Estimating multivariate GARCH and Stochastic Correlation models equation by equation

Christian Francq

CREST and Université de Lille (EQUIPPE)

Jean-Michel Zakoïan

CREST and Université de Lille (EQUIPPE)

October 31, 2014

Abstract. This paper investigates the estimation of a wide class of multivariate volatility models. Instead of estimating a m-multivariate volatility model, a much simpler and numerically efficient method consists in estimating m univariate GARCH-type models Equation by Equation (EbE) in the first step, and a correlation matrix in the second step. Strong consistency and asymptotic normality (CAN) of the EbE estimator are established in a very general framework, including Dynamic Conditional Correlation (DCC) models. The EbE estimator can be used to test the restrictions imposed by a particular MGARCH specification. For general Constant Conditional Correlation (CCC) models, we obtain the CAN of the two-step estimator. Comparisons with the global method, in which the model parameters are estimated in one step, are provided. Monte-Carlo experiments and applications to financial series illustrate the interest of the approach.

Keywords: Constant conditional correlation, Dynamic conditional correlation, Markov switching correlation matrix, Multivariate GARCH specification testing, Quasi maximum likelihood estimation.

Address for correspondence: Jean-Michel Zakoïan, CREST, 15 Boulevard Gabriel Péri, 92245 Malakoff cedex.

Email: zakoian@ensae.fr

1. Introduction

Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models have featured prominently in the analysis of financial time series. The last twenty years have witnessed significant research devoted to the multivariate extension of the concepts and models initially developed for univariate GARCH. Among the numerous specifications of multivariate GARCH (MGARCH) models, the most popular seem to be the Constant Conditional Correlations (CCC) model introduced by Bollerslev (1990) and extended by Jeantheau (1998), the BEKK model developed by Baba, Engle, Kraft and Kroner, in a preliminary version of Engle and Kroner (1995), and the Dynamic Conditional Correlations (DCC) models proposed by Tse and Tsui (2002) and Engle (2002). Reviews on the rapidly changing literature on MGARCH are Bauwens, Laurent and Rombouts (2006), Silvennoinen and Teräsvirta (2009), Francq and Zakoïan (2010, Chapter 11), Bauwens, Hafner and Laurent (2012), Tsay (2014, Chapter 7).

The complexity of MGARCH models has been a major obstacle to their use in applied works. Indeed, in asset pricing applications or portfolio management, cross-sections of hundreds of stocks are common. However, as the dimension of the cross section increases, the number of parameters can become very large in MGARCH models, making estimation increasingly cumbersome. This "dimensionality curse" is general in multivariate time series, but is particularly problematic in GARCH models. The reason is that the parameters of interest are involved in the conditional variance matrix, which has to be inverted in Gaussian likelihood-based estimation methods.

Existing approaches to alleviate the dimensionality curse rely on either constraining the structure of the model in order to reduce the number of parameters, or using an alternative estimation criterion. Examples of models belonging to the first category are the Factor ARCH models of Engle, Ng and Rotschild (1990), the Generalized Orthogonal GARCH model of van der Weide (2002), and the Generalized Orthogonal Factor GARCH model of Lanne and Saikkonen (2007). The second strategy was advocated by Engle, Shephard and Sheppard (2008), who suggested to use a composite likelihood instead of the usual quasi-likelihood. An approach combining the two concepts, reduction of the parameter dimension and use of a partial likelihood, was recently proposed by Engle and Kelly (2012) who introduced the Dynamic Equicorrelation (DECO) model.

A solution to the high-dimension problem which does not preclude a high-dimensional parameter set relies on two steps. In the first stage, univariate GARCH models are estimated for each individual series, *equation by equation*, and in the second stage, standardized residuals are used to estimate the parameters of the dynamic correlation. This approach, initially proposed by Engle and Sheppard (2001) and Engle (2002) in the context of DCC models, was advocated by Pelletier (2006) for regime-switching dynamic correlation models, by Aielli (2013) for DCC models, and it was used in several empirical studies (see e.g. Hafner and Reznikova (2012), Sucarrat, Grønneberg and Escribano (2013) for recent references). However, the statistical properties of such two-step estimators have not been established[1].

The first goal of the present paper is to develop asymptotic results for the Equation-by-Equation (EbE) estimator *of the volatility parameters*, based on the Quasi-Maximum Likelihood (QML). Our framework for the individual volatilities specification is extremely general. First, the conditional variance of component k is a parametric function of the past of *all components* of the vector of returns. This allows to capture serial dependencies between components, that do not appear in the conditional correlation matrix. Second, the volatilities, being specified as any parametric functions of the past returns, are able to accommodate leverage-effects or any other type of "nonlinearity". One issue of interest, as far as the asymptotic theory is concerned, is whether individual estimation of the conditional variances necessarily entails an efficiency loss with respect to a global QML method which estimates them jointly.

Apart from the numerical simplicity, one advantage of this approach is that the derivation of EbE estimators (EbEE) is independent from the specification of a conditional correlation matrix. It can therefore be employed for CCC as well as for DCC GARCH models, leading to the same estimators of the individual volatilities. It can also be used for multivariate models that are not GARCH. We consider a class of *Stochastic Correlation* (SC) models which have the same multiplicative form as GARCH-type models, except that the correlation matrix is not a measurable function of the past observations. The term stochastic correlation obviously refers to the class of Stochastic Volatility models, which differ from GARCH by the fact that the volatility depends on unobservable stochastic factors.

Another aim of this paper is to provide asymptotic results for the second step of the

[1]See the recent survey by Caporin and McAleer (2012) for a discussion of the existence, or the absence, of asymptotic results for multivariate GARCH models.

two-stage approach, that is the estimation of a time-varying correlation matrix using the standardized returns obtained in the first step. At this stage, a specification of the conditional correlation dynamic is required. For CCC models, the constant conditional correlation matrix can be estimated by the empirical correlation matrix of the EbEE residuals. In this article, we derive asymptotic results for this estimator, which can be seen as an extension of the two-step estimator proposed by Engle and Sheppard (2001) in the case where the individual volatilities have pure GARCH forms with iid innovations. For some DCC and SC models, the structure of the time-varying correlation can also be estimated.

The paper is organized as follows. Section 2 presents the assumptions and notations for the class of multivariate processes studied in this article. Such assumptions are discussed under different specifications of the correlation matrix $\boldsymbol{R}_t$. In Section 3, we study the estimation of the volatility parameters without any assumption on $\boldsymbol{R}_t$. Particular parameterizations are discussed in Section 4. Section 5 develops the two-step estimation method when the correlation matrix $\boldsymbol{R}_t$ is parameterized. Numerical illustrations are presented in Section 6. Section 7 concludes. The most technical assumptions and the proofs of the main theorems are collected in the Appendix. Due to space restrictions, several proofs, along with additional numerical illustrations, are included in a supplementary file.

2. Models and assumptions

Let $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \cdots, \epsilon_{mt})'$ be a $\mathbb{R}^m$-valued process and let $\mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}$ be the σ-field generated by $\{\boldsymbol{\epsilon}_u, u < t\}$. Assume

$$E(\boldsymbol{\epsilon}_t \mid \mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}) = \mathbf{0}, \qquad \mathrm{Var}(\boldsymbol{\epsilon}_t \mid \mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}) = \boldsymbol{H}_t \quad \text{exists and is positive definite.} \tag{2.1}$$

Denoting by σ_{kt}^2 the diagonal elements of $\boldsymbol{H}_t$, that is the variances of the components of $\boldsymbol{\epsilon}_t$ conditional on $\mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}$, we introduce the vector

$$\boldsymbol{\eta}_t^* = \boldsymbol{D}_t^{-1}\boldsymbol{\epsilon}_t = (\epsilon_{1t}/\sigma_{1t}, \dots, \epsilon_{mt}/\sigma_{mt})' \quad \text{where } \boldsymbol{D}_t = \mathrm{diag}(\sigma_{1t}, \dots, \sigma_{mt}).$$

By (2.1) we have, $E(\boldsymbol{\eta}_t^* \mid \mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}) = \mathbf{0}$ and the conditional correlation matrix of $\boldsymbol{\epsilon}_t$ is given by

$$\boldsymbol{R}_t = \mathrm{Var}(\boldsymbol{\eta}_t^* \mid \mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}) = \boldsymbol{D}_t^{-1}\boldsymbol{H}_t\boldsymbol{D}_t^{-1}. \tag{2.2}$$

It follows that, for $k = 1, \dots, m$,

$$E(\eta_{kt}^* \mid \mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}) = 0, \qquad \mathrm{Var}(\eta_{kt}^* \mid \mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}) = 1. \tag{2.3}$$

Introducing the vector $\boldsymbol{\eta}_t$ such that $\boldsymbol{\eta}_t^* = \boldsymbol{R}_t^{1/2}\boldsymbol{\eta}_t$, the previous equations can be summarized as follows. The square root has to be understood in the sense of the Cholesky factorization, that is, $\boldsymbol{R}_t^{1/2}(\boldsymbol{R}_t^{1/2})' = \boldsymbol{R}_t$ and $\boldsymbol{H}_t^{1/2}(\boldsymbol{H}_t^{1/2})' = \boldsymbol{H}_t$.

Assumptions and notations: *The $\mathbb{R}^m$-valued process $(\boldsymbol{\epsilon}_t)$ satisfies*

$$\begin{cases} \boldsymbol{\epsilon}_t &= \boldsymbol{H}_t^{1/2}\boldsymbol{\eta}_t, \qquad E(\boldsymbol{\eta}_t \mid \mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}) = \boldsymbol{0}, \quad Var(\boldsymbol{\eta}_t \mid \mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}) = \boldsymbol{I}_m, \\ \\ \boldsymbol{H}_t &= \boldsymbol{H}(\boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \ldots) = \boldsymbol{D}_t\boldsymbol{R}_t\boldsymbol{D}_t, \end{cases} \tag{2.4}$$

where $\boldsymbol{H}_t$ is positive definite, $\boldsymbol{D}_t = \{diag(\boldsymbol{H}_t)\}^{1/2}$ and $\boldsymbol{R}_t = Corr(\boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_t \mid \mathcal{F}_{t-1}^{\boldsymbol{\epsilon}})$.

We assume that the conditional variance of the k-th component of $\boldsymbol{\epsilon}_t$ is parameterized by some parameter $\boldsymbol{\theta}_0^{(k)} \in \mathbb{R}^{d_k}$, so that

$$\begin{cases} \epsilon_{kt} &= \sigma_{kt}\eta_{kt}^*, \\ \sigma_{kt} &= \sigma_k(\boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \ldots; \boldsymbol{\theta}_0^{(k)}), \end{cases} \tag{2.5}$$

where σ_k is a positive function. In view of (2.3), the process $(\boldsymbol{\eta}_t^*)$ can be called the vector of EbE innovations of $(\boldsymbol{\epsilon}_t)$.

Remark 2.1. In Model (2.4)-(2.5), the volatility of any component of $\boldsymbol{\epsilon}_t$ is allowed to depend on the past values of all components. This assumption represents an extension of the classical set up of univariate GARCH models and, for this reason, Model (2.5) can be referred to as an *augmented GARCH* model in the terminology of Hörmann (2008). This extension is firstly motivated by the sake of generality: it seems very restrictive to assume that the conditional variance of a component is not influenced by the past of other components. On the other hand, the EbE estimation approach of the paper makes this extension amenable to statistical inference, without causing an explosion in the number of parameters. For instance, if the individual volatilities have GARCH(1,1)-type dynamics,

$$\sigma_{kt}^2 = \omega_k + \sum_{\ell=1}^{m} \alpha_{k,\ell}\epsilon_{\ell,t-1}^2 + \beta_k\sigma_{k,t-1}^2, \quad \omega_k > 0, \alpha_{k,\ell} \geq 0, \beta_k \geq 0, \tag{2.6}$$

increasing by K the number of components entails an additional number of K parameters by equation. Finally, this extension allows to tackle the problem of asynchronous data by allowing each conditional variance to depend on the most recent observations. See Section 6.2.1 for more details on this issue.

Remark 2.2. A variety of parametric forms of function $\boldsymbol{H}$ has been introduced in the literature. In particular, a standard specification of $\boldsymbol{D}_t$ is, in vector form, given by

$$\boldsymbol{h}_t = \boldsymbol{\omega} + \sum_{i=1}^{q} \boldsymbol{A}_i \underline{\boldsymbol{\epsilon}}_{t-i} + \sum_{j=1}^{p} \boldsymbol{B}_j \boldsymbol{h}_{t-j} \qquad (2.7)$$

where $\boldsymbol{h}_t = \left(\sigma_{1t}^2, \cdots, \sigma_{mt}^2\right)'$, $\underline{\boldsymbol{\epsilon}}_t = \left(\epsilon_{1t}^2, \cdots, \epsilon_{mt}^2\right)'$, $\boldsymbol{A}_i$ and $\boldsymbol{B}_j$ are $m \times m$ matrices with positive coefficients and $\boldsymbol{\omega} = (\omega_1, \cdots, \omega_m)'$ is a vector of strictly positive coefficients. When $p = q = 1$ and $\boldsymbol{B}_1$ is diagonal, the individual volatilities satisfy a dynamic of the form (2.6).

Remark 2.3. The positivity of the function σ_k generally entails restrictions on the parameter values which cannot be made explicit under the general formulation. For particular models such constraints can be explicited, as in (2.6). Note that the EbE innovations η_{kt}^* are not iid in general, and thus (2.5) is not a Data Generating Process (DGP).

We now consider two classes of DGP satisfying the previous assumptions.

2.1. GARCH-type models

Consider a GARCH process, defined as a non anticipative[2] solution of

$$\boldsymbol{\epsilon}_t = \boldsymbol{D}_t \boldsymbol{R}_t^{1/2} \boldsymbol{\eta}_t, \quad \text{where } (\boldsymbol{\eta}_t) \text{ is an iid sequence.} \qquad (2.8)$$

Obviously, $(\boldsymbol{\epsilon}_t)$ thus satisfies (2.4). In this paper, we will distinguish CCC models, for which

$$\boldsymbol{R}_t = \boldsymbol{R} \text{ is a constant correlation matrix,} \qquad (2.9)$$

from DCC models where $\boldsymbol{R}_t$ is a non constant function of the past of $\boldsymbol{\epsilon}_t$, that is,

$$\boldsymbol{R}_t = \boldsymbol{R}(\boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \ldots) \neq \boldsymbol{R}.$$

Note that in the case of CCC models, the sequence $(\boldsymbol{\eta}_t^*)$ is iid which is generally not the case for DCC models. In the econometric literature, CCC models are generally introduced under the specification (2.7) of the individual conditional variances[3]. To avoid confusion, we will refer to (2.7)-(2.9) as the CCC-GARCH(p, q) model .

[2]that is $\boldsymbol{\epsilon}_t \in \mathcal{F}_t^{\boldsymbol{\eta}}$, the σ-field generated by $\{\boldsymbol{\eta}_u, u \leq t\}$.

[3]Bollerslev (1990) introduced this model in the case of diagonal matrices $\boldsymbol{A}_i$ and $\boldsymbol{B}_j$. Ling and McAleer (2003) proved the asymptotic properties of a general version of this model (without any diagonality assumption) subsequently called the *Extended* CCC model by He and Teräsvirta (2004).

2.2. Stochastic Correlation Models

To obtain a DGP satisfying (2.4), an alternative to GARCH-type models is to introduce correlation matrices that are not only function of the past but also depend on some latent process (Δ_t). More precisely, let

$$\boldsymbol{\epsilon}_t = \boldsymbol{D}_t \boldsymbol{R}_t^{*1/2} \boldsymbol{\xi}_t, \tag{2.10}$$

where $(\boldsymbol{\xi}_t)$ is an iid $(\mathbf{0}, \boldsymbol{I}_m)$ sequence and

$$\boldsymbol{R}_t^* = \boldsymbol{R}^*(\boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \dots, \Delta_t), \qquad \Delta_t \notin \mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}. \tag{2.11}$$

By analogy with the so-called Stochastic Volatility models, in which the volatility is not a measurable function of the past observables, we can call model (2.10)-(2.11) a *Stochastic Correlation* (SC) model. For this model, the individual volatilities σ_{kt}, as given by (2.5), are of GARCH-type, while the correlations between components in $\boldsymbol{R}_t^*$ are not. In this context, a non anticipative solution of the model is such that $\boldsymbol{\epsilon}_t \in \mathcal{F}_t^{\boldsymbol{\xi},\Delta}$, the σ-field generated by $\{\boldsymbol{\xi}_u, \Delta_u, u \leq t\}$. Assuming that

$$(\boldsymbol{\epsilon}_t) \text{ is a non anticipative solution and } \boldsymbol{\xi}_t \text{ is independent from } \mathcal{F}_t^{\Delta}, \tag{2.12}$$

the σ-field generated by $\{\Delta_u, u \leq t\}$, we have $E(\boldsymbol{\epsilon}_t \mid \mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}) = \mathbf{0}$, and

$$\boldsymbol{H}_t \quad = \quad \mathrm{Var}(\boldsymbol{\epsilon}_t \mid \mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}) = \boldsymbol{D}_t E(\boldsymbol{R}_t^{*1/2} \boldsymbol{\xi}_t \boldsymbol{\xi}_t' \boldsymbol{R}_t^{*'1/2} \mid \mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}) \boldsymbol{D}_t = \boldsymbol{D}_t E(\boldsymbol{R}_t^* \mid \mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}) \boldsymbol{D}_t,$$

using the fact that $E(\boldsymbol{\xi}_t \boldsymbol{\xi}_t') = \boldsymbol{I}_m$. Note that $\boldsymbol{R}_t = E(\boldsymbol{R}_t^* \mid \mathcal{F}_{t-1}^{\boldsymbol{\epsilon}})$.

Therefore, SC models (2.10)-(2.12) satisfy Assumptions (2.4). Note that the three innovations sequences are linked by

$$\boldsymbol{\eta}_t^* = \boldsymbol{R}_t^{*1/2} \boldsymbol{\xi}_t = \boldsymbol{R}_t^{1/2} \boldsymbol{\eta}_t.$$

3. Equation-by-equation estimation of volatility parameters in MGARCH models

In this section, we are interested in estimating the conditional variance of each component of $\boldsymbol{\epsilon}_t$ satisfying (2.4). In other words, we study the estimation of the parameter $\boldsymbol{\theta}_0^{(k)}$ in the augmented GARCH model (2.5), under (2.3), for $k = 1, \dots, m$.

To estimate $\boldsymbol{\theta}_0^{(k)}$ we will use the Gaussian QML, which is the most widely used estimation method for univariate GARCH models, but other methods could be considered as well (for instance the LAD method or the weighted QML studied by Ling (2007), the non

Gaussian QML studied by Berkes and Horváth (2004)). In view of Remarks 2.1 and 2.3, the augmented GARCH model (2.5) is not, in general, a univariate GARCH and we cannot directly rely on existing results for its estimation.

Given observations $\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_n$, and arbitrary initial values $\tilde{\boldsymbol{\epsilon}}_i$ for $i \leq 0$, we define $\tilde{\sigma}_{kt}(\boldsymbol{\theta}^{(k)}) = \sigma_k(\boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \dots, \boldsymbol{\epsilon}_1, \tilde{\boldsymbol{\epsilon}}_0, \tilde{\boldsymbol{\epsilon}}_{-1}, \dots; \boldsymbol{\theta}^{(k)})$ for $k = 1, \dots, m$ and $\boldsymbol{\theta}^{(k)} \in \boldsymbol{\Theta}_k$, assuming that $\boldsymbol{\Theta}_k$ is a compact parameter set and $\boldsymbol{\theta}_0^{(k)} \in \boldsymbol{\Theta}_k$. This random variable will be approximated by $\sigma_{kt}(\boldsymbol{\theta}^{(k)}) = \sigma_k(\boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \dots; \boldsymbol{\theta}^{(k)})$.

Let $\hat{\boldsymbol{\theta}}_n^{(k)}$ denote the equation-by-equation estimator of $\boldsymbol{\theta}_0^{(k)}$:

$$\hat{\boldsymbol{\theta}}_n^{(k)} = \arg \min_{\boldsymbol{\theta}^{(k)} \in \boldsymbol{\Theta}^{(k)}} \tilde{Q}_n^{(k)}(\boldsymbol{\theta}^{(k)}), \quad \tilde{Q}_n^{(k)}(\boldsymbol{\theta}^{(k)}) = \frac{1}{n} \sum_{t=1}^{n} \log \tilde{\sigma}_{kt}^2 \left(\boldsymbol{\theta}^{(k)}\right) + \frac{\epsilon_{kt}^2}{\tilde{\sigma}_{kt}^2 \left(\boldsymbol{\theta}^{(k)}\right)}.$$

3.1. Consistency and asymptotic normality of the EbEE

We make the following assumption on the process $(\boldsymbol{\epsilon}_t)$.

A1: $(\boldsymbol{\epsilon}_t)$ is a strictly stationary and ergodic process satisfying (2.4), with $E|\epsilon_{kt}|^s < \infty$ for some $s > 0$. Moreover, $E \log \sigma_{kt}^2 < \infty$.

This assumption will be made more explicit for specific models in Section 4 (see also Theorem 2.1 and Corollary 2.2 in Francq and Zakoïan (2012)). Technical assumptions on the function σ_k are relegated to Appendix A. We also assume the existence of a lower bound, ensuring that the criterion be well defined for any parameter value.

A4: we have $\sigma_{kt}(\cdot) > \underline{\omega}$ for some $\underline{\omega} > 0$.

Assumptions **A4-A6** are required for the consistency. To prove the asymptotic normality, we need to assume

A7: $\boldsymbol{\theta}_0^{(k)}$ belongs to the interior of $\boldsymbol{\Theta}^{(k)}$,

A8: $E\left|\eta_{kt}^*\right|^{4(1+\delta)} < \infty$, for some $\delta > 0$,

and some additional technical assumptions **A9-A12**.

THEOREM 3.1. *If* **A1** *and* **A4**-**A6** *hold, the EbEE of* $\boldsymbol{\theta}_0^{(k)}$ *in the augmented GARCH model* (2.5) *is strongly consistent*

$$\hat{\boldsymbol{\theta}}_n^{(k)} \to \boldsymbol{\theta}_0^{(k)}, \quad a.s. \quad as\ n \to \infty. \tag{3.1}$$

If, in addition, **A7**-**A12** *hold, then*

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n^{(k)} - \boldsymbol{\theta}_0^{(k)}\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left\{0, \boldsymbol{J}_{kk}^{-1}\boldsymbol{I}_{kk}\boldsymbol{J}_{kk}^{-1}\right\}, \tag{3.2}$$

where

$$\boldsymbol{I}_{kk} = E\left(\{\eta_{kt}^{*4} - 1\}\boldsymbol{d}_{kt}\boldsymbol{d}_{kt}'\right), \quad \boldsymbol{J}_{kk} = E\left(\boldsymbol{d}_{kt}\boldsymbol{d}_{kt}'\right), \quad \boldsymbol{d}_{kt} = \frac{1}{\sigma_{kt}^2}\frac{\partial\sigma_{kt}^2(\boldsymbol{\theta}_0^{(k)})}{\partial\boldsymbol{\theta}^{(k)}}.$$

Note that the sequence $(\boldsymbol{\eta}_t)$ in (2.4) is not assumed to be iid. It is only assumed to be a conditionally homoscedastic martingale difference, as in Bollerslev and Wooldridge (1992), which allows us to encompass SC models. The analogous of this result was established, in the case of semi-strong univariate GARCH(p, q) models, by Escanciano (2009) as an extension of Berkes, Horváth and Kokoszka (2003) and Francq and Zakoïan (2004).

3.2. Comparison with the theoretical QML estimator

A question of interest is whether the EbEE approach necessarily entails an efficiency loss (the price paid for its simplicity) with respect to a QML method in which the volatility parameters are jointly estimated. To be able to write the global quasi-likelihood, it is necessary to specify the conditional correlation matrix. Because we wish to compare the estimators of the volatility parameters, we consider the global QML estimator (QMLE) of $\boldsymbol{\theta}_0$ based on the assumption that the matrix $\boldsymbol{R}_t$ is constant and is known[4].

A theoretical QMLE of $\boldsymbol{\theta}_0$ is defined as any measurable solution $\hat{\boldsymbol{\theta}}_n^{QML}$ of

$$\hat{\boldsymbol{\theta}}_n^{QML} = \arg\min_{\boldsymbol{\theta}\in\Theta} n^{-1}\sum_{t=1}^{n}\tilde{\ell}_t(\boldsymbol{\theta}), \qquad \tilde{\ell}_t(\boldsymbol{\theta}) = \boldsymbol{\epsilon}_t'\tilde{\boldsymbol{H}}_t^{-1}\boldsymbol{\epsilon}_t + \log|\tilde{\boldsymbol{H}}_t|,$$

where $\tilde{\boldsymbol{H}}_t = \tilde{\boldsymbol{D}}_t\boldsymbol{R}\tilde{\boldsymbol{D}}_t$ and $\tilde{\boldsymbol{D}}_t = \mathrm{diag}(\tilde{\sigma}_{1t}(\boldsymbol{\theta}^{(1)}), \ldots, \tilde{\sigma}_{mt}(\boldsymbol{\theta}^{(m)}))$. Let $\boldsymbol{R}^{-1} = (r_{k\ell}^*)$. Let the $d \times d$ matrix $\boldsymbol{M} = (\boldsymbol{M}_{k\ell})$ where

$$\boldsymbol{M}_{k\ell} = \tau_{k\ell}\boldsymbol{J}_{k\ell} - \sum_{i,j=1}^{m}\xi_{ki}\xi_{j\ell}(\kappa_{ij} - 1)\boldsymbol{J}_{ki}\boldsymbol{J}_{ii}^{-1}\boldsymbol{J}_{ij}\boldsymbol{J}_{jj}^{-1}\boldsymbol{J}_{j\ell},$$

and

$$\kappa_{k\ell} = E\left(\eta_{kt}^{*2}\eta_{\ell t}^{*2}\right), \quad \xi_{k\ell} = \frac{1}{2}\begin{cases} r_{kk}^* + 1 & \text{if} \quad k = \ell \\ r_{k\ell}r_{k\ell}^* & \text{if} \quad k \neq \ell \end{cases}, \quad \tau_{k\ell} = \sum_{i,j=1}^{m} r_{ki}^* r_{\ell j}^* E(\eta_{kt}^*\eta_{it}^*\eta_{jt}^*\eta_{\ell t}^*) - 1.$$

[4]We can thus call this estimator theoretical QMLE, or infeasible QMLE.

Proposition 3.1. *Under the assumptions of Theorem 3.1, the QMLE of the volatility parameters, assuming that $\boldsymbol{R}_t = \boldsymbol{R}$ is known, is asymptotically more efficient (resp. less efficient) than the EbEE if and only if $\boldsymbol{M}$ is negative definite (resp. positive definite).*

When $\boldsymbol{R}$ is the identity matrix, $\boldsymbol{M} = \boldsymbol{0}$ and the two methods are equivalent, producing the same estimators. In practical implementation of the QML, the matrix $\boldsymbol{R}$ has to be estimated, which may lower the accuracy of the volatility parameters estimators. It is interesting to note that the QMLE is not always asymptotically more efficient than the EbEE, even in the favorable situation where $\boldsymbol{R}$ is known (which has no consequence for the EbEE). Calculations reported in the supplementary file show that the EbEE may be asymptotically superior to the QMLE when the distribution of $(\boldsymbol{\eta}_t^*)$ is sufficiently far from the Gaussian.

3.3. Asymptotic results for strong augmented GARCH models

The asymptotic distribution of the EbEE can be simplified under the assumption that

$$\eta_{kt}^* \text{ is independent from } \mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}. \tag{3.3}$$

Moreover, **A8** can be replaced by the weaker assumption

A8*: $E\left|\eta_{kt}^*\right|^4 < \infty$,

and the technical assumptions **A10** on the volatility function can be slightly weakened (see **A10*** in Appendix A). The asymptotic distribution of the EbEE is modified as follows.

Theorem 3.2. *Under* (3.3) *and the assumptions of Theorem 3.1, with* **A8** *replaced by* **A8*** *and* **A10** *replaced by* **A10****, we have*

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n^{(k)} - \boldsymbol{\theta}_0^{(k)}\right) \quad \stackrel{\mathcal{L}}{\to} \quad \mathcal{N}\left\{0, (E\eta_{kt}^{*4} - 1)\boldsymbol{J}_{kk}^{-1}\right\}.$$

It can be noted that (3.3) is always satisfied in the CCC case, that is, under (2.8) and $\boldsymbol{R}_t = \boldsymbol{R}$. The next result shows that (3.3) can be satisfied for other GARCH-type models.

Proposition 3.2. *Assume that the distribution of $\boldsymbol{\eta}_t$ is spherical in Model* (2.8)*. Then* (3.3) *is satisfied. Moreover, (η_{kt}^*) is an iid (0,1) sequence.*

Remark 3.1. It is worth noting that, under the assumptions of Proposition 3.2, the process $(\boldsymbol{\eta}_t^*)$ is neither independent nor identically distributed in general (even if its components are iid). To see this, consider for example, for $\lambda_1, \lambda_2 \in \mathbb{R}$ and for $k \neq \ell$,

$$\lambda_1 \eta_{kt}^* + \lambda_2 \eta_{\ell t}^* \stackrel{d}{=} \|(\lambda_1 \boldsymbol{e}_k' + \lambda_2 \boldsymbol{e}_\ell')\boldsymbol{R}_t^{1/2}\|\eta_{1t} = \{\lambda_1^2 + \lambda_2^2 + 2\lambda_1\lambda_2 \boldsymbol{R}_t(k,\ell)\}^{1/2}\eta_{1t},$$

conditionally on $\mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}$, where $\boldsymbol{e}_k$ denotes the k-th column of $\boldsymbol{I}_m$. The variable in the right-hand side of the latter equality is in general non independent of the past values of $\boldsymbol{\eta}_t^*$.

Because SC models (2.10)-(2.12) satisfy Assumptions (2.4), the volatility parameters $\boldsymbol{\theta}_0^{(k)}$ can be estimated equation by equation, and Theorem 3.1 applies. We show in the supplementary file that (3.3) holds for SC models if the distribution of $\boldsymbol{\xi}_t$ is spherical, the sequences $(\boldsymbol{\Delta}_t)$ and $(\boldsymbol{\xi}_t)$ are independent, and $\mathcal{F}_{t-1}^{\boldsymbol{\epsilon}} = \mathcal{F}_{t-1}^{\boldsymbol{\eta}^*}$.

3.4. Adding an intercept

We consider an extension of Model (2.4) in which an intercept is included. The $\mathbb{R}^m$-valued process $(\boldsymbol{y}_t)$ is supposed to satisfy

$$\begin{cases} \boldsymbol{y}_t &= \boldsymbol{\mu}_0 + \boldsymbol{H}_t^{1/2}\boldsymbol{\eta}_t, \qquad E(\boldsymbol{\eta}_t \mid \mathcal{F}_{t-1}^{\boldsymbol{y}}) = \boldsymbol{0}, \quad \mathrm{Var}(\boldsymbol{\eta}_t \mid \mathcal{F}_{t-1}^{\boldsymbol{y}}) = \boldsymbol{I}_m, \\ \boldsymbol{H}_t &= \boldsymbol{H}(\boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \dots) = \boldsymbol{D}_t\boldsymbol{R}_t\boldsymbol{D}_t, \end{cases}$$

where $\boldsymbol{\mu}_0 = (\mu_0^{(1)}, \dots, \mu_0^{(m)})' \in \mathbb{R}^m$, $\boldsymbol{D}_t = \{\mathrm{diag}(\boldsymbol{H}_t)\}^{1/2}$ and $\boldsymbol{R}_t = \mathrm{Corr}(\boldsymbol{y}_t, \boldsymbol{y}_t \mid \mathcal{F}_{t-1}^{\boldsymbol{y}})$.

Letting $\boldsymbol{\eta}_t^* = \boldsymbol{D}_t^{-1}(\boldsymbol{y}_t - \boldsymbol{\mu}_0)$, we get

$$\begin{cases} y_{kt} &= \mu_0^{(k)} + \sigma_{kt}\eta_{kt}^*, \\ \sigma_{kt} &= \sigma_k(\boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \dots; \boldsymbol{\theta}_0^{(k)}), \end{cases} \tag{3.5}$$

and we study the estimation of the parameter $\boldsymbol{\gamma}_0^{(k)} = (\mu_0^{(k)}, \boldsymbol{\theta}_0^{(k)'})'$ in Model (3.5), under

$$E(\eta_{kt}^* \mid \mathcal{F}_{t-1}^{\boldsymbol{y}}) = 0, \qquad \mathrm{Var}(\eta_{kt}^* \mid \mathcal{F}_{t-1}^{\boldsymbol{y}}) = 1.$$

Given observations $\boldsymbol{y}_1, \dots, \boldsymbol{y}_n$, and arbitrary initial values $\tilde{\boldsymbol{y}}_i$ for $i \leq 0$, we define $\tilde{\sigma}_{kt}(\boldsymbol{\theta}^{(k)}) = \sigma_k(\boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \dots, \boldsymbol{y}_1, \tilde{\boldsymbol{y}}_0, \tilde{\boldsymbol{y}}_{-1}, \dots; \boldsymbol{\theta}^{(k)})$ for $k = 1, \dots, m$. Let also $\sigma_{kt}(\boldsymbol{\theta}^{(k)}) = \sigma_k(\boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \dots; \boldsymbol{\theta}^{(k)})$. Let $\hat{\boldsymbol{\gamma}}_n^{(k)} = (\hat{\mu}_n^{(k)}, \hat{\boldsymbol{\theta}}_n^{(k)'})'$ denote the EbEE of $\boldsymbol{\gamma}_0^{(k)}$:

$$\hat{\boldsymbol{\gamma}}_n^{(k)} = \arg\min_{\boldsymbol{\gamma}^{(k)} \in M^{(k)} \times \boldsymbol{\Theta}^{(k)}} \tilde{Q}_n^{(k)}(\boldsymbol{\gamma}^{(k)}), \quad \tilde{Q}_n^{(k)}(\boldsymbol{\gamma}^{(k)}) = \frac{1}{n}\sum_{t=1}^n \log \tilde{\sigma}_{kt}^2\left(\boldsymbol{\theta}^{(k)}\right) + \frac{\{\epsilon_{kt}(\mu^{(k)})\}^2}{\tilde{\sigma}_{kt}^2\left(\boldsymbol{\theta}^{(k)}\right)},$$

where $\epsilon_{kt}(\mu^{(k)}) = y_{kt} - \mu^{(k)}$ and $M^{(k)}$ is a compact subset of $\mathbb{R}$. Let $\boldsymbol{\epsilon}_t = y_t - \boldsymbol{\mu}_0$.

THEOREM 3.3. *If* **A1** *and* **A4**-**A6** *hold, then* $\hat{\gamma}_n^{(k)} \to \gamma_0^{(k)}$, *a.s.* *as* $n \to \infty$. *If, in addition,* **A7**-**A12** *hold and* $\mu_0^{(k)}$ *belongs to the interior of* $M^{(k)}$, *then*

$$\sqrt{n}\left(\hat{\gamma}_n^{(k)} - \gamma_0^{(k)}\right) \xrightarrow{\mathcal{L}} \mathcal{N}\{0, \boldsymbol{\Upsilon}\}, \qquad \textit{where}$$

$$\boldsymbol{\Upsilon} = \begin{pmatrix} \left\{E\left(\frac{1}{\sigma_{kt}^2}\right)\right\}^{-1} & -\left\{E\left(\frac{1}{\sigma_{kt}^2}\right)\right\}^{-1} E\left(\eta_{kt}^{*3}\frac{1}{\sigma_{kt}}\boldsymbol{d}_{kt}'\right)\boldsymbol{J}_{kk}^{-1} \\ -\left\{E\left(\frac{1}{\sigma_{kt}^2}\right)\right\}^{-1}\boldsymbol{J}_{kk}^{-1}E\left(\eta_{kt}^{*3}\frac{1}{\sigma_{kt}}\boldsymbol{d}_{kt}\right) & \boldsymbol{J}_{kk}^{-1}\boldsymbol{I}_{kk}\boldsymbol{J}_{kk}^{-1} \end{pmatrix}.$$

It is interesting to note that, despite the presence of an intercept, the asymptotic variance of $\hat{\boldsymbol{\theta}}_n^{(k)}$ is the same as in Theorem 3.1. Note also that $\hat{\mu}_n^{(k)}$ and $\hat{\boldsymbol{\theta}}_n^{(k)}$ are not asymptotically independent in general. A case where the asymptotic independence holds is when (3.3) holds and $E(\eta_{kt}^{*3}) = 0$.

3.5. *The case of standard GARCH volatilities*

The EbE approach is particularly suited for the specification (2.7) with diagonal matrices $\boldsymbol{B}_j$, for which each component of $\boldsymbol{\theta}_0$ is only involved in one volatility equation. If in addition the matrices $\boldsymbol{A}_i$ are diagonal, more primitive assumptions can be given in Theorem 3.1. Thus, suppose that, for $k = 1, \ldots, m$ and for some nonnegative integers p_k, q_k,

$$\sigma_{kt}^2 = \omega_{0k} + \sum_{i=1}^{q_k}\alpha_{0ki}\epsilon_{k,t-i}^2 + \sum_{j=1}^{p_k}\beta_{0kj}\sigma_{k,t-j}^2, \quad \omega_{0k} > 0, \alpha_{0ki} \geq 0, \beta_{0kj} \geq 0,$$

which can be equivalently written as $\sigma_{kt}^2 = \omega_{0k} + \sum_{i=1}^{r_k} a_i(\eta_{k,t-i}^*)\sigma_{k,t-i}^2$, where $r_k = \max(p_k, q_k)$ and $a_i(z) = \alpha_{0ki}z^2 + \beta_{0ki}$. Let $\gamma(\mathbf{A_k})$ denote the top Lyapunov exponent, whose existence is guaranteed by Assumption **A1*** below, associated with the sequence

$$\boldsymbol{A}_{k,t} = \begin{pmatrix} a_1(\eta_{k,t-1}^*) & \ldots & a_{r_k}(\eta_{k,t-r_k}^*) \\ \boldsymbol{I}_{r_k-1} & & \boldsymbol{0} \end{pmatrix}.$$

Let $\boldsymbol{\theta}_0^{(k)} = (\omega_{0k}, \alpha_{0k1}, \ldots, \alpha_{0kq_k}, \beta_{0k1}, \ldots, \beta_{0kp_k})' \in \boldsymbol{\Theta}_k$ where $\boldsymbol{\Theta}_k$ is compact subset of $(0, +\infty) \times (0, +\infty]^{p_k+q_k}$. We make the following assumptions which do not impose the strict stationarity of the full vector $\boldsymbol{\epsilon}_t$ but, instead, the strict stationarity of the k-th component.

A1***:** (η_{kt}^*) is a strictly stationary and ergodic process with a non degenerate distribution, such that $E\log^+|\eta_{kt}^*| < \infty$.

A2***:** $\gamma(\mathbf{A_k}) < 0$ and $\forall \boldsymbol{\theta}_k \in \boldsymbol{\Theta}_k, \quad \sum_{j=1}^{p_k}\beta_{kj} < 1$.

A3*: if $p_k > 0$, the polynomials $\mathcal{A}_{\boldsymbol{\theta}_0^{(k)}}(z) = \sum_{i=1}^{q_k} \alpha_{0ki} z^i$ and $\mathcal{B}_{\boldsymbol{\theta}_0^{(k)}}(z) = \sum_{j=1}^{p_k} \beta_{0kj} z^j$ do not have common roots, $\mathcal{A}_{\boldsymbol{\theta}_0^{(k)}}(1) \neq 0$, and $\alpha_{0q_k} + \beta_{0p_k} \neq 0$.

A4*: $E|\epsilon_{kt}|^s < \infty$ for some $s > 0$.

PROPOSITION 3.3. *If* **A1***-**A4*** *hold, then* $\hat{\boldsymbol{\theta}}_n^{(k)} \to \boldsymbol{\theta}_0^{(k)}$, *a.s. as* $n \to \infty$. *If, in addition,* **A7**-**A8** *hold,* $\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n^{(k)} - \boldsymbol{\theta}_0^{(k)}\right)$ *satisfies* (3.2).

4. Inference in particular MGARCH models based on the EbE approach

Theorem 3.1 can be used for estimating the individual conditional variances in particular classes of MGARCH models. It can also be used for testing their adequacy, preliminary to their estimation. Indeed, most commonly used MGARCH specifications imply strong restrictions on the volatility of the individual components. We focus on the classes of DCC-GARCH and BEKK models.

4.1. Estimating the conditional variances in DCC-GARCH models

DCC-GARCH models are generally used under the assumption that the diagonal elements of $\boldsymbol{D}_t$ follow univariate GARCH(1,1) models, that is,

$$\sigma_{kt}^2 = \omega_k + \alpha_k \epsilon_{k,t-1}^2 + \beta_k \sigma_{k,t-1}^2, \quad \omega_k > 0, \alpha_k \geq 0, \beta_k \geq 0. \tag{4.1}$$

In the so-called corrected DCC (cDCC) of Aielli (2013)[5], the conditional correlation matrix is modelled as a function of the past standardized returns as

$$\boldsymbol{R}_t = \boldsymbol{Q}_t^{*-1/2} \boldsymbol{Q}_t \boldsymbol{Q}_t^{*-1/2}, \quad \boldsymbol{Q}_t = (1-\alpha-\beta)\boldsymbol{S} + \alpha \boldsymbol{Q}_{t-1}^{*1/2} \boldsymbol{\eta}_{t-1}^* \boldsymbol{\eta}_{t-1}^{*'} \boldsymbol{Q}_{t-1}^{*1/2} + \beta \boldsymbol{Q}_{t-1}, \tag{4.2}$$

where $\alpha, \beta \geq 0, \alpha + \beta < 1$, $\boldsymbol{S}$ is a correlation matrix, and $\boldsymbol{Q}_t^*$ is the diagonal matrix with the same diagonal elements as $\boldsymbol{Q}_t$. No formally established asymptotic results exist for the full estimation of the DCC and cDCC models. The strong consistency and asymptotic normality

[5]In the original DCC model of Engle (2002), the dynamics of $\boldsymbol{Q}_t$ is given by

$$\boldsymbol{Q}_t = (1-\alpha-\beta)\boldsymbol{S} + \alpha \boldsymbol{\eta}_{t-1}^* \boldsymbol{\eta}_{t-1}^{*'} + \beta \boldsymbol{Q}_{t-1}.$$

Aielli (2013) pointed out that the commonly used estimator of $\boldsymbol{S}$ defined as the sample second moment of the standardized returns is not consistent in this formulation. Stationarity conditions for DCC models have been recently established by Fermanian and Malongo (2014).

of the EbEE of $\boldsymbol{\theta}_0^{(k)} = (\omega_k, \alpha_k, \beta_k)'$ in (4.1) could be obtained by applying Proposition 3.3. We establish them under more explicit conditions in the following theorem.

THEOREM 4.1. *Assume that $\alpha+\beta<1$, $\alpha_\ell+\beta_\ell<1$, and either $\alpha_\ell\beta_\ell>0$ or $\beta_\ell=0$, for $\ell=1,\dots,m$. Let $\boldsymbol{\eta}_1$ admit, with respect to the Lebesgue measure on $\mathbb{R}^m$, a positive density around 0. Suppose that $\boldsymbol{\theta}_0^{(k)} \in \boldsymbol{\Theta}_k$ where $\boldsymbol{\Theta}_k$ is any compact subset of $(0,\infty)\times[0,\infty)\times[0,1)$. Then $\hat{\boldsymbol{\theta}}_n^{(k)} \to \boldsymbol{\theta}_0^{(k)}$, a.s. as $n\to\infty$. If, in addition, $\boldsymbol{\theta}_0^{(k)}$ is an interior point of $\boldsymbol{\Theta}_k$, and $E\|\boldsymbol{\eta}_t\|^{4(1+\delta)}<\infty$, for some $\delta>0$, then the sequence $\sqrt{n}(\hat{\boldsymbol{\theta}}_n^{(k)}-\boldsymbol{\theta}_0^{(k)})$ is asymptotically normally distributed.*

4.2. Estimating semi-diagonal BEKK models

Consider a BEKK-GARCH(p,q) model given by

$$\boldsymbol{\epsilon}_t = \boldsymbol{H}_t^{1/2}\boldsymbol{\eta}_t, \qquad \boldsymbol{H}_t = \boldsymbol{\Omega}_0 + \sum_{i=1}^q \boldsymbol{A}_{0i}\boldsymbol{\epsilon}_{t-i}\boldsymbol{\epsilon}_{t-i}'\boldsymbol{A}_{0i}' + \sum_{j=1}^p \boldsymbol{B}_{0j}\boldsymbol{H}_{t-j}\boldsymbol{B}_{0j}, \tag{4.3}$$

where $(\boldsymbol{\eta}_t)$ is an iid $\mathbb{R}^m$-valued centered sequence with $E\boldsymbol{\eta}_t\boldsymbol{\eta}_t' = \boldsymbol{I}_m$, $\boldsymbol{A}_{0i} = (a_{ik\ell})_{1\le k,\ell\le m}$, $\boldsymbol{B}_{0j} = \mathrm{diag}(b_{j1},\dots,b_{jm})$, and $\boldsymbol{\Omega}_0 = (\omega_{k\ell})_{1\le k,\ell\le m}$ is a positive definite $m\times m$ matrix. In this model, which can be called "semi-diagonal" (as opposed to the diagonal BEKK in which both the $\boldsymbol{B}_{0j}$ and $\boldsymbol{A}_{0i}$ are diagonal matrices), the conditional variance of any return may depend on the past of all returns. The k-th diagonal entry of $\boldsymbol{H}_t$ satisfies a stochastic recurrence equation of the form

$$h_{kk,t} = \omega_{kk} + \sum_{i=1}^q \left(\sum_{\ell=1}^m a_{ik\ell}\epsilon_{\ell,t-i}\right)^2 + \sum_{j=1}^p b_{jk}^2 h_{kk,t-j}. \tag{4.4}$$

Let $\boldsymbol{\theta}_0^{(k)} = (\omega_{kk}, \boldsymbol{a}_{1k}', \dots, \boldsymbol{a}_{qk}', \boldsymbol{b}_k)'$ for $k=1,\dots,m$, where $\boldsymbol{a}_{ik}'$ denotes the k-th row of the matrix $\boldsymbol{A}_{0i}$, and $\boldsymbol{b}_k = (b_{1k}^2,\dots,b_{pk}^2)$. It is clear that an identifiability restriction is needed, $h_{kk,t}$ being invariant to a change of sign of the k-th row of any matrix $\boldsymbol{A}_i$. For simplicity, we therefore assume that $a_{ik1}>0$ for $i=1,\dots,q$. Let $\boldsymbol{\theta}^{(k)} = (\boldsymbol{\theta}_i^{(k)}) \in \mathbb{R}^{1+mq+p}$ denote a generic parameter value. The parameter space $\boldsymbol{\Theta}_k$ is any compact subset of

$$\left\{\boldsymbol{\theta}^{(k)} \mid \theta_1^{(k)}>0, \theta_2^{(k)}, \theta_{m+2}^{(k)}, \dots, \theta_{(q-1)m+2}^{(k)} > 0, \theta_{1+mq+1}^{(k)}, \dots, \theta_{1+mq+p}^{(k)} \ge 0, \sum_{j=1}^p \theta_{1+mq+j}^{(k)} < 1\right\}.$$

Let

$$\boldsymbol{A}_0 = \sum_{i=1}^q \boldsymbol{H}_m(\boldsymbol{A}_{0i}\otimes\boldsymbol{A}_{0i})\boldsymbol{K}_m', \quad \boldsymbol{B}_0 = \sum_{j=1}^p \boldsymbol{H}_m(\boldsymbol{B}_{0j}\otimes\boldsymbol{B}_{0j})\boldsymbol{K}_m'$$

where $\otimes$ is the Kronecker product and $\boldsymbol{H}_m$ and $\boldsymbol{K}_m$ are the usual elimination and duplication matrices[6].

THEOREM 4.2. *Let the spectral radius of $\boldsymbol{A}_0 + \boldsymbol{B}_0$ be less than 1. Let $\boldsymbol{\eta}_1$ admit, with respect to the Lebesgue measure on $\mathbb{R}^m$, a positive density around 0, and suppose that $E|\eta_{kt}|^{4(1+\delta)} < \infty$, and $E\|\boldsymbol{\epsilon}_t\|^{4(1+1/\delta)} < \infty$ for some $\delta > 0$. Suppose that $\boldsymbol{\theta}_0^{(k)} \in \boldsymbol{\Theta}_k$. Then $\hat{\boldsymbol{\theta}}_n^{(k)} \to \boldsymbol{\theta}_0^{(k)}$, a.s. as $n \to \infty$. If, in addition, $\boldsymbol{\theta}_0^{(k)}$ is an interior point of $\boldsymbol{\Theta}_k$, the sequence $\sqrt{n}(\hat{\boldsymbol{\theta}}_n^{(k)} - \boldsymbol{\theta}_0^{(k)})$ is asymptotically normally distributed.*

Full BEKK models are generally considered as unfeasible for large cross-sectional dimensions (see for instance Laurent, Rombouts and Violante (2012)) and practitioners focus on diagonal, or even scalar, models. It follows from Theorem 4.2 that if the matrix $\boldsymbol{\Omega}_0$ is diagonal[7], the semi-diagonal BEKK-GARCH(p,q) model (4.3) can be fully estimated by successively applying the EbEE to each equation. Indeed, each parameter of the model appears in one, and only one, equation. For the general BEKK (without assuming diagonality of the matrices $\boldsymbol{B}_{0j}$), the asymptotic properties of the QML method were derived by Comte and Lieberman (2003), though under some high-level assumptions[8].

4.3. Testing adequacy of BEKK models

Equation (4.4) can be viewed as a restricted form, implied by the BEKK model, of a more general volatility specification. Testing for such a restriction in this more general framework can thus be used to check the validity of the BEKK specification. For ease of presentation, we focus on the case $m = 2$ and $p = q = 1$. Letting $\boldsymbol{\theta}_0^{(k)} = (\omega_{kk}, a_{1k}^2, 2a_{1k}a_{2k}, a_{2k}^2, b_{1k}^2)'$ for $k = 1, 2$, the validity of Model (4.3) can be studied by estimating Model (2.5) for each component of $\boldsymbol{\epsilon}_t$ with, in view of (4.4) for $m = 2$ and $p = q = 1$,

$$\sigma_{kt}^2 = \theta_{01}^{(k)} + \theta_{02}^{(k)}\epsilon_{1,t-1}^2 + \theta_{03}^{(k)}\epsilon_{1,t-1}\epsilon_{2,t-1} + \theta_{04}^{(k)}\epsilon_{2,t-1}^2 + \theta_{05}^{(k)}\sigma_{k,t-1}^2, \quad k = 1, 2, \quad (4.5)$$

[6] $\boldsymbol{H}_m$ and $\boldsymbol{K}_m$ are $\frac{m(m+1)}{2} \times m^2$ matrices such that $\boldsymbol{H}_m\boldsymbol{K}_m' = \boldsymbol{I}_{m(m+1)/2}$ and $\text{vec}(\boldsymbol{A}) = \boldsymbol{K}_m'\text{vech}(\boldsymbol{A})$, $\text{vech}(\boldsymbol{A}) = \boldsymbol{H}_m\text{vec}(\boldsymbol{A})$ for any symmetric $m \times m$ matrix $\boldsymbol{A}$.

[7] Or, more generally, if it is parameterized in function of its diagonal entries

[8] In particular, the model was assumed to be identifiable and the existence of eighth-order moments was required for $\boldsymbol{\epsilon}_t$. On the other hand, Avarucci, Beutner and Zaffaroni (2013) showed that for the BEKK, the finiteness of the variance of the scores requires at least the existence of second-order moments of the observable process.

under the positivity constraints $\theta_{01}^{(k)} > 0, \quad \theta_{0i}^{(k)} \geq 0, i = 2,4,5$. The restrictions implied by the BEEK-GARCH(1,1) model (4.3) are of the form:

$$H_0^{(k)}: \qquad |\theta_{03}^{(k)}| = 2\sqrt{\theta_{02}^{(k)}\theta_{04}^{(k)}}, \quad k = 1,2.$$

Let

$$\boldsymbol{\Theta}^{(k)} = \boldsymbol{\Theta}_k^* \cap \left\{\boldsymbol{\theta}^{(k)}; |\theta_3^{(k)}| \in \left[0, 2\sqrt{\theta_2^{(k)}\theta_4^{(k)}}\right]\right\},$$

where $\boldsymbol{\Theta}_k^*$ is a compact subset of $\{\theta_1^{(k)} > 0, \theta_i^{(k)} \geq 0$, for $i = 2,4$ and $\theta_5^{(k)} \in [0,1)\}$. Note that, under $H_0^{(k)}$, the true parameter value is at the boundary of the parameter set.

PROPOSITION 4.1. *Let the spectral radius of $\boldsymbol{A} + \boldsymbol{B}$ be less than 1. Let $\boldsymbol{\eta}_1$ admit, with respect to the Lebesgue measure on $\mathbb{R}^2$, a positive density around 0, and suppose that $E|\eta_{kt}|^{4(1+\delta)} < \infty$, for $k = 1,2$ and some $\delta > 0$. Let $\boldsymbol{\theta}_0^{(k)}$ belong to the interior of $\boldsymbol{\Theta}_k^*$ for $k = 1,2$.*

Let $(\boldsymbol{\epsilon}_t)$ be the strictly stationary solution of Model (4.3). Let the Wald statistic for the hypothesis $H_0^{(k)}$,

$$\boldsymbol{W}_n^{(k)} = \frac{n\left\{\hat{\theta}_{n3}^{(k)^2} - 4\hat{\theta}_{n2}^{(k)}\hat{\theta}_{n4}^{(k)}\right\}^2}{\boldsymbol{X}_n'\hat{\boldsymbol{J}}_{kk}^{-1}\hat{\boldsymbol{I}}_{kk}\hat{\boldsymbol{J}}_{kk}^{-1}\boldsymbol{X}_n}, \qquad \text{where} \quad \hat{\boldsymbol{\theta}}_n^{(k)} = (\hat{\theta}_{n1}^{(k)}, \ldots, \hat{\theta}_{n5}^{(k)})',$$

$\boldsymbol{X}_n = \left(0, 4\hat{\theta}_{n4}^{(k)}, -2\hat{\theta}_{n3}^{(k)}, 4\hat{\theta}_{n2}^{(k)}, 0\right)'$, $\hat{\eta}_{kt}^* = \epsilon_{kt}/\tilde{\sigma}_{kt}(\hat{\boldsymbol{\theta}}_n^{(k)})$ *and*

$$\hat{\boldsymbol{J}}_{kk} = \frac{1}{n}\sum_{t=1}^n \hat{\boldsymbol{d}}_{kt}\hat{\boldsymbol{d}}_{kt}', \quad \hat{\boldsymbol{I}}_{kk} = \frac{1}{n}\sum_{t=1}^n \{\hat{\eta}_{kt}^{*4} - 1\}\hat{\boldsymbol{d}}_{kt}\hat{\boldsymbol{d}}_{kt}', \quad \hat{\boldsymbol{d}}_{kt} = \frac{1}{\tilde{\sigma}_{kt}^2(\hat{\boldsymbol{\theta}}_n)}\frac{\partial\tilde{\sigma}_{kt}^2(\hat{\boldsymbol{\theta}}_n^{(k)})}{\partial\boldsymbol{\theta}^{(k)}}.$$

Then, $\boldsymbol{W}_n^{(k)}$ asymptotically follows a mixture of the χ^2 distribution with one degree of freedom and the Dirac measure at 0:

$$\boldsymbol{W}_n^{(k)} \xrightarrow{\mathcal{L}} \frac{1}{2}\chi^2(1) + \frac{1}{2}\delta_0, \qquad \text{as } n \to \infty.$$

In view of this result, testing $H_0^{(k)}$ at the asymptotic level $\alpha \in (0, 1/2)$ can thus be achieved by using the critical region $\{\boldsymbol{W}_n^{(k)} > \chi_{1-2\alpha}^2(1)\}$.

5. Estimating conditional and stochastic correlation matrices

Having estimated the individual conditional variances of a vector $(\boldsymbol{\epsilon}_t)$ satisfying (2.4) in a first step, it is generally of interest to estimate the complete conditional variance matrix $\boldsymbol{H}_t$, which thus reduces to estimating the conditional correlation $\boldsymbol{R}_t$.

Suppose that matrix $\boldsymbol{R}_t$ is parameterized by some parameter $\boldsymbol{\rho}_0 \in \mathbb{R}^r$, together with the volatility parameter $\boldsymbol{\theta}_0$, as

$$\boldsymbol{R}_t = \boldsymbol{R}(\boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \dots; \boldsymbol{\theta}_0, \boldsymbol{\rho}_0) = \boldsymbol{\mathcal{R}}(\boldsymbol{\eta}^*_{t-1}, \boldsymbol{\eta}^*_{t-2}, \dots; \boldsymbol{\rho}_0).$$

Let $\Lambda \subset \mathbb{R}^r$ denote a parameter set such that $\boldsymbol{\rho}_0 \in \Lambda$. If the $\boldsymbol{\eta}^*_t$ were observed, in view of (2.2) a QMLE of $\boldsymbol{\rho}_0$ would be obtained as any measurable solution of

$$\arg\min_{\boldsymbol{\rho} \in \Lambda} n^{-1} \sum_{t=1}^{n} \boldsymbol{\eta}^{*\prime}_t \tilde{\boldsymbol{R}}_t^{-1} \boldsymbol{\eta}^*_t + \log |\tilde{\boldsymbol{R}}_t|,$$

where, introducing initial values $\tilde{\boldsymbol{\eta}}^*_i$ for $i \le 0$, $\tilde{\boldsymbol{R}}_t = \boldsymbol{\mathcal{R}}(\boldsymbol{\eta}^*_{t-1}, \boldsymbol{\eta}^*_{t-2}, \dots, \tilde{\boldsymbol{\eta}}^*_0, \tilde{\boldsymbol{\eta}}^*_{-1}, \dots; \boldsymbol{\rho})$.

We therefore consider the two-step estimation method of the parameters of Model (2.4).

(a) **First step**: EbE estimation of the volatility parameters $\boldsymbol{\theta}_0^{(k)}$ and extraction of the vectors of residuals $\hat{\boldsymbol{\eta}}^*_t = (\hat{\eta}^*_{1t}, \dots, \hat{\eta}^*_{mt})'$ where $\hat{\eta}^*_{kt} = \tilde{\sigma}^{-1}_{kt}(\hat{\boldsymbol{\theta}}^{(k)})\epsilon_{kt}$;

(b) **Second step**: QML estimation of the conditional correlation matrix $\boldsymbol{\rho}_0$ by EbE, as a solution of

$$\arg\min_{\boldsymbol{\rho} \in \Lambda} n^{-1} \sum_{t=1}^{n} \hat{\boldsymbol{\eta}}^{*\prime}_t \tilde{\boldsymbol{R}}_t^{-1} \hat{\boldsymbol{\eta}}^*_t + \log |\tilde{\boldsymbol{R}}_t|,$$

where $\tilde{\boldsymbol{R}}_t = \boldsymbol{\mathcal{R}}(\hat{\boldsymbol{\eta}}^*_{t-1}, \hat{\boldsymbol{\eta}}^*_{t-2}, \dots, \hat{\boldsymbol{\eta}}^*_1, \tilde{\boldsymbol{\eta}}^*_0, \tilde{\boldsymbol{\eta}}^*_{-1}, \dots; \boldsymbol{\rho})$.

We will establish the asymptotic properties of this approach in the case where $\boldsymbol{R}_t$ is constant, that is for Model (2.8)-(2.9).The case of the classical CCC-GARCH(p, q) models will be considered in Section 6.1.1.

5.1. Estimating general CCC models

Let

$$\boldsymbol{\rho} = (R_{21}, \dots, R_{m1}, R_{32}, \dots, R_{m2}, \dots, R_{m,m-1})' = \mathrm{vech}^0(\boldsymbol{R}),$$

denoting by vech^0 the operator which stacks the sub-diagonal elements (excluding the diagonal) of a matrix. The global parameter is denoted

$$\boldsymbol{\vartheta} = (\boldsymbol{\theta}^{(1)'}, \dots, \boldsymbol{\theta}^{(m)'}, \boldsymbol{\rho}')' := (\boldsymbol{\theta}', \boldsymbol{\rho}')' \in \mathbb{R}^d \times [-1, 1]^{m(m-1)/2}, \quad d = \sum_{k=1}^{m} d_k,$$

and it belongs to the compact parameter set $\boldsymbol{\Theta} = \prod_{k=1}^{m} \boldsymbol{\Theta}_k \times [-1,1]^{m(m-1)/2}$. The second-step estimator of the constant correlation matrix $\boldsymbol{R}_t$ is given by $\hat{\boldsymbol{R}}_n = \frac{1}{n}\sum_{t=1}^{n} \hat{\boldsymbol{\eta}}_t^* \left(\hat{\boldsymbol{\eta}}_t^*\right)'$. Let $\hat{\boldsymbol{\vartheta}}_n = \left(\hat{\boldsymbol{\theta}}_n' := (\hat{\boldsymbol{\theta}}_n^{(1)'}, \dots, \hat{\boldsymbol{\theta}}_n^{(m)'}), \hat{\boldsymbol{\rho}}_n'\right)'$, where $\hat{\boldsymbol{\rho}}_n = \mathrm{vech}^0(\hat{\boldsymbol{R}}_n)$.

Theorem 5.1. *For the CCC model* (2.8)-(2.9)*, if* **A1**-**A6** *hold, then*

$$\hat{\boldsymbol{\vartheta}}_n \to \boldsymbol{\vartheta}_0, \quad a.s. \quad as\ n \to \infty.$$

For the asymptotic normality, we introduce the following notations. Let the $d \times d$ matrix $\boldsymbol{J}^* = ((\kappa_{k\ell} - 1)\boldsymbol{J}_{k\ell})$ for $k, \ell = 1, \dots, m$, and $\boldsymbol{J}_{k\ell} = E\left(\boldsymbol{d}_{kt}\boldsymbol{d}_{\ell t}'\right)$. Let, for $\boldsymbol{J}_0 = \mathrm{diag}(\boldsymbol{J}_{11}, \dots, \boldsymbol{J}_{mm})$ in bloc-matrix notation,

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = \boldsymbol{J}_0^{-1}\boldsymbol{J}^*\boldsymbol{J}_0^{-1} = \left((\kappa_{k\ell} - 1)\boldsymbol{J}_{kk}^{-1}\boldsymbol{J}_{k\ell}\boldsymbol{J}_{\ell\ell}^{-1}\right).$$

Let also $\boldsymbol{d}_t = (\boldsymbol{d}_{1t}', \dots, \boldsymbol{d}_{mt}')' \in \mathbb{R}^d$, $\boldsymbol{\Omega}_k = E\boldsymbol{d}_{kt}$ and $\boldsymbol{\Omega} = (\boldsymbol{\Omega}_1', \dots, \boldsymbol{\Omega}_m')' \in \mathbb{R}^d$. Let $\boldsymbol{\Gamma} = \mathrm{var}\left(\mathrm{vech}^0\left\{\boldsymbol{\eta}_t^*\left(\boldsymbol{\eta}_t^*\right)'\right\}\right)$. For $\boldsymbol{x} \in \mathbb{R}^m$, let the $d \times d$ matrices $\boldsymbol{F}(\boldsymbol{x}) = \mathrm{diag}\{(1 - x_1^2)\boldsymbol{j}_1, \dots, (1 - x_m^2)\boldsymbol{j}_m\}$, where $\boldsymbol{j}_k = (1, \dots, 1) \in \mathbb{R}^{d_k}$, and $\boldsymbol{A}_{k\ell} = E\{\eta_{kt}^*\eta_{\ell t}^*\boldsymbol{F}(\boldsymbol{\eta}_t^*)\}$. Let, for $k, \ell = 2, \dots, m$, the $d \times d$ matrix $\mathbf{M}_{k,\ell-1} = \mathrm{diag}\left(\mathbf{M}_{k,\ell-1}^{(1)}, \dots, \mathbf{M}_{k,\ell-1}^{(m)}\right)$ where

$$\mathbf{M}_{k,\ell-1}^{(i)} = \begin{cases} \mathbf{0}_{d_i \times d_i} & \text{if } i \neq k \quad \text{and} \quad i \neq \ell \\ R_{k,\ell-1}\boldsymbol{I}_{d_i} & \text{otherwise.} \end{cases}$$

Let the $d \times dm(m-1)/2$ matrices $\boldsymbol{A} = (\boldsymbol{A}_{21} \dots \boldsymbol{A}_{m1} \ \boldsymbol{A}_{32} \dots \boldsymbol{A}_{m,m-1})$ and $\boldsymbol{M} = (\boldsymbol{M}_{21} \dots \boldsymbol{M}_{m1} \ \boldsymbol{M}_{32} \dots \boldsymbol{M}_{m,m-1})$. Let the $d \times m(m-1)/2$ matrices

$$\boldsymbol{L} = \boldsymbol{A}(\boldsymbol{I}_{m(m-1)/2} \otimes \boldsymbol{\Omega}), \qquad \boldsymbol{\Lambda} = \boldsymbol{M}(\boldsymbol{I}_{m(m-1)/2} \otimes \boldsymbol{\Omega}).$$

Let

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}\boldsymbol{\rho}} = -\frac{1}{2}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\boldsymbol{\Lambda} - \boldsymbol{J}_0^{-1}\boldsymbol{L}, \quad \boldsymbol{\Sigma}_{\boldsymbol{\rho}} = \frac{1}{4}\boldsymbol{\Lambda}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\boldsymbol{\Lambda} + \frac{1}{2}\left(\boldsymbol{\Lambda}'\boldsymbol{J}_0^{-1}\boldsymbol{L} + \boldsymbol{L}'\boldsymbol{J}_0^{-1}\boldsymbol{\Lambda}\right) + \boldsymbol{\Gamma}.$$

We need an additional assumption.

A13: The distribution of $\mathrm{vech}(\boldsymbol{\eta}_t\boldsymbol{\eta}_t')$ is not supported on an hyperplane.

Theorem 5.2. *For the CCC model* (2.8)-(2.9)*, if* **A1**-**A13** *hold, for* $k = 1, \dots, m$*, and* $\boldsymbol{\rho}_0 \in (-1,1)^{m(m-1)/2}$*, then*

$$\begin{pmatrix} \sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \\ \sqrt{n}(\hat{\boldsymbol{\rho}}_n - \boldsymbol{\rho}_0) \end{pmatrix} \stackrel{\mathcal{L}}{\to} \mathcal{N}\left\{0, \boldsymbol{\Sigma} := \begin{pmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\theta}} & \boldsymbol{\Sigma}_{\boldsymbol{\theta}\boldsymbol{\rho}} \\ \boldsymbol{\Sigma}_{\boldsymbol{\theta}\boldsymbol{\rho}}' & \boldsymbol{\Sigma}_{\boldsymbol{\rho}} \end{pmatrix}\right\},$$

and $\boldsymbol{\Sigma}$ *is a non-singular matrix.*

Remark 5.1. Even though the components of $\boldsymbol{\theta}_0$ are estimated independently, the components $\hat{\boldsymbol{\theta}}_n^{(k)}$ of $\hat{\boldsymbol{\theta}}_n$ are asymptotically non independent in general. More precisely, it can be seen that $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ is bloc diagonal if $\mathrm{Cov}(\eta_{kt}^{*2}, \eta_{\ell t}^{*2}) = 0$ for any $k \neq \ell$.

Remark 5.2. In the asymptotic variance $\boldsymbol{\Sigma}_{\boldsymbol{\rho}}$ of $\hat{\boldsymbol{\rho}}_n$, the first two matrices in the sum reflect the effect of the estimation of $\boldsymbol{\theta}_0$, while the remaining matrix, $\boldsymbol{\Gamma}$, is independent of $\boldsymbol{\theta}_0$. A limit case is when the components of $\boldsymbol{\eta}_t^*$ are serially independent, that is when $\boldsymbol{\eta}_t^* = \boldsymbol{\eta}_t$ and $\boldsymbol{R}$ is the identity matrix. Then, straightforward computation shows that $\boldsymbol{L} = \boldsymbol{\Lambda} = \boldsymbol{0}$ and thus, in bloc-matrix notation,

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\theta}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{m(m-1)/2} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_{\boldsymbol{\theta}} = \mathrm{diag}((\kappa_{11}-1)\boldsymbol{J}_{11}^{-1}, \ldots, (\kappa_{mm}-1)\boldsymbol{J}_{mm}^{-1}).$$

Remark 5.3. It is worthnoting that all the matrices involved in the asymptotic covariance matrix $\boldsymbol{\Sigma}$ take the form of expectations. A simple estimator of $\boldsymbol{\Sigma}$ is thus obtained by replacing those expectations by their sample counterparts. For instance, it can be shown that a consistent estimator of $\boldsymbol{A}_{k\ell}$ is

$$\hat{\boldsymbol{A}}_{k\ell} = \frac{1}{n}\sum_{t=1}^{n} \hat{\eta}_{kt}^* \hat{\eta}_{\ell t}^* F(\hat{\boldsymbol{\eta}}_t^*).$$

Remark 5.4. In financial applications, the different returns are generally not available over the same time horizons. Discarding dates for which at least one return is not available may entail a severe sample size reduction. Instead, the correlations can be estimated by considering the returns pairwise (with different sample lengths for different pairs). Such estimators of the correlations are consistent, even if the estimated global correlation matrix may not be positive definite. This approach will be used in the empirical section.

5.2. *Estimating DCC models*

The asymptotic properties of the first-step EbEE were established in Theorem 4.1 for diagonal first-order DCC models. The second step can be applied for estimating $\boldsymbol{\rho} = (\alpha, \beta, (\mathrm{vech}^0(\boldsymbol{S}))')'$ in Model (4.2). The matrices $\tilde{\boldsymbol{R}}_t$ involved in the second step are obtained as $\tilde{\boldsymbol{R}}_t = \tilde{\boldsymbol{Q}}_t^{*-1/2}(\boldsymbol{\rho})\tilde{\boldsymbol{Q}}_t(\boldsymbol{\rho})\tilde{\boldsymbol{Q}}_t^{*-1/2}(\boldsymbol{\rho})$, where the $\tilde{\boldsymbol{Q}}_t(\boldsymbol{\rho})$ are computed recursively as

$$\tilde{\boldsymbol{Q}}_t(\boldsymbol{\rho}) = (1-\alpha-\beta)\boldsymbol{S} + \alpha\tilde{\boldsymbol{Q}}_{t-1}^{*1/2}(\boldsymbol{\rho})\hat{\boldsymbol{\eta}}_{t-1}^*\hat{\boldsymbol{\eta}}_{t-1}^{*'}\tilde{\boldsymbol{Q}}_{t-1}^{*1/2}(\boldsymbol{\rho}) + \beta\tilde{\boldsymbol{Q}}_{t-1}(\boldsymbol{\rho}), \quad t \geq 1,$$

with initial value $\tilde{\boldsymbol{Q}}_0(\boldsymbol{\rho}) = \boldsymbol{S}$. The asymptotic properties of the second-step EbEE are an open issue.

5.3. Estimating stochastic correlations driven by a hidden Markov chain

A natural extension of the CCC model is obtained by allowing the matrix $\boldsymbol{R}_t^*$ to be driven by a Markov chain. This extension was advocated by Pelletier (2006) who interprets it as a "midpoint between the CCC model of Bollerslev (1990) and models such as the DCC of Engle (2002) where the correlations change every period." Assume that $(\boldsymbol{\epsilon}_t)$ is generated by Model (2.10) with

$$\boldsymbol{R}_t^* = \boldsymbol{R}^*(\Delta_t), \text{ where } (\Delta_t) \text{ is a Markov chain on } \mathcal{E} = \{1, \dots, N\}. \tag{5.2}$$

Note that the Markov chain is not observed but the number of states, N, is assumed to be known. Denoting by $p(i,j) = P(\Delta_t = j \mid \Delta_{t-1} = i)$ the transition probabilities of the Markov chain, the parameter vector is now

$$\begin{aligned} \boldsymbol{\zeta} &= (\boldsymbol{\theta}^{(1)'}, \dots, \boldsymbol{\theta}^{(m)'}, \boldsymbol{\rho}'(1), \dots, \boldsymbol{\rho}'(N), \boldsymbol{p}')' \\ &:= (\boldsymbol{\theta}', \boldsymbol{\rho}', \boldsymbol{p}')' \in \mathbb{R}^d \times [-1,1]^{Nm(m-1)/2} \times [0,1]^{N(N-1)}, \end{aligned}$$

where $\boldsymbol{p} = (p(1,2), p(1,3), \dots, p(1,N), p(2,2), \dots, p(N,N))'$ and $\boldsymbol{\rho}(i) = \text{vech}^0\{\boldsymbol{R}^*(i)\}$ for $i = 1, \dots, N$.

The full maximum likelihood method is generally intractable, in particular when the regimes are not Markovian (that is, when the conditional variances σ_{kt}^2 do not depend on a finite number of past values of $\boldsymbol{\epsilon}_t$). However, a two-step approach can be followed: having estimated $\boldsymbol{\theta}_0$ in the first step, we may apply the maximum likelihood (for a given distribution of the iid process) on the standardized residuals to estimate the remaining parameters, $\boldsymbol{\rho}_0$ and $\boldsymbol{p}_0$, in a second step. This procedure will be illustrated on exchange rates series in the empirical section, the asymptotic properties being left for future research.

6. Numerical Illustrations

The first part of the section will be devoted to Monte-Carlo experiments aiming at studying the performance of the EbE approach in finite sample. Real data examples will be presented in the second part.

6.1. Monte-Carlo study

We will first illustrate the gains in computation time brought by the two-step EbE approach, by comparison with the usual Full QML (FQML) in which all the parameters are estimated in one step. We will also investigate, for CCC and DCC models, whether the gains in numerical complexity have a price in terms of finite-sample accuracy.

6.1.1. *Time complexity and accuracy comparison of the EbEE and the full QMLE*

Let us compare the computation cost of the EbEE with that of the FQMLE in the case of a diagonal CCC-GARCH$(1,1)$ model of dimension m, that is, under the specification (2.7) with $p = q = 1$ and diagonal matrices $\boldsymbol{A}_1$ and $\boldsymbol{B}_1$. EbEE of all the model parameters requires m estimations of univariate GARCH-type models with 3 parameters, plus the computation of the empirical correlation of the EbE residuals. The full QMLE requires the optimization of a function of $3m + m(m-1)/2$ parameters. Because the time complexity of an optimization generally grows rapidly with the dimension of the objective function, the full QMLE should be much more costly than the EbEE in terms of computation time. The two estimators were fitted on simulations of length $n = 2000$ of the CCC-GARCH$(1,1)$ model (2.7) with $\boldsymbol{A}_1 = 0.05\boldsymbol{I}_m$ and $\boldsymbol{B}_1 = 0.9\boldsymbol{I}_m$ (such values are close to those generally fitted to real series). The correlation matrix used for the simulations is $\boldsymbol{R} = \boldsymbol{I}_m$, but the $m(m-1)/2$ subdiagonal terms of $\boldsymbol{R}$ were estimated, together with the $3m$ other parameters of the model. The distribution of $\boldsymbol{\eta}_t$ is Gaussian, which has little impact on the computation times, but should give an advantage to the FQMLE (which is then the MLE) in terms of accuracy. Table 1 compares the effective computation times required by the two estimators as a function of the dimension m. As expected, the comparison of the CPU's is clearly in favor of the EbEE. Note that these computation times have been obtained using a single processor. Since the EbEE is clearly easily parallelizable (using one processor for each of the m optimizations), the advantage of the EbEE should be even more pronounced with a multiprocessing implementation. Table 1 also compares the relative efficiencies of the EbEE with respect to the FQMLE. To this aim, we first computed the approximated information matrix $\boldsymbol{J}_n = -\frac{1}{2n}\frac{\partial^2}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\sum_{t=1}^n \boldsymbol{\epsilon}_t'\boldsymbol{H}_t^{-1}\boldsymbol{\epsilon}_t + \log|\boldsymbol{H}_t|$. Note that when $(\boldsymbol{\eta}_t)$ is Gaussian and when $\hat{\boldsymbol{\vartheta}}_{ML}$ is the (Q)MLE, then $n(\hat{\boldsymbol{\vartheta}}_{ML} - \boldsymbol{\vartheta}_0)'\boldsymbol{J}_n(\hat{\boldsymbol{\vartheta}}_{ML} - \boldsymbol{\vartheta}_0)$ follows asymptotically a χ^2 distribution. More generally, the quadratic form $n(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0)'\boldsymbol{J}_n(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0)$ can serve as a measure of accuracy of an estimator $\hat{\boldsymbol{\vartheta}}_n$ (the Euclidean distance, obtained by replacing $\boldsymbol{J}_n$ by the identity matrix, has the drawback of being scale dependent). The relative efficiency

Table 1. Computation time of the two estimators (CPU time in seconds) and Relative Efficiency (RE) of the EbEE with respect to the FQMLE (NA means "Not Available" because of the impossibility to compute the FQMLE) for m-dimensional CCC-GARCH(1,1) models.

Dim. m	2	3	4	5	6	7	8	9
Nb. of param.	7	12	18	25	33	42	52	63
CPU for EbEE	0.57	0.88	1.18	1.31	1.52	1.85	2.04	2.37
CPU for FQMLE	32.49	100.78	123.33	215.38	317.85	617.33	876.52	1113.68
ratio of CPU	57.00	114.52	104.52	164.41	209.11	333.69	429.67	469.91
RE	0.96	1.00	0.99	0.97	0.99	0.99	0.97	1.00
Dim. m	10	11	12	50	100	200	400	800
Nb. of param.	75	88	102	1375	5250	20500	81000	322000
CPU for EbEE	2.82	2.98	3.49	13.67	27.89	56.58	110.00	226.32
CPU for FQMLE	1292.34	1520.60	1986.38	NA	NA	NA	NA	NA
ratio of CPU	458.28	510.27	569.16	NA	NA	NA	NA	NA
RE	102.42	304.36	14.22	NA	NA	NA	NA	NA

(RE) displayed in Table 1 is equal to

$$RE = \frac{(\hat{\boldsymbol{\vartheta}}_{EbEE} - \boldsymbol{\vartheta}_0)' \boldsymbol{J}_n (\hat{\boldsymbol{\vartheta}}_{EbEE} - \boldsymbol{\vartheta}_0)}{(\hat{\boldsymbol{\vartheta}}_{QMLE} - \boldsymbol{\vartheta}_0)' \boldsymbol{J}_n (\hat{\boldsymbol{\vartheta}}_{QMLE} - \boldsymbol{\vartheta}_0)}$$

where $\hat{\boldsymbol{\vartheta}}_{EbEE}$ and $\hat{\boldsymbol{\vartheta}}_{QMLE}$ denote respectively the EbEE and FQMLE. Because the computation time of the FQMLE is enormous when m is large, the RE and CPU times are only computed on 1 simulation, but they are typical of what is generally observed. When $m \leq 9$, the accuracies are very similar, with a slight advantage to the FQMLE (which corresponds here to the MLE). When the number of parameters becomes too large ($m > 9$) the computation time of the FQMLE becomes prohibitive, and more importantly the optimization fails to give a reasonable value of $\hat{\boldsymbol{\vartheta}}_{QMLE}$ (see the RE for $m \geq 10$).

6.1.2. *Estimating a DCC model by two-step EbE and by FQML*

We now compare the standard one-step FQMLE with the two step method described in Section 5 in the case of a bivariate cDCC-GARCH(1,1) model defined by (2.7) and (4.2), with a full matrix $\boldsymbol{A} = \boldsymbol{A}_1$ and a diagonal matrix $\boldsymbol{B} = \boldsymbol{B}_1$. The value of $\boldsymbol{\vartheta}_0$ is given in the first column of Table 2, and $(\boldsymbol{\eta}_t)$ is an iid sequence distributed as a Student distribution with $\nu = 7$ degrees of freedom, standardized in such a way that $\text{Var}(\boldsymbol{\eta}_t) = \boldsymbol{I}_2$. Note that our Monte Carlo experiment is restricted to a bivariate model because the computation time of the FQMLE is too demanding when $m > 2$.[9] Table 2 summarizes the distribution of the two estimators over 100 independent simulations of the length $n = 1000$ of the model. The EbEE is remarkably more accurate than the FQMLE, whatever the parameter. The FQMLE produces more outliers (such as for example $\hat{b} = 0$) than the EbEE, the Root Mean Square Errors (RMSE) of estimation are much smaller for EbEE than for the FQMLE, and the interquartile range is also in clear favor of the EbEE. One difficulty encountered in the implementation of the two estimators is that the constraints $\rho(\boldsymbol{B}) < 1$ and $\beta < 1$ are not sufficient to ensure the non explosiveness of $\boldsymbol{Q}_t(\boldsymbol{\vartheta})$ as $t \to \infty$. The problem seems to be more severe for the FQMLE than for the EbEE, which may explain the surprisingly poor performance of the FQMLE compared to the EbEE.

6.2. Empirical examples

6.2.1. Dealing with missing or asynchronous data

One problem encountered in modelling multivariate financial series is that the different return components may not be available over the same time horizon. An obvious solution is to discard the dates for which at least one return is missing but this may entail serious information losses. More sophisticated approaches are based on a reconstruction of the missing data (for instance using the Kalman filter). Another issue with financial returns is the lack of synchronicity. For daily returns, the time of measurement is typically the closing time, which can be very different for series across different markets entering in the construction of portfolios. Different techniques of synchronization have been proposed. For instance Audrino and Bühlmann (2004) developed a procedure for the CCC-GARCH(1,1) model. However, the need to choose an auxiliary model for predicting the missing observations may be found unsatisfactory.

The EbE procedure has interest for both issues, missing data and asynchronicity. First, the estimation of a given equation generally does not require observability of the whole

[9] Results reported in the supplementary file illustrate the ability of the EbEE to estimate the individual volatilities of a cDCC for larger dimensions ($m > 2$).

Table 2. Empirical distributions of the EbEE and QMLE over 100 replications for a bivariate DCC-GARCH(1,1) of length $n = 1000$.

	true val.	estim.	bias	RMSE	min	Q_1	Q_2	Q_3	max
$\boldsymbol{\omega}$	0.01	EbEE	0.037	0.134	0.000	0.008	0.014	0.022	0.749
		QMLE	0.116	0.239	0.000	0.009	0.017	0.071	0.808
	0.01	EbEE	0.040	0.159	0.000	0.008	0.014	0.025	0.947
		QMLE	0.104	0.229	0.000	0.009	0.019	0.053	0.788
$\boldsymbol{A}$	0.025	EbEE	-0.001	0.017	0.000	0.014	0.020	0.033	0.114
		QMLE	0.037	0.105	0.000	0.013	0.025	0.042	0.404
	0.025	EbEE	0.005	0.028	0.000	0.017	0.026	0.039	0.237
		QMLE	0.046	0.109	0.000	0.018	0.032	0.057	0.398
	0.025	EbEE	0.011	0.023	0.000	0.025	0.031	0.041	0.150
		QMLE	0.048	0.108	0.000	0.024	0.034	0.056	0.390
	0.025	EbEE	0.000	0.019	0.000	0.012	0.024	0.036	0.116
		QMLE	0.040	0.107	0.000	0.014	0.027	0.044	0.378
diag($\boldsymbol{B}$)	0.94	EbEE	-0.058	0.194	0.000	0.909	0.932	0.944	0.976
		QMLE	-0.157	0.319	0.000	0.823	0.926	0.944	0.972
	0.94	EbEE	-0.049	0.193	0.000	0.912	0.934	0.948	1.001
		QMLE	-0.147	0.309	0.000	0.838	0.925	0.943	0.987
$\boldsymbol{S}[1,2]$	0.3	EbEE	-0.001	0.137	-0.020	0.215	0.308	0.396	0.610
		QMLE	0.024	0.222	-0.624	0.206	0.336	0.428	0.900
α	0.4	EbEE	0.002	0.015	0.009	0.032	0.043	0.051	0.093
		QMLE	0.017	0.048	0.000	0.033	0.046	0.061	0.352
β	0.95	EbEE	-0.013	0.028	0.853	0.923	0.936	0.955	0.983
		QMLE	-0.055	0.172	0.000	0.905	0.931	0.951	0.991

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

returns over the entire period. This is in particular the case for diagonal models. Moreover, the estimation of the correlation matrix in CCC models can be achieved by considering the returns pairwise (see Remark 5.4). The missing data issue is illustrated in the supplementary file. Concerning asynchronicity, we propose the following illustration based on world stock market indices.

At the opening of the New York stock exchange, investors have knowledge of the closing price at the Tokyo stock exchange. It is thus possible to use e.g. the squared return of the Nikkei 225 of day t (say Nik_t) to predict the squared return of the SP500 at the same date (say SP_t). Since Nik_t conveys more recent information than SP_{t-1}, it is reasonable to think that it may appear significantly in the volatility of the SP500 at time t. Modeling the individual volatilities by augmented GARCH models is a convenient way to tackle the problem. For simplicity, we considered only four indices: the SP500 (closing price at around 21 GMT), the CAC and FTSE (closing price at 16:30) and the Nikkei (closing price at 6). As a function of the most recent available returns and a feedback mechanism, the fitted individual volatilities can be written, with obvious notations, as

$$\sigma^2_{\text{SP}_t} = \underset{(0.008)}{0.039} + \underset{(0.013)}{0.064}\ \text{SP}_{t-1} + \underset{(0.009)}{0.038}\ \text{CAC}_t + \underset{(0.020)}{0.187}\ \text{FTSE}_t + \underset{(0.003)}{0.000}\ \text{Nik}_t + \underset{(0.024)}{0.660}\sigma^2_{\text{SP}_{t-1}}$$

$$\sigma^2_{\text{CAC}_t} = \underset{(0.010)}{0.042} + \underset{(0.014)}{0.050}\ \text{SP}_{t-1} + \underset{(0.012)}{0.064}\ \text{CAC}_{t-1} + \underset{(0.018)}{0.036}\ \text{FTSE}_{t-1} + \underset{(0.004)}{0.015}\ \text{Nik}_t + \underset{(0.018)}{0.844}\sigma^2_{\text{CAC}_{t-1}}$$

$$\sigma^2_{\text{FTSE}_t} = \underset{(0.004)}{0.013} + \underset{(0.007)}{0.039}\ \text{SP}_{t-1} + \underset{(0.004)}{0.000}\ \text{CAC}_{t-1} + \underset{(0.0010)}{0.071}\ \text{FTSE}_{t-1} + \underset{(0.002)}{0.006}\ \text{Nik}_t + \underset{(0.013)}{0.869}\sigma^2_{\text{CAC}_{t-1}}$$

$$\sigma^2_{\text{Nik}_t} = \underset{(0.015)}{0.068} + \underset{(0.016)}{0.055}\ \text{SP}_{t-1} + \underset{(0.011)}{0.006}\ \text{CAC}_{t-1} + \underset{(0.019)}{0.010}\ \text{FTSE}_{t-1} + \underset{(0.014)}{0.108}\ \text{Nik}_{t-1} + \underset{(0.019)}{0.826}\sigma^2_{\text{CAC}_{t-1}}$$

where the estimated standard deviations, obtained from Theorem 3.1, are given into brackets. It is seen that, for instance, the FTSE at time t has strong influence on the volatility of the SP500 at the same date (but a few hours later). Thus, by taking into account the availability of the most recent observations the model reveals spillover effects between series.

6.2.2. SC models for exchange rates

In this section, we will illustrate the interest of the EbE approach for SC models. We consider returns series of the daily exchange rates of the Canadian Dollar (CAD), the Swiss Franc (CHF), the Chinese Yuan (CNY), the British Pound (GBP), the Japanese Yen (JPY) and the American Dollar (USD) with respect to the Euro. The observations have been downloaded from the website `http://www.ecb.int/home/html/index.en.html`, and cover the period from January 14, 2000 to May 16, 2013, which corresponds to 2081 observations. On these 6 series, we fitted a CCC-GARCH(1,1) model of the form

$$\underline{h}_t = \underline{\omega} + \mathbf{A}\underline{\epsilon}_{t-1} + \mathbf{B}\underline{h}_{t-1}$$

where $\mathbf{B}$ is diagonal. This assumption allows to fit the model equation by equation. The estimated values of $\mathbf{A}$ and $\mathbf{B}$ are

$$\hat{\mathbf{A}} = \begin{pmatrix} \underset{(0.010)}{0.029} & \underset{(0.003)}{0.002} & \underset{(0.040)}{0.015} & \underset{(0.013)}{0.012} & \underset{(0.003)}{0.003} & \underset{(0.038)}{0.000} \\ \underset{(0.002)}{0.000} & \underset{(0.023)}{0.136} & \underset{(0.004)}{0.000} & \underset{(0.003)}{0.003} & \underset{(0.001)}{0.000} & \underset{(0.003)}{0.000} \\ \underset{(0.005)}{0.000} & \underset{(0.002)}{0.002} & \underset{(0.028)}{0.031} & \underset{(0.007)}{0.008} & \underset{(0.002)}{0.002} & \underset{(0.027)}{0.001} \\ \underset{(0.004)}{0.006} & \underset{(0.002)}{0.001} & \underset{(0.020)}{0.004} & \underset{(0.012)}{0.041} & \underset{(0.002)}{0.006} & \underset{(0.019)}{0.000} \\ \underset{(0.012)}{0.017} & \underset{(0.005)}{0.003} & \underset{(0.054)}{0.000} & \underset{(0.016)}{0.002} & \underset{(0.012)}{0.061} & \underset{(0.052)}{0.000} \\ \underset{(0.005)}{0.000} & \underset{(0.002)}{0.003} & \underset{(0.028)}{0.024} & \underset{(0.007)}{0.007} & \underset{(0.002)}{0.002} & \underset{(0.028)}{0.008} \end{pmatrix}, \quad \hat{\mathbf{B}} = \mathrm{diag} \begin{pmatrix} \underset{(0.022)}{0.92} \\ \underset{(0.017)}{0.88} \\ \underset{(0.010)}{0.95} \\ \underset{(0.015)}{0.93} \\ \underset{(0.014)}{0.93} \\ \underset{(0.009)}{0.96} \end{pmatrix},$$

and the estimation of the correlation matrix $\boldsymbol{R}$ is

$$\hat{\boldsymbol{R}} = \begin{pmatrix} 1.00 & \underset{(0.026)}{0.00} & \underset{(0.039)}{0.46} & \underset{(0.031)}{0.39} & \underset{(0.034)}{0.17} & \underset{(0.032)}{0.47} \\ 0.00 & 1.00 & \underset{(0.040)}{0.14} & \underset{(0.027)}{0.12} & \underset{(0.043)}{0.42} & \underset{(0.045)}{0.13} \\ 0.46 & 0.14 & 1.00 & \underset{(0.033)}{0.44} & \underset{(0.039)}{0.58} & \underset{(0.031)}{0.98} \\ 0.39 & 0.12 & 0.44 & 1.00 & \underset{(0.071)}{0.26} & \underset{(0.040)}{0.45} \\ 0.17 & 0.42 & 0.58 & 0.26 & 1.00 & \underset{(0.044)}{0.57} \\ 0.47 & 0.13 & 0.98 & 0.45 & 0.57 & 1.00 \end{pmatrix} \begin{matrix} \text{CAD} \\ \text{CHF} \\ \text{CNY} \\ \text{GBP} \\ \text{JPY} \\ \text{USD} \end{matrix}$$

The estimated standard deviations of the estimators were obtained from Theorem 5.2 and are displayed into brackets. It can be noted that the different exchange rates are mainly linked by the strong cross correlations of the residuals, which can be interpreted as an effect of instantaneous causality between the squared returns. By contrast, in view of the (almost) diagonal form of $\hat{\mathbf{A}}$, the volatility of a given exchange rate is mainly explained by its own past returns. A noticeable exception is the volatility of the USD which shows more sensitivity to the variations of the CNY than to its own variations. These two exchange rates are also strongly related by the correlation (0.98) between their rescaled residuals.

We now relax the constant correlation assumption (2.9) by considering a DCC matrix $\boldsymbol{R}_t^*$ of the form (5.2) with $N = 2$ regimes. The estimates of the GARCH(1,1) parameters are unchanged, but the estimated CCC matrix $\hat{\boldsymbol{R}}$ is replaced by the estimates $\hat{\boldsymbol{R}}^*(1)$ and $\hat{\boldsymbol{R}}^*(2)$ of the correlation matrix in each of the two regimes, respectively given by

$$\begin{pmatrix} 1.00 & \underset{(0.15)}{0.38} & \underset{(0.06)}{0.71} & \underset{(0.14)}{0.69} & \underset{(0.12)}{0.58} & \underset{(0.06)}{0.72} \\ 0.38 & 1.00 & \underset{(0.14}{0.59} & \underset{(0.11)}{0.52} & \underset{(0.06)}{0.66} & \underset{(0.14)}{0.59} \\ 0.71 & 0.59 & 1.00 & \underset{(0.13)}{0.81} & \underset{(0.10)}{0.89} & \underset{(0.00}{0.99} \\ 0.69 & 0.52 & 0.81 & 1.00 & \underset{(0.15)}{0.76} & \underset{(0.14)}{0.82} \\ 0.58 & 0.66 & 0.89 & 0.76 & 1.00 & \underset{(0.10)}{0.90} \\ 0.72 & 0.59 & 0.99 & 0.82 & 0.90 & 1.00 \end{pmatrix} \text{ and } \begin{pmatrix} 1.00 & \underset{(0.04)}{-0.04} & \underset{(0.03)}{0.42} & \underset{(0.03)}{0.34} & \underset{(0.04)}{0.10} & \underset{(0.03)}{0.43} \\ -0.04 & 1.00 & \underset{(0.04)}{0.08} & \underset{(0.04)}{0.08} & \underset{(0.03)}{0.39} & \underset{(0.04)}{0.07} \\ 0.42 & 0.08 & 1.00 & \underset{(0.04)}{0.38} & \underset{(0.03)}{0.52} & \underset{(0.00)}{0.98} \\ 0.34 & 0.08 & 0.38 & 1.00 & \underset{(0.05)}{0.18} & \underset{(0.04)}{0.38} \\ 0.10 & 0.39 & 0.52 & 0.18 & 1.00 & \underset{(0.03)}{0.51} \\ 0.43 & 0.07 & 0.98 & 0.38 & 0.51 & 1.00 \end{pmatrix}.$$

The estimated standard deviations of the estimators, in parentheses, are obtained by taking the empirical standard deviations of the estimates of $N = 100$ independent simulations of the DCC model that have been fitted on the real data set.

The transition probabilities of the Markov chain are estimated by $\hat{p}(1,1) = 0.826$, $\hat{p}(1,2) = 0.174$, $\hat{p}(2,1) = 0.039$ and $\hat{p}(2,2) = 0.961$, with respective estimated standard deviations 0.036, 0.036, 0.013 and 0.013. This corresponds to regimes with relative frequencies $\hat{P}(\Delta_t = 1) = 0.18$ and $\hat{P}(\Delta_t = 2) = 0.82$. The second regime being the most frequent, it is not surprising to observe that $\hat{\boldsymbol{R}}^*(2)$ and $\hat{\boldsymbol{R}}$ are close. It seems however that the introduction of two regimes is relevant. Indeed, the less frequent regime is characterized by significantly more correlated residuals. Figure 1 illustrates the high positive correlation between the GBP and JPY residuals when the most probable regime is the first one (left figure). Examen of the filtered probabilities (see the supplementary file for a graph) shows that the regime with the highest residual correlations (*i.e.* regime 1) is often more plausible when the volatilities are high.

6.2.3. *Bivariate BEKK for exchange rates?*

Finally, we tested the adequacy of bivariate BEKK models on the same exchange rates series, using Proposition 4.1. For each pair of exchange rates, we estimated Model (4.5) and we tested the restrictions $H_0^{(1)}$ and $H_0^{(2)}$ that are satisfied when the DGP is the BEKK-GARCH(1,1) model (4.3). Table 3 shows that, for 12 bivariate series over 15, either $H_0^{(1)}$ or $H_0^{(2)}$ is clearly rejected, which invalidates the adequacy of the bivariate BEKK model for

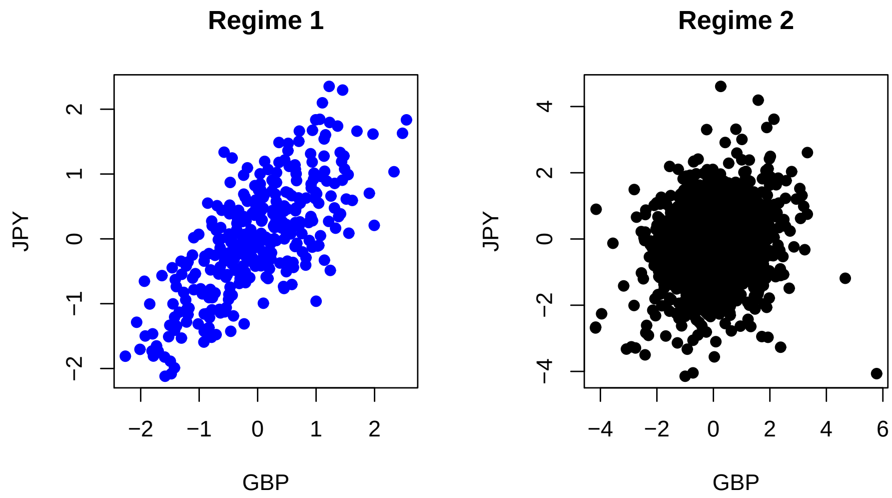

Figure 1. GBP and JPY residuals as function of the most probable regime

the 12 pairs. Using the Bonferroni correction, one can indeed reject the model at significant level less than α if one of the two hypothesis $H_0^{(k)}$ is rejected at the level $\alpha/2$. This does not mean that a global BEKK model would be rejected for the vector of 6 series. An extension of Proposition 4.1 for larger m would allow to perform a test but such an extension is left for future research.

Table 3. For each pair of exchange rates: p-values of the tests of the null hypotheses $H_0^{(1)}$ and $H_0^{(2)}$ implied by the bivariate BEKK-GARCH(1,1) model. Gray cells contain p-values less than 2.5%.

	CAD		CHF		CNY		GBP		JPY	
	$H_0^{(1)}$	$H_0^{(2)}$	$H_0^{(1)}$	$H_0^{(2)}$	$H_0^{(1)}$	$H_0^{(2)}$	$H_0^{(1)}$	$H_0^{(2)}$	$H_0^{(1)}$	$H_0^{(2)}$
CHF	0.000	0.163								
CNY	0.120	0.015	0.122	0.500						
GBP	0.012	0.023	0.128	0.000	0.005	0.100				
JPY	0.007	0.006	0.500	0.500	0.500	0.087	0.050	0.000		
USD	0.500	0.021	0.114	0.000	0.500	0.381	0.068	0.000	0.102	0.000

7. Conclusion

EbE estimation of MGARCH models is a standard method used in applied works to alleviate the computational burden implied by large cross-sectional dimensions. In this study, we established asymptotic properties of the EbEE of the individual conditional variances, under general assumptions on their parameterization. Unexpectedly, we found that such EbE estimators may be superior to the QMLE in terms of asymptotic accuracy. Our framework covers the most widely used MGARCH models in financial applications. For semi-diagonal BEKK models and DCC models, the asymptotic results were shown to hold under explicit conditions. In the former case, we explained how to test the constraints implied by the BEKK specification. For CCC models (including the standard CCC-GARCH(p, q) model) we proved the consistency and the joint asymptotic normality of the EbE volatility and correlation matrix estimators.

The main motivation for using an EbE approach in applications is the important gains in computation time, and our simulation experiments confirmed that such gains can be huge. For moderate dimensions the global QML estimator can even be unfeasible, while we did not encounter such difficulties with the EbE approach. Our experiments revealed that the EbE estimator may be superior to the QMLE in terms of accuracy, not only for the volatility parameters but also for the parameters of a DCC specification of the conditional correlation. For real series, the separate estimation of the volatilities allows to handle, without discarding too many data, series that are not available at the same date, or at the same hour for daily returns. Stochastic correlation models, in which the correlation matrix is not only driven by the past but also by a latent variable, can also be handled by this approach. On exchange rates data, we found evidence of a two-regime Markov-switching stochastic correlation. The asymptotic properties of the estimators of the correlations and transition probabilities are an area for future research.

Appendix

A. Technical assumptions

We make the following assumptions on the volatility function.

A2: for any real sequence $(e_i)_{i\geq 1}$, the function $\boldsymbol{\theta}^{(k)} \mapsto \sigma_k(e_1, e_2, \dots; \boldsymbol{\theta}^{(k)})$ is continuous and there exists a measurable function $K : \mathbb{R}^\infty \mapsto (0, \infty)$ such that

$$|\sigma_k(e_1, e_2, \dots; \boldsymbol{\theta}^{(k)}) - \sigma_k(e_1, e_2, \dots; \boldsymbol{\theta}_0^{(k)})| \leq K(e_1, \dots)\|\boldsymbol{\theta}^{(k)} - \boldsymbol{\theta}_0^{(k)}\|,$$

and

$$E\left(\frac{K(\boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \dots)}{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}\right)^2 < \infty.$$

A3: there exists a neighborhood $\mathcal{V}(\boldsymbol{\theta}_0^{(k)})$ of $\boldsymbol{\theta}_0^{(k)}$ such that

$$E \sup_{\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \left(\frac{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}{\sigma_{kt}(\boldsymbol{\theta}^{(k)})}\right)^2 < \infty.$$

A5: we have $\sigma_{kt}(\boldsymbol{\theta}_0^{(k)}) = \sigma_{kt}(\boldsymbol{\theta}^{(k)})$ $a.s.$ iff $\boldsymbol{\theta}^{(k)} = \boldsymbol{\theta}_0^{(k)}$.

The next assumption allows to show that initial values have no effect on the asymptotic properties of the estimator of $\boldsymbol{\theta}_0^{(k)}$. Let $\Delta_{kt}(\boldsymbol{\theta}^{(k)}) = \tilde{\sigma}_{kt}(\boldsymbol{\theta}^{(k)}) - \sigma_{kt}(\boldsymbol{\theta}^{(k)})$, $a_t = \sup_k \sup_{\boldsymbol{\theta}^{(k)} \in \boldsymbol{\Theta}^{(k)}} |\Delta_{kt}(\boldsymbol{\theta}^{(k)})|$. Let C and ρ be generic constants with $C > 0$ and $0 < \rho < 1$. The "constant" C is allowed to depend on variables anterior to $t = 0$.

A6: We have $a_t \leq C\rho^t$, a.s.

To derive the asymptotic distribution of $\hat{\boldsymbol{\theta}}_n$, the following additional assumptions are considered.

A9: for any real sequence $(e_i)_{i\geq 1}$, the function $\boldsymbol{\theta}^{(k)} \mapsto \sigma_k(e_1, e_2, \dots; \boldsymbol{\theta}^{(k)})$ has continuous second-order derivatives;

A10: there exists a neighborhood $\mathcal{V}(\boldsymbol{\theta}_0^{(k)})$ of $\boldsymbol{\theta}_0^{(k)}$ such that

$$\sup_{\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \left\| \frac{1}{\sigma_{kt}(\boldsymbol{\theta}^{(k)})} \frac{\partial \sigma_{kt}(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)}} \right\|^{4(1+\frac{1}{\delta})}, \quad \sup_{\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \left\| \frac{1}{\sigma_{kt}(\boldsymbol{\theta}^{(k)})} \frac{\partial^2 \sigma_{kt}(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)} \partial \boldsymbol{\theta}^{(k)\prime}} \right\|^{2(1+\frac{1}{\delta})},$$

$$\sup_{\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \left| \frac{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}{\sigma_{kt}(\boldsymbol{\theta}^{(k)})} \right|^4,$$

have finite expectations.

The next assumption is introduced to handle initial values.

A11: We have

$$b_t := \sup_k \sup_{\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \left\| \frac{\partial \Delta_{kt}(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)}} \right\| \leq C\rho^t, \quad a.s.$$

The next assumption will be used to show the invertibility of the asymptotic covariance matrix.

A12: For $k = 1, \ldots, m$ and for any $\boldsymbol{x} \in \mathbb{R}^{d_k}$, we have: $\quad \boldsymbol{x}' \frac{\partial \sigma_{kt}^2(\boldsymbol{\theta}_0^{(k)})}{\partial \boldsymbol{\theta}^{(k)}} = 0, \ a.s. \quad \Rightarrow \quad \boldsymbol{x} = 0.$

The next assumption is used in Theorem 3.2.

A10*: there exists a neighborhood $\mathcal{V}(\boldsymbol{\theta}_0^{(k)})$ of $\boldsymbol{\theta}_0^{(k)}$ such that

$$\sup_{\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \left\| \frac{1}{\sigma_{kt}(\boldsymbol{\theta}^{(k)})} \frac{\partial \sigma_{kt}(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)}} \right\|^4, \quad \sup_{\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \left\| \frac{1}{\sigma_{kt}(\boldsymbol{\theta}^{(k)})} \frac{\partial^2 \sigma_{kt}(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)} \partial \boldsymbol{\theta}^{(k)\prime}} \right\|^2,$$

$$\sup_{\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \left| \frac{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}{\sigma_{kt}(\boldsymbol{\theta}^{(k)})} \right|^4,$$

have finite expectations.

B. Proofs

To save space, the proofs of Proposition 3.1, Theorem 3.3, Proposition 4.1 and Theorem 5.1 are displayed in the supplementary file.

B.1. Proof of Theorem 3.1

Because the proof of the consistency follows along the same lines as that of Theorem 7.1 in Francq and Zakoïan (2010) we omit details (see the supplementary file). To prove the asymptotic normality, define $\tilde{\ell}_{kt}$ as ℓ_{kt}, with σ_{kt} replaced by $\tilde{\sigma}_{kt}$. The proof relies on a set of preliminary results.

$i)\ E\left\|\frac{\partial \ell_{kt}(\boldsymbol{\theta}_0^{(k)})}{\partial \boldsymbol{\theta}^{(k)}}\frac{\partial \ell_{kt}(\boldsymbol{\theta}_0^{(k)})}{\partial \boldsymbol{\theta}^{(k)\prime}}\right\|<\infty, \quad E\left\|\frac{\partial^2 \ell_{kt}(\boldsymbol{\theta}_0^{(k)})}{\partial \boldsymbol{\theta}^{(k)}\partial \boldsymbol{\theta}^{(k)\prime}}\right\|<\infty,$

$ii)$ There exists a neighbourhood $\mathcal{V}(\boldsymbol{\theta}_0^{(k)})$ of $\boldsymbol{\theta}_0^{(k)}$ such that

$$\sup_{\boldsymbol{\theta}^{(k)}\in\mathcal{V}(\boldsymbol{\theta}_0^{(k)})}\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^n \frac{\partial \ell_{kt}(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)}}-\frac{\partial \tilde{\ell}_{kt}(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)}}\right\|\to 0\ ,$$

$iii)\ \frac{1}{n}\sum_{t=1}^n \frac{\partial^2 \ell_{kt}(\boldsymbol{\theta}_n^{(k)})}{\partial \boldsymbol{\theta}^{(k)}\partial \boldsymbol{\theta}^{(k)\prime}}\to \boldsymbol{J}_{kk}\ ,\quad$ a.s. for any $\boldsymbol{\theta}_n^{(k)}$ between $\hat{\boldsymbol{\theta}}_n^{(k)}$ and $\boldsymbol{\theta}_0^{(k)}$,

$iv)$ $\boldsymbol{J}_{kk}$ is non singular,

$v)\ \frac{1}{\sqrt{n}}\sum_{t=1}^n \frac{\partial \ell_{kt}(\boldsymbol{\theta}_0^{(k)})}{\partial \boldsymbol{\theta}^{(k)}}\xrightarrow{\mathcal{L}}\mathcal{N}\left(\mathbf{0},\boldsymbol{I}_{kk}\right).$

Note that

$$\begin{aligned}
\frac{\partial \ell_{kt}(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)}} &= \left\{1-\frac{\epsilon_{kt}^2}{\sigma_{kt}^2}\right\}\left\{\frac{2}{\sigma_{kt}}\frac{\partial \sigma_{kt}}{\partial \boldsymbol{\theta}^{(k)}}\right\},\\
\frac{\partial^2 \ell_{kt}(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)}\partial \boldsymbol{\theta}^{(k)\prime}} &= \left\{1-\frac{\epsilon_{kt}^2}{\sigma_{kt}^2}\right\}\left\{\frac{2}{\sigma_{kt}}\frac{\partial^2 \sigma_{kt}}{\partial \boldsymbol{\theta}^{(k)}\partial \boldsymbol{\theta}^{(k)\prime}}\right\}\\
&\quad +2\left\{3\frac{\epsilon_{kt}^2}{\sigma_{kt}^2}-1\right\}\left\{\frac{1}{\sigma_{kt}}\frac{\partial \sigma_{kt}}{\partial \boldsymbol{\theta}^{(k)}}\right\}\left\{\frac{1}{\sigma_{kt}}\frac{\partial \sigma_{kt}}{\partial \boldsymbol{\theta}^{(k)\prime}}\right\}. \qquad \text{(B.1)}
\end{aligned}$$

Let $\|\cdot\|_r$ denote the L^r norm, for $r\geq 1$, on the space of real random variables. We have, by the Hölder inequality,

$$\left\|\left(1-\eta_{kt}^{*2}\right)\frac{1}{\sigma_{kt}}\frac{\partial \sigma_{kt}}{\partial \boldsymbol{\theta}^{(k)}}(\boldsymbol{\theta}_0^{(k)})\right\|_2 \quad\leq\quad \|1-\eta_{kt}^{*2}\|_{2(\delta+1)}\left\|\frac{1}{\sigma_{kt}}\frac{\partial \sigma_{kt}}{\partial \boldsymbol{\theta}^{(k)}}(\boldsymbol{\theta}_0^{(k)})\right\|_{2(1+1/\delta)},$$

which is finite by Assumptions **A8** and **A10**. The first result in i) follows. The second result can be shown similarly.

Now, turning to ii), we have

$$\begin{aligned}
&\left\|\frac{\partial \ell_{kt}(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)}}-\frac{\partial \tilde{\ell}_{kt}(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)}}\right\|\\
=&\ \left\|\left\{\frac{\epsilon_{kt}^2}{\tilde{\sigma}_{kt}^2}-\frac{\epsilon_{kt}^2}{\sigma_{kt}^2}\right\}\left\{\frac{2}{\sigma_{kt}}\frac{\partial \sigma_{kt}}{\partial \boldsymbol{\theta}^{(k)}}\right\}+2\left\{1-\frac{\epsilon_{kt}^2}{\tilde{\sigma}_{kt}^2}\right\}\left\{\frac{1}{\sigma_{kt}}-\frac{1}{\tilde{\sigma}_{kt}}\right\}\left\{\frac{\partial \sigma_{kt}}{\partial \boldsymbol{\theta}^{(k)}}\right\}\right.\\
&\left.+\left\{1-\frac{\epsilon_{kt}^2}{\tilde{\sigma}_{kt}^2}\right\}\left\{\frac{2}{\tilde{\sigma}_{kt}}\right\}\left\{\frac{\partial \sigma_{kt}}{\partial \boldsymbol{\theta}^{(k)}}-\frac{\partial \tilde{\sigma}_{kt}}{\partial \boldsymbol{\theta}^{(k)}}\right\}\right\|(\boldsymbol{\theta}^{(k)})\leq C\rho^t u_t,
\end{aligned}$$

where

$$u_t=(1+\eta_{kt}^{*2})\left(1+\sup_{\boldsymbol{\theta}^{(k)}\in\mathcal{V}(\boldsymbol{\theta}_0^{(k)})}\left\|\frac{1}{\sigma_{kt}}\frac{\partial \sigma_{kt}}{\partial \boldsymbol{\theta}^{(k)}}(\boldsymbol{\theta}^{(k)})\right\|\right)\left(1+\sup_{\boldsymbol{\theta}^{(k)}\in\mathcal{V}(\boldsymbol{\theta}_0^{(k)})}\left|\frac{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}{\sigma_{kt}(\boldsymbol{\theta}^{(k)})}\right|^2\right),$$

as a consequence of Assumptions **A4, A6** and **A11**. We have $E|u_t| < \infty$ by Assumption **A10** and the Cauchy-Schwarz inequality. Thus $C\sum_{t=1}^{n} \rho^t u_t$ is bounded a.s., which entails ii).

To prove iii), by Exercise 7.9 in Francq and Zakoïan (2010), it will be sufficient to establish that for any $\varepsilon > 0$, there exists a neighborhood $\mathcal{V}(\boldsymbol{\theta}_0^{(k)})$ of $\boldsymbol{\theta}_0^{(k)}$ such that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \sup_{\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \left\| \frac{\partial^2 \ell_{kt}(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)} \partial \boldsymbol{\theta}^{(k)\prime}} - \frac{\partial^2 \ell_{kt}(\boldsymbol{\theta}_0^{(k)})}{\partial \boldsymbol{\theta}^{(k)} \partial \boldsymbol{\theta}^{(k)\prime}} \right\| \leq \varepsilon \quad a.s. \tag{B.2}$$

By the ergodic theorem, the limit in the left-hand side is equal to

$$E \sup_{\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \left\| \frac{\partial^2 \ell_{kt}(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)} \partial \boldsymbol{\theta}^{(k)\prime}} - \frac{\partial^2 \ell_{kt}(\boldsymbol{\theta}_0^{(k)})}{\partial \boldsymbol{\theta}^{(k)} \partial \boldsymbol{\theta}^{(k)\prime}} \right\|$$

provided that this expectation is finite. In view of **A9**, the conclusion will follow by the dominated convergence theorem: the latter expectation tends to zero when the neighborhood $\mathcal{V}(\boldsymbol{\theta}_0^{(k)})$ shrinks to the singleton $\{\boldsymbol{\theta}_0^{(k)}\}$. To complete the proof of iii), it thus remains to show that

$$E \sup_{\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \left\| \frac{\partial^2 \ell_{kt}(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)} \partial \boldsymbol{\theta}^{(k)\prime}} \right\| < \infty. \tag{B.3}$$

Let us consider the first product in the right-hand side of (B.1). We have, by the Hölder inequality,

$$\begin{aligned} &E \sup_{\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \left\| \left\{ 1 - \frac{\epsilon_{kt}^2}{\sigma_{kt}^2} \right\} \left\{ \frac{1}{\sigma_{kt}} \frac{\partial^2 \sigma_{kt}}{\partial \boldsymbol{\theta}^{(k)} \partial \boldsymbol{\theta}^{(k)\prime}} \right\} \right\| \\ \leq\ & \left\{ 1 + \|\eta_{kt}^{*2}\|_{2(1+\delta)} \left\| \sup_{\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \frac{\sigma_{kt}^2(\boldsymbol{\theta}_0^{(k)})}{\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})} \right\|_2 \right\} \left\| \sup_{\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \left\| \frac{1}{\sigma_{kt}} \frac{\partial^2 \sigma_{kt}}{\partial \boldsymbol{\theta}^{(k)} \partial \boldsymbol{\theta}^{(k)\prime}} (\boldsymbol{\theta}^{(k)}) \right\| \right\|_{2(1+1/\delta)}, \end{aligned}$$

which is finite by Assumptions **A8** and **A10**. The second product in the right-hand side of (B.1)) can be handled similarly. Thus iii) is established.

The invertibility of $\boldsymbol{J}_{kk}$ is a straightforward consequence of **A12**. Now

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial \ell_{kt}(\boldsymbol{\theta}_0^{(k)})}{\partial \boldsymbol{\theta}^{(k)}} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{1 - \eta_{kt}^{*2}\} \boldsymbol{d}_{kt},$$

and v) follows from the Central Limit Theorem of Billingsley (1961) for ergodic, stationary and square integrable martingale differences. Indeed, the square integrability follows from Hölder's inequality,

$$E\left(\{1 - \eta_{kt}^{*2}\}^2 \|\boldsymbol{d}_{kt} \boldsymbol{d}_{kt}'\|\right) \leq \|(1 - \eta_{kt}^{*2})^2\|_{1+\delta} \|\boldsymbol{d}_{kt} \boldsymbol{d}_{kt}'\|_{1+1/\delta},$$

and Assumptions **A8** and **A10**. Moreover, $(\boldsymbol{\eta}_t^*)$ is strictly stationary and ergodic as a function of the process $(\boldsymbol{\epsilon}_t)$.

We are now in a position to complete the proof of Theorem 3.1. Since $\hat{\boldsymbol{\theta}}_n^{(k)}$ converges to $\boldsymbol{\theta}_0^{(k)}$, which stands in the interior of the parameter space by **A7**, the derivative of the criterion $\tilde{Q}_n^{(k)}$ is equal to zero at $\hat{\boldsymbol{\theta}}_n^{(k)}$. In view of point ii), we thus have by a Taylor expansion of $Q_n^{(k)}$ at $\boldsymbol{\theta}_0^{(k)}$,

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n^{(k)}-\boldsymbol{\theta}_0^{(k)}\right) \overset{o_P(1)}{=} -\left(\frac{1}{n}\sum_{t=1}^n \frac{\partial^2 \ell_{kt}(\boldsymbol{\theta}_{ij}^{*(k)})}{\partial\theta_i^{(k)}\partial\theta_j^{(k)}}\right)^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^n\frac{\partial}{\partial\boldsymbol{\theta}^{(k)}}\ell_{kt}(\boldsymbol{\theta}_0^{(k)})$$

where the $\boldsymbol{\theta}_{ij}^{*(k)}$'s are between $\hat{\boldsymbol{\theta}}_n^{(k)}$ and $\boldsymbol{\theta}_0^{(k)}$. The conclusion follows from the intermediate results i)-v). □

B.2. Proof of Theorem 3.2

Note that under the independence assumption (3.3),

$$E|\eta_{kt}^*|^s<\infty \quad \text{and} \quad E|\epsilon_{kt}|^s=E|\sigma_{kt}|^sE|\eta_{kt}^*|^s<\infty$$

imply $E|\sigma_{kt}|^s<\infty$. Therefore the condition $E\log\sigma_{kt}^2<\infty$ can be omitted in **A1**.

The proof of the asymptotic normality relies on the same steps *i)-v)* as in the proof of the second part of Theorem 3.1, except that one can replace $\boldsymbol{I}_{kk}$ by $(E\eta_{k1}^4-1)\boldsymbol{J}_{kk}$, and the assumptions **A8** and **A10** by **A8*** and **A10***. In particular, to show (B.3), note that, by the Cauchy-Schwarz inequality,

$$\begin{aligned}&E\sup_{\boldsymbol{\theta}^{(k)}\in\mathcal{V}(\boldsymbol{\theta}_0^{(k)})}\left\|\left\{1-\frac{\epsilon_{kt}^2}{\sigma_{kt}^2}\right\}\left\{\frac{1}{\sigma_{kt}}\frac{\partial^2\sigma_{kt}}{\partial\boldsymbol{\theta}^{(k)}\partial\boldsymbol{\theta}^{(k)\prime}}\right\}\right\|\\ \le\;&\left\{1+\|\eta_{kt}^*\|_2\left\|\sup_{\boldsymbol{\theta}^{(k)}\in\mathcal{V}(\boldsymbol{\theta}_0^{(k)})}\frac{\sigma_{kt}^2(\boldsymbol{\theta}_0^{(k)})}{\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})}\right\|_2\right\}\left\|\sup_{\boldsymbol{\theta}^{(k)}\in\mathcal{V}(\boldsymbol{\theta}_0^{(k)})}\left\|\frac{1}{\sigma_{kt}}\frac{\partial^2\sigma_{kt}}{\partial\boldsymbol{\theta}^{(k)}\partial\boldsymbol{\theta}^{(k)\prime}}(\boldsymbol{\theta}^{(k)})\right\|\right\|_2,\end{aligned}$$

which is finite under Assumptions **A8*** and **A10***. □

B.3. Proof of Proposition 3.2

Recall that for any spherically distributed variable $\boldsymbol{X}=(X_1,\dots,X_m)'$, we have $\boldsymbol{\lambda}'\boldsymbol{X}\overset{d}{=}\|\boldsymbol{\lambda}\|X_1$ for any $\boldsymbol{\lambda}\in\mathbb{R}^m$, where $\overset{d}{=}$ stands for equality in distribution and $\|\cdot\|$ denotes the Euclidian norm on $\mathbb{R}^m$. Letting $\boldsymbol{e}_k$ the k-th column of $\boldsymbol{I}_m$, we have

$$\eta_{kt}^*=\boldsymbol{e}_k'\boldsymbol{R}_t^{1/2}\boldsymbol{\eta}_t\overset{d}{=}\|\boldsymbol{e}_k'\boldsymbol{R}_t^{1/2}\|\eta_1=\eta_1 \tag{B.4}$$

conditionally to $\mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}$, and thus unconditionally. Since the conditional and unconditional distributions of η_{kt}^* coincide, (3.3) holds. We also note that $\boldsymbol{\eta}_t^*$ belongs to $\mathcal{F}_t^{\boldsymbol{\epsilon}}$. The independence of the sequence (η_{kt}^*) follows. □

B.4. Proof of Proposition 3.3

In view of the Markov vector representation of (3.5), $\boldsymbol{h}_t = \boldsymbol{\omega} + \boldsymbol{A}_k(\eta_{kt}^*)\boldsymbol{h}_{t-1}$, with $\underline{\omega} = (\omega_{0k}, 0, \dots, 0)'$, $\boldsymbol{h}_{kt} = (\sigma_{kt}^2, \dots, \sigma_{k,t-r+1}^2)'$, we have

$$\boldsymbol{h}_{kt} = \boldsymbol{\omega} + \sum_{i=1}^{\infty} \boldsymbol{A}_k(\eta_{kt}^*) \dots \boldsymbol{A}_k(\eta_{k,t-i+1}^*)\boldsymbol{\omega}.$$

The existence of the infinite sum is justified by the assumption $\gamma(\mathbf{A_k}) < 0$ (see, for instance, Francq and Zakoian (Section 2.2.2, 2010)). The strict stationarity of (ϵ_{kt}) follows. The asymptotic properties of the EbEE estimator thus follow from Francq and Zakoïan (2004) and Escanciano (2009).

B.5. Proof of Theorem 4.1

The proof consists in verifying the conditions required in Theorem 3.1. By Boussama, Fuchs and Stelzer (2011) and Aielli (2013), the conditions of the theorem ensure the existence of a strictly stationary, non anticipative and ergodic solution $[\mathrm{vech}(\boldsymbol{R}_t)', \boldsymbol{\eta}_t^{*'}]'$, that is a measurable function of $\{\boldsymbol{\eta}_{t-u}, u \geq 0\}$, to the equations $\boldsymbol{\eta}_t^* = \boldsymbol{R}_t^{1/2}\boldsymbol{\eta}_t$ and (4.2). Now, in view of $\epsilon_{\ell t} = \sigma_{\ell t}\eta_{\ell t}^*$, we have

$$\sigma_{\ell t}^2 = \omega_\ell + a_\ell(\eta_{\ell,t-1}^*)\sigma_{\ell,t-1}^2 = \omega_\ell \left(1 + \sum_{s=1}^{\infty} a_\ell(\eta_{\ell,t-1}^*) \dots a_\ell(\eta_{\ell,t-s}^*)\right), \quad a.s.$$

where $a_\ell(x) = \alpha_\ell x^2 + \beta_\ell$. The a.s. convergence follows from the Cauchy rule for positive terms series, which can be applied because

$$E \log a_\ell(\eta_{\ell,t}^*) \leq \log E a_\ell(\eta_{\ell,t}^*) = \log(\alpha_\ell + \beta_\ell) < 0.$$

It follows that $\boldsymbol{\epsilon}_t$ is a measurable function of $\{\boldsymbol{\eta}_{t-u}, u \geq 0\}$. Using the second equality in (2.3), we have by the law of iterated conditional expectations $E a_\ell(\eta_{\ell,t-1}^*) \dots a_\ell(\eta_{\ell,t-s}^*) = (\alpha_\ell + \beta_\ell)^s$, from which we deduce that $E\sigma_{\ell t}^2 < \infty$. Thus **A1** is satisfied. We note that

$$\eta_{k,t}^{*2} \leq (\boldsymbol{\eta}_t^*)'\boldsymbol{\eta}_t^* = (\boldsymbol{\eta}_t)'\boldsymbol{R}_t\boldsymbol{\eta}_t \leq \sum_{\ell,\ell'} |\eta_{\ell,t}\eta_{\ell',t}| = \left(\sum_{\ell,\ell'} |\eta_{\ell,t}|\right)^2,$$

from which **A8** follows, using $E\|\boldsymbol{\eta}_t\|^{4(1+\delta)} < \infty$. The other assumptions required to apply Theorem 3.1 can be shown to hold as in standard GARCH(1,1) models (see for instance Francq and Zakoïan (2004)).

B.6. Proof of Theorem 4.2

The existence of a (unique) ergodic, non anticipative, strictly and second-order stationary solution $(\boldsymbol{\epsilon}_t)$ of Model (4.3), under the conditions given in the corollary, follows from Boussama, Fuchs and Stelzer (2011), Theorem 2.4. Thus **A1** holds with $s = 2$. By Proposition 4.5 of the same article, if the spectral radius of $\boldsymbol{A}_0 + \boldsymbol{B}_0$ is less than 1, the spectral radius of $\sum_{j=1}^{p} \boldsymbol{H}_m(\boldsymbol{B}_{0j} \otimes \boldsymbol{B}_{0j})\boldsymbol{K}_m'$ is also less than 1. The latter matrix being diagonal, it can be seen that this entails $\sum_{j=1}^{p} b_{jk}^2 < 1$. Thus, under the strict stationarity condition, it is always possible to choose the compact set $\boldsymbol{\Theta}_k$ so that it contains the true parameter value. Assumption **A4** is satisfied by definition of $\boldsymbol{\Theta}^{(k)}$.

Now we turn to **A5**. Suppose $\sigma_t(\boldsymbol{\theta}_0^{(k)}) = \sigma_t(\boldsymbol{\theta}^{(k)})$, a.s. The polynomial $\mathcal{B}_{\boldsymbol{\theta}^{(k)}}(L)$ being invertible for any $\boldsymbol{\theta}^{(k)} \in \boldsymbol{\Theta}^{(k)}$, we have

$$\begin{aligned} & \mathcal{B}^{-1}_{\boldsymbol{\theta}_0^{(k)}}(L) \sum_{i=1}^{q} \left(\sum_{\ell=1}^{m} \theta^{(k)}_{0,\ell+1+m(i-1)} \epsilon_{\ell,t-i} \right)^2 - \mathcal{B}^{-1}_{\boldsymbol{\theta}^{(k)}}(L) \sum_{i=1}^{q} \left(\sum_{\ell=1}^{m} \theta^{(k)}_{\ell+1+m(i-1)} \epsilon_{\ell,t-i} \right)^2 \\ = \; & \mathcal{B}^{-1}_{\boldsymbol{\theta}^{(k)}}(1)\theta_1^{(k)} - \mathcal{B}^{-1}_{\boldsymbol{\theta}_0^{(k)}}(1)\theta_{01}^{(k)}, \qquad a.s. \end{aligned}$$

Then there exists some variables $c_{t-2}, a_{\ell,\ell',t-2}, \ell, \ell' = 1, \dots m$ belonging to the past of $\boldsymbol{\eta}_{t-1}$ such that

$$c_{t-2} + \sum_{\ell,\ell'=1}^{m} a_{\ell,\ell',t-2}\eta_{\ell,t-1}\eta_{\ell',t-1} = 0.$$

Therefore, if those variables are not all equal to zero, the distribution of $\boldsymbol{\eta}_t$ conditional to the past is degenerate. Since $\boldsymbol{\eta}_t$ is independent from the past, this means that the unconditional distribution of $\boldsymbol{\eta}_t$ is degenerate, in contradiction with the existence of a density around zero. Thus $c_{t-2} = a_{1,1,t-2} = \cdots = a_{m,m,t-2} = 0$, from which we deduce, by iterating the same argument, that $\boldsymbol{\theta}^{(k)} = \boldsymbol{\theta}_0^{(k)}$. Therefore, **A5** is verified. We omit the proof of **A6**, which can be done following the lines of proof of (4.6) in Francq and Zakoïan (2004). Hence the proof of the consistency of $\hat{\boldsymbol{\theta}}_n^{(k)}$ is complete.

Now we turn to the asymptotic normality. Assumption **A7** holds by assumption, and **A8** can be verified by the arguments used in the proof of Theorem 4.1. **A9** is obviously satisfied. The proofs of **A10-A11** being similar to those made for standard GARCH models in Francq and Zakoïan (2004), they will be omitted. To establish **A12**, let $\boldsymbol{x} \in \mathbb{R}^{1+p+qm}$ such that $\boldsymbol{x}' \frac{\partial \sigma^2_{kt}(\boldsymbol{\theta}_0^{(k)})}{\partial \boldsymbol{\theta}^{(k)}} = 0,\ a.s.$ It follows that

$$x_1 + \sum_{i=1}^{q} 2 \left(\sum_{\ell=1}^{m} a_{ik\ell} \epsilon_{\ell,t-i} \right) \boldsymbol{x}' \begin{pmatrix} \mathbf{0}_{1+m(i-1)\times 1} \\ \boldsymbol{\epsilon}_{t-i} \\ \mathbf{0}_{p+m(q-i)\times 1} \end{pmatrix}$$
$$+ \sum_{j=1}^{p} x_{1+qm+j} h_{kk,t-j} + \sum_{j=1}^{p} b_{jk}^2 \boldsymbol{x}' \frac{\partial \sigma^2_{k,t-j}(\boldsymbol{\theta}_0^{(k)})}{\partial \boldsymbol{\theta}^{(k)}} = 0, \qquad a.s.$$

Note that the latter sum is equal to zero by stationarity. We thus have

$$x_1 + 2 \left(\sum_{\ell=1}^{m} a_{1k\ell} \epsilon_{\ell,t-1} \right) \left(\sum_{\ell=1}^{m} x_{1+\ell} \epsilon_{\ell,t-1} \right) + z_{t-2} = 0, \qquad a.s.$$

where z_{t-2} is a variable belonging to the past of $\boldsymbol{\epsilon}_{t-1}$. The arguments given for the proof of **A5** allow to conclude that $x_1 = \cdots = x_{m+1} = 0$. By iterating the argument we get $\boldsymbol{x} = 0$. The asymptotic normality follows from Theorem 3.1. □

B.7. Proof of Theorem 5.2

Let

$$\dot{\boldsymbol{\ell}}_t(\boldsymbol{\theta}) = \left(\frac{\partial}{\partial \boldsymbol{\theta}^{(1)'}} \ell_{1t}(\boldsymbol{\theta}^{(1)}), \ldots, \frac{\partial}{\partial \boldsymbol{\theta}^{(m)'}} \ell_{mt}(\boldsymbol{\theta}^{(m)}) \right)'.$$

For $\boldsymbol{\theta}^{(k)} \in \boldsymbol{\Theta}^{(k)}$ let $\tilde{\eta}^*_{kt}(\boldsymbol{\theta}^{(k)}) = \tilde{\sigma}^{-1}_{kt}(\boldsymbol{\theta}^{(k)})\epsilon_{kt}$ and $\eta^*_{kt}(\boldsymbol{\theta}^{(k)}) = \sigma^{-1}_{kt}(\boldsymbol{\theta}^{(k)})\epsilon_{kt}$. The proof relies on a set of preliminary results.

$$i)\ E \left\| \dot{\boldsymbol{\ell}}_t(\boldsymbol{\theta}_0) \dot{\boldsymbol{\ell}}_t'(\boldsymbol{\theta}_0) \right\| < \infty,$$

$$ii)\ \sup_{\boldsymbol{\theta} \in \mathcal{V}(\boldsymbol{\theta}_0)} \left\| \frac{1}{n} \sum_{t=1}^{n} \frac{\partial \mathrm{vech}^0 \left\{ \boldsymbol{\eta}_t^* (\boldsymbol{\eta}_t^*)' \right\}}{\partial \boldsymbol{\theta}'} - \frac{\partial \mathrm{vech}^0 \left\{ \tilde{\boldsymbol{\eta}}_t^* (\tilde{\boldsymbol{\eta}}_t^*)' \right\}}{\partial \boldsymbol{\theta}'} \right\| \to 0\ , \quad \text{in probability,}$$

$$iii) \frac{1}{n} \sum_{t=1}^{n} \left(\frac{\partial \mathrm{vech}^0 \left\{ \boldsymbol{\eta}_t^* (\boldsymbol{\eta}_t^*)' \right\}}{\partial \boldsymbol{\theta}'} \right)_{\boldsymbol{\theta}_n} \to -\frac{1}{2} \boldsymbol{\Lambda}', \quad \text{a.s.} \quad \text{for any } \boldsymbol{\theta}_n \text{ between } \hat{\boldsymbol{\theta}}_n \text{ and } \boldsymbol{\theta}_0,$$

$$iv)\ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \begin{pmatrix} \dot{\boldsymbol{\ell}}_t(\boldsymbol{\theta}_0) \\ \mathrm{vech}^0 \left\{ \boldsymbol{\eta}_t^* (\boldsymbol{\eta}_t^*)' - \boldsymbol{R} \right\} \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N} \left(\mathbf{0}, \begin{pmatrix} \boldsymbol{J}^* & \boldsymbol{L} \\ \boldsymbol{L}' & \boldsymbol{\Gamma} \end{pmatrix} \right),$$

Point *i)* follows from the arguments given to prove i) in the proof of the asymptotic normality of $\hat{\boldsymbol{\theta}}_n$ (Theorem 3.1). Point ii) is equivalent to

$$\sup_{\boldsymbol{\theta} \in \mathcal{V}(\boldsymbol{\theta}_0)} \left\| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial \left(\eta^*_{kt} \eta^*_{\ell t} \right)}{\partial \boldsymbol{\theta}'}(\boldsymbol{\theta}) - \frac{\partial \left(\tilde{\eta}^*_{kt} \tilde{\eta}^*_{\ell t} \right)}{\partial \boldsymbol{\theta}'}(\boldsymbol{\theta}) \right\| \to 0\ , \quad \text{in probability.}$$

In view of

$$\frac{\partial}{\partial \boldsymbol{\theta}'}\left\{\eta_{kt}^{*}\eta_{\ell t}^{*}\right\}(\boldsymbol{\theta}) = -\frac{\epsilon_{kt}}{\sigma_{kt}(\boldsymbol{\theta}^{(k)})}\frac{\epsilon_{\ell t}}{\sigma_{\ell t}(\boldsymbol{\theta}^{(\ell)})}\left(\frac{1}{\sigma_{kt}}\frac{\partial \sigma_{kt}(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}'} + \frac{1}{\sigma_{\ell t}}\frac{\partial \sigma_{\ell t}(\boldsymbol{\theta}^{(\ell)})}{\partial \boldsymbol{\theta}'}\right),$$

and the same equality for $\partial\left\{\tilde{\eta}_{kt}^{*}\tilde{\eta}_{\ell t}^{*}\right\}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'$, with σ_{kt} and $\sigma_{\ell t}$ replaced by $\tilde{\sigma}_{kt}$ and $\tilde{\sigma}_{\ell t}$, the conclusion follows by the arguments used to establish ii) in the proof of the asymptotic normality of $\hat{\boldsymbol{\theta}}_n$.

Now we turn to iii). Note that

$$\frac{\partial}{\partial \boldsymbol{\theta}'}\left\{\eta_{kt}^{*}\eta_{\ell t}^{*}\right\}(\boldsymbol{\theta}_0) = -\eta_{kt}^{*}\eta_{\ell t}^{*}\left(\frac{1}{\sigma_{kt}}\frac{\partial \sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}{\partial \boldsymbol{\theta}'} + \frac{1}{\sigma_{\ell t}}\frac{\partial \sigma_{\ell t}(\boldsymbol{\theta}_0^{(\ell)})}{\partial \boldsymbol{\theta}'}\right).$$

Thus, letting $d_{(k)} = \sum_{i=k+1}^{m} d_i$ and ${}_{(k)}d = \sum_{i=1}^{k-1} d_i$, with obvious conventions when $k=1$ or $k=m$,

$$E\left(\frac{\partial}{\partial \boldsymbol{\theta}'}\left\{\eta_{kt}^{*}\eta_{\ell t}^{*}\right\}(\boldsymbol{\theta}_0)\right) = -\frac{1}{2}R_{k\ell}[(\mathbf{0}_{1\times {}_{(k)}d}\ \boldsymbol{\Omega}_k'\ \mathbf{0}_{1\times d_{(k)}}) + (\mathbf{0}_{1\times {}_{(\ell)}d}\ \boldsymbol{\Omega}_\ell'\ \mathbf{0}_{1\times d_{(\ell)}})]$$

Therefore, we have

$$E\left(\frac{\partial}{\partial \boldsymbol{\theta}'}\left(\mathrm{vech}^0\left\{\boldsymbol{\eta}_t^{*}\left(\boldsymbol{\eta}_t^{*}\right)'\right\}\right)_{\boldsymbol{\theta}_0}\right) = -\frac{1}{2}\begin{pmatrix}\boldsymbol{\Omega}'\mathbf{M}_{21}\\ \boldsymbol{\Omega}'\mathbf{M}_{31}\\ \vdots\\ \boldsymbol{\Omega}'\mathbf{M}_{m,m-1}\end{pmatrix} = -\frac{1}{2}\boldsymbol{\Lambda}'.$$

By the law of large numbers, it follows that

$$\frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial \mathrm{vech}^0\left\{\boldsymbol{\eta}_t^{*}\left(\boldsymbol{\eta}_t^{*}\right)'\right\}}{\partial \boldsymbol{\theta}'}\right)_{\boldsymbol{\theta}_0} \to -\frac{1}{2}\boldsymbol{\Lambda}', \quad \text{a.s.}$$

To complete the proof of iii), we will show that similarly to (B.2), for any $\varepsilon > 0$, there exists a neighborhood $\mathcal{V}(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ such that, almost surely,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)}\left\|\left(\frac{\partial \mathrm{vech}^0\left\{\boldsymbol{\eta}_t^{*}\left(\boldsymbol{\eta}_t^{*}\right)'\right\}}{\partial \boldsymbol{\theta}'}\right)_{\boldsymbol{\theta}} - \left(\frac{\partial \mathrm{vech}^0\left\{\boldsymbol{\eta}_t^{*}\left(\boldsymbol{\eta}_t^{*}\right)'\right\}}{\partial \boldsymbol{\theta}'}\right)_{\boldsymbol{\theta}_0}\right\| \le \varepsilon.$$

The latter convergence is equivalent to

$$\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\sup_{\boldsymbol{\theta}^{(k)}\in\mathcal{V}(\boldsymbol{\theta}_0^{(k)})}\sup_{\boldsymbol{\theta}^{(\ell)}\in\mathcal{V}(\boldsymbol{\theta}_0^{(\ell)})}\left\|\frac{\epsilon_{kt}}{\sigma_{kt}(\boldsymbol{\theta}^{(k)})}\frac{\epsilon_{\ell t}}{\sigma_{\ell t}(\boldsymbol{\theta}^{(\ell)})}\frac{1}{\sigma_{kt}(\boldsymbol{\theta}^{(k)})}\frac{\partial \sigma_{kt}(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}'}\right.$$
$$\left. - \frac{\epsilon_{kt}}{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}\frac{\epsilon_{\ell t}}{\sigma_{\ell t}(\boldsymbol{\theta}_0^{(\ell)})}\frac{1}{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}\frac{\partial \sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}{\partial \boldsymbol{\theta}'}\right\| \le \varepsilon, \quad a.s. \qquad \text{(B.5)}$$

for any $k, \ell = 1, \dots, m$. By the arguments used to prove iii) in the proof of the asymptotic normality of $\hat{\boldsymbol{\theta}}_n$, we have

$$E \sup_{\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \sup_{\boldsymbol{\theta}^{(\ell)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(\ell)})} \left\| \frac{\epsilon_{kt}}{\sigma_{kt}(\boldsymbol{\theta}^{(k)})} \frac{\epsilon_{\ell t}}{\sigma_{\ell t}(\boldsymbol{\theta}^{(\ell)})} \frac{1}{\sigma_{kt}(\boldsymbol{\theta}^{(k)})} \frac{\partial \sigma_{kt}(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}'} \right\| < \infty,$$

from which (B.5) follows. Thus, iii) is established.

It remains to show iv). We note that

$$\boldsymbol{Z}_t := \begin{pmatrix} \dot{\boldsymbol{\ell}}_t(\boldsymbol{\theta}_0) \\ \text{vech}^0 \left\{ \boldsymbol{\eta}_t^* \left(\boldsymbol{\eta}_t^* \right)' - \boldsymbol{R} \right\} \end{pmatrix} = \begin{pmatrix} F(\boldsymbol{\eta}_t^*) \boldsymbol{d}_t \\ \text{vech}^0 \left\{ \boldsymbol{\eta}_t^* \left(\boldsymbol{\eta}_t^* \right)' - \boldsymbol{R} \right\} \end{pmatrix}$$

is measurable with respect to the σ-field $\mathcal{F}_t$ generated by $\{\boldsymbol{\eta}_u^*, u \le t\}$. We have, using the independence of the sequence $(\boldsymbol{\eta}_t^*)$ under (2.9),

$$\begin{aligned} \text{Var}(F(\boldsymbol{\eta}_t^*)\boldsymbol{d}_t) &= E\{F(\boldsymbol{\eta}_t^*)E(\boldsymbol{d}_t\boldsymbol{d}_t')F(\boldsymbol{\eta}_t^*)\} = \boldsymbol{J}^*, \\ \text{Cov}\left[F(\boldsymbol{\eta}_t^*)\boldsymbol{d}_t, \text{vech}^0 \left\{ \boldsymbol{\eta}_t^* \left(\boldsymbol{\eta}_t^* \right)' \right\}\right] &= E\left\{F(\boldsymbol{\eta}_t^*)\boldsymbol{\Omega} \left[\text{vech}^0 \left\{ \boldsymbol{\eta}_t^* \left(\boldsymbol{\eta}_t^* \right)' \right\}\right]'\right\} = \boldsymbol{L}. \end{aligned}$$

Thus, $\forall \lambda \in \mathbb{R}^{d+m(m-1)/2}$, the sequence $\{\lambda' \boldsymbol{Z}_t, \mathcal{F}_t\}_t$ is an ergodic, stationary and square integrable martingale difference. The conclusion follows from the central limit theorem of Billingsley (1961).

We are now in a position to complete the proof of Theorem 5.2. Since $\hat{\boldsymbol{\theta}}_n^{(k)}$ converges to $\boldsymbol{\theta}_0^{(k)}$, which stands in the interior of the parameter space by **A7**, the derivative of the criterion $\tilde{Q}_n^{(k)}$ is equal to zero at $\hat{\boldsymbol{\theta}}_n^{(k)}$. In view of point ii), we thus have by a Taylor expansion of $Q_n^{(k)}$ at $\boldsymbol{\theta}_0^{(k)}$,

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n^{(k)} - \boldsymbol{\theta}_0^{(k)}\right) \overset{o_P(1)}{=} -\left(\frac{1}{n}\sum_{t=1}^n \frac{\partial^2 \ell_{kt}(\boldsymbol{\theta}_{ij}^{*(k)})}{\partial \theta_i^{(k)} \partial \theta_j^{(k)}}\right)^{-1} \frac{1}{\sqrt{n}}\sum_{t=1}^n \frac{\partial}{\partial \boldsymbol{\theta}^{(k)}} \ell_{kt}(\boldsymbol{\theta}_0^{(k)})$$

where the $\boldsymbol{\theta}_{ij}^{*(k)}$'s are between $\hat{\boldsymbol{\theta}}_n^{(k)}$ and $\boldsymbol{\theta}_0^{(k)}$. Thus we have, using iii) and iv),

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \overset{o_P(1)}{=} -\boldsymbol{J}_0^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^n \dot{\boldsymbol{\ell}}_t(\boldsymbol{\theta}_0).$$

Another Taylor expansion around $\boldsymbol{\theta}_0$ yields,

$$\begin{aligned} &\sqrt{n}(\hat{\boldsymbol{\rho}}_n - \boldsymbol{\rho}_0) \\ = \ &\frac{1}{\sqrt{n}}\sum_{t=1}^n \text{vech}^0 \left\{ \boldsymbol{\eta}_t^* \left(\boldsymbol{\eta}_t^* \right)' - \boldsymbol{R} \right\} + \frac{1}{n}\sum_{t=1}^n \frac{\partial}{\partial \boldsymbol{\theta}'} \left(\text{vech}^0 \left\{ \tilde{\boldsymbol{\eta}}_t^* \left(\tilde{\boldsymbol{\eta}}_t^* \right)' \right\} \right)_{\tilde{\boldsymbol{\theta}}_n} \sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right), \end{aligned}$$

where $\tilde{\boldsymbol{\theta}}_n$ is between $\hat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}_0$, and

$$\tilde{\boldsymbol{\eta}}_t^* = \tilde{\boldsymbol{\eta}}_t^*(\boldsymbol{\theta}) = \tilde{\boldsymbol{D}}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\epsilon}_t \quad \text{and} \quad \tilde{\boldsymbol{D}}_t(\boldsymbol{\theta}) = \text{diag}\{\tilde{\sigma}_{1t}(\boldsymbol{\theta}^{(1)}), \dots, \tilde{\sigma}_{mt}(\boldsymbol{\theta}^{(m)})\}.$$

It follows that, using v) and vi), denoting by $\boldsymbol{I}$ the identity matrix of size $m(m-1)/2$ and by $\mathbf{0}$ is null matrix of size $d \times m(m-1)/2$,

$$\begin{pmatrix} \sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \\ \sqrt{n}(\hat{\boldsymbol{\rho}}_n - \boldsymbol{\rho}_0) \end{pmatrix} \overset{o_P(1)}{=} \begin{pmatrix} -\boldsymbol{J}_0^{-1} & \mathbf{0} \\ \frac{1}{2}\boldsymbol{\Lambda}'\boldsymbol{J}_0^{-1} & \boldsymbol{I} \end{pmatrix} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \boldsymbol{Z}_t. \tag{B.6}$$

The asymptotic distribution of Theorem 5.2 thus follows from iv).

It remains to establish that $\boldsymbol{\Sigma}$ is non singular. By (B.6), it suffices to show that $\text{Var}(\boldsymbol{Z}_t)$ is nonsingular. We will show that for any $\boldsymbol{x} = (\boldsymbol{x}_i) \in \mathbb{R}^d$, where $\boldsymbol{x}_i \in \mathbb{R}^{d_i}$, for any $\boldsymbol{y} = (y_{k\ell}) \in \mathbb{R}^{m(m-1)/2}$ and any $c \in \mathbb{R}$,

$$\boldsymbol{x}'\dot{\boldsymbol{\ell}}_t(\boldsymbol{\theta}_0) + \boldsymbol{y}'\text{vech}^0\left\{\boldsymbol{\eta}_t^*\left(\boldsymbol{\eta}_t^*\right)' - \boldsymbol{R}\right\} = c, \ a.s. \quad \Rightarrow \quad \boldsymbol{x} = 0 \text{ and } \boldsymbol{y} = 0. \tag{B.7}$$

Assume that the left-hand side of (B.7) holds. Then we have

$$\sum_{i=1}^{m} \left(1 - \eta_{it}^{*2}\right) z_{i,t-1} + \sum_{k \neq \ell} y_{k\ell}\left(\eta_{kt}^*\eta_{\ell t}^* - R_{k\ell}\right) = c$$

where $z_{i,t-1} = \frac{1}{\sigma_{it}^2}\boldsymbol{x}_i'\frac{\partial \sigma_{it}^2(\boldsymbol{\theta}_0^{(i)})}{\partial \boldsymbol{\theta}^{(i)}}$. It follows that $\boldsymbol{a}_{t-1}'\text{vech}\left\{\boldsymbol{\eta}_t^*\left(\boldsymbol{\eta}_t^*\right)'\right\} = b_{t-1}$ for some vector $\boldsymbol{a}_{t-1}$ and some scalar b_{t-1} belonging to the past. By **A13** and (2.9) we must have $\boldsymbol{a}_{t-1} = \mathbf{0}$ and $b_{t-1} = 0$. Noting that the $z_{i,t-1}$ are components of $\boldsymbol{a}_{t-1}$, we must have $z_{i,t-1} = 0$ for $i = 1, \dots, m$, in contradiction with **A12** unless if $\boldsymbol{x} = \mathbf{0}$. It is then straightforward to show that $\boldsymbol{y} = 0$ and the proof is complete. □

References

Aielli, G.P. (2013) Dynamic conditional correlation: on properties and estimation. *Journal of Business & Economic Statistics* 31, 282–299.

Audrino, F. and P. Bühlmann (2004) Synchronizing multivariate financial time series. *The Journal of Risk* 6, 81–106.

Avarucci, M., Beutner, E. and P. Zaffaroni (2013) On moment conditions for quasi-maximum likelihood estimation of multivariate ARCH models. *Econometric Theory* 29, 545–566.

Bauwens, L., Hafner, C. and S. Laurent (2012) *Handbook of volatility models and their applications.* Wiley.

Bauwens, L., Laurent, S. and J.V.K. Rombouts (2006) Multivariate GARCH models: a survey. *Journal of Applied Econometrics* 21, 79–109.

Berkes, I. and L. Horváth (2004) The efficiency of the estimators of the parameters in GARCH processes. *The Annals of Statistics* 32, 633–655.

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.

Bollerslev, T. (1990) Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *Review of Economics and Statistics* 72, 498–505.

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., Fuchs, F. and R. Stelzer (2011) Stationarity and Geometric Ergodicity of BEKK Multivariate GARCH Models. *Stochastic Processes and their Applications* 121, 2331–2360.

Caporin, M. and M. McAleer (2012) Do we really need both BEKK and DCC? A tale of two multivariate GARCH models. *Journal of Economic Surveys* 26, 736–751.

Comte, F. and O. Lieberman (2003) Asymptotic Theory for Multivariate GARCH Processes. *Journal of Multivariate Analysis* 84, 61–84.

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. and B. Kelly (2012) Dynamic equicorrelation. *Journal of Business & Economic Statistics* 30, 212–228.

Engle, R.F. and K. Kroner (1995) Multivariate simultaneous GARCH. *Econometric Theory* 11, 122–150.

Engle, R.F. and K. Sheppard (2001) Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH, University of California San Diego, Discussion paper.

Engle, R.F., Ng, V.K. and M. Rotschild (1990) Asset pricing with a factor ARCH covariance structure: empirical estimates for treasury bills. *Journal of Econometrics* 45, 213-238.

Engle, R.F., Shephard, N. and K. Sheppard (2008) Fitting vast dimensional time-varying covariance models. NYU Working Paper FIN-08-009.

Escanciano, J.C. (2009) Quasi-maximum likelihood estimation of semi-strong GARCH models. *Econometric Theory* 25, 561–570.

Fermanian, J.D. and H. Malongo (2014) Asymptotic theory of DCC models. Unpublished document.

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 (2010) *GARCH models: structure, statistical inference and financial applications.* Chichester: John Wiley.

Francq, C. and J-M. Zakoïan (2012) QML estimation of a class of multivariate asymmetric GARCH models. *Econometric Theory* 28, 179–206.

Hafner, C. and O. Reznikova (2012) On the estimation of dynamic correlation models. *Computational Statistics and Data Analysis* 56, 3533–3545.

He, C. and T. Teräsvirta (2004) An extended constant conditional correlation GARCH model and its fourth-moment structure. *Econometric Theory* 20, 904–926.

Hörmann, S. (2008) Augmented GARCH sequences: Dependence structure and asymptotics. *Bernoulli* 14, 543–561.

Jeantheau, T. (1998) Strong consistency of estimators for multivariate ARCH models. *Econometric Theory* 14, 70–86.

Lanne, M. and P. Saikkonen (2007) A multivariate generalized orthogonal factor GARCH model. *Journal of Business & Economic Statistics* 25, 61–75.

Laurent, S., Rombouts, J.V.K. and F. Violante On the forecasting accuracy of multivariate GARCH models. *Journal of Applied Econometrics* 27, 934–955.

Ling, S. (2007) Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models. *Journal of Econometrics* 140, 849–873.

Ling, S. and M. McAleer (2003) Asymptotic theory for a vector ARMA-GARCH model. *Econometric Theory* 19, 280–310.

Pelletier, D. (2006) Regime switching for dynamic correlations. *Journal of Econometrics* 131, 445–473.

Silvennoinen, A. and T. Teräsvirta (2009) Multivariate GARCH models. *Handbook of Financial Time Series* T.G. Andersen, R.A. Davis, J-P. Kreiss and T. Mikosch, eds. New York: Springer.

Sucarrat, G., Grønneberg, S. and A. Escribano (2013) Estimation and inference in univariate and multivariate LOG-GARCH-X models when the conditional density is unknown. MPRA Discussion Paper No. 49344.

Tsay, R.S. (2014) Multivariate time series: with R and financial applications. John Wiley.

Tse, Y.K. and A. Tsui (2002) A multivariate GARCH model with time-varying correlations. *Journal of Business and Economic Statistics* 20, 351–362.

van der Weide, R. (2002) GO-GARCH: A multivariate generalized orthogonal GARCH model. *Journal of Applied Econometrics* 17, 549–564.

Estimating multivariate GARCH and Stochastic Correlation models equation by equation: complementary results

A. For Section 3

A.1. Proof of the consistency in Theorem 3.1

Let

$$Q_n^{(k)}(\boldsymbol{\theta}^{(k)}) = \frac{1}{n}\sum_{t=1}^{n} \log \sigma_{kt}^2\left(\boldsymbol{\theta}^{(k)}\right) + \frac{\epsilon_{kt}^2}{\sigma_{kt}^2\left(\boldsymbol{\theta}^{(k)}\right)} := \frac{1}{n}\sum_{t=1}^{n} \ell_{kt}(\boldsymbol{\theta}^{(k)}).$$

The strong consistency of $\hat{\boldsymbol{\theta}}_n^{(k)}$ is a consequence of the following intermediate results:

$$i)\ \lim_{n\to\infty} \sup_{\boldsymbol{\theta}^{(k)}\in\boldsymbol{\Theta}^{(k)}} |Q_n^{(k)}(\boldsymbol{\theta}^{(k)}) - \tilde{Q}_n^{(k)}(\boldsymbol{\theta}^{(k)})| = 0\ , \quad a.s.,$$

$$ii)\ \mathbb{E}|\ell_{k,1}(\boldsymbol{\theta}_0^{(k)})| < \infty, \quad \text{and if } \boldsymbol{\theta}^{(k)} \neq \boldsymbol{\theta}_0^{(k)}\ , \quad \mathbb{E}\ell_{k,1}(\boldsymbol{\theta}_0^{(k)}) < \mathbb{E}\ell_{k,1}(\boldsymbol{\theta}^{(k)})\ ,$$

$iii)$ any $\boldsymbol{\theta}^{(k)} \neq \boldsymbol{\theta}_0^{(k)}$ has a neighborhood $V(\boldsymbol{\theta}^{(k)})$ such that

$$\liminf_{n\to\infty} \inf_{\boldsymbol{\theta}^*\in V(\boldsymbol{\theta}^{(k)})} \tilde{Q}_n^{(k)}(\boldsymbol{\theta}^*) > \limsup_{n\to\infty} \tilde{Q}_n^{(k)}(\boldsymbol{\theta}_0^{(k)})\ , \quad a.s.$$

The proof follows along the same lines as the proof of Theorem 7.1 in Francq and Zakoïan (2010). It is easy to see that i) follows from **A4**, **A6** and the existence of $E|\epsilon_{kt}|^s$. Now, in view of (2.3), we have

$$E\ell_{kt}(\boldsymbol{\theta}^{(k)}) = E\left\{\frac{\sigma_{kt}^2(\boldsymbol{\theta}_0^{(k)})\eta_{kt}^{*2}}{\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})} + \log\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})\right\} = E\frac{\sigma_{kt}^2(\boldsymbol{\theta}_0^{(k)})}{\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})} + E\log\sigma_{kt}^2(\boldsymbol{\theta}^{(k)}).$$

Since $E\log\sigma_{kt}^2 < \infty$, we have $E\ell_{kt}(\boldsymbol{\theta}_0^{(k)}) < \infty$, whereas $E\ell_{kt}(\boldsymbol{\theta}^{(k)}) > -\infty$, for any $\boldsymbol{\theta}^{(k)} \in \boldsymbol{\Theta}^{(k)}$, by **A4**. Using the elementary inequality $\log x \leq x - 1$ and **A5**, ii) follows. The last point follows from the ergodic theorem, which can be applied for any $\boldsymbol{\theta}^{(k)} \in \boldsymbol{\Theta}^{(k)}$ to the sequence $\inf_{\boldsymbol{\theta}_*\in V(\boldsymbol{\theta}^{(k)})\cap\boldsymbol{\theta}^{(k)}} \ell_{kt}(\boldsymbol{\theta}_*)$, which is strictly stationary and ergodic under **A1** and admits an expectation in $(-\infty, \infty]$.

A.2. Proof of Proposition 3.1

The asymptotic properties of the theoretical QMLE of $\boldsymbol{\theta}_0$ can be established following the lines of proof of Francq and Zakoïan (2012). Details will be omitted. It can be shown that

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n^{QML} - \boldsymbol{\theta}_0\right) \overset{\mathcal{L}}{\to} \mathcal{N}\left\{0, \boldsymbol{J}_{QML}^{-1}\boldsymbol{I}_{QML}\boldsymbol{J}_{QML}^{-1}\right\},$$

where

$$\boldsymbol{I}_{QML} \quad = \quad E\left(\frac{\partial \ell_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\frac{\partial \ell_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}\right), \qquad \boldsymbol{J}_{QML} = E\left(\frac{\partial^2 \ell_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right),$$

with $\ell_t(\boldsymbol{\theta}) = \boldsymbol{\epsilon}_t'\boldsymbol{D}_t^{-1}\boldsymbol{R}^{-1}\boldsymbol{D}_t^{-1}\boldsymbol{\epsilon}_t + \log|\boldsymbol{D}_t\boldsymbol{R}\boldsymbol{D}_t|$ and $\boldsymbol{D}_t = \mathrm{diag}(\sigma_{1t}(\boldsymbol{\theta}^{(1)}),\dots,\sigma_{mt}(\boldsymbol{\theta}^{(m)}))$. Letting $u_{kt} = 1-\sum_{i=1}^m r_{ki}^*\eta_{kt}^*\eta_{it}^*$, for $k=1,\dots,m$, we find that

$$\frac{\partial \ell_t(\boldsymbol{\theta}_0)}{\partial \theta} = \begin{pmatrix} u_{1t}\boldsymbol{I}_{d_1} & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{0} & \ddots & & \vdots \\ \vdots & & \ddots & \boldsymbol{0} \\ \boldsymbol{0} & \dots & \boldsymbol{0} & u_{mt}\boldsymbol{I}_{d_m}\end{pmatrix}\boldsymbol{d}_t,$$

denoting by $\boldsymbol{I}_d$ the identity matrix of size d. It follows that, in bloc-matrix notation,

$$\boldsymbol{I}_{QML} \quad = \quad (\tau_{k\ell}J_{k\ell}).$$

Turning to the second-order derivatives, we note that for any components θ_i, θ_j of $\boldsymbol{\theta}$,

$$\frac{\partial^2 \ell_t(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j} = \boldsymbol{\epsilon}_t'\frac{\partial^2 \boldsymbol{D}_t^{-1}}{\partial\theta_i\partial\theta_j}\boldsymbol{R}^{-1}\boldsymbol{D}_t^{-1}\boldsymbol{\epsilon}_t + \boldsymbol{\epsilon}_t'\boldsymbol{D}_t^{-1}\boldsymbol{R}^{-1}\frac{\partial^2 \boldsymbol{D}_t^{-1}}{\partial\theta_i\partial\theta_j}\boldsymbol{\epsilon}_t + \frac{\partial^2 \log|\boldsymbol{D}_t^2|}{\partial\theta_i\partial\theta_j}.$$

We first consider the derivatives with respect to the first two components of $\boldsymbol{\theta}^{(1)}$. Write $\boldsymbol{d}_{1t}' = (d_{1it})_{i=1,\dots,d_1}$. We have, for $i=1,2$

$$\begin{aligned}
&\frac{\partial^2 \ell_t(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j} \\
= &\left(d_{1it}d_{1jt} - \frac{2}{\sigma_{1t}}\frac{\partial^2\sigma_{1t}}{\partial\theta_i\partial\theta_j}\right)\left(\sum_{k=1}^m r_{1k}^*\frac{\epsilon_{1t}}{\sigma_{1t}}\frac{\epsilon_{kt}}{\sigma_{kt}}\right) + \frac{\epsilon_{1t}^2}{2\sigma_{1t}^2}d_{1it}d_{1jt}r_{11}^* + \left(\frac{2}{\sigma_{1t}}\frac{\partial^2\sigma_{1t}}{\partial\theta_i\partial\theta_j} - \frac{1}{2}d_{1it}d_{1jt}\right) \\
= &\ d_{1it}d_{1jt}\left(\sum_{k=1}^m r_{1k}^*\frac{\epsilon_{1t}}{\sigma_{1t}}\frac{\epsilon_{kt}}{\sigma_{kt}} + \frac{r_{11}^*\epsilon_{1t}^2}{2\sigma_{1t}^2} - \frac{1}{2}\right) + \frac{2}{\sigma_{1t}}\frac{\partial^2\sigma_{1t}}{\partial\theta_i\partial\theta_j}\left(1 - \frac{1}{\sigma_{1t}}\sum_{k=1}^m\frac{r_{1k}^*}{\sigma_{kt}}\epsilon_{1t}\epsilon_{kt}\right).
\end{aligned}$$

Hence

$$\frac{\partial^2 \ell_t(\boldsymbol{\theta}_0)}{\partial\theta_i\partial\theta_j} \quad = \quad d_{1it}d_{1jt}\left(\sum_{k=1}^m r_{1k}^*\eta_{1t}^*\eta_{kt}^* + \frac{1}{2}(r_{11}^*\eta_{1t}^{*2}-1)\right) + \frac{2}{\sigma_{1t}}\frac{\partial^2\sigma_{1t}}{\partial\theta_i\partial\theta_j}\left(1-\sum_{k=1}^m r_{1k}^*\eta_{1t}^*\eta_{kt}^*\right),$$

and thus

$$\begin{aligned}
E\left(\frac{\partial^2 \ell_t(\boldsymbol{\theta}_0)}{\partial\theta_i\partial\theta_j}\right) \quad &= \quad E(d_{1it}d_{1jt})\left(\sum_{k=1}^m r_{1k}^*r_{1k} + \frac{1}{2}(r_{11}^*-1)\right) + E\left(\frac{2}{\sigma_{1t}}\frac{\partial^2\sigma_{1t}}{\partial\theta_i\partial\theta_j}\right)\left(1-\sum_{k=1}^m r_{1k}^*r_{1k}\right) \\
&= \quad \frac{1}{2}E(d_{1it}d_{1jt})\,(r_{11}^*+1).
\end{aligned}$$

It can similarly be shown that, for $k = 1, \ldots, m$,

$$E\left(\frac{\partial^2 \ell_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^{(k)} \partial \boldsymbol{\theta}^{(k)\prime}}\right) = \frac{1}{2} E(\boldsymbol{d}_{kt} \boldsymbol{d}_{kt}^{\prime}) \left(r_{kk}^* + 1\right).$$

and for $\ell \neq k$,

$$E\left(\frac{\partial^2 \ell_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^{(k)} \partial \boldsymbol{\theta}^{(\ell)\prime}}\right) = \frac{1}{2} E(\boldsymbol{d}_{kt} \boldsymbol{d}_{\ell t}^{\prime}) r_{k\ell} r_{k\ell}^*.$$

Finally,

$$\boldsymbol{J}_{QML} = \left(\xi_{k\ell} J_{k\ell}\right).$$

Now by Theorem 5.2, the asymptotic distribution of the EbEE is given by

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \stackrel{\mathcal{L}}{\to} \mathcal{N}\left\{0, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}\right\},$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = \left((\kappa_{k\ell} - 1)\boldsymbol{J}_{kk}^{-1} \boldsymbol{J}_{k\ell} \boldsymbol{J}_{\ell\ell}^{-1}\right)$. In view of (A.1) and (A.2), the QMLE is asymptotically more efficient than the EbEE iff $\boldsymbol{\Sigma}_{\boldsymbol{\theta}} \succ \boldsymbol{J}_{QML}^{-1} \boldsymbol{I}_{QML} \boldsymbol{J}_{QML}^{-1}$, in the sense of positive definite matrices, or equivalently iff $\boldsymbol{J}_{QML} \boldsymbol{\Sigma}_{\boldsymbol{\theta}} \boldsymbol{J}_{QML} \succ \boldsymbol{I}_{QML}$. The conclusion straightforwardly follows.

A.3. Comparisons of the EbEE and the QMLE

We will show that the EbEE may be asymptotically superior to the QMLE when the distribution of $(\boldsymbol{\eta}_t^*)$ is sufficiently far from the Gaussian. To see this, consider the particular case where the only unknown coefficients are the parameters of the first volatility, $\boldsymbol{\theta}_0^{(1)}$. We find that

$$\boldsymbol{I}_{QML} = \tau_{11} J_{11}, \quad \boldsymbol{J}_{QML} = \xi_{11} J_{11}, \quad \boldsymbol{\Sigma}_{\boldsymbol{\theta}} = (\kappa_{11} - 1) J_{11}^{-1}.$$

Then, an adaptation of the proof of Proposition 3.1 shows that, to estimate $\boldsymbol{\theta}_0^{(1)}$, the QMLE is asymptotically less efficient than the EbEE if and only if

$$\tau_{11} - \xi_{11}^2(\kappa_{11} - 1) > 0. \tag{A.3}$$

Let us now show that in the Gaussian case, $\tau_{11} - \xi_{11}^2(\kappa_{11} - 1) < 0$, meaning that the theoretical QMLE is more efficient than the EbEE. First note that $\kappa_{ij} = 1 + 2r_{ij}^2$. Thus we have to show that $\tau_{11} < 2$, or equivalently that

$$\sum_{i,j=1}^{m} r_{1i}^* r_{1j}^* E(\eta_{1t}^{*2} \eta_{it}^* \eta_{jt}^*) = E\left\{\eta_{1t}^{*2}\left(\sum_{i=1}^{m} r_{1i}^* \eta_{it}^*\right)^2\right\} < 3.$$

By the Cauchy-Schwarz inequality, it suffices to show that

$$E\left(\sum_{i=1}^{m} r_{1i}^{*}\eta_{it}^{*}\right)^{4} < 3.$$

The variable inside the parentheses follows a centered Gaussian distribution with variance r_{11}^{*}, so the conclusion follows. It is not surprising that the theoretical QMLE be more efficient than the EbEE as the QMLE coincides with the MLE in this case. We now describe situations where the converse holds true. When $m = 2$, letting $\rho = r_{12}$, condition (A.3) writes

$$E\left\{\left(\eta_{1t}^{*} - \rho\eta_{2t}^{*}\right)\eta_{1t}^{*}\right\}^{2} > \frac{(2-\rho^{2})^{2}}{4}\left(E\eta_{1t}^{*4} - 1\right) + (1-\rho^{2})^{2}. \tag{A.4}$$

A particular case where condition (A.4) holds is by choosing (i) any value $\rho \in (-1, 1), \rho \neq 0$, (ii) for η_{1t}^{*} a distribution such that $E(\eta_{1t}^{*}) = 0$, $E(\eta_{1t}^{*2}) = 1$ and

$$E(\eta_{1t}^{*4}) < 2 - \frac{\rho^{2}}{4 - 3\rho^{2}}$$

and (iii) by setting $\eta_{2t}^{*} = \rho\eta_{1t}^{*} + u_t$ where u_t is any variable independent of η_{1t}^{*}, with $E(u_t) = 0$, $E(u_t^2) = 1 - \rho^2$ and $E(u_t^4) < \infty$. It can be seen that the latter conditions, ensuring the asymptotic superiority of the EbEE over the QMLE, imply a thin tail for the distribution of η_{1t}^{*}. Now choose for η_{2t}^{*} any distribution with $E(\eta_{2t}^{*}) = 0$, $E(\eta_{2t}^{*2}) = 1$ and $E(\eta_{2t}^{*4}) < \infty$. Let $\rho \in (-1, 1), \rho \neq 0$, and let v_t be any variable independent of η_{2t}^{*}, with $E(u_t) = 0$, $E(u_t^2) = 1 - \rho^2$ and a fourth-order moment satisfying

$$\rho^{2}(4-\rho^{2})Ev_t^{4} > (2-\rho^{2})^{2}\left\{6\rho^{2}(1-\rho^{2}) + \rho^{4}E\eta_{2t}^{*4} - 1\right\} - 4(1-\rho^{2})(2\rho^{2}-1),$$

or equivalently

$$(4-\rho^{2})Ev_t^{4} > (2-\rho^{2})^{2}\rho^{2}(E\eta_{2t}^{*4} - 1) + 16 - 37\rho^{2} + 24\rho^{4} - 5\rho^{6}.$$

Then condition (A.4) is satisfied in this case, showing that the EbEE may also be asymptotically superior to the QMLE when the distributions of the innovations have fat/moderate tails.

A.4. Estimating conditional variances in SC models

Because SC models (2.10)-(2.12) satisfy Assumptions (2.4), the volatility parameters $\boldsymbol{\theta}_0^{(k)}$ can be estimated equation by equation, and Theorem 3.1 applies.

We now discuss conditions under which (3.3) holds, in which case the asymptotic covariance matrix of the EbEE simplifies as in Theorem 3.2. The next result shows that when the correlation matrix $\boldsymbol{R}_t^*$ is a function of the latent process $(\boldsymbol{\Delta}_t)$ and when the distribution of $\boldsymbol{\xi}_t$ is spherical, a slightly weaker condition than (3.3) holds. Let $\mathcal{F}_{t-1}^{\boldsymbol{\eta}^*}$ be the σ-field generated by $\{\boldsymbol{\eta}_u^*, u < t\}$.

Proposition A.1. *Assume that the distribution of $\boldsymbol{\xi}_t$ is spherical and that the sequences $(\boldsymbol{\Delta}_t)$ and $(\boldsymbol{\xi}_t)$ are independent. Then, the SC model (2.10)-(2.12) with $\boldsymbol{R}_t^* = \boldsymbol{R}^*(\Delta_t)$ satisfies*

$$\eta_{kt}^* \textit{ is independent from } \mathcal{F}_{t-1}^{\boldsymbol{\eta}^*}. \tag{A.5}$$

Moreover, (η_{kt}^*) is an iid (0,1) sequence.

Proof. Recall that for any spherically distributed variable $\boldsymbol{X} = (X_1, \dots, X_m)'$, we have $\boldsymbol{\lambda}'\boldsymbol{X} \stackrel{d}{=} \|\boldsymbol{\lambda}\| X_1$ for any $\boldsymbol{\lambda} \in \mathbb{R}^m$, where $\stackrel{d}{=}$ stands for equality in distribution and $\|\cdot\|$ denotes the Euclidian norm on $\mathbb{R}^m$. Letting $\boldsymbol{e}_k$ the k-th column of $\boldsymbol{I}_m$, we have

$$\eta_{kt}^* = \boldsymbol{e}_k' \boldsymbol{R}_t^{*1/2} \boldsymbol{\xi}_t \stackrel{d}{=} \|\boldsymbol{e}_k' \boldsymbol{R}_t^{*1/2}\| \xi_1 = \xi_1 \tag{A.6}$$

conditionally to $\boldsymbol{R}_t^*$, and thus unconditionally.

Now for any $x, y \in \mathbb{R}$, using successively the independence between $\boldsymbol{\xi}_t$ et $\boldsymbol{\xi}_{t-1}$ and the independence between $(\boldsymbol{R}_t^*)$ and $(\boldsymbol{\xi}_t)$, for $k, \ell = 1, \dots, m$,

$$\begin{aligned} P(\eta_{kt}^* < x, \eta_{\ell,t-1}^* < y \mid \boldsymbol{R}_t^*, \boldsymbol{R}_{t-1}^*) &= P(\eta_{kt}^* < x \mid \boldsymbol{R}_t^*, \boldsymbol{R}_{t-1}^*) P(\eta_{\ell,t-1}^* < y \mid \boldsymbol{R}_t^*, \boldsymbol{R}_{t-1}^*) \\ &= P(\eta_{kt}^* < x \mid \boldsymbol{R}_t^*) P(\eta_{\ell,t-1}^* < y \mid \boldsymbol{R}_{t-1}^*) \\ &= P(\eta_{kt}^* < x) P(\eta_{\ell,t-1}^* < y), \end{aligned}$$

the last equality following from (A.6). We similarly prove that for any positive integer j

$$P(\eta_{k_1 t}^* < x_1, \dots, \eta_{k_j, t-j+1}^* < x_j) = \prod_{i=1}^{j} P(\eta_{k_i, t-i+1}^* < x_i)$$

for all sequences (k_i) and (x_i). The conclusion follows. □

Remark A.1. It is worth noting that, under the assumptions of Proposition A.1, the process $(\boldsymbol{\eta}_t^*)$ is neither independent nor identically distributed in general (even if its components are iid). To see this, consider for example, for $\lambda_1, \lambda_2 \in \mathbb{R}$ and for $k \neq \ell$,

$$\lambda_1 \eta_{kt}^* + \lambda_2 \eta_{\ell t}^* \stackrel{d}{=} \|(\lambda_1 \boldsymbol{e}_k' + \lambda_2 \boldsymbol{e}_\ell') \boldsymbol{R}_t^{*1/2}\| \xi_1 = \{\lambda_1^2 + \lambda_2^2 + 2\lambda_1\lambda_2 \boldsymbol{R}_t^*(k, \ell)\}^{1/2} \xi_1,$$

conditionally on $\boldsymbol{R}_t^*$, where $\boldsymbol{e}_k$ denotes the k-th column of $\boldsymbol{I}_m$. The variable in the right-hand side of the latter equality is in general non independent of the past values of $\boldsymbol{\eta}_t^*$, and may also not be stationary (except when $\boldsymbol{R}_t^*(k,\ell)$ is stationary).

Since $\boldsymbol{\eta}_t^* = \boldsymbol{D}_t^{-1}\boldsymbol{\epsilon}_t$ with $\boldsymbol{D}_t \in \mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}$, it is clear that $\mathcal{F}_{t-1}^{\boldsymbol{\eta}^*} \subset \mathcal{F}_{t-1}^{\boldsymbol{\epsilon}}$. Therefore (3.3) entails (A.5). Conversely, the equation $\boldsymbol{\epsilon}_t = \boldsymbol{D}_t\boldsymbol{\eta}_t^*$ can be viewed as a GARCH-type model with non iid innovations $(\boldsymbol{\eta}_t^*)$. Under appropriate assumptions on the GARCH recursion defined by $\boldsymbol{D}_t$, the model has a solution of the form $\boldsymbol{\epsilon}_t = \varphi(\boldsymbol{\eta}_t^*, \boldsymbol{\eta}_{t-1}^*, \dots)$ for some measurable function φ. In such a case (3.3) and (A.5) are equivalent, since we have

$$\mathcal{F}_{t-1}^{\boldsymbol{\epsilon}} = \mathcal{F}_{t-1}^{\boldsymbol{\eta}^*}. \tag{A.7}$$

This is illustrated in the following example.

Example A.1 (Information sets). Consider the multivariate stationary ARCH(1) model, in which the diagonal elements of $\boldsymbol{H}_t$ have the form

$$\sigma_{it}^2 = \omega_i + \sum_{j=1}^{m} \alpha_{ij}\epsilon_{j,t-1}^2, \qquad \omega_i > 0,\ \alpha_{ij} \geq 0, \quad i,j = 1,\dots,m.$$

Let $\boldsymbol{h}_t = (\sigma_{1t}^2, \dots, \sigma_{mt}^2)'$ and $\boldsymbol{\omega} = (\omega_1, \dots, \omega_m)'$. We have

$$\boldsymbol{h}_t = \boldsymbol{\omega} + \mathbf{A}(\boldsymbol{\eta}_{t-1}^*)\boldsymbol{h}_{t-1},$$

where $\mathbf{A}(\boldsymbol{\eta}_{t-1}^*) = (\alpha_{ij}\eta_{j,t-1}^{*2})_{i,j}$. It follows that

$$\boldsymbol{h}_t = \left(\boldsymbol{I}_m + \sum_{k=1}^{\infty} \mathbf{A}(\boldsymbol{\eta}_{t-1}^*) \dots \mathbf{A}(\boldsymbol{\eta}_{t-k}^*)\right)\boldsymbol{\omega}. \tag{A.8}$$

Under **A1**, the infinite sum is well-defined and is finite componentwise. Otherwise, the norm of $\boldsymbol{h}_t$ would not be finite with probability 1, and this would contradict the strict stationarity of $\boldsymbol{\epsilon}_t$. In view of (A.8), the σ-fields of $\boldsymbol{\epsilon}$ and $\boldsymbol{\eta}^*$ coincide, in the sense of (A.7).

A straightforward consequence of Proposition A.1 and Theorems 3.1-3.2 is the next result.

Corollary A.1. *For Model* (2.10)-(2.12)*, we have strong consistency of* $\hat{\boldsymbol{\theta}}_n^{(k)}$ *under* **A1** *and* **A4**-**A6**. *If, in addition, the distribution of* $\boldsymbol{\xi}_t$ *is spherical, the sequences* $(\boldsymbol{\Delta}_t)$ *and* $(\boldsymbol{\xi}_t)$ *are independent,* $\mathcal{F}_{t-1}^{\boldsymbol{\epsilon}} = \mathcal{F}_{t-1}^{\boldsymbol{\eta}^*}$*, and* **A7**, **A8***, **A9**, **A10***, **A11**, **A12** *hold, the asymptotic normality in* (3.4) *holds.*

A.5. Proof of Theorem 3.3

The proof of the strong consistency is very similar to that of Theorem 3.1, therefore is it omitted. To establish the asymptotic normality, we use the following derivatives

$$\begin{aligned}
\frac{\partial \ell_{kt}(\boldsymbol{\gamma}^{(k)})}{\partial \boldsymbol{\gamma}^{(k)}} &= \left\{1-\frac{\epsilon_{kt}^2}{\sigma_{kt}^2}\right\}\left\{\frac{2}{\sigma_{kt}}\frac{\partial \sigma_{kt}}{\partial \boldsymbol{\gamma}^{(k)}}\right\}+\frac{2\epsilon_{kt}}{\sigma_{kt}^2}\frac{\partial \epsilon_{kt}}{\partial \boldsymbol{\gamma}^{(k)}},\\
\frac{\partial^2 \ell_{kt}(\boldsymbol{\gamma}^{(k)})}{\partial \boldsymbol{\gamma}^{(k)}\partial \boldsymbol{\gamma}^{(k)\prime}} &= \left\{1-\frac{\epsilon_{kt}^2}{\sigma_{kt}^2}\right\}\left\{\frac{2}{\sigma_{kt}}\frac{\partial^2 \sigma_{kt}}{\partial \boldsymbol{\gamma}^{(k)}\partial \boldsymbol{\gamma}^{(k)\prime}}\right\}\\
&\quad +2\left\{3\frac{\epsilon_{kt}^2}{\sigma_{kt}^2}-1\right\}\left\{\frac{1}{\sigma_{kt}}\frac{\partial \sigma_{kt}}{\partial \boldsymbol{\gamma}^{(k)}}\right\}\left\{\frac{1}{\sigma_{kt}}\frac{\partial \sigma_{kt}}{\partial \boldsymbol{\gamma}^{(k)\prime}}\right\}+\frac{2}{\sigma_{kt}^2}\frac{\partial \epsilon_{kt}}{\partial \boldsymbol{\gamma}^{(k)}}\frac{\partial \epsilon_{kt}}{\partial \boldsymbol{\gamma}^{(k)\prime}}\\
&\quad -\frac{4\epsilon_{kt}}{\sigma_{kt}^2}\left(\frac{\partial \epsilon_{kt}}{\partial \boldsymbol{\gamma}^{(k)}}\frac{1}{\sigma_{kt}}\frac{\partial \sigma_{kt}}{\partial \boldsymbol{\gamma}^{(k)\prime}}+\frac{1}{\sigma_{kt}}\frac{\partial \sigma_{kt}}{\partial \boldsymbol{\gamma}^{(k)}}\frac{\partial \epsilon_{kt}}{\partial \boldsymbol{\gamma}^{(k)\prime}}\right)
\end{aligned}$$

and we prove the following intermediate results

$i)$ $E\left\|\frac{\partial \ell_{kt}(\boldsymbol{\gamma}_0^{(k)})}{\partial \boldsymbol{\gamma}^{(k)}}\frac{\partial \ell_{kt}(\boldsymbol{\gamma}_0^{(k)})}{\partial \boldsymbol{\gamma}^{(k)\prime}}\right\|<\infty, \quad E\left\|\frac{\partial^2 \ell_{kt}(\boldsymbol{\gamma}_0^{(k)})}{\partial \boldsymbol{\gamma}^{(k)}\partial \boldsymbol{\gamma}^{(k)\prime}}\right\|<\infty,$

$ii)$ There exists a neighbourhood $\mathcal{V}(\boldsymbol{\gamma}_0^{(k)})$ of $\boldsymbol{\gamma}_0^{(k)}$ such that

$$\sup_{\boldsymbol{\gamma}^{(k)}\in\mathcal{V}(\boldsymbol{\gamma}_0^{(k)})}\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^n \frac{\partial \ell_{kt}(\boldsymbol{\gamma}^{(k)})}{\partial \boldsymbol{\gamma}^{(k)}}-\frac{\partial \tilde{\ell}_{kt}(\boldsymbol{\gamma}^{(k)})}{\partial \boldsymbol{\gamma}^{(k)}}\right\|\to 0\,,$$

$iii)$ $\frac{1}{n}\sum_{t=1}^n \frac{\partial^2 \ell_{kt}(\boldsymbol{\gamma}_n^{(k)})}{\partial \boldsymbol{\gamma}^{(k)}\partial \boldsymbol{\gamma}^{(k)\prime}}\to \boldsymbol{J}_{kk}^*\,,$ a.s. for any $\boldsymbol{\gamma}_n^{(k)}$ between $\hat{\boldsymbol{\gamma}}_n^{(k)}$ and $\boldsymbol{\gamma}_0^{(k)}$,

$iv)$ $\boldsymbol{J}_{kk}^*$ is non singular,

$v)$ $\frac{1}{\sqrt{n}}\sum_{t=1}^n \frac{\partial \ell_{kt}(\boldsymbol{\gamma}_0^{(k)})}{\partial \boldsymbol{\gamma}^{(k)}}\xrightarrow{\mathcal{L}}\mathcal{N}\left(\mathbf{0},\boldsymbol{I}_{kk}^*\right),$

where

$$\boldsymbol{J}_{kk}^*=E\left(\boldsymbol{d}_{kt}^*\boldsymbol{d}_{kt}^{*\prime}+2\boldsymbol{s}_{kt}\boldsymbol{s}_{kt}^{\prime}\right),\quad \boldsymbol{d}_{kt}^*=\frac{1}{\sigma_{kt}^2}\frac{\partial \sigma_{kt}^2(\boldsymbol{\theta}_0^{(k)})}{\partial \boldsymbol{\gamma}^{(k)}},\quad \boldsymbol{s}_{kt}=\frac{1}{\sigma_{kt}}\boldsymbol{e}_1^{(k)},$$

$$\boldsymbol{I}_{kk}^*=E\left\{\{\eta_{kt}^{*4}-1\}\boldsymbol{d}_{kt}^*\boldsymbol{d}_{kt}^{*\prime}+4\boldsymbol{s}_{kt}\boldsymbol{s}_{kt}^{\prime}-2\eta_{kt}^{*3}\left(\boldsymbol{d}_{kt}^*\boldsymbol{s}_{kt}^{\prime}+\boldsymbol{s}_{kt}\boldsymbol{d}_{kt}^{*\prime}\right)\right\},$$

and $\boldsymbol{e}_1^{(k)}=(1,0,\dots,0)'\in\mathbb{R}^{d_k+1}$.

We only indicate here the differences with the proof of Theorem 3.1. For i) we use

$$\left\|\frac{\epsilon_{kt}}{\sigma_{kt}^2}\frac{\partial \epsilon_{kt}}{\partial \boldsymbol{\gamma}^{(k)}}(\boldsymbol{\gamma}_0^{(k)})\right\|_2=\left\|\frac{\eta_{kt}^*}{\sigma_{kt}}(-1,0,\dots,0)'\right\|_2<\infty.$$

Now, turning to ii), we have

$$\begin{aligned}
&\left\| \frac{\partial \ell_{kt}(\boldsymbol{\gamma}^{(k)})}{\partial \boldsymbol{\gamma}^{(k)}} - \frac{\partial \tilde{\ell}_{kt}(\boldsymbol{\gamma}^{(k)})}{\partial \boldsymbol{\gamma}^{(k)}} \right\| \\
= \; &\left\| \left\{ \frac{\epsilon_{kt}^2}{\tilde{\sigma}_{kt}^2} - \frac{\epsilon_{kt}^2}{\sigma_{kt}^2} \right\} \left\{ \frac{2}{\sigma_{kt}} \frac{\partial \sigma_{kt}}{\partial \boldsymbol{\gamma}^{(k)}} \right\} + 2 \left\{ 1 - \frac{\epsilon_{kt}^2}{\tilde{\sigma}_{kt}^2} \right\} \left\{ \frac{1}{\sigma_{kt}} - \frac{1}{\tilde{\sigma}_{kt}} \right\} \left\{ \frac{\partial \sigma_{kt}}{\partial \boldsymbol{\gamma}^{(k)}} \right\} \right. \\
&+ \left\{ 1 - \frac{\epsilon_{kt}^2}{\tilde{\sigma}_{kt}^2} \right\} \left\{ \frac{2}{\tilde{\sigma}_{kt}} \right\} \left\{ \frac{\partial \sigma_{kt}}{\partial \boldsymbol{\gamma}^{(k)}} - \frac{\partial \tilde{\sigma}_{kt}}{\partial \boldsymbol{\gamma}^{(k)}} \right\} \\
&\left. + 2\epsilon_{kt} \left\{ \frac{1}{\sigma_{kt}^2} - \frac{1}{\tilde{\sigma}_{kt}^2} \right\} \frac{\partial \epsilon_{kt}}{\partial \boldsymbol{\gamma}^{(k)}} \right\| (\boldsymbol{\gamma}^{(k)}) \leq C \rho^t u_t,
\end{aligned}$$

where

$$u_t = (1 + |\eta_{kt}^*| + \eta_{kt}^{*2}) \left(1 + \sup_{\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \left\| \frac{1}{\sigma_{kt}} \frac{\partial \sigma_{kt}}{\partial \boldsymbol{\theta}^{(k)}} (\boldsymbol{\theta}^{(k)}) \right\| \right) \left(1 + \sup_{\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \left| \frac{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}{\sigma_{kt}(\boldsymbol{\theta}^{(k)})} \right|^2 \right).$$

The proof of iii) follows by arguments already used, as well as the proof of v). To show the invertibility of J_{kk}^*, we note that for any $\boldsymbol{x} = (x_1, \tilde{x}')' \in \mathbb{R}^{d_k+1}$, with $x_1 \in \mathbb{R}$, we have

$$\boldsymbol{x}' J_{kk}^* \boldsymbol{x} = 2x_1^2 E\left(\frac{1}{\sigma_{kt}^2} \right) + \tilde{x}' E(\boldsymbol{d}_{kt} \boldsymbol{d}_{kt}') \tilde{x}.$$

Thus, $\boldsymbol{x}' J_{kk}^* \boldsymbol{x} = 0$ implies $x_1 = 0$ and, by **A12**, $\tilde{x} = \mathbf{0}$.

Thus, the intermediate results are established and it follows that

$$\sqrt{n} \left(\hat{\gamma}_n^{(k)} - \gamma_0^{(k)} \right) \quad \stackrel{\mathcal{L}}{\to} \quad \mathcal{N} \left\{ 0, (\boldsymbol{J}_{kk}^*)^{-1} \boldsymbol{I}_{kk}^* (\boldsymbol{J}_{kk}^*)^{-1} \right\},$$

Noting that

$$\boldsymbol{J}_{kk}^* = \begin{pmatrix} 2E\left(\frac{1}{\sigma_{kt}^2}\right) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{J}_{kk} \end{pmatrix},$$

straightforward computation shows that $(\boldsymbol{J}_{kk}^*)^{-1} \boldsymbol{I}_{kk}^* (\boldsymbol{J}_{kk}^*)^{-1} = \boldsymbol{\Upsilon}$ and the proof of Theorem 3.3 is complete.

B. Proof of Proposition 4.1

The proof of Proposition 4.1 relies on two lemmas. The first one shows that $\hat{\boldsymbol{\theta}}_n^{(k)}$ is a consistent estimator of $\boldsymbol{\theta}_0^{(k)}$.

LEMMA B.1. *Let the assumptions of Proposition 4.1 be satisfied. Then*

$$\hat{\boldsymbol{\theta}}_n^{(k)} \to \boldsymbol{\theta}_0^{(k)}, \quad a.s. \quad as\ n \to \infty.$$

Proof: It consists in verifying the conditions required in Theorem 3.1 for the convergence in (3.1).

The existence of a (unique) ergodic, non anticipative, strictly and second-order stationary solution $(\boldsymbol{\epsilon}_t)$ of Model (4.3), under the conditions given in the corollary, follows from Boussama , Fuchs and Stelzer (2011), Theorem 2.4. Thus **A1** holds with $s = 2$.

Recall that $\theta_5^{(k)} \in (0,1)$ for all $\boldsymbol{\theta}^{(k)} \in \boldsymbol{\Theta}^{(k)}$. Straightforward calculation shows that

$$\begin{aligned} & |\sigma_{kt}^2(\boldsymbol{\theta}^{(k)}) - \sigma_{kt}^2(\boldsymbol{\theta}_0^{(k)})| \\ \leq \quad & K\|\boldsymbol{\theta}^{(k)} - \boldsymbol{\theta}_0^{(k)}\| \sum_{i\geq 0} \left(\{\theta_{05}^{(k)}\}^i + \{\theta_5^{(k)}\}^i\right) (\epsilon_{1,t-i-1}^2 + |\epsilon_{1,t-1}\epsilon_{2,t-1}| + \epsilon_{2,t-i-1}^2). \end{aligned}$$

It follows, using the fact that $\boldsymbol{\epsilon}_t$ belongs to L^2, that **A2** is satisfied. We similarly show that **A3** holds true, and **A4** is satisfied by definition of $\boldsymbol{\Theta}^{(k)}$.

Now we turn to **A5**. Suppose $\sigma_t(\boldsymbol{\theta}_0^{(k)}) = \sigma_t(\boldsymbol{\theta}^{(k)})$, that is

$$\begin{aligned} & \theta_{01}^{(k)} + \theta_{02}^{(k)}\epsilon_{1,t-1}^2 + \theta_{03}^{(k)}\epsilon_{1,t-1}\epsilon_{2,t-1} + \theta_{04}^{(k)}\epsilon_{2,t-1}^2 + \theta_{05}^{(k)}\sigma_{t-1}^2 \\ = \quad & \theta_1^{(k)} + \theta_2^{(k)}\epsilon_{1,t-1}^2 + \theta_3^{(k)}\epsilon_{1,t-1}\epsilon_{2,t-1} + \theta_4^{(k)}\epsilon_{2,t-1}^2 + \theta_5^{(k)}\sigma_{t-1}^2. \end{aligned}$$

Then there exists some non zero variables $a_{t-2}, b_{t-2}, c_{t-2}, d_{t-2}$ belonging to the past of $\boldsymbol{\eta}_{t-1}$ such that

$$a_{t-2} + b_{t-2}\eta_{1,t-1}^2 + c_{t-2}\eta_{1,t-1}\eta_{2,t-1} + d_{t-2}\eta_{2,t-1}^2 = 0.$$

Therefore, the distribution of $\boldsymbol{\eta}_t$ conditional to the past is degenerate. Since $\boldsymbol{\eta}_t$ is independent from the past, this means that the unconditional distribution of $\boldsymbol{\eta}_t$ is degenerate, in contradiction with the existence of a density around zero. Thus $a_{t-2} = b_{t-2} = c_{t-2} = d_{t-2} = 0$, from which we deduce that $\boldsymbol{\theta}^{(k)} = \boldsymbol{\theta}_0^{(k)}$. Therefore, **A5** is verified. □

Now we turn to the asymptotic distribution. Assumption **A7** being in failure, we cannot use Theorem 3.2 to derive the asymptotic distribution of $\hat{\boldsymbol{\theta}}_n^{(k)}$. It will be more convenient to work with a reparameterization. Consider the transformation defined by $\boldsymbol{\Theta}^{(k)} \mapsto \boldsymbol{\Psi}^{(k)} = H(\boldsymbol{\Theta}^{(k)}) : \boldsymbol{x} = (x_1, x_2, x_3, x_4, x_5)' \mapsto H(\boldsymbol{x}) = (x_1, x_2, 4x_2x_4 - x_3^2, x_4, x_5)'$. Write $\boldsymbol{\psi} = H(\boldsymbol{\theta})$. The following lemma derives the asymptotic distribution of $\hat{\boldsymbol{\psi}}_n^{(k)} = H(\hat{\boldsymbol{\theta}}_n^{(k)})$. Let $\boldsymbol{\Lambda} = \mathbb{R}^2 \times (0,\infty) \times \mathbb{R}^2$.

Lemma B.2. *Let the assumptions of Proposition 4.1 be satisfied. Then*

$$\sqrt{n}(\hat{\boldsymbol{\psi}}_n^{(k)} - \boldsymbol{\psi}_0^{(k)}) \quad \overset{\mathcal{L}}{\to} \quad \boldsymbol{\lambda}^{\boldsymbol{\Lambda}} := \arg\inf_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda}} \{\boldsymbol{\lambda} - \boldsymbol{Z}\}'\dot{\boldsymbol{H}}_k^{-1}\boldsymbol{J}_{kk}(\dot{\boldsymbol{H}}_k^{-1})'\{\boldsymbol{\lambda} - \boldsymbol{Z}\}$$

where $\boldsymbol{Z} \sim \mathcal{N}\left\{0, \dot{\boldsymbol{H}}_k' \boldsymbol{J}_{kk}^{-1} \boldsymbol{I}_{kk} \boldsymbol{J}_{kk}^{-1} \dot{\boldsymbol{H}}_k\right\}$, *with* $\dot{\boldsymbol{H}}_k' = \frac{\partial H}{\partial \boldsymbol{\theta}'}(\boldsymbol{\theta}_0^{(k)})$.

Proof: Note that, because $H_0^{(k)}$ is satisfied for the BEKK-GARCH(1,1) model, the third component of $\boldsymbol{\psi}_0^{(k)}$ is equal to zero, the other ones being strictly positive. We follow the lines of proof of Theorem 2 in Francq and Zakoian (2007). First note that the matrix $\dot{\boldsymbol{H}}_k$ is non-singular. Note also that, $\boldsymbol{\Lambda}$ being a convex cone, $\boldsymbol{\lambda}^{\boldsymbol{\Lambda}}$ is uniquely determined.

Except **A7**, the assumptions of Theorem 3.2 are satisfied. For instance, the verification of **A12** is achieved by the same arguments as those used for **A5**. For brevity, we do not detail the verification of all the assumptions. It follows in particular that J_{kk} is non singular.

A Taylor expansion of $H(\hat{\boldsymbol{\theta}}_n^{(k)})$ around $\boldsymbol{\theta}_0^{(k)}$ yields,

$$\sqrt{n}\left\{\hat{\boldsymbol{\psi}}_n^{(k)} - \boldsymbol{\psi}_0^{(k)}\right\} \overset{o_P(1)}{=} \dot{\boldsymbol{H}}_k' \sqrt{n}(\hat{\boldsymbol{\theta}}_n^{(k)} - \boldsymbol{\theta}_0^{(k)}),$$

using the convergence established in Lemma B.1 and the continuity of $\partial H / \partial \boldsymbol{\theta}$ (the notation $a_n \overset{o_P(1)}{=} b_n$ stands for sequences (a_n) and (b_n) such that $a_n - b_n$ converges to zero in probability). Now let

$$\boldsymbol{Z}_n = -\dot{\boldsymbol{H}}_k' \boldsymbol{J}_{kk}^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (1 - \eta_{kt}^{*2}) \boldsymbol{d}_{kt}.$$

Note that we do not have equality (up to $o_P(1)$ terms) between $\boldsymbol{Z}_n$ and the left-hand side of (B.2) because, under $H_0^{(k)}$, the third component of this vector is a nonnegative random variable. This is not the case of $\boldsymbol{Z}_n$ which, by Theorem 3.2, converges in distribution to $\boldsymbol{Z}$.

We will establish that

$$\sqrt{n}\left\{\hat{\boldsymbol{\psi}}_n^{(k)} - \boldsymbol{\psi}_0^{(k)}\right\} \overset{o_P(1)}{=} \boldsymbol{\lambda}_n^{\boldsymbol{\Lambda}}$$

where $\boldsymbol{\lambda}_n^{\boldsymbol{\Lambda}} = \arg\inf_{\boldsymbol{\lambda} \in \boldsymbol{\Lambda}} \{\boldsymbol{\lambda} - \boldsymbol{Z}_n\}' \dot{\boldsymbol{H}}_k^{-1} \boldsymbol{J}_{kk} (\dot{\boldsymbol{H}}_k^{-1})' \{\boldsymbol{\lambda} - \boldsymbol{Z}_n\}$. Note that $\boldsymbol{\lambda}_n^{\boldsymbol{\Lambda}}$ can be interpreted as the orthogonal projection of $\boldsymbol{Z}_n$ on $\boldsymbol{\Lambda}$ for the inner product $< x, y >_{\dot{\boldsymbol{H}}_k^{-1} \boldsymbol{J}_{kk} (\dot{\boldsymbol{H}}_k^{-1})'} = x' \dot{\boldsymbol{H}}_k^{-1} \boldsymbol{J}_{kk} (\dot{\boldsymbol{H}}_k^{-1})' y$. We also introduce the orthogonal projection of $\boldsymbol{Z}_n$ on $\sqrt{n}(\boldsymbol{\Psi}^{(k)} - \boldsymbol{\psi}_0^{(k)})$, defined by

$$\tilde{\boldsymbol{\psi}}_n^{(k)} = \arg \inf_{\boldsymbol{\psi}^{(k)} \in \boldsymbol{\Psi}^{(k)}} \|\boldsymbol{Z}_n - \sqrt{n}(\boldsymbol{\psi}^{(k)} - \boldsymbol{\psi}_0^{(k)})\|_{\dot{\boldsymbol{H}}_k^{-1} \boldsymbol{J}_{kk} (\dot{\boldsymbol{H}}_k^{-1})'}.$$

Because $\sqrt{n}(\boldsymbol{\Psi}^{(k)} - \boldsymbol{\psi}_0^{(k)})$ increases to $\boldsymbol{\Lambda}$, it can be noted that the variables $\boldsymbol{\lambda}_n^{\boldsymbol{\Lambda}}$ and $\sqrt{n}\left\{\tilde{\boldsymbol{\psi}}_n^{(k)} - \boldsymbol{\psi}_0^{(k)}\right\}$ are equal for n sufficiently large.

A Taylor expansion of the quasi-likelihood function yields

$$\begin{aligned}
&\tilde{Q}_n^{(k)}(\boldsymbol{\theta}^{(k)}) - \tilde{Q}_n^{(k)}(\boldsymbol{\theta}_0^{(k)}) \\
= &\frac{\partial \tilde{Q}_n^{(k)}(\boldsymbol{\theta}_0^{(k)})}{\partial \boldsymbol{\theta}^{(k)\prime}}(\boldsymbol{\theta}^{(k)} - \boldsymbol{\theta}_0^{(k)}) + \frac{1}{2}(\boldsymbol{\theta}^{(k)} - \boldsymbol{\theta}_0^{(k)})' \left[\frac{\partial^2 \tilde{Q}_n^{(k)}(\boldsymbol{\theta}_0^{(k)})}{\partial \boldsymbol{\theta}^{(k)} \partial \boldsymbol{\theta}^{(k)\prime}}\right](\boldsymbol{\theta}^{(k)} - \boldsymbol{\theta}_0^{(k)}) + R_n(\boldsymbol{\theta}^{(k)}) \\
= &-\frac{1}{2n}\boldsymbol{Z}_n'\dot{\boldsymbol{H}}_k^{-1}\boldsymbol{J}_{kk}\sqrt{n}(\boldsymbol{\theta}^{(k)} - \boldsymbol{\theta}_0^{(k)}) - \frac{1}{2n}\sqrt{n}(\boldsymbol{\theta}^{(k)} - \boldsymbol{\theta}_0^{(k)})'\boldsymbol{J}_{kk}(\dot{\boldsymbol{H}}_k')^{-1}\boldsymbol{Z}_n \\
&+\frac{1}{2}(\boldsymbol{\theta}^{(k)} - \boldsymbol{\theta}_0^{(k)})'\boldsymbol{J}_{kk}(\boldsymbol{\theta}^{(k)} - \boldsymbol{\theta}_0^{(k)}) + R_n(\boldsymbol{\theta}^{(k)}) + R_n^*(\boldsymbol{\theta}^{(k)}) \\
= &\frac{1}{2n}\|(\dot{\boldsymbol{H}}_k')^{-1}\boldsymbol{Z}_n - \sqrt{n}(\boldsymbol{\theta}^{(k)} - \boldsymbol{\theta}_0^{(k)})\|_{\boldsymbol{J}_{kk}}^2 - \frac{1}{2n}\boldsymbol{Z}_n'\dot{\boldsymbol{H}}_k^{-1}\boldsymbol{J}_{kk}(\dot{\boldsymbol{H}}_k^{-1})'\boldsymbol{Z}_n \\
&+R_n(\boldsymbol{\theta}^{(k)}) + R_n^*(\boldsymbol{\theta}^{(k)}) \\
= &\frac{1}{2n}\|\boldsymbol{Z}_n - \sqrt{n}(\boldsymbol{\psi}^{(k)} - \boldsymbol{\psi}_0^{(k)})\|_{\dot{\boldsymbol{H}}_k^{-1}\boldsymbol{J}_{kk}(\dot{\boldsymbol{H}}_k^{-1})'}^2 - \frac{1}{2n}\boldsymbol{Z}_n'\dot{\boldsymbol{H}}_k^{-1}\boldsymbol{J}_{kk}(\dot{\boldsymbol{H}}_k^{-1})'\boldsymbol{Z}_n \\
&+R_n(\boldsymbol{\theta}^{(k)}) + R_n^*(\boldsymbol{\theta}^{(k)}).
\end{aligned}$$

Following the lines of proof of Theorem 2 in Francq and Zakoian (2007), it can be shown that

i) $\sqrt{n}(\tilde{\boldsymbol{\psi}}_n^{(k)} - \boldsymbol{\psi}_0^{(k)}) = O_P(1)$,

ii) $\sqrt{n}(\hat{\boldsymbol{\psi}}_n^{(k)} - \boldsymbol{\psi}_0^{(k)}) = O_P(1)$,

iii) for any sequence $(\boldsymbol{\theta}_n)$ such that $\sqrt{n}(\boldsymbol{\theta}_n^{(k)} - \boldsymbol{\theta}_0^{(k)}) = O_P(1)$,

$$R_n(\boldsymbol{\theta}_n^{(k)}) = o_P(n^{-1}), \quad R_n^*(\boldsymbol{\theta}_n^{(k)}) = o_P(n^{-1}),$$

iv) $\|\boldsymbol{Z}_n - \sqrt{n}\left\{\hat{\boldsymbol{\psi}}_n^{(k)} - \boldsymbol{\psi}_0^{(k)}\right\}\|_{\dot{\boldsymbol{H}}_k^{-1}\boldsymbol{J}_{kk}(\dot{\boldsymbol{H}}_k^{-1})'}^2 \overset{o_P(1)}{=} \|\boldsymbol{Z}_n - \boldsymbol{\lambda}_n^{\boldsymbol{\Lambda}}\|_{\dot{\boldsymbol{H}}_k^{-1}\boldsymbol{J}_{kk}(\dot{\boldsymbol{H}}_k^{-1})'}^2$,

v) $\sqrt{n}\left\{\hat{\boldsymbol{\psi}}_n^{(k)} - \boldsymbol{\psi}_0^{(k)}\right\} \overset{o_P(1)}{=} \boldsymbol{\lambda}_n^{\boldsymbol{\Lambda}}$,

vi) $\boldsymbol{\lambda}_n^{\boldsymbol{\Lambda}} \overset{\mathcal{L}}{\to} \boldsymbol{\lambda}^{\boldsymbol{\Lambda}}$.

We omit the proof of these steps, which relies on arguments already given. The proof of Lemma B.2 then follows from v) and vi). □

Now we complete the proof of Proposition 4.1. Note that, from Example 8.2 in Francq and Zakoïan (2010), the third component of $\boldsymbol{\lambda}$ is the positive part, Z_3^+ say, of the third component of $\boldsymbol{Z}$. It follows that, letting $\boldsymbol{e}_3 = (0,0,1,0,0)'$,

$$\boldsymbol{e}_3'\sqrt{n}(\hat{\boldsymbol{\psi}}_n^{(k)} - \boldsymbol{\psi}_0^{(k)}) = \boldsymbol{e}_3'\sqrt{n}\hat{\boldsymbol{\psi}}_n^{(k)} \overset{\mathcal{L}}{\to} \boldsymbol{e}_3'\boldsymbol{\lambda}^{\boldsymbol{\Lambda}} = Z_3^+, \qquad Z_3 \sim \mathcal{N}\left\{0, \boldsymbol{e}_3'\dot{\boldsymbol{H}}_k'\boldsymbol{J}_{kk}^{-1}\boldsymbol{I}_{kk}\boldsymbol{J}_{kk}^{-1}\dot{\boldsymbol{H}}_k\boldsymbol{e}_3\right\}.$$

Noting that $\boldsymbol{e}_3'\dot{\boldsymbol{H}}_k' = \left(0, 4\theta_{04}^{(k)}, -2\theta_{03}^{(k)}, 4\theta_{02}^{(k)}, 0\right)$, the conclusion straightforwardly follows from the consistency of $\boldsymbol{X}_n$, $\hat{\boldsymbol{J}}_{kk}$ and $\hat{\boldsymbol{I}}_{kk}$ to $\boldsymbol{e}_3'\dot{\boldsymbol{H}}_k$, $\boldsymbol{J}_{kk}$ and $\boldsymbol{I}_{kk}$ respectively. □

C. Proof of Theorem 5.1

The consistency of $\hat{\boldsymbol{\theta}}_n$ follows from Theorem 3.1. It suffices to prove the consistency of $\hat{\boldsymbol{\rho}}_n$.

Let vec denote the operator that stacks the columns of a matrix. Let $\boldsymbol{K}_m^0$ denote a $m(m-1)/2 \times m^2$ matrix such that for any symmetric $m \times m$ matrix $\boldsymbol{A}$, $\boldsymbol{K}_m^0 \mathrm{vec}(\boldsymbol{A}) = \mathrm{vech}^0(\boldsymbol{A})$. We have

$$\hat{\boldsymbol{\rho}}_n = \frac{1}{n}\sum_{t=1}^n \boldsymbol{K}_m^0 \left(\hat{\boldsymbol{\eta}}_t^* \otimes \hat{\boldsymbol{\eta}}_t^*\right).$$

Letting

$$\boldsymbol{\rho}_n = \frac{1}{n}\sum_{t=1}^n \boldsymbol{K}_m^0 \left(\boldsymbol{\eta}_t^* \otimes \boldsymbol{\eta}_t^*\right),$$

we have

$$\|\hat{\boldsymbol{\rho}}_n - \boldsymbol{\rho}_n\| \le \frac{C}{n}\sum_{t=1}^n \|\hat{\boldsymbol{\eta}}_t^* - \boldsymbol{\eta}_t^*\|(\|\boldsymbol{\eta}_t^*\| + \|\hat{\boldsymbol{\eta}}_t^* - \boldsymbol{\eta}_t^*\|).$$

Now, using **A2** and **A4**,

$$\begin{aligned}
\|\hat{\boldsymbol{\eta}}_t^* - \boldsymbol{\eta}_t^*\| &\le C\sum_{k=1}^m \frac{|\sigma_{kt}(\boldsymbol{\theta}_0^{(k)}) - \tilde{\sigma}_{kt}(\hat{\boldsymbol{\theta}}_n^{(k)})|}{\tilde{\sigma}_{kt}(\hat{\boldsymbol{\theta}}_n^{(k)})}|\eta_{kt}^*| \\
&\le C\sum_{k=1}^m \frac{|\sigma_{kt}(\boldsymbol{\theta}_0^{(k)}) - \sigma_{kt}(\hat{\boldsymbol{\theta}}_n^{(k)})| + |\sigma_{kt}(\hat{\boldsymbol{\theta}}_n^{(k)}) - \tilde{\sigma}_{kt}(\hat{\boldsymbol{\theta}}_n^{(k)})|}{\tilde{\sigma}_{kt}(\hat{\boldsymbol{\theta}}_n^{(k)})}|\eta_{kt}^*| \\
&\le C\sum_{k=1}^m \left(\frac{|\sigma_{kt}(\boldsymbol{\theta}_0^{(k)}) - \sigma_{kt}(\hat{\boldsymbol{\theta}}_n^{(k)})|}{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}\frac{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}{\sigma_{kt}(\hat{\boldsymbol{\theta}}_n^{(k)})}\frac{\sigma_{kt}(\hat{\boldsymbol{\theta}}_n^{(k)})}{\tilde{\sigma}_{kt}(\hat{\boldsymbol{\theta}}_n^{(k)})} + a_t\right)|\eta_{kt}^*| \\
&\le C\sum_{k=1}^m \left(\frac{K(\boldsymbol{\epsilon}_{t-1},\dots)\|\hat{\boldsymbol{\theta}}_n^{(k)} - \boldsymbol{\theta}_0^{(k)}\|}{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}\frac{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}{\sigma_{kt}(\hat{\boldsymbol{\theta}}_n^{(k)})}(1+a_t) + a_t\right)|\eta_{kt}^*|.
\end{aligned}$$

We thus have, by **A6**, for n large enough such that $\hat{\boldsymbol{\theta}}_n^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})$,

$$\begin{aligned}
\|\hat{\boldsymbol{\rho}}_n - \boldsymbol{\rho}_n\| &\le \|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|\frac{C}{n}\sum_{t=1}^n \|\boldsymbol{\eta}_t^*\|^2 \sum_{k=1}^m \frac{K(\boldsymbol{\epsilon}_{t-1},\dots)}{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})} \sup_{\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \frac{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}{\sigma_{kt}(\boldsymbol{\theta}^{(k)})} \\
&\quad + \frac{C}{n}\sum_{t=1}^n \rho^t \|\boldsymbol{\eta}_t^*\|^2 + \frac{C}{n}\sum_{t=1}^n \|\hat{\boldsymbol{\eta}}_t^* - \boldsymbol{\eta}_t^*\|^2 := S_{n1} + S_{n2} + S_{n3}.
\end{aligned}$$

We have, using again the independence between $\boldsymbol{\eta}_t^*$ and $\{\boldsymbol{\epsilon}_u, u < t\}$ under (2.9),

$$\begin{aligned}
& E\left(\|\boldsymbol{\eta}_t^*\|^2 \sum_{k=1}^m \frac{K(\boldsymbol{\epsilon}_{t-1},\ldots)}{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})} \sup_{\boldsymbol{\theta}^{(k)}\in\mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \frac{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}{\sigma_{kt}(\boldsymbol{\theta}^{(k)})} \right) \\
= \; & E\|\boldsymbol{\eta}_t^*\|^2 \sum_{k=1}^m E\left(\frac{K(\boldsymbol{\epsilon}_{t-1},\ldots)}{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})} \sup_{\boldsymbol{\theta}^{(k)}\in\mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \frac{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}{\sigma_{kt}(\boldsymbol{\theta}^{(k)})} \right) \\
\leq \; & E\|\boldsymbol{\eta}_t^*\|^2 \sum_{k=1}^m \left\| \frac{K(\boldsymbol{\epsilon}_{t-1},\ldots)}{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})} \right\|_2 \left\| \sup_{\boldsymbol{\theta}^{(k)}\in\mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \frac{\sigma_{kt}(\boldsymbol{\theta}_0^{(k)})}{\sigma_{kt}(\boldsymbol{\theta}^{(k)})} \right\|_2 < \infty,
\end{aligned}$$

using the Cauchy-Schwarz inequality. The last inequality is a consequence of Assumptions **A2-A3**. It follows that S_{n1} is the product of $\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|$ which converges to zero a.s., by Theorem 3.1, and a term which is bounded a.s. by the ergodic theorem. Thus $S_{n1} \to 0$ a.s. We similarly show that $S_{n2} \to 0$ and $S_{n3} \to 0$ a.s. Because $\boldsymbol{\eta}_t^* = \boldsymbol{R}^{1/2}\boldsymbol{\eta}_t$, the sequence $(\boldsymbol{\eta}_t^*)$ is iid . We thus have $\boldsymbol{\rho}_n \to \boldsymbol{\rho}_0$ by the strong law of large numbers. □

D. Numerical results

D.1. Estimating individual volatilities of a DCC by EbEE

In order to illustrate the ability of the EbEE to estimate the individual volatilities of a cDCC, we made the following Monte Carlo experiment. We first simulated an iid sequence $(\boldsymbol{\eta}_t)$, with $m = 4$ independent components, distributed as a Student distribution with $\nu = 7$ degrees of freedom, standardized in such a way that $\mathrm{Var}(\boldsymbol{\eta}_t) = \boldsymbol{I}_m$. We then simulated using (4.2) the sequence of correlations $\boldsymbol{R}_t$ and the sequence of innovations $\boldsymbol{\eta}_t^* = \boldsymbol{R}_t^{1/2}\boldsymbol{\eta}_t$, where $\boldsymbol{S}$ is the Toepliz correlation matrix with element 0.3^i on the i-th subdiagonal, $\alpha = 0.04$ and $\beta = 0.95$ (these values have been used for Figure 1 of Aielli, 2013). For these recursions, we took the initial values $\boldsymbol{Q}_0 = \boldsymbol{S}$ and $\boldsymbol{\eta}_0^* = \boldsymbol{0}$. We then generated the sequence $\boldsymbol{\epsilon}_t = \boldsymbol{D}_t\boldsymbol{\eta}_t^*$, where the elements of $\boldsymbol{D}_t$ are obtained from (2.7) with $p = q = 1$, $\boldsymbol{\omega} = (0.01, \cdots, 0.01)'$, $\boldsymbol{A} = \boldsymbol{A}_1$ the full $m \times m$ matrix with elements 0.02 and $\boldsymbol{B} = \boldsymbol{B}_1 = \mathrm{diag}(0.91, \ldots, 0.91)$. We discarded the first 500 simulated values to attenuate the effect of the initial values.

Table 4 displays the EbE estimates of the volatility coefficients $\boldsymbol{\omega}$, $\boldsymbol{A}$ and $\boldsymbol{B}$, as well as the variance of the EbE innovations $\boldsymbol{R}^* = \mathrm{Var}(\boldsymbol{\eta}_1^*)$, over 100 independent replications of length $n = 2000$ of the DCC model. It can be seen that the estimation bias is very small. We also checked that the Root Mean Square Errors (RMSE) decrease when the sample size

Table 4. Averaged EbEE over 100 replications of the cDCC model with $m = 4$ (standard deviations in small font).

$\boldsymbol{\omega}$	$\boldsymbol{A}$				diag($\boldsymbol{B}$)	$\boldsymbol{R}^*$			
0.013	0.017	0.021	0.020	0.021	0.905	1.000	0.276	0.087	0.023
0.008	0.013	0.011	0.011	0.012	0.024	–	0.096	0.103	0.096
0.014	0.020	0.018	0.022	0.021	0.902	0.278	1.000	0.268	0.086
0.010	0.011	0.012	0.013	0.013	0.025	0.096	–	0.087	0.111
0.013	0.022	0.022	0.016	0.021	0.906	0.087	0.268	1.000	0.299
0.010	0.012	0.011	0.011	0.014	0.027	0.103	0.087	–	0.098
0.018	0.021	0.022	0.021	0.018	0.899	0.023	0.086	0.299	1.000
0.039	0.011	0.011	0.011	0.011	0.064	0.096	0.111	0.098	–

Table 5. As Table4 but for Engle's DCC formulation ($i.e.$ without Aielli's correction).

$\boldsymbol{\omega}$	$\boldsymbol{A}$				diag($\boldsymbol{B}$)	$\boldsymbol{R}^*$			
0.013	0.019	0.022	0.019	0.021	0.904	1.000	0.272	0.065	0.035
0.008	0.014	0.012	0.011	0.011	0.021	–	0.086	0.100	0.097
0.015	0.020	0.018	0.021	0.021	0.902	0.272	1.000	0.281	0.081
0.010	0.012	0.013	0.011	0.011	0.027	0.086	–	0.093	0.101
0.012	0.021	0.020	0.016	0.019	0.910	0.065	0.281	1.000	0.262
0.007	0.012	0.012	0.011	0.011	0.020	0.100	0.093	–	0.096
0.013	0.022	0.023	0.021	0.018	0.907	0.035	0.081	0.262	1.000
0.010	0.012	0.012	0.012	0.011	0.027	0.097	0.101	0.096	–

n increases, and that they are not too sensitive to the nuisance parameters α, β and $\boldsymbol{S}$ involved in the sequence $(\boldsymbol{R}_t)$ of DCC matrices.

Table 5 concerns the same experiments but for Engle's DCC. The results are very similar.

D.2. Filtered probabilities for the SC model of Section 6.2.2

Figure 2 provides the filtered probabilities of the two regimes, for the SC model of exchange rates. It is seen that the regime with the highest residual correlations is often more plausible when the volatilities are large.

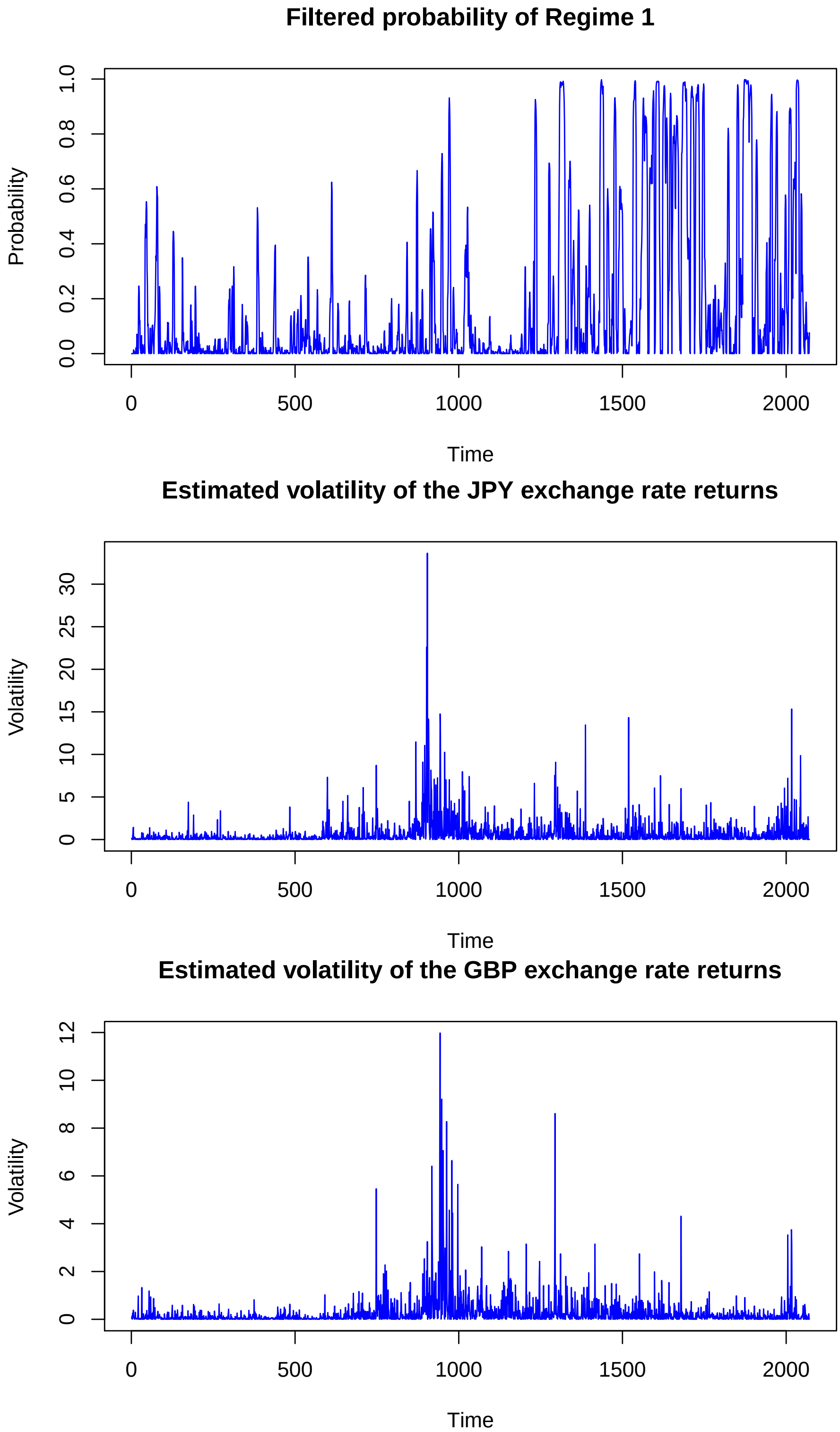

Figure 2. Filtered probability of Regime 1, and estimated volatilities of the GBP and JPY exchange rate returns

D.3. An application to world stock market indices

From the Yahoo Finance Website `http://finance.yahoo.com/`, we downloaded the whole set of the major World indices. We kept for these series the names given by Yahoo. We took the daily data available over the period from 1990-01-01 to 2013-04-22, and we eliminated a few series with too few observations. We then obtained a total number of 25 series: 5 for Americas, 11 for Asia-Pacific, 8 for Europe and 1 for Middle East. Because some series do not cover the entire period and the working days are not the same for all the financial markets, the number n of observations varies a lot, from $n = 2157$ for the series "NZ50" to $n = 6040$ for "AEX.AS". We corrected the "MERV" series for the stock spilt that occurred in Brazil on 1997-03-11, and we started at 1990-08-02 for the series "GD.AT" because of the presence of unexpected variations before this date. On each of the 25 series, we fitted PGARCH(1,1) models of the form

$$\begin{cases} \epsilon_t = \sigma_t \eta_t \\ \sigma_t^\delta = \omega + \alpha_+(\epsilon_{t-1}^+)^\delta + \alpha_-(-\epsilon_{t-1}^-)^\delta + \beta\sigma_{t-1}^\delta \end{cases} \tag{D.1}$$

where $x^+ = \max(x, 0)$, $x^- = \min(x, 0)$, $\alpha_+ \geq 0$, $\alpha_- \geq 0$, $\beta \in [0, 1)$, $\omega > 0$, and $\delta > 0$. As shown by Hamadeh and Zakoian (2011), the effective estimation of the parameter δ is an issue. The quasi-likelihood in the direction of δ being often relatively flat, the QML estimation of this parameter is imprecise and considerably slows down the optimization procedure. For this reason we decided to perform the QML optimization on only 4 values of this parameter: $\delta \in \{0.5, 1, 1.5, 2\}$. For each of the 4 values of δ, the remainder parameter $\boldsymbol{\theta} = (\omega, \alpha_+, \alpha_-, \beta)'$ is estimated by QML. Following the (quasi-)likelihood principle, the selected values of δ and the final estimated value of $\boldsymbol{\theta}$ maximize the QML over the 4 optimizations.

Table 6 displays the estimated PGARCH(1,1) models for each series, the estimated standard deviation into parentheses, and the selected value of δ in the last column. For all series, one can see a strong leverage effect ($\alpha_- > \alpha_+$) which means that negative returns tend to have an higher impact on the future volatility than positive returns of the same magnitude.

Table 7 gives an empirical estimate $\hat{\boldsymbol{R}}$ of the correlation matrix $\boldsymbol{R}$ of the residuals of the 25 PGARCH(1,1) equations. Because there are numerous missing values, due to the fact that the series are not always observed at the same dates, we used the `R` function `cor()` with

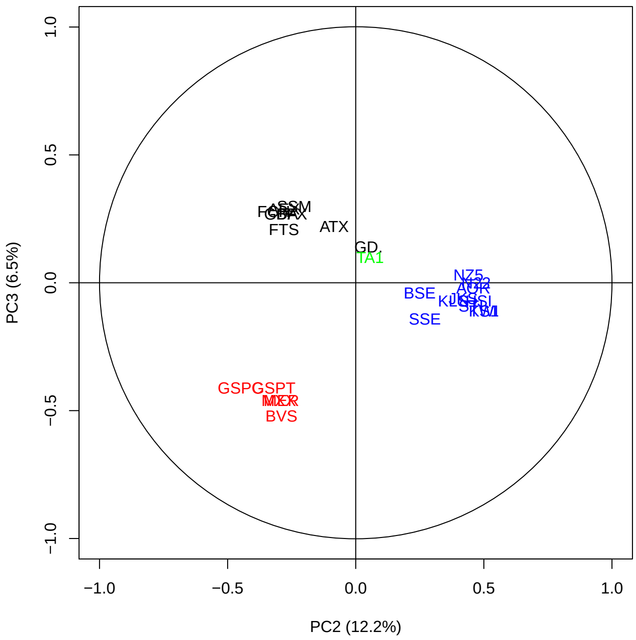

Figure 3. Factorial plan PC2-PC3.

the option "use=pairwise.complete.obs", which means that the correlation between each pair of variables is computed using all complete pairs of observations on those variables.

A principal component analysis (PCA) has been performed on the matrix $\hat{\boldsymbol{R}}$. The percentage of variance explained by the first four principal components are respectively 34.6%, 12.2%, 6.5% and 3.8%. Table 8 gives the so-called loading matrix, that is the correlation between the variables and the factors. From this table, it is clear that the first principal component PC1 is a scaling factor. PC1 is negatively correlated with all the series of returns. Noting that, in (D.1), the signs of ϵ_t and η_t are the same, the PC1 factor thus opposes the days where the markets are globally profitable to days where the markets go down. Therefore, we can interpret PC1 as the global trend of the World markets (with the negative sign for PC1 when the returns are globally positive). The second factor PC2 opposes the American and European to the Asian markets, whereas PC3 opposes the European and American markets (see Figure 3 for a graphical illustration). These relationships are certainly related to the opening hours of the different markets.

Additional References

Francq, C. and J-M. Zakoïan (2007) Quasi-maximum likelihood estimation in GARCH pro-

Table 6. PGARCH(1,1) models fitted by EbEE on daily returns of the major World stock indices. The estimated standard deviation are displayed into parentheses. The last column gives the selected value of the power δ.

	$\widehat{\omega}$	$\widehat{\alpha}_+$	$\widehat{\alpha}_-$	$\widehat{\beta}$	$\widehat{\delta}$
MERV	0.151 (0.002)	0.063 (0.002)	0.151 (0.001)	0.858 (0.004)	2
BVSP	0.077 (0.001)	0.068 (0.001)	0.138 (0.002)	0.884 (0.002)	2
GSPTSE	0.012 (0.009)	0.046 (0.002)	0.109 (0.004)	0.926 (0.007)	1
MXX	0.032 (0.003)	0.044 (0.001)	0.167 (0.002)	0.896 (0.004)	1.5
GSPC	0.016 (0.006)	0.000 (0.002)	0.134 (0.003)	0.927 (0.004)	1.5
AORD	0.023 (0.007)	0.030 (0.002)	0.131 (0.003)	0.910 (0.006)	1
SSEC	0.031 (0.010)	0.082 (0.004)	0.123 (0.003)	0.904 (0.012)	1
HSI	0.029 (0.008)	0.049 (0.003)	0.120 (0.003)	0.916 (0.009)	1
BSESN	0.055 (0.004)	0.062 (0.003)	0.179 (0.002)	0.872 (0.005)	1.5
JKSE	0.063 (0.005)	0.096 (0.002)	0.190 (0.001)	0.856 (0.005)	1.5
KLSE	0.087 (0.022)	0.071 (0.002)	0.157 (0.001)	0.835 (0.014)	2
N225	0.044 (0.004)	0.038 (0.003)	0.148 (0.002)	0.898 (0.006)	1
NZ50	0.018 (0.019)	0.044 (0.006)	0.120 (0.004)	0.898 (0.010)	1.5
STI	0.027 (0.011)	0.078 (0.001)	0.178 (0.001)	0.876 (0.005)	1.5
KS11	0.017 (0.009)	0.049 (0.001)	0.121 (0.004)	0.923 (0.008)	1.5
TWII	0.028 (0.012)	0.041 (0.004)	0.123 (0.003)	0.918 (0.010)	1
ATX	0.030 (0.005)	0.050 (0.002)	0.137 (0.003)	0.902 (0.007)	1
BFX	0.027 (0.005)	0.028 (0.002)	0.154 (0.003)	0.898 (0.005)	1.5
FCHI	0.026 (0.008)	0.014 (0.003)	0.112 (0.004)	0.931 (0.009)	1
GDAXI	0.028 (0.010)	0.022 (0.003)	0.114 (0.006)	0.926 (0.011)	1
AEX.AS	0.019 (0.005)	0.030 (0.002)	0.130 (0.002)	0.917 (0.005)	1.5
SSMI	0.038 (0.008)	0.024 (0.003)	0.145 (0.004)	0.897 (0.008)	1
FTSE	0.015 (0.010)	0.017 (0.003)	0.111 (0.003)	0.935 (0.008)	1
GD.AT	0.045 (0.001)	0.104 (0.002)	0.157 (0.001)	0.865 (0.004)	2
TA100	0.088 (0.007)	0.057 (0.002)	0.178 (0.001)	0.854 (0.007)	1.5

Table 7. Correlation matrix estimate $\hat{R}$

	MER	BVS	GST	MXX	GSC	AOR	SSE	HSI	BSE	JKS	KLS	N22	NZ5
MERV	1.00												
BVSP	0.53	1.00											
GSPT	0.47	0.48	1.00										
MXX	0.47	0.52	0.48	1.00									
GSPC	0.48	0.52	0.67	0.55	1.00								
AORD	0.17	0.17	0.21	0.17	0.12	1.00							
SSEC	0.06	0.08	0.08	0.06	0.02	0.18	1.00						
HSI	0.21	0.19	0.22	0.21	0.14	0.49	0.28	1.00					
BSES	0.17	0.19	0.21	0.20	0.15	0.31	0.14	0.40	1.00				
JKSE	0.15	0.15	0.14	0.15	0.08	0.36	0.15	0.43	0.31	1.00			
KLSE	0.10	0.10	0.11	0.12	0.06	0.28	0.14	0.36	0.19	0.32	1.00		
N225	0.11	0.13	0.19	0.12	0.12	0.46	0.16	0.44	0.27	0.34	0.28	1.00	
NZ50	0.09	0.06	0.10	0.09	0.04	0.48	0.16	0.31	0.21	0.29	0.22	0.38	1.00
STI	0.22	0.20	0.22	0.20	0.16	0.44	0.18	0.56	0.38	0.44	0.39	0.40	0.32
KS11	0.15	0.20	0.20	0.20	0.15	0.49	0.16	0.55	0.33	0.36	0.27	0.54	0.32
TWII	0.13	0.14	0.15	0.13	0.10	0.41	0.18	0.47	0.27	0.33	0.27	0.44	0.31
ATX	0.31	0.27	0.33	0.30	0.30	0.32	0.12	0.33	0.27	0.28	0.19	0.27	0.22
BFX	0.35	0.33	0.40	0.36	0.42	0.30	0.09	0.31	0.27	0.24	0.17	0.25	0.20
FCHI	0.37	0.36	0.44	0.39	0.47	0.26	0.06	0.31	0.28	0.21	0.15	0.26	0.17
GDAX	0.36	0.37	0.44	0.38	0.47	0.30	0.07	0.34	0.28	0.21	0.16	0.27	0.16
AEX	0.37	0.36	0.45	0.39	0.45	0.31	0.06	0.35	0.29	0.22	0.18	0.28	0.18
SSMI	0.33	0.31	0.39	0.35	0.41	0.29	0.05	0.31	0.27	0.23	0.16	0.27	0.19
FTSE	0.38	0.37	0.46	0.39	0.47	0.28	0.06	0.32	0.29	0.22	0.17	0.27	0.18
GD	0.19	0.18	0.20	0.19	0.16	0.21	0.07	0.24	0.26	0.20	0.14	0.19	0.17
TA10	0.24	0.24	0.27	0.26	0.23	0.33	0.06	0.36	0.28	0.24	0.18	0.29	0.18

	STI	KS1	TWI	ATX	BFX	FCH	GDA	AEX	SSM	FTS	GD	TA1
STI	1.00											
KS11	0.50	1.00										
TWII	0.45	0.51	1.00									
ATX	0.32	0.28	0.23	1.00								
BFX	0.30	0.25	0.19	0.56	1.00							
FCHI	0.30	0.26	0.20	0.55	0.71	1.00						
GDAX	0.31	0.27	0.20	0.59	0.70	0.79	1.00					
AEX	0.33	0.28	0.22	0.58	0.74	0.82	0.79	1.00				
SSMI	0.30	0.26	0.21	0.52	0.66	0.72	0.72	0.74	1.00			
FTSE	0.31	0.27	0.19	0.54	0.66	0.77	0.70	0.76	0.69	1.00		
GD	0.25	0.27	0.21	0.32	0.34	0.34	0.33	0.33	0.32	0.30	1.00	
TA10	0.36	0.28	0.25	0.38	0.39	0.42	0.40	0.41	0.40	0.40	0.33	1.00

Table 8. Correlations between the variables and the first 3 factors of the PCA

	PC1	PC2	PC3		PC1	PC2	PC3
MER	-0.52	-0.29	-0.46	STI	-0.58	0.45	-0.09
BVS	-0.52	-0.29	-0.52	KS1	-0.55	0.50	-0.11
GSPT	-0.59	-0.32	-0.41	TWI	-0.46	0.50	-0.11
MXX	-0.54	-0.30	-0.46	ATX	-0.68	-0.08	0.22
GSPC	-0.56	-0.45	-0.41	BFX	-0.75	-0.25	0.27
AOR	-0.55	0.46	-0.02	FCH	-0.79	-0.32	0.28
SSE	-0.19	0.27	-0.14	GDA	-0.79	-0.29	0.27
HSI	-0.60	0.48	-0.07	AEX	-0.81	-0.28	0.29
BSE	-0.48	0.25	-0.04	SSM	-0.75	-0.24	0.30
JKS	-0.45	0.42	-0.06	FTS	-0.78	-0.28	0.21
KLS	-0.35	0.38	-0.07	GD.	-0.46	0.05	0.14
N22	-0.50	0.47	-0.00	TA1	-0.57	0.06	0.10
NZ5	-0.37	0.44	0.03				

cesses when some coefficients are equal to zero. *Stochastic Processes and Their Applications* 117, 1265–1284

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.