the standard GARCH without covariates and are omitted. Here I only thus show the point ii).

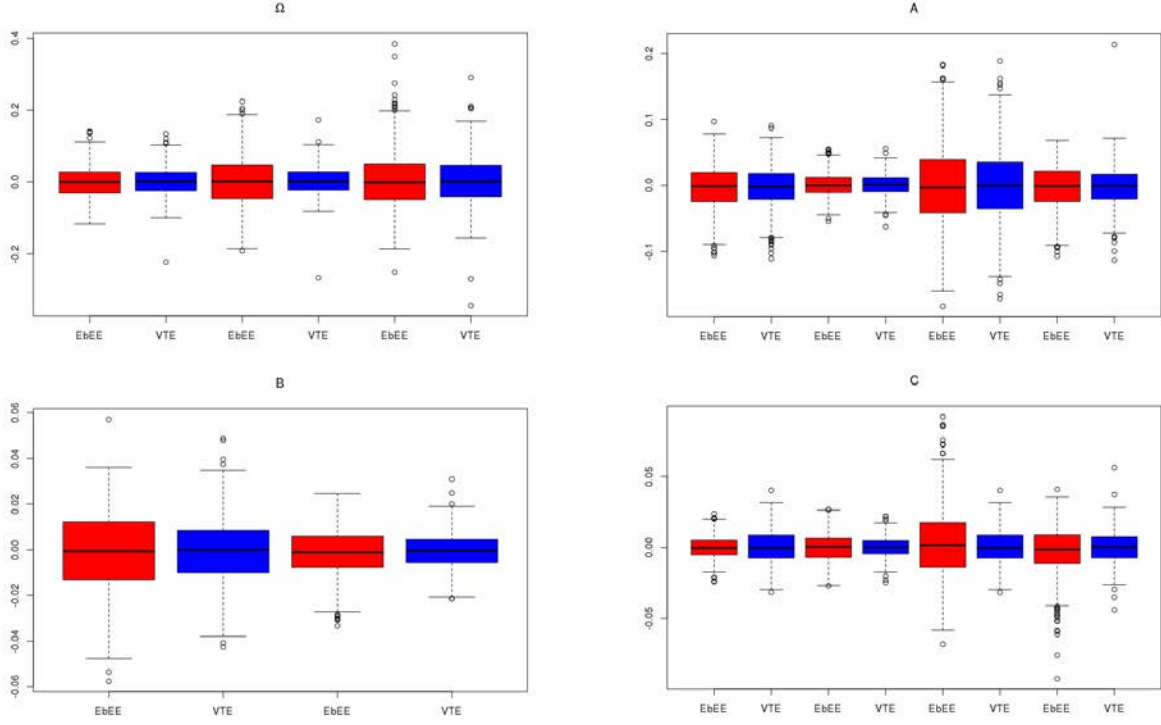


FIGURE 4.5 – Boxplots of 500 estimation errors for the EbEE and VTE.

Let L be the back-shift operator, *i.e.* $L(u_t) = u_{t-1}$. By Assumption **C5**, the polynomial $1 - b_k^2 L$ is invertible for any $\theta^{(k)} \in \Theta^{(k)}$. Assume that $\sigma_{kt}^2(\theta_0^{(k)}) = \sigma_{kt}^2(\theta^{(k)})$ a.s. We have

$$\begin{aligned}
& (1 - (b_k^0)^2 L)^{-1} \left(\sum_{\ell=1}^m a_{k\ell}^0 \varepsilon_{\ell,t-1} \right)^2 - (1 - b_k^2 L)^{-1} \left(\sum_{\ell=1}^m a_{k\ell} \varepsilon_{\ell,t-1} \right)^2 \\
& + (1 - (b_k^0)^2 L)^{-1} \left(\sum_{s=1}^r c_{ks}^0 x_{s,t-1} \right)^2 - (1 - b_k^2 L)^{-1} \left(\sum_{s=1}^r c_{ks} x_{s,t-1} \right)^2 \\
& = (1 - b_k^2)^{-1} \omega_{kk} - (1 - (b_k^0)^2)^{-1} \omega_{kk}^0 \quad \text{a.s.}
\end{aligned}$$

Then

$$\sum_{i=0}^{\infty} \sum_{\ell, \ell'=1}^m \bar{a}_{i\ell\ell'}^{(k)} \varepsilon_{\ell, t-i-1} \varepsilon_{\ell', t-i-1} + \sum_{j=0}^{\infty} \sum_{s, s'=1}^r \bar{c}_{jss'}^{(k)} x_{s, t-j-1} x_{s', t-j-1} = c \quad \text{a.s.} \quad (4.21)$$

where $\bar{a}_{i\ell\ell'}^{(k)} = (b_k^0)^{2i} a_{k\ell}^0 a_{k\ell'}^0 - (b_k)^{2i} a_{k\ell} a_{k\ell'}$, $\bar{c}_{jss'}^{(k)} = (b_k^0)^{2i} c_{ks}^0 c_{ks'}^0 - (b_k)^{2i} c_{ks} c_{ks'}$ and $c = (1 - (b_k^0)^2)^{-1} \omega_{kk} - (1 - b_k^2)^{-1} \omega_{kk}^0$. If $b_k^0 \neq b_k$ or there exists ℓ^* such that $a_{k\ell^*}^0 \neq a_{k\ell^*}$ then $\varepsilon_{\ell^*, t-1}^2$ is a linear combination of the $x_{s,u} x_{s',u}$, $s, s' = 1, \dots, r, u < t$, the $\varepsilon_{\ell, v} \varepsilon_{\ell', v}$, $\ell, \ell' = 1, \dots, m, v < t-1$ and the $\varepsilon_{\ell, t-1} \varepsilon_{\ell', t-1}$, $(\ell, \ell') \neq (\ell^*, \ell^*)$ which is impossible by Assumption **C7**. Therefore $b_k^0 = b_k$ and $a_{k\ell}^0 = a_{k\ell}$

for all $\ell = 1, \dots, m$ and (4.21) becomes

$$\sum_{j=0}^{\infty} \sum_{s,s'=1}^r \bar{c}_{jss'}^{(k)} x_{s,t-j-1} x_{s',t-j-1} = c \quad \text{a.s.}$$

Similarly, if there exists s^* such that $c_{ks^*}^0 \neq c_{ks^*}$, $x_{s^*,t-1}^2$ is a linear combination of $x_{u,t-j} x_{u',t-j}$, $u, u' = 1, \dots, r, j > 1$ and $x_{s,t-1} x_{s',t-1}$, $(s, s') \neq (s^*, s^*)$ which contradicts Assumption C8. Therefore we have $c_{ks}^0 = c_{ks}$ for all $s = 1, \dots, r$. Hence ii) is proved. \square

For the proof the asymptotic distribution in the Theorem 4.2, we need the following lemmas

Lemma 4.1. *Under assumptions of Theorem 4.2, there exists a neighborhood $\mathcal{V}(\theta_0^{(k)})$ of $\theta_0^{(k)}$ such that*

$$E \left\{ \sup_{\theta^{(k)} \in \mathcal{V}(\theta_0^{(k)})} \left\| \frac{1}{\sigma_{kt}^2(\theta^{(k)})} \frac{\partial \sigma_{kt}^2(\theta^{(k)})}{\partial \theta^{(k)}} \right\|^{4(1+1/\delta)} \right\} < \infty, \quad \text{for some } \delta > 0, \quad (4.22)$$

$$E \left\{ \sup_{\theta^{(k)} \in \mathcal{V}(\theta_0^{(k)})} \left\| \frac{1}{\sigma_{kt}^2(\theta^{(k)})} \frac{\partial^2 \sigma_{kt}^2(\theta^{(k)})}{\partial \theta^{(k)} \partial \theta^{(k)'}} \right\|^{2(1+1/\delta)} \right\} < \infty, \quad \text{for some } \delta > 0, \quad (4.23)$$

$$E \left\{ \sup_{\theta^{(k)} \in \mathcal{V}(\theta_0^{(k)})} \left| \frac{\sigma_{kt}^2(\theta_0^{(k)})}{\sigma_{kt}^2(\theta^{(k)})} \right|^s \right\} < \infty, \quad \text{for any } s > 0. \quad (4.24)$$

Proof of Lemma 4.1

Iteratively using the volatility equation in (4.4), we obtain

$$\sigma_{kt}^2(\theta^{(k)}) = \sum_{j=0}^{\infty} b_k^{2j} \left\{ \omega_{kk} + \sum_{\ell, \ell'=1}^m a_{k\ell} a_{k\ell'} \varepsilon_{\ell, t-j-1} \varepsilon_{\ell', t-j-1} + \sum_{s, s'=1}^r c_{ks} c_{ks'} x_{s, t-j-1} x_{s', t-j-1} \right\}. \quad (4.25)$$

Derive (4.25) with respect to $\theta^{(k)}$, we get

$$\begin{aligned} \frac{\partial \sigma_{kt}^2(\theta^{(k)})}{\partial \omega_{kk}} &= \sum_{j=0}^{\infty} b_k^{2j} = \frac{1}{1 - b_k^2}, \\ \frac{\partial \sigma_{kt}^2(\theta^{(k)})}{\partial a_{k\ell}} &= 2 \sum_{j=0}^{\infty} b_k^{2j} \sum_{\ell'=1}^m a_{k\ell'} \varepsilon_{\ell, t-j-1} \varepsilon_{\ell', t-j-1}, \\ \frac{\partial \sigma_{kt}^2(\theta^{(k)})}{\partial b_k^2} &= \sum_{j=1}^{\infty} j b_k^{2(j-1)} \left\{ \omega_{kk} + \sum_{\ell, \ell'=1}^m a_{k\ell} a_{k\ell'} \varepsilon_{\ell, t-j-1} \varepsilon_{\ell', t-j-1} + \sum_{s, s'=1}^r c_{ks} c_{ks'} x_{s, t-j-1} x_{s', t-j-1} \right\}, \\ \frac{\partial \sigma_{kt}^2(\theta^{(k)})}{\partial c_{ks}} &= 2 \sum_{j=0}^{\infty} b_k^{2j} \sum_{s'=1}^r c_{ks'} x_{s, t-j-1} x_{s', t-j-1} \end{aligned}$$

Similar expressions hold for the second order derivatives. Noting that

$$\underline{\omega}_{kk} := \inf_{\boldsymbol{\theta}^{(k)} \in \Theta^{(k)}} \sigma_{kt}^2 > 0.$$

Using the moment conditions **C10** and (4.20), we obtain (4.22) and (4.23).

The moment condition (4.24) will be showed even if some components of $\underline{\mathbf{a}}_k^0$ or $\underline{\mathbf{c}}_k^0$ are zero. Indeed, there exists a neighborhood $\mathcal{V}(\boldsymbol{\theta}_0^{(k)})$ such that for all $\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})$

$$\frac{\sigma_{kt}^2(\boldsymbol{\theta}_0^{(k)})}{\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})} \leq K + K \sum_{j=0}^{\infty} \frac{(b_k^0)^{2j}}{b_k^{2j}} \left\{ \frac{\left(\sum_{\substack{\ell=1 \\ a_{k\ell}^0 \neq 0}}^m \frac{a_{k\ell}}{\sqrt{\omega_{kk}}} \varepsilon_{\ell,t-j-1} \right)^2}{1 + \left(\sum_{\substack{\ell=1 \\ a_{k\ell}^0 \neq 0}}^m \frac{a_{k\ell}}{\sqrt{\omega_{kk}}} \varepsilon_{\ell,t-j-1} \right)^2} + \frac{\left(\sum_{\substack{s=1 \\ c_{ks}^0 \neq 0}}^r \frac{c_{ks}}{\sqrt{\omega_{kk}}} x_{s,t-j-1} \right)^2}{1 + \left(\sum_{\substack{s=1 \\ c_{ks}^0 \neq 0}}^r \frac{c_{ks}}{\sqrt{\omega_{kk}}} x_{s,t-j-1} \right)^2} \right\}.$$

For all $\delta > 0$, there exists $\mathcal{V}(\boldsymbol{\theta}_0^{(k)})$ such that $b_k^0 \leq (1 + \delta)b_k$ for all $\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})$. It follows that for all $\delta > 0$ and $u \in (0, 1)$, using the inequality $z/(1+z) \leq z^u$ for all $z \geq 0$, there exists $\mathcal{V}(\boldsymbol{\theta}_0^{(k)})$ such that

$$\sup_{\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \frac{\sigma_{kt}^2(\boldsymbol{\theta}_0^{(k)})}{\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})} \leq K + K \sum_{j=0}^{\infty} (1 + \delta)^j \rho^{ju} \left\{ \left(\sum_{\substack{\ell=1 \\ a_{k\ell}^0 \neq 0}}^m \varepsilon_{\ell,t-j-1} \right)^{2u} + \left(\sum_{\substack{s=1 \\ c_{ks}^0 \neq 0}}^r x_{s,t-j-1} \right)^{2u} \right\}.$$

Denote by $\|\cdot\|_d$ the L^d norm, for $d \geq 1$, on the space of real random variables. Using the Minkowski inequality and choosing u such that $E\|\varepsilon_1\|^{2us} < \infty$ and $E\|x_1\|^{2us} < \infty$ and choosing, for instance, $\delta = \frac{1 - \rho^u}{2\rho^u}$ and by Assumption **C6**, we have

$$\left\| \sup_{\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \frac{\sigma_{kt}^2(\boldsymbol{\theta}_0^{(k)})}{\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})} \right\|_s \leq K + K \sum_{j=0}^{\infty} (1 + \delta)^j \rho^{ju} \left\{ \sum_{\substack{\ell=1 \\ a_{k\ell}^0 \neq 0}}^m \|\varepsilon_{\ell,t-j-1}\|_{2us}^{2u} + \sum_{\substack{s=1 \\ c_{ks}^0 \neq 0}}^r \|x_{s,t-j-1}\|_{2us}^{2u} \right\} < \infty.$$

(4.24) is thus obtained. □

Lemma 4.2. *Under assumptions of Theorem 4.2, for all t , we have*

$$i) \ E \sup_{\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \left\| \frac{\partial^2 \ell_{kt}(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)} \partial \boldsymbol{\theta}^{(k)'}} \right\| < \infty, \text{ for some neighborhood } \mathcal{V}(\boldsymbol{\theta}_0^{(k)}) \text{ of } \boldsymbol{\theta}_0^{(k)}.$$

- ii) $\frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\partial^2 \ell_{kt}(\bar{\theta}_n^{(k)})}{\partial \theta^{(k)} \partial \theta^{(k)'}} \rightarrow \mathbf{J}_{kk}$, *a.s.* for any $\bar{\theta}_n^{(k)}$ between $\hat{\theta}_n^{(k)}$ and $\theta_0^{(k)}$.
 iii) \mathbf{J}_{kk} is non singular.

Proof of Lemma 4.2

The derivatives of $\ell_{kt}(\theta^{(k)})$ is given by

$$\frac{\partial \ell_{kt}(\theta^{(k)})}{\partial \theta^{(k)}} = \left\{ 1 - \frac{\varepsilon_{kt}^2}{\sigma_{kt}^2(\theta^{(k)})} \right\} \left\{ \frac{1}{\sigma_{kt}^2(\theta^{(k)})} \frac{\partial \sigma_{kt}^2(\theta^{(k)})}{\partial \theta^{(k)}} \right\} \quad (4.26)$$

and

$$\begin{aligned} \frac{\partial^2 \ell_{kt}(\theta^{(k)})}{\partial \theta^{(k)} \partial \theta^{(k)'}} &= \left\{ 1 - \frac{\varepsilon_{kt}^2}{\sigma_{kt}^2(\theta^{(k)})} \right\} \left\{ \frac{1}{\sigma_{kt}^2(\theta^{(k)})} \frac{\partial^2 \sigma_{kt}^2(\theta^{(k)})}{\partial \theta^{(k)} \partial \theta^{(k)'}} \right\} \\ &+ \left\{ 2 \frac{\varepsilon_{kt}^2}{\sigma_{kt}^2(\theta^{(k)})} - 1 \right\} \left\{ \frac{1}{\sigma_{kt}^2(\theta^{(k)})} \frac{\partial \sigma_{kt}^2(\theta^{(k)})}{\partial \theta^{(k)}} \right\} \left\{ \frac{1}{\sigma_{kt}^2(\theta^{(k)})} \frac{\partial \sigma_{kt}^2(\theta^{(k)})}{\partial \theta^{(k)'}} \right\}. \end{aligned} \quad (4.27)$$

Using the triangle inequality, i) is obtained by showing the existence of the expectations of the two terms in the right-hand side of (4.27). Let us consider the first one. We have, by the Holder inequality,

$$\begin{aligned} E \sup_{\theta^{(k)} \in \mathcal{V}(\theta_0^{(k)})} &\left\| \left\{ 1 - \frac{\varepsilon_{kt}^2}{\sigma_{kt}^2(\theta^{(k)})} \right\} \left\{ \frac{1}{\sigma_{kt}^2(\theta^{(k)})} \frac{\partial^2 \sigma_{kt}^2(\theta^{(k)})}{\partial \theta^{(k)} \partial \theta^{(k)'}} \right\} \right\| \\ &\leq \left\{ 1 + \|\eta_{kt}^{*2}\|_{2(\delta+1)} \left\| \sup_{\theta^{(k)} \in \mathcal{V}(\theta_0^{(k)})} \frac{\sigma_{kt}^2(\theta_0^{(k)})}{\sigma_{kt}^2(\theta^{(k)})} \right\|_2 \right\} \left\| \sup_{\theta^{(k)} \in \mathcal{V}(\theta_0^{(k)})} \left\| \frac{1}{\sigma_{kt}^2(\theta^{(k)})} \frac{\partial^2 \sigma_{kt}^2(\theta^{(k)})}{\partial \theta^{(k)} \partial \theta^{(k)'}} \right\| \right\|_{2(1+1/\delta)} \end{aligned}$$

which is finite by (4.23) and (4.24). The second product in the right-hand side of (4.27) can be handled similarly using (4.22).

To prove ii), by Exercise 7.9 in Francq and Zakoian (2010), it will be sufficient to establish that for any $\epsilon > 0$, there exists a neighborhood $\mathcal{V}(\theta_0^{(k)})$ of $\theta_0^{(k)}$ such that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \sup_{\theta^{(k)} \in \mathcal{V}(\theta_0^{(k)})} \left\| \frac{\partial^2 \ell_{kt}(\theta^{(k)})}{\partial \theta^{(k)} \partial \theta^{(k)'}} - \frac{\partial^2 \ell_{kt}(\theta_0^{(k)})}{\partial \theta^{(k)} \partial \theta^{(k)'}} \right\| \leq \epsilon \quad a.s. \quad (4.28)$$

By the ergodic theorem, the limit in the left-hand side is equal to

$$E \sup_{\theta^{(k)} \in \mathcal{V}(\theta_0^{(k)})} \left\| \frac{\partial^2 \ell_{kt}(\theta^{(k)})}{\partial \theta^{(k)} \partial \theta^{(k)'}} - \frac{\partial^2 \ell_{kt}(\theta_0^{(k)})}{\partial \theta^{(k)} \partial \theta^{(k)'}} \right\|$$

that is finite by i). This expectation tends to zero when the neighborhood $\mathcal{V}(\theta_0^{(k)})$ shrinks to singleton

$\{\boldsymbol{\theta}_0^{(k)}\}$. The point ii) is thus proved.

We now turn show the invertibility of \mathbf{J}_{kk} . Assume that \mathbf{J}_{kk} is singular. Then there exists a vector $\boldsymbol{\pi} \in \mathbb{R}^{2+m+r}$ such that $\boldsymbol{\pi}' \frac{\partial \sigma_{kt}^2(\boldsymbol{\theta}_0^{(k)})}{\partial \boldsymbol{\theta}^{(k)}} = 0$ a.s. It follows that

$$\begin{aligned} & \pi_1 + 2 \left(\sum_{\ell=1}^m a_{k\ell}^0 \varepsilon_{\ell,t-1} \right) \boldsymbol{\pi}' \begin{pmatrix} 0 \\ \boldsymbol{\varepsilon}_{t-1} \\ \mathbf{0}_{(1+r) \times 1} \end{pmatrix} + \pi_{2+m} \sigma_{k,t-1}^2(\boldsymbol{\theta}_0^{(k)}) \\ & + 2 \left(\sum_{s=1}^r c_{ks}^0 x_{s,t-1} \right) \boldsymbol{\pi}' \begin{pmatrix} \mathbf{0}_{(2+m) \times 1} \\ \mathbf{x}_{t-1} \end{pmatrix} + b_k^{02} \boldsymbol{\pi}' \frac{\partial \sigma_{k,t-1}^2(\boldsymbol{\theta}_0^{(k)})}{\partial \boldsymbol{\theta}^{(k)}} = 0, \quad \text{a.s.} \end{aligned}$$

Note that the last term in the left-hand side of this equation is equal to zero by the stationarity. We thus have

$$\pi_1 + 2 \sum_{\ell,\ell'=1}^m a_{k\ell}^0 \pi_{1+\ell'} \varepsilon_{\ell,t-1} \varepsilon_{\ell',t-1} + \pi_{2+m} \sigma_{k,t-1}^2(\boldsymbol{\theta}_0^{(k)}) + 2 \sum_{s,s'=1}^r c_{ks}^0 \pi_{2+m+s'} x_{s,t-1} x_{s',t-1} = 0 \text{ a.s.}$$

By the similar arguments used to show ii) in the proof of Theorem 4.1, we can conclude that $\boldsymbol{\pi} = 0$. \square

Proof of Theorem 4.2

Since $\widehat{\boldsymbol{\theta}}_n^{(k)}$ strongly converges to $\boldsymbol{\theta}_0^{(k)}$ which belongs to the interior of the parameter space $\Theta^{(k)}$, the derivative of the criterion $\widetilde{Q}_n^{(k)}$ is equal to zero at $\widehat{\boldsymbol{\theta}}_n^{(k)}$. Let $\mathbf{J}_{kkn}(\boldsymbol{\theta}^{(k)}) = \frac{1}{n} \sum_{t=1}^n \frac{\partial^2 \ell_{kt}(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)} \partial \boldsymbol{\theta}^{(k)'}}$. Applying a Taylor expansion for $Q_n^{(k)}$ at $\boldsymbol{\theta}_0^{(k)}$ and the mean-value theorem gives

$$0 = \frac{1}{n} \sum_{t=1}^n \frac{\ell_{kt}(\boldsymbol{\theta}_0^{(k)})}{\partial \boldsymbol{\theta}^{(k)}} + \mathbf{J}_{kkn}(\bar{\boldsymbol{\theta}}_n^{(k)}) \left(\widehat{\boldsymbol{\theta}}_n^{(k)} - \boldsymbol{\theta}_0^{(k)} \right), \quad (4.29)$$

where $\bar{\boldsymbol{\theta}}_n^{(k)}$ is between $\widehat{\boldsymbol{\theta}}_n^{(k)}$ and $\boldsymbol{\theta}_0^{(k)}$. By the points ii) and iii) in Lemma ?? and the consistency of $\widehat{\boldsymbol{\theta}}_n^{(k)}$, the matrix $\mathbf{J}_{kkn}(\bar{\boldsymbol{\theta}}_n^{(k)})$ is a.s. invertible for sufficiently large n . Hence multiplying by \sqrt{n} and solving for $\sqrt{n} \left(\widehat{\boldsymbol{\theta}}_n^{(k)} - \boldsymbol{\theta}_0^{(k)} \right)$ gives

$$\sqrt{n} \left(\widehat{\boldsymbol{\theta}}_n^{(k)} - \boldsymbol{\theta}_0^{(k)} \right) \stackrel{op(1)}{=} -\mathbf{J}_{kkn}^{-1}(\bar{\boldsymbol{\theta}}_n^{(k)}) \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\ell_{kt}(\boldsymbol{\theta}_0^{(k)})}{\partial \boldsymbol{\theta}^{(k)}}. \quad (4.30)$$

Let $\dot{\ell}_t(\boldsymbol{\theta}) = \left(\frac{\partial \ell_{1t}(\boldsymbol{\theta}^{(1)})}{\partial \boldsymbol{\theta}^{(1)'}} , \dots , \frac{\partial \ell_{mt}(\boldsymbol{\theta}^{(m)})}{\partial \boldsymbol{\theta}^{(m)'}} \right)'$. Collecting all these Taylor expansions, we have

$$\sqrt{n} \left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \right) \stackrel{o_p(1)}{=} \mathbf{J}^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^n \dot{\ell}_t(\boldsymbol{\theta}_0).$$

Note that $\boldsymbol{\eta}_t^* = \mathbf{D}_{0t}^{-1} \mathbf{H}_{0t}^{1/2} \boldsymbol{\eta}_t$. From (4.26), we then have

$$\begin{aligned} \dot{\ell}_t(\boldsymbol{\theta}_0) &= -\boldsymbol{\Delta}_t T_m \text{vec} \left(\boldsymbol{\eta}_t^* \boldsymbol{\eta}_t^{*'} - \mathbf{D}_{0t}^{-1} \mathbf{H}_{0t} \mathbf{D}_{0t}^{-1} \right) \\ &= -\boldsymbol{\Delta}_t T_m \left(\mathbf{D}_{0t}^{-1} \mathbf{H}_{0t}^{1/2} \right)^{\otimes 2} \text{vec} \left(\boldsymbol{\eta}_t \boldsymbol{\eta}_t' - \mathbf{I}_m \right). \end{aligned}$$

Then, we get

$$\sqrt{n} \left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \right) \stackrel{o_p(1)}{=} -\mathbf{J}^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^n \boldsymbol{\Delta}_t T_m \left(\mathbf{D}_{0t}^{-1} \mathbf{H}_{0t}^{1/2} \right)^{\otimes 2} \text{vec} \left(\boldsymbol{\eta}_t \boldsymbol{\eta}_t' - \mathbf{I}_m \right). \quad (4.31)$$

We now introduce the martingale difference

$$\boldsymbol{\nu}_t = \text{vec}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') - \text{vec}(\mathbf{H}_{0t}) = \left(\mathbf{H}_{0t}^{1/2} \otimes \mathbf{H}_{0t}^{1/2} \right) \text{vec}(\boldsymbol{\eta}_t \boldsymbol{\eta}_t' - \mathbf{I}_m).$$

In the representation of $\text{vec}(\mathbf{H}_{0t})$ obtained from (4.2), we replace $\text{vec}(\mathbf{H}_{0t})$ by $\text{vec}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') - \boldsymbol{\nu}_t$. Then, we get

$$\begin{aligned} \text{vec}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' - E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t')) &= (\mathbf{A}_0^{\otimes 2} + \mathbf{B}_0^{\otimes 2}) \text{vec}(\boldsymbol{\varepsilon}_{t-1} \boldsymbol{\varepsilon}_{t-1}' - E(\boldsymbol{\varepsilon}_{t-1} \boldsymbol{\varepsilon}_{t-1}')) \\ &\quad + \mathbf{C}_0^{\otimes 2} \text{vec}(\mathbf{x}_{t-1} \mathbf{x}_{t-1}' - E\mathbf{x}_{t-1} \mathbf{x}_{t-1}') + (\boldsymbol{\nu}_t - \mathbf{B}_0^{\otimes 2} \boldsymbol{\nu}_{t-1}). \end{aligned}$$

Note that under assumption **C5**, the matrix $\mathbf{I}_{m^2} - \mathbf{A}_0^{\otimes 2} - \mathbf{B}_0^{\otimes 2}$ is invertible. Taking the average of the two sides of the equality for $t = 1, \dots, n$ gives

$$\begin{aligned} \widehat{\boldsymbol{\gamma}}_{\boldsymbol{\varepsilon},n} - \boldsymbol{\gamma}_{\boldsymbol{\varepsilon},0} &= L_m (\mathbf{I}_{m^2} - \mathbf{A}_0^{\otimes 2} - \mathbf{B}_0^{\otimes 2})^{-1} (\mathbf{I}_{m^2} - \mathbf{B}_0^{\otimes 2}) \frac{1}{n} \sum_{t=1}^n \boldsymbol{\nu}_t \\ &\quad + L_m (\mathbf{I}_{m^2} - \mathbf{A}_0^{\otimes 2} - \mathbf{B}_0^{\otimes 2})^{-1} \mathbf{C}_0^{\otimes 2} D_r (\widehat{\boldsymbol{\gamma}}_{\mathbf{x},n} - \boldsymbol{\gamma}_{\mathbf{x},0}) + o_p(1), \quad a.s. \end{aligned}$$

We then have

$$\begin{pmatrix} \sqrt{n} \left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \right) \\ \sqrt{n} \left(\widehat{\boldsymbol{\gamma}}_{\boldsymbol{\varepsilon},n} - \boldsymbol{\gamma}_{\boldsymbol{\varepsilon}} \right) \\ \sqrt{n} \left(\widehat{\boldsymbol{\gamma}}_{\mathbf{x},n} - \boldsymbol{\gamma}_{\mathbf{x}} \right) \end{pmatrix} \stackrel{o_p(1)}{=} \begin{pmatrix} -\mathbf{J}^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{N}_1 & \mathbf{N}_2 \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{r(r+1)/2} \end{pmatrix} \frac{1}{\sqrt{n}} \begin{pmatrix} \sum_{t=1}^n \boldsymbol{\Upsilon}_{0t} \text{vec} \left(\boldsymbol{\eta}_t \boldsymbol{\eta}_t' - \mathbf{I}_m \right) \\ \sum_{t=1}^n \text{vech}(\mathbf{x}_t \mathbf{x}_t' - E(\mathbf{x}_t \mathbf{x}_t)) \end{pmatrix}.$$

The arguments for establishing the limiting distribution of $\frac{1}{\sqrt{n}} \begin{pmatrix} \sum_{t=1}^n \mathbf{r}_{0t} \text{vec}(\boldsymbol{\eta}_t \boldsymbol{\eta}'_t - I_m) \\ \sum_{t=1}^n \text{vech}(\mathbf{x}_t \mathbf{x}'_t - E(\mathbf{x}_t \mathbf{x}_t)) \end{pmatrix}$ being very similar to Lemma 6 in Thieu (2016), I just give a sketch of proof. For any $s > 0$, the k -th individual volatility in (4.4) can be written $\sigma_{kt}^2 = \underline{\sigma}_{kts}^2 + \bar{\sigma}_{kts}^2$, where

$$\begin{aligned} \sigma_{kts}^2 &= \sum_{j=0}^s b_k^{2j} \left\{ \omega_{kk} + \left(\sum_{\ell=1}^m a_{k\ell} \varepsilon_{\ell, t-j-1} \right)^2 + \left(\sum_{s=1}^r c_{ks} x_{s, t-j-1} \right)^2 \right\} \\ \bar{\sigma}_{kts}^2 &= \sum_{j=s+1}^{\infty} b_k^{2j} \left\{ \omega_{kk} + \left(\sum_{\ell=1}^m a_{k\ell} \varepsilon_{\ell, t-j-1} \right)^2 + \left(\sum_{s=1}^r c_{ks} x_{s, t-j-1} \right)^2 \right\}. \end{aligned}$$

Then we can write $\mathbf{r}_{0t} \text{vec}(\boldsymbol{\eta}_t \boldsymbol{\eta}'_t - I_m) = \underline{\mathbf{Y}}_{t,s} + \mathbf{Y}_{t,s}^*$, where

$$\underline{\mathbf{Y}}_{t,s} = \begin{pmatrix} \underline{\Delta}_{ts} T_m \left(\underline{\mathbf{D}}_{0ts}^{-1} \underline{\mathbf{H}}_{0t,s}^{1/2} \right) \otimes \left(\underline{\mathbf{D}}_{0ts}^{-1} \underline{\mathbf{H}}_{0t,s}^{1/2} \right) \\ \underline{\mathbf{H}}_{0t,s}^{1/2} \otimes \underline{\mathbf{H}}_{0t,s}^{1/2} \end{pmatrix} \text{vec}(\boldsymbol{\eta}_t \boldsymbol{\eta}'_t - I_m),$$

with $\underline{\mathbf{D}}_{0ts} = \text{diag}(\underline{\sigma}_{1ts}, \dots, \underline{\sigma}_{mts})$ and $\underline{\Delta}_{ts} = \text{diag}(\underline{\Delta}_{1ts}, \dots, \underline{\Delta}_{mts})$, $\underline{\Delta}_{kts} = \frac{1}{\underline{\sigma}_{kts}^2} \frac{\partial \sigma_{kts}^2(\boldsymbol{\theta}_0^{(k)})}{\partial \boldsymbol{\theta}^{(k)}}$, for $k = 1, \dots, m$ and $\mathbf{Y}_{t,s}^*$ is stationary and centered process satisfying

$$\lim_{s \rightarrow \infty} \limsup_{n \rightarrow \infty} P \left(\left\| n^{-1/2} \sum_{t=1}^n \mathbf{Y}_{t,s}^* \right\| > \epsilon \right) = 0.$$

Under Assumption C11 and using the same argument as in the reference, we have for some ν and $\delta > 0$

$$\left\| \left(\underline{\mathbf{H}}_{0t,s}^{1/2} \otimes \underline{\mathbf{H}}_{0t,s}^{1/2} \right) \text{vec}(\boldsymbol{\eta}_t \boldsymbol{\eta}'_t - I_m) \right\|_{2+\nu} < \infty.$$

Note that $\mathbf{D}_{0t}^{-1} \mathbf{H}_{0t} \mathbf{D}_{0t}^{-1}$ is the conditional correlation matrix of $\boldsymbol{\varepsilon}_t$ and

$$\mathbf{D}_{0t}^{-1} \mathbf{H}_{0t} \mathbf{D}_{0t}^{-1} = \left(\mathbf{D}_{0t}^{-1} \mathbf{H}_{0t}^{1/2} \right) \left(\mathbf{D}_{0t}^{-1} \mathbf{H}_{0t}^{1/2} \right)'$$

All the elements of the matrix $\mathbf{D}_{0t}^{-1} \mathbf{H}_{0t}^{1/2}$ are thus smaller than one. It is thus easy to show that all the ones of the matrix $\left(\underline{\mathbf{D}}_{0t,s}^{-1} \underline{\mathbf{H}}_{0t,s}^{1/2} \right) \otimes \left(\underline{\mathbf{D}}_{0t,s}^{-1} \underline{\mathbf{H}}_{0t,s}^{1/2} \right)$ are also smaller than one. Using the Holder

inequality, we then have

$$\begin{aligned}
& \left\| \underline{\Delta}_{ts} T_m \left(\underline{D}_{0t,s}^{-1} \underline{H}_{0t,s}^{1/2} \right) \otimes \left(\underline{D}_{0t,s}^{-1} \underline{H}_{0t,s}^{1/2} \right) \text{vec}(\boldsymbol{\eta}_t \boldsymbol{\eta}'_t - I_m) \right\|_{2+\nu} \\
& \leq \|\underline{\Delta}_{ts}\|_{(2+\nu)(1+1/\delta)} \left\| T_m \left(\underline{D}_{0t,s}^{-1} \underline{H}_{0t,s}^{1/2} \right) \otimes \left(\underline{D}_{0t,s}^{-1} \underline{H}_{0t,s}^{1/2} \right) \text{vec}(\boldsymbol{\eta}_t \boldsymbol{\eta}'_t - I_m) \right\|_{(2+\nu)(1+\delta)} \\
& \leq K \|\underline{\Delta}_{ts}\|_{(2+\nu)(1+1/\delta)} \left\| \text{vec}(\boldsymbol{\eta}_t \boldsymbol{\eta}'_t - I_m) \right\|_{(2+\nu)(1+\delta)} \\
& < \infty,
\end{aligned}$$

where the last inequality follows from (4.20) and Assumption C11. It implies that $\|\underline{\mathbf{Y}}_{t,s}\|_{2+\nu} < \infty$. The process $(\underline{\mathbf{Y}}_{t,s})_t$, for s fixed, is thus strongly mixing under Assumption C11. Applying the central limit theorem of Herrndorf (1984), we get

$$\frac{1}{\sqrt{n}} \left(\begin{array}{c} \sum_{t=1}^n \mathbf{Y}_{0t} \text{vec}(\boldsymbol{\eta}_t \boldsymbol{\eta}'_t - I_m) \\ \sum_{t=1}^n \text{vech}(\mathbf{x}_t \mathbf{x}'_t - E(\mathbf{x}_t \mathbf{x}_t)) \end{array} \right) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma}).$$

The asymptotic distribution of Theorem 4.2 thus follows from the Slutsky theorem. \square

The proof of Theorem 4.3

The strong consistency of $\widehat{\boldsymbol{\xi}}_n$ is obviously obtained.

Now we turn to its asymptotic normality.

Using the elementary relation $\text{vec}(ABC) = (C' \otimes A)\text{vec}(B)$, we have

$$\begin{aligned}
\text{vec}(\widehat{\mathbf{A}}_n \widehat{\boldsymbol{\Sigma}}_{\varepsilon n} \widehat{\mathbf{A}}'_n - \mathbf{A}_0 \boldsymbol{\Sigma}_{\varepsilon 0} \mathbf{A}'_0) &= \text{vec}(\widehat{\mathbf{A}}_n \widehat{\boldsymbol{\Sigma}}_{\varepsilon n} (\widehat{\mathbf{A}}'_n - \mathbf{A}'_0)) + \text{vec}(\widehat{\mathbf{A}}_n (\widehat{\boldsymbol{\Sigma}}_{\varepsilon n} - \boldsymbol{\Sigma}_{\varepsilon}) \mathbf{A}'_0) \\
&\quad + \text{vec}((\widehat{\mathbf{A}}_n - \mathbf{A}_0) \boldsymbol{\Sigma}_{\varepsilon} \mathbf{A}'_0) \\
&= \left\{ I_m \otimes (\widehat{\mathbf{A}}_n \widehat{\boldsymbol{\Sigma}}_{\varepsilon n}) + ((\mathbf{A}_0 \boldsymbol{\Sigma}_{\varepsilon}) \otimes I_m) M_{mm} \right\} \text{vec}(\widehat{\mathbf{A}}'_n - \mathbf{A}'_0) \\
&\quad + (\mathbf{A}_0 \otimes \widehat{\mathbf{A}}_n) D_m \text{vech}(\widehat{\boldsymbol{\Sigma}}_{\varepsilon n} - \boldsymbol{\Sigma}_{\varepsilon}).
\end{aligned}$$

Similarly, we also get

$$\begin{aligned}
\text{vec}(\widehat{\mathbf{B}}_n \widehat{\boldsymbol{\Sigma}}_{\varepsilon n} \widehat{\mathbf{B}}_n - \mathbf{B}_0 \boldsymbol{\Sigma}_{\varepsilon} \mathbf{B}_0) &= \left\{ I_m \otimes (\widehat{\mathbf{B}}_n \widehat{\boldsymbol{\Sigma}}_{\varepsilon n}) + ((\mathbf{B}_0 \boldsymbol{\Sigma}_{\varepsilon}) \otimes I_m) \right\} \text{vec}(\widehat{\mathbf{B}}_n - \mathbf{B}_0) \\
&\quad + (\mathbf{B}_0 \otimes \widehat{\mathbf{B}}_n) D_m \text{vech}(\widehat{\boldsymbol{\Sigma}}_{\varepsilon n} - \boldsymbol{\Sigma}_{\varepsilon})
\end{aligned}$$

and

$$\begin{aligned}
\text{vec}(\widehat{\mathbf{C}}_n \widehat{\boldsymbol{\Sigma}}_{\mathbf{x}n} \widehat{\mathbf{C}}'_n - \mathbf{C}_0 \boldsymbol{\Sigma}_{\mathbf{x}0} \mathbf{C}'_0) &= \left\{ I_m \otimes (\widehat{\mathbf{C}}_n \widehat{\boldsymbol{\Sigma}}_{\mathbf{x}n}) + ((\mathbf{C}_0 \boldsymbol{\Sigma}_{\mathbf{x}}) \otimes I_m) M_{mr} \right\} \text{vec}(\widehat{\mathbf{C}}'_n - \mathbf{C}'_0) \\
&\quad + (\mathbf{C}_0 \otimes \widehat{\mathbf{C}}_n) D_r \text{vech}(\widehat{\boldsymbol{\Sigma}}_{\mathbf{x}n} - \boldsymbol{\Sigma}_{\mathbf{x}}).
\end{aligned}$$

Let

$$\begin{aligned}\widehat{\mathbf{A}}_n^* &= -P_m \left\{ I_m \otimes (\widehat{\mathbf{A}}_n \widehat{\boldsymbol{\Sigma}}_{\varepsilon n}) + ((\mathbf{A}_0 \boldsymbol{\Sigma}_{\varepsilon 0}) \otimes I_m) M_{mm} \right\}, \\ \widehat{\mathbf{B}}_n^* &= -P_m \left\{ I_m \otimes (\widehat{\mathbf{B}}_n \widehat{\boldsymbol{\Sigma}}_{\varepsilon n}) + (\mathbf{B}_0 \boldsymbol{\Sigma}_{\varepsilon 0}) \otimes I_m \right\}, \\ \widehat{\mathbf{C}}_n^* &= -P_m \left\{ I_m \otimes (\widehat{\mathbf{C}}_n \widehat{\boldsymbol{\Sigma}}_{\mathbf{x}n}) + ((\mathbf{C}_0 \boldsymbol{\Sigma}_{\mathbf{x}0}) \otimes I_m) M_{mr} \right\}, \\ \widehat{\mathbf{E}}_n^* &= P_m (I_{m^2} - \mathbf{B}_0 \otimes \widehat{\mathbf{B}}_n - \mathbf{A}_0 \otimes \widehat{\mathbf{A}}_n) D_m, \quad \widehat{\mathbf{X}}_n^* = -P_m (\mathbf{C}_0 \otimes \widehat{\mathbf{C}}_n) D_r.\end{aligned}$$

We then have

$$\sqrt{n}(\widehat{\boldsymbol{\omega}}_n - \boldsymbol{\omega}_0) = \begin{pmatrix} \widehat{\mathbf{A}}_n^* & \widehat{\mathbf{B}}_n^* & \widehat{\mathbf{C}}_n^* & \widehat{\mathbf{E}}_n^* & \widehat{\mathbf{X}}_n^* \end{pmatrix} \sqrt{n} \begin{pmatrix} \widehat{\mathbf{a}}_n - \mathbf{a}_0 \\ \widehat{\mathbf{b}}_n - \mathbf{b}_0 \\ \widehat{\mathbf{c}}_n - \mathbf{c}_0 \\ \widehat{\boldsymbol{\gamma}}_{\varepsilon n} - \boldsymbol{\gamma}_{\varepsilon 0} \\ \widehat{\boldsymbol{\gamma}}_{\mathbf{x}n} - \boldsymbol{\gamma}_{\mathbf{x}0} \end{pmatrix}, \quad (4.32)$$

where $\widehat{\mathbf{a}}_n = \text{vec}(\widehat{\mathbf{A}}_n')$, $\widehat{\mathbf{b}}_n = \text{vec}(\widehat{\mathbf{B}}_n)$ and $\widehat{\mathbf{c}}_n = \text{vec}(\widehat{\mathbf{C}}_n')$ are the estimators of $\mathbf{a}_0 = \text{vec}(\mathbf{A}_0')$, $\mathbf{b}_0 = \text{vec}(\mathbf{B}_0)$ and $\mathbf{c}_0 = \text{vec}(\mathbf{C}_0')$ respectively. Note that

$$\begin{pmatrix} \widehat{\mathbf{a}}_n - \mathbf{a}_0 \\ \widehat{\mathbf{b}}_n - \mathbf{b}_0 \\ \widehat{\mathbf{c}}_n - \mathbf{c}_0 \\ \widehat{\boldsymbol{\gamma}}_{\varepsilon n} - \boldsymbol{\gamma}_{\varepsilon 0} \\ \widehat{\boldsymbol{\gamma}}_{\mathbf{x}n} - \boldsymbol{\gamma}_{\mathbf{x}0} \end{pmatrix} = \boldsymbol{\Psi} \begin{pmatrix} \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \\ \widehat{\boldsymbol{\gamma}}_{\varepsilon n} - \boldsymbol{\gamma}_{\varepsilon 0} \\ \widehat{\boldsymbol{\gamma}}_{\mathbf{x}n} - \boldsymbol{\gamma}_{\mathbf{x}0} \end{pmatrix}, \quad (4.33)$$

where $\boldsymbol{\Psi}$ is a $d_1 \times d_2$ matrix, $d_1 = 2m^2 + mr + \frac{m(m+1)}{2} + \frac{r(r+1)}{2}$ and $d_2 = md + \frac{m(m+1)}{2} + \frac{r(r+1)}{2}$, given by, for $k = 1, \dots, m$,

$$\begin{aligned}\boldsymbol{\Psi}[(k-1)m+1 : km, (k-1)(m+r+2)+2 : (k-1)(m+r+2)+m+1] &= I_m, \\ \boldsymbol{\Psi}[m^2 + (k-1)m+k, k(m+2) + (k-1)r] &= 1, \\ \boldsymbol{\Psi}[2m^2 + (k-1)r+1 : m^2 + m+kr, k(m+2) + (k-1)r+1 : k(m+r+2)] &= I_r, \\ \boldsymbol{\Psi}[2m^2 + mr+1 : d_1, md+1 : d_2] &= I_{\frac{m(m+1)}{2} + \frac{r(r+1)}{2}}\end{aligned}$$

and the others entries are zero. We then have

$$\sqrt{n} \begin{pmatrix} \widehat{\boldsymbol{\omega}}_n - \boldsymbol{\omega}_0 \\ \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \end{pmatrix} = \widehat{\boldsymbol{\Omega}}_n \sqrt{n} \begin{pmatrix} \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \\ \widehat{\boldsymbol{\gamma}}_{\varepsilon n} - \boldsymbol{\gamma}_{\varepsilon 0} \\ \widehat{\boldsymbol{\gamma}}_{\mathbf{x}n} - \boldsymbol{\gamma}_{\mathbf{x}0} \end{pmatrix}, \quad (4.34)$$

where $\widehat{\Omega}_n = \begin{pmatrix} \widehat{\Omega}_{1n} \\ \Omega_2 \end{pmatrix}$ with $\widehat{\Omega}_{1n} = \begin{pmatrix} \widehat{A}_n^* & \widehat{B}_n^* & \widehat{C}_n^* & \widehat{E}_n^* & \widehat{X}_n^* \end{pmatrix} \Psi$. Note that $\widehat{\Omega}_n$ is strongly consistent estimator of Ω . By applying the Slutsky theorem, the asymptotic normality in Theorem 4.3 is thus obtained. \square

References

- Bollerslev, T.** (1990) Modelling the coherence in short-run nominal exchange rates : a multivariate generalized ARCH model. *The review of economics and statistics* 495–505.
- Bollerslev, T. and J.M. Wooldridge** (1992) Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric reviews* 11, 143–172.
- Boussama, F., F. Fuchs and R. Stelzer** (2011) Stationarity and geometric ergodicity of BEKK multivariate GARCH models. *Stochastic Processes and their Applications* 121, 2331–2360.
- Comte, F. and O. Lieberman** (2003) Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis* 84, 61–84.
- Engle, R.F., V.K. Ng and K.F. Kroner** (1990) Asset pricing with a factor-ARCH covariance structure : Empirical estimates for treasury bills. *Journal of Econometrics* 45, 213–237.
- Engle, R.F. and K.F. Kroner** (1995) Multivariate simultaneous generalized ARCH. *Econometric theory* 11, 122–150.
- Engle, R.F. and Sheppard, K.** (2001) Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. *National Bureau of Economic Research*.
- Engle, R.F.** (2002) Dynamic conditional correlation : A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics* 20, 339–350.
- Engle, R.** (2009) *Anticipating correlations : a new paradigm for risk management*. Princeton University Press.
- Francq, C. and J-M. Zakoïan** (1998) Estimating linear representations of nonlinear processes. *Journal of Statistical Planning and Inference* 68, 145–165.
- Francq, C. and J-M. Zakoïan** (2000) Covariance matrix estimation for estimators of mixing weak ARMA models. *Journal of Statistical Planning and Inference* 83, 369–394.
- Francq, C. and J-M. Zakoïan** (2010) *GARCH Models : Structure, Statistical Inference and Financial Applications*. John Wiley.
- Francq, C. and Thieu, L.Q.** (2015) Qml inference for volatility models with covariates. *University Library of Munich, Germany*, No. 63198.

- Francq, C. and G. Sucarrat** (2016) Equation-by-Equation Estimation of a Multivariate Log-GARCH-X Model of Financial Returns. *Journal of the Royal Statistical Society : Series B* 78, 613–635, <http://onlinelibrary.wiley.com/doi/10.1111/rssb.12126/epdf>
- Hafner, C.M. and A. Preminger** (2009) On asymptotic theory for multivariate GARCH models. *Journal of Multivariate Analysis* 100, 2044–2054.
- Han, H. and D. Kristensen** (2014) Asymptotic Theory for the QMLE in GARCH-X Models With Stationary and Nonstationary Covariates. *Journal of Business & Economic Statistics* 32, 416–429.
- Herrndorf, N.** (1984) A functional central limit theorem for weakly dependent sequences of random variables. *The Annals of Probability*, 141–153.
- Pedersen, R.S. and A. Rahbek** (2014) Multivariate variance targeting in the BEKK-GARCH model. *The Econometrics Journal* 17, 24–55.
- Phillips, P.C.B., Y. Sun and S. Jin** (2003) Consistent HAC estimation and robust regression testing using sharp origin kernels with no truncation. *Discussion paper, Yale University*.
- Shephard, N. and Sheppard, K.** (2010) Realizing the future : forecasting with high frequency based volatility (HEAVY) models. *Journal of Applied Econometrics* 25, 197–231.
- Silvennoinen, A. and T. Teräsvirta** (2009) Modeling multivariate autoregressive conditional heteroskedasticity with the double smooth transition conditional correlation GARCH model. *Journal of Financial Econometrics* 7, 373–411.

Conclusion et perspectives

L'objectif principal de ce mémoire était d'étudier les propriétés probabilistes et statistiques de modèles conditionnellement hétéroscédastiques augmentés par des variables explicatives exogènes. Pour prendre en compte des effets exogènes mesurés à l'aide de variables additionnelles, nous avons fait le choix dans cette thèse d'utiliser la famille de modèles univariés et multivariés de type GARCH. Résumons ici, les principaux résultats et présentons comment ils pourraient être développés et étendus.

Le premier chapitre a présenté une vaste classe de modèles APARCH demi-forts avec variables exogènes. Nous nous sommes intéressés à cette classe de modèles essentiellement parce que cette famille est très générale puisqu'elle contient les GARCH ordinaires, les TGARCH, ou encore les Log-GARCH. Les principales hypothèses imposées sur les variables exogènes sont la stationnarité et la non-colinéarité entre les variables explicatives. Notre principale contribution de ce chapitre est de développer le comportement asymétrique du QMV des estimateurs sans supposer que le vrai paramètre appartient à l'intérieur de l'espace des paramètres et que les innovations sont i.i.d. Du point de vue pratique cela permet de choisir la covariable significative. En effet, à l'aide des propriétés asymptotiques des estimateurs lorsque le paramètre se situe sur le bord de l'espace des paramètres, nous avons pu construire des tests de la nullité d'un seul ou de plusieurs coefficients. Nos résultats asymptotiques sont obtenus sous des conditions qui ne sont que légèrement plus fortes que ces conditions de moments optimaux, ce qui étend et améliore les résultats existant pour les modèles GARCH sans covariables. Nous pensons que cet article apporte une contribution théorique nouvelle et importante sur la théorie statistique autour de ce modèle.

Après avoir étudié l'influence des variables exogènes sur la volatilité de rendement d'un actif nous continuons à prendre en compte des effets exogènes de variables additionnelles sur la covariance conditionnelle des rendements d'un groupe d'actifs. Dans le troisième chapitre, en adoptant la suggestion de [Engle and Kroner \(1995\)](#), nous nous sommes concentrés sur le modèle BEKK(1,1)-X. Nous avons établi le comportement asymptotique d'estimateurs du VT des paramètres sans imposer la condition iid sur les innovations. Au meilleur de nos connaissances, dans la littérature tous les résultats théoriques de modèles BEKK sans covariables ont été établis sous l'hypothèse iid des innovations. En plus il n'existait pas de recherche sur le modèle BEKK(1,1)-X. Nous pensons que cette

étude est nouvelle et a une contribution dans la domaine de recherche des modèles conditionnellement hétéroscédastiques multivariés. Il serait intéressant de consacrer une étude statistique basée sur des tests d'adéquation pour tester la signification des variables exogènes.

Dans le dernier chapitre de ce manuscrit, nous avons continué à étudier le modèle BEKK(1,1)-X. Les éléments diagonaux de la matrice d'interception et tous les autres paramètres sont estimés par l'approche EbE. Nous avons donné l'estimateur des éléments restants de la matrice d'interception. En effet [Francq and Zakoïan \(2016\)](#) ont développé la méthode EbE et appliqué au modèle BEKK diagonal sans covariables. Cependant il n'a étudié que le comportement de l'estimateur EbE sous l'hypothèse que l'innovation est iid. Dans ce chapitre, nous avons étendu leur résultat pour l'estimation EbE du modèle BEKK(1,1)-X sans imposer l'hypothèse iid sur l'innovation. En plus, nous avons établi la consistance et la normalité asymptotique d'estimateurs des BEKK-X paramètres. Cette étude pourrait être complétée à l'avenir par construire des tests d'adéquation du modèle ainsi des tests des variables exogènes pertinentes.

Pour résumer, ce projet de thèse a été de fournir un travail théorique qui doit permettre d'étudier l'inférence de modèles de la volatilité avec variables exogènes et de développer des tests la signification des variables explicatives. Le travail présenté dans ce manuscrit a également apporté une contribution théorique pour les modèles du type BEKK(1,1)-X. Toutefois, une thèse n'étant jamais terminée. Nous continuons de développer des tests expliqués ci-dessus et étudier l'influence de variables exogènes dans des autres modèles conditionnellement hétéroscédastiques.

Bibliographie

- Andrews, D.W.** (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.
- Andrews, D.W.** (1999) Estimation when a parameter is on a boundary. *Econometrica* 67, 1341–1383.
- Andrews, D.W.** (2001) Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica* 69, 683–734.
- Baillie, R.** (1996) Long memory processes and fractional integration in econometrics. *Journal of Econometrics* 73, 5–59.
- Berkes, I., Horváth, L. and P. Kokoszka** (2003) GARCH processes : structure and estimation. *Bernoulli* 9, 201–227.
- Billingsley, P.** (1961) The Lindeberg-Levy theorem for martingales. *Proceedings of the American Mathematical Society* 12, 788–792.
- Billingsley, P.** (1995) *Probability and measure*. John Wiley.
- Bollerslev, T.** (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.
- Bollerslev, T., R.F. Engle and J.M. Wooldridge** (1988) A capital asset pricing model with time-varying covariances. *The Journal of Political Economy*, 116–131.
- Bollerslev, T.** (1990) Modelling the coherence in short-run nominal exchange rates : a multivariate generalized ARCH model. *The review of economics and statistics* 495–505.
- Bollerslev, T. and J.M. Wooldridge** (1992) Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric reviews* 11, 143–172.
- Bollerslev, T.** (2008) Glossary to ARCH (GARCH). *Volatility and Time Series Econometrics : Essays in Honor of Robert F. Engle* (eds. T. Bollerslev, J. R. Russell and M. Watson), Oxford University Press, Oxford, UK.
- Bouissou, M.B., J-J. Laffont and Q.H. Vuong** (1986) Tests of noncausality under Markov assumptions for qualitative panel data. *Econometrica : Journal of the Econometric Society*, 395–414.

- Bougerol, P. and N. Picard** (1992a) Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics* 52, 115–127.
- Bougerol, P. and N. Picard** (1992b) Strict stationarity of generalized autoregressive processes. *Annals of Probability* 20, 1714–1729.
- Boussama, F., F. Fuchs and R. Stelzer** (2011) Stationarity and geometric ergodicity of BEKK multivariate GARCH models. *Stochastic Processes and their Applications* 121, 2331–2360.
- Brandt, A.** (1986) The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Advance in Applied Probability* 18, 221–254.
- Cakmakli, C. and D.J.C Van Dijk** (2010) Getting the most out of macroeconomic information for predicting stock returns and volatility.
- Christiansen, C., M. Schmeling and A. Schrimpf** (2012) A comprehensive look at financial volatility prediction by economic variables. *Journal of Applied Econometrics* 27, 956–977.
- Comte, F. and O. Lieberman** (2003) Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis* 84, 61–84.
- Ding, Z., C. Granger and R.F. Engle** (1993) A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1, 83–106.
- Engle, R.F., Hendry, D.F. and J.F. Richard** (1983) Exogeneity. *Econometrica* 51, 277–304.
- Engle, R.F., V.K. Ng and K.F. Kroner** (1990) Asset pricing with a factor-ARCH covariance structure : Empirical estimates for treasury bills. *Journal of Econometrics* 45, 213–237.
- Engle, R.F. and K.F. Kroner** (1995) Multivariate simultaneous generalized ARCH. *Econometric theory* 11, 122–150.
- Engle, R.F. and A.J. Patton** (2001) What good is a volatility model. *Quantitative finance* 1, 237–245.
- Engle, R.F. and K. Sheppar** (2001) Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. *National Bureau of Economic Research*.
- Engle, R.F.** (2002) Dynamic conditional correlation : A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics* 20, 339–350.
- Engle, R.F.** (2009) *Anticipating correlations : a new paradigm for risk management*. Princeton University Press.
- Escanciano, J.C.** (2009) Quasi-maximum likelihood estimation of semi-strong GARCH models. *Econometric Theory* 25, 561–570.
- Fisher, T.J. and C.M. Gallagher** (2012) New weighted portmanteau statistics for time series goodness of fit testing. *Journal of the American Statistical Association* 107, 777–787.

- Florens, J.-P. and M. Mouchart** (1982) A note on noncausality. *Econometrica : Journal of the Econometric Society*, 583–591.
- Francq, C. and J.-M. Zakoïan** (1998) Estimating linear representations of nonlinear processes. *Journal of Statistical Planning and Inference* 68, 145–165.
- Francq, C. and J.-M. Zakoïan** (2000) Covariance matrix estimation for estimators of mixing weak ARMA models. *Journal of Statistical Planning and Inference* 83, 369–394.
- Francq, C. and J.-M. Zakoïan** (2004) Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* 10, 605–637.
- Francq, C., R. Roy and J.-M. Zakoïan** (2005) Diagnostic checking in ARMA models with uncorrelated errors. *Journal of the American Statistical Association* 100, 532–544.
- Francq, C. and J.-M. Zakoïan** (2007) Quasi-Maximum Likelihood Estimation in GARCH Processes when some coefficients are equal to zero. *Stochastic Processes and their Applications* 117, 1265–1284.
- Francq, C. and J.-M. Zakoïan** (2009a) Bartlett’s formula for a general class of nonlinear processes. *Journal of Time Series Analysis* 30, 449–465.
- Francq, C. and J.-M. Zakoïan** (2009b) Testing the nullity of GARCH coefficients : correction of the standard tests and relative efficiency comparisons. *Journal of the American Statistical Association* 104, 313–324.
- Francq, C. and J.-M. Zakoïan** (2010) *GARCH Models : Structure, Statistical Inference and Financial Applications*. John Wiley.
- Francq, C., L. Horváth and J.-M. Zakoïan** (2011) Merits and drawbacks of variance targeting in GARCH models. *Journal of Financial Econometrics* 9, 619–656.
- Francq, C., L. Horváth and J.-M. Zakoïan** (2014) Variance targeting estimation of multivariate GARCH models. *Journal of Financial Econometrics*, https://mpra.ub.uni-muenchen.de/57794/1/MPRA_paper_57794.pdf
- Francq, C. and Thieu, L.Q.** (2015) Qml inference for volatility models with covariates. *University Library of Munich, Germany*, No. 63198.
- Francq, C. and G. Sucarrat** (2015) Equation-by-Equation Estimation of a Multivariate Log-GARCH-X Model of Financial Returns.
- Fuertes, A.M., M. Izzeldin and E. Kalotychou** (2009) On forecasting daily stock volatility : the role of intraday information and market conditions. *International Journal of Forecasting* 25, 259–281.
- Glosten, L.R., R. Jaganathan and D. Runkle** (1993) On the relation between the expected values and the volatility of the nominal excess return on stocks. *Journal of Finance* 48, 1779–1801.

- Guo, S., Ling, S., and Zhu, K.** (2014) Factor double autoregressive models with application to simultaneous causality testing. *Journal of Statistical Planning and Inference* 148, 82–94.
- Hafner, C.M. and A. Preminger** (2009) On asymptotic theory for multivariate GARCH models. *Journal of Multivariate Analysis* 100, 2044–2054.
- Hall, P. and Q. Yao** (2003) Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica* 71, 285–317.
- Hamadeh, T. and J-M. Zakoïan** (2011) Asymptotic properties of LS and QML estimators for a class of nonlinear GARCH processes. *Journal of Statistical Planning and Inference* 141, 488–507.
- Han, H. and D. Kristensen** (2014) Asymptotic Theory for the QMLE in GARCH-X Models With Stationary and Nonstationary Covariates. *Journal of Business & Economic Statistics* 32, 416–429.
- Herrndorf, N.** (1984) A functional central limit theorem for weakly dependent sequences of random variables. *The Annals of Probability*, 141–153.
- Kanas, A.** (1984) Modelling the risk-return relation for the SP100 : The role of VIX. *Economic Modelling* 29, 795–809.
- Kenneth S.M.** (1981) On the Inverse of the Sum of Matrices. *Mathematics Magazine*. <http://www.jstor.org/stable/2690437>, 54, 67–72
- Koopmans, T.C. and O. Reiersol** (1950) The identification of structural characteristics. *The Annals of Mathematical Statistics* 21, 165–181.
- Kristensen, D. and A. Rahbek** (2005) Asymptotics of the QMLE for a class of ARCH(q) models. *Econometric Theory* 21, 946–961.
- Laurent, S., C. Lecourt and F.C. Palm** (2014) Testing for jumps in conditionally Gaussian ARMA-GARCH models, a robust approach. *Computational Statistics & Data Analysis*. <http://dx.doi.org/10.1016/j.csda.2014.05.015>
- Li, W.K. and T.K. Mak** (1994) On the Squared Residual Autocorrelations in Non-Linear Time Series With Conditional Heteroskedasticity. *Journal of Time Series Analysis* 15, 627–636.
- Ling, S.** (2007) Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models. *Journal of Econometrics* 140, 849–873.
- Mezrich, J. and R. F. Engle** (1996) GARCH for groups. *Risk* 9, 36–40.
- Nelson, D. B.** (1991) Conditional heteroskedasticity in asset returns : A new approach. *Econometrica* 59, 347–370.
- Newey, W.K. and K.D. West** (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix. *Econometrica* 55, 703–708.

- Nijman, T. and E. Sentana** (1996) Marginalization and contemporaneous aggregation in multivariate GARCH processes. *Journal of Econometrics* 71, 71–87.
- Pan, J., H. Wang and H. Tong** (2008) Estimation and tests for power-transformed and threshold GARCH models. *Journal of Econometrics* 142, 352–378.
- Pedersen, R.S. and A. Rahbek** (2014) Multivariate variance targeting in the BEKK-GARCH model. *The Econometrics Journal* 17, 24–55.
- Pedersen, R. S.** (2015) Inference and Testing on the Boundary in Extended Constant Conditional Correlation GARCH Models. Available at SSRN : <http://dx.doi.org/10.2139/ssrn.2656058>
- Phillips, P.C.B., Y. Sun and S. Jin** (2003) Consistent HAC estimation and robust regression testing using sharp origin kernels with no truncation. *Discussion paper, Yale University*.
- Silvennoinen, A. and T. Teräsvirta** (2009) Modeling multivariate autoregressive conditional heteroskedasticity with the double smooth transition conditional correlation GARCH model. *Journal of Financial Econometrics* 7, 373–411.
- Sucarrat, G. and A. Escribano** (2010) The Power Log-GARCH Model. *Working document*, Economic Series 10-13, University Carlos III, Madrid.
- Sucarrat, G., S. Grønneberg and Á. Escribano** (2016) Estimation and Inference in Univariate and Multivariate Log-GARCH-X Models When the Conditional Density is Unknown. *Computational Statistics & Data Analysis* 100, 582-594.
- Taylor, S.J.** (1986) *Modelling Financial Time Series*, New-York : Wiley.
- Thieu, L.Q.** (2016) *Variance targeting estimation of the BEKK-X model*, Working paper.
- Wintenberger, O.** (2013) Continuous invertibility and stable QML estimation of the EGARCH(1,1) model. *Scandinavian Journal of Statistics* 40, 846–867.
- Zakoïan, J-M.** (1994) Threshold heteroskedastic models. *Journal of Economic Dynamics and Control* 18, 931–955.