# Estimation risk for the VaR of portfolios driven by semi-parametric multivariate models

Christian Francq[*] and Jean-Michel Zakoïan[†]

December 29, 2017

**Abstract**

Joint estimation of market and estimation risks in portfolios is investigated, when the individual returns follow a semi-parametric multivariate dynamic model and the asset composition is time-varying. Under ellipticity of the conditional distribution, asymptotic theory for the estimation of the conditional Value-at-Risk (VaR) is developed. An alternative method - the Filtered Historical Simulation - which does not rely on ellipticity, is also studied. Asymptotic confidence intervals for the conditional VaR, which allow to simultaneously quantify the market and estimation risks, are derived. The particular case of minimum variance portfolios is analyzed in more detail. Potential usefulness, feasibility and drawbacks of the two approaches are illustrated via Monte-Carlo experiments and an empirical study based on stock returns.

*JEL Classification:* C13, C31 and C58.

*Keywords:* Confidence Intervals for VaR, Dynamic Portfolio, Elliptical Distribution, Filtered Historical Simulation, Minimum Variance Portfolio, Model Risk, Multivariate GARCH.

[*]CREST and University of Lille, BP 60149, 59653 Villeneuve d'Ascq cedex, France. E-Mail: christian.francq@univ-lille3.fr

[†]Corresponding author: Jean-Michel Zakoïan, University of Lille and CREST, 15 boulevard Gabriel Péri, 92245 Malakoff Cedex, France. E-mail: zakoian@ensae.fr, Phone number: 33.1.41.17.77.25.

# 1 Introduction

A large strand of the recent literature on quantitative risk management has been concerned with risk aggregation (see for instance Embrechts and Puccetti (2010) and the references therein). For a vector of one-period profit-and-loss random variables $\boldsymbol{y} = (y_1, \ldots, y_m)'$, risk aggregation concerns the risk implied by an aggregate financial position defined as a real-valued function of $\boldsymbol{y}$. For instance, under the terms of Basel II, banks often measure the risk of a vector $\boldsymbol{y}$ of financial positions by the Value-at-Risk (VaR) of $\boldsymbol{a}'\boldsymbol{y} = a_1y_1 + \cdots + a_my_m$ where the $a_i$'s define the composition of a portfolio. Exact calculation of the risk associated with an aggregate position can represent a difficult task, as it requires knowledge of the joint distribution of the components of $\boldsymbol{y}$.

It is even more difficult, in a dynamic framework, to evaluate the *conditional risk* of a portfolio of assets or returns. The current regulatory framework for banking supervision (Basel II and Basel III), allows large international banks to develop internal models for the calculation of risk capital. The so-called advanced approaches are based on conditional distributions, that is, conditional on the past, rather than marginal ones. The superiority of dynamic approaches over static methods based on marginal distributions has been demonstrated empirically, for instance in Kuester, Mittnik and Paolella (2006). The dynamics is not only present in the returns, $\boldsymbol{y}_t$ instead of $\boldsymbol{y}$, but also in the weights of the portfolio, $\boldsymbol{a}_{t-1}$ instead of $\boldsymbol{a}$. Such weights can be both time-varying and stochastic: the notation $\boldsymbol{a}_{t-1}$ highlights the fact that investors may rebalance their portfolios at time $t$ using, in particular, the information contained in the historical prices.

To evaluate the conditional VaR of a portfolio, whose returns are defined by $r_t = \boldsymbol{a}'_{t-1}\boldsymbol{y}_t$, this paper focuses on multivariate semi-parametric approaches. Multivariate approaches are based on a time series model for the vector $\boldsymbol{y}_t$, instead of a univariate model for $r_t$. As emphasized by Rombouts and Verbeek (2009), the advantage of multivariate approaches is to "take into account the dynamic interrelationships between the portfolio components, while the model underlying the VaR calculations is independent of the portfolio composition". Indeed, the multivariate approach is particularly relevant if the VaR has to be computed for a large number of portfolio compositions $\boldsymbol{a}_{t-1}$. Moreover, semi-parametric methods allow for more flexibility than fully parametric methods relying on a complete specification of the conditional distribution of $\boldsymbol{y}_t$.

To our knowledge, the asymptotic properties of VaR estimators in a dynamic multivariate semi-parametric framework are unknown. It seems however important to evaluate the accuracy of risk estimators. Estimation risk refers to the uncertainty implied by statistical procedures in the implementation of risk measures. Uncertainty affects the estimation of risk measures, as well as the backtesting procedures used to assess the validity of risk measures. The new regulatory frameworks require that financial institutions take estimation risk into account (see e.g. Farkas, Fringuellotti and Tunaru (2016) and the references therein). The econometric literature devoted to the estimation risk in dynamic models is scant. [1] Christoffersen and Gonçalves (2005), and Spierdijk (2014) used resampling techniques to account for parameter estimation uncertainty in univariate dynamic models. Escanciano and Olmo (2010, 2011) proposed corrections of the standard backtesting procedures in presence of estimation risk (and also of model risk). Gouriéroux and Zakoïan (2013) showed that estimation induces an asymptotic bias in the coverage probabilities and proposed a corrected VaR. Francq and Zakoïan (2015) introduced the notion of risk parameter and derived asymptotic confidence intervals for the conditional VaR of univariate returns.

The first aim of this paper is to study the asymptotic properties of two multivariate semi-parametric approaches for estimating the conditional VaR of a portfolio of risk factors (returns). One approach for estimating conditional VaR's requires sphericity of the innovations distribution. An alternative approach, known as the Filtered Historical Simulation (FHS) method in the literature (see Barone-Adesi, Giannopoulos and Vosper (1999), Mancini and Trojani (2011) and the references therein), is assumption-free on the innovations distribution. The second aim is to provide methods based on the asymptotic theory or resampling schemes for constructing confidence intervals for the conditional VaR of portfolios. Such confidence intervals are in particular useful to visualize simultaneously the estimation and financial risks. As far as we know, our paper is the first one to study the asymptotic accuracy of conditional VaR estimators in a semi-parametric multivariate framework.

The rest of this paper is organized as follows. Section 2 presents the general framework. Section 3 is devoted to the asymptotic properties of the estimators of the conditional VaR under the sphericity assumption. This assumption also allows us to extend the concept of risk parameter to multivariate semi-parametric models. Section 4 gives the asymptotic properties of the FHS method, which relaxes the sphericity assumption. A numerical illustration and an empirical study based on stock returns are proposed in Section 5. Section 6 concludes. Complementary results and proofs are collected in the Appendix.

[1] For i.i.d. data, the literature is more voluminous, see Farkas, Fringuellotti and Tunaru (2016) for a recent reference.

# 2 Model and conditional VaR

Let $\boldsymbol{p}_t = (p_{1t}, \ldots, p_{mt})'$ denote the vector of prices of $m$ assets at time $t$. Let $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{mt})'$ be the corresponding vector of log-returns, with $y_{it} = \log(p_{it}/p_{i,t-1})$ for $i = 1, \ldots, m$.

Consider a portfolio of the $m$ assets, whose return is given by

$$r_t = \sum_{i=1}^{m} a_{i,t-1} y_{it} = \mathbf{a}'_{t-1} \boldsymbol{y}_t, \tag{2.1}$$

where $\mathbf{a}_{t-1} = (a_{1,t-1}, \ldots, a_{m,t-1})'$ is the vector of portfolio weights for the $m$ assets. Such weights are assumed to be stochastic and measurable with respect to some information set $\mathcal{I}_{t-1}$ containing the past prices (and possibly other variables). A portfolio is called *crystallized* when the number $\mu_i$ of units of each asset $i$ is time independent. For such a portfolio we have $a_{i,t-1} = \frac{\mu_i p_{i,t-1}}{\sum_{j=1}^{m} \mu_j p_{j,t-1}}$.

The *conditional* VaR of the portfolio's return process $(r_t)$ at risk level $\alpha \in (0,1)$, denoted by $\mathrm{VaR}_{t-1}^{(\alpha)}(r_t)$, is defined by

$$P_{t-1}\left[r_t < -\mathrm{VaR}_{t-1}^{(\alpha)}(r_t)\right] = \alpha, \tag{2.2}$$

where $P_{t-1}$ denotes the historical distribution conditional on $\mathcal{I}_{t-1}$.[2] More generally, we denote by $\mathrm{VaR}_{t-1}^{(\alpha)}(z_t)$ the conditional VaR of $z_t$ given $\mathcal{I}_{t-1}$, and by $\mathrm{VaR}^{(\alpha)}(z)$ the marginal VaR of a stationary process $(z_t)$.

Consider a general multivariate model for the vector of log-returns

$$\boldsymbol{y}_t = \boldsymbol{m}_t(\boldsymbol{\theta}_0) + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t = \boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)\boldsymbol{\eta}_t, \tag{2.3}$$

where $(\boldsymbol{\eta}_t)$ is a sequence of independent and identically distributed (iid) $\mathbb{R}^m$-valued variables with zero mean and identity covariance matrix; the $m \times m$ non-singular matrix $\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)$ and the $m \times 1$ vector $\boldsymbol{m}_t(\boldsymbol{\theta}_0)$ are specified as functions depending on the infinite past of $\boldsymbol{y}_t$ and parameterized by a $d$-dimensional parameter $\boldsymbol{\theta}_0$:

$$\boldsymbol{m}_t(\boldsymbol{\theta}_0) = \boldsymbol{m}(\boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \ldots, \boldsymbol{\theta}_0), \qquad \boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0) = \boldsymbol{\Sigma}(\boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \ldots, \boldsymbol{\theta}_0). \tag{2.4}$$

For the sake of generality, we do not consider a particular specification for the conditional mean $\boldsymbol{m}_t$ and the conditional variance $\boldsymbol{H}_t(\boldsymbol{\theta}_0) := \boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}'_t(\boldsymbol{\theta}_0)$,[3] but we assume

[2] In this formula, we assumed for simplicity that the conditional cdf of $r_t$ is continuous and strictly increasing.

[3] The most widely used specifications of Multivariate GARCH (MGARCH) models are discussed in Bauwens, Laurent and Rombouts (2006), Silvennoinen and Teräsvirta (2009), Francq and Zakoïan (2010, Chapter 11), Bauwens, Hafner and Laurent (2012), Tsay (2014, Chapter 7). Model (2.3)-(2.4) also includes multivariate extensions of the double-autoregressive models studied by Ling (2004).

**A1:** $(\boldsymbol{y}_t)$ is a strictly stationary solution of Model (2.3)-(2.4), and $\boldsymbol{\eta}_t$ is independent from $\mathcal{I}_{t-1}$.

This assumption will be made explicit for particular classes of MGARCH models satisfying Model (2.3)-(2.4).

Under (2.3)-(2.4), the portfolio's return defined in (2.1) satisfies

$$r_t = \mathbf{a}'_{t-1}\boldsymbol{m}_t(\boldsymbol{\theta}_0) + \mathbf{a}'_{t-1}\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)\boldsymbol{\eta}_t, \tag{2.5}$$

from which it follows that the portfolio's conditional VaR at level $\alpha$ is given by[4]

$$\mathrm{VaR}^{(\alpha)}_{t-1}(r_t) = -\mathbf{a}'_{t-1}\boldsymbol{m}_t(\boldsymbol{\theta}_0) + \mathrm{VaR}^{(\alpha)}_{t-1}\left(\mathbf{a}'_{t-1}\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)\boldsymbol{\eta}_t\right). \tag{2.6}$$

The VaR formula can be simplified if the errors $\boldsymbol{\eta}_t$ have a spherical distribution, that is, $\boldsymbol{P}\boldsymbol{\eta}_t$ and $\boldsymbol{\eta}_t$ have the same distribution for any orthogonal matrix $\boldsymbol{P}$. Ellipticity of the conditional distribution of $\boldsymbol{y}_t$ is equivalent to

**A2:** for any non-random vector $\boldsymbol{\lambda} \in \mathbb{R}^m$, $\boldsymbol{\lambda}'\boldsymbol{\eta}_t \overset{d}{=} \|\boldsymbol{\lambda}\|\eta_{1t}$,

where $\|\cdot\|$ denotes the euclidian norm on $\mathbb{R}^m$, $\eta_{it}$ denotes the $i$-th component of $\boldsymbol{\eta}_t$, and $\overset{d}{=}$ stands for the equality in distribution. [5]

**Remark 2.1 (Restrictiveness of the sphericity assumption)** Assumption **A2** entails that though not independent (except in the Gaussian case), the components of $\boldsymbol{\eta}_t$ have the same symmetric distribution. This assumption is commonly used in finance and econometrics (see for instance Sentana (2003), Fiorentini and Sentana (2016)). The importance of the class of spherical - and more generally elliptical - distributions to risk management is discussed in Bradley and Taqqu (2002). Examples of spherical distribution are the Gaussian $\mathcal{N}(\mathbf{0}, \boldsymbol{I}_m)$ distribution, and the standard multivariate Student distribution (see McNeil, Frey and Embrechts (2005) for details on spherical distributions). In fact, most parametric approaches for VaR estimation assume a spherical Gaussian or Student error distribution, which is very restrictive in terms of kurtosis (for the Gaussian distribution) and more generally on the tails of the distribution. By contrast, Assumption **A2** does not constrain (apart from symmetry) the size of the tails. It should also be noted that, while **A2** entails

[4] The presence of the sign "$-$" in this formula comes from the fact that the VaR is defined in terms of returns instead of loss variables.

[5] Note that, in **A2**, the Euclidian norm cannot be replaced by any other norm $N(\cdot)$ under the assumption of unit covariance matrix for $\boldsymbol{\eta}_t$. Indeed, if $\boldsymbol{\lambda}'\boldsymbol{\eta}_t \overset{d}{=} N(\boldsymbol{\lambda})\eta_{1t}$, we have $\mathrm{Var}(\boldsymbol{\lambda}'\boldsymbol{\eta}_t) = \boldsymbol{\lambda}'\boldsymbol{\lambda} = N(\boldsymbol{\lambda})^2\mathrm{Var}(\eta_{1t}) = N(\boldsymbol{\lambda})^2$.

that the components of $\boldsymbol{\eta}_t$ have the same symmetric distributions, this does not hold in general neither for the marginal nor for the conditional distribution of $\boldsymbol{y}_t$. In particular, this assumption is compatible with the usual leverage effect observed on financial returns. In other words, the elliptical model may incorporate asymmetric conditional second moments.

Under the sphericity assumption **A2** we have

$$\mathrm{VaR}_{t-1}^{(\alpha)}(r_t) = -\mathbf{a}_{t-1}'\boldsymbol{m}_t(\boldsymbol{\theta}_0) + \left\|\mathbf{a}_{t-1}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)\right\| \mathrm{VaR}^{(\alpha)}(\eta), \tag{2.7}$$

where $\mathrm{VaR}^{(\alpha)}(\eta)$ is the (marginal) VaR of $\eta_{1t}$.

**Remark 2.2 (Usefulness of sphericity for the VaR)** By contrast with formula (2.6), the interest of (2.7) is to relate the VaR of any portfolio to the first two conditional moments of the portfolio's return $r_t$, and to a simple characteristic of the innovations distribution. Under the sphericity assumption, the VaR is indeed a function of three ingredients: the mean-volatility parameter, the quantile of the errors and the portfolio's composition. In other words,

$$\mathrm{VaR}_{t-1}^{(\alpha)}(r_t) = F(\boldsymbol{\theta}_0, \mathrm{VaR}^{(\alpha)}(\eta); \mathbf{a}_{t-1}'), \tag{2.8}$$

where the first two components have to be estimated, while the third one is chosen by the risk manager. Such a decomposition does not hold in (2.6), which requires estimating a conditional quantile for any choice of the portfolio's composition (see Section 4). As we will see in Section 3.2, for most time series models (2.8) can even be reduced to a formula of the form

$$\mathrm{VaR}_{t-1}^{(\alpha)}(r_t) = F^*(\boldsymbol{\theta}_0^{(\alpha)}; \mathbf{a}_{t-1}'), \tag{2.9}$$

with a new parameter $\boldsymbol{\theta}_0^{(\alpha)}$ of the same dimension as $\boldsymbol{\theta}_0$, henceforth called conditional VaR parameter. These simplifications, (2.8) and (2.9), of the general VaR formula (2.6) have obvious interest for risk management, in particular when several portfolios based on the same risk factors have to be managed simultaneously.

# 3 VaR estimation under conditional ellipticity

Formula (2.7) is well-known in the literature dealing with theoretical properties of VaR (see for instance McNeil et al., 2005), but its econometric implications have been surprisingly overlooked. We now consider the statistical implementation of this formula.

Under the sphericity assumption **A2**, a natural strategy for estimating the conditional VaR of a portfolio is to estimate $\boldsymbol{\theta}_0$ by some consistent estimator $\widehat{\boldsymbol{\theta}}_n$ in a first step, to extract the residuals and to estimate $\mathrm{VaR}^{(\alpha)}(\eta)$ in a second step. For the first step, we will consider a general estimator satisfying regularity conditions. For the second step, the sphericity assumption will allow us to interpret $\mathrm{VaR}^{(\alpha)}(\eta)$ as the $(1-2\alpha)$-quantile $\xi_{1-2\alpha}$ of the absolute residuals, and to estimate this quantile by an empirical quantile using all the components of the first-step residuals.

Let $\Theta$ denote the parameter space, and assume $\boldsymbol{\theta}_0 \in \Theta$. Let $\widehat{\boldsymbol{\theta}}_n$ denote an estimator of parameter $\boldsymbol{\theta}_0$, obtained from observations $\boldsymbol{y}_1, \dots, \boldsymbol{y}_n$ and initial values $\widetilde{\boldsymbol{y}}_0, \widetilde{\boldsymbol{y}}_{-1}, \dots$. The vector of residuals is defined by $\widehat{\boldsymbol{\eta}}_t = \widetilde{\boldsymbol{\Sigma}}_t^{-1}(\widehat{\boldsymbol{\theta}}_n)\{\boldsymbol{y}_t - \widetilde{\boldsymbol{m}}_t(\widehat{\boldsymbol{\theta}}_n)\}) = (\widehat{\eta}_{1t}, \dots, \widehat{\eta}_{mt})'$, where $\widetilde{\boldsymbol{m}}_t(\boldsymbol{\theta}) = \boldsymbol{m}(\boldsymbol{y}_{t-1}, \dots, \boldsymbol{y}_1, \widetilde{\boldsymbol{y}}_0, \widetilde{\boldsymbol{y}}_{-1}, \dots, \boldsymbol{\theta})$, $\widetilde{\boldsymbol{\Sigma}}_t(\boldsymbol{\theta}) = \boldsymbol{\Sigma}(\boldsymbol{y}_{t-1}, \dots, \boldsymbol{y}_1, \widetilde{\boldsymbol{y}}_0, \widetilde{\boldsymbol{y}}_{-1}, \dots, \boldsymbol{\theta})$, for $t \geq 1$ and $\boldsymbol{\theta} \in \Theta$. For $\alpha \in (0,1)$, let $q_\alpha(S)$ denote the $\alpha$-quantile of a finite set $S \subset \mathbb{R}$. In view of (2.7), under the conditional ellipticity/sphericity assumption, an estimator of the conditional VaR at level $\alpha$ is

$$\widehat{\mathrm{VaR}}_{S,t-1}^{(\alpha)}(r_t) = -\mathbf{a}_{t-1}' \widetilde{\boldsymbol{m}}_t(\widehat{\boldsymbol{\theta}}_n) + \|\mathbf{a}_{t-1}' \widetilde{\boldsymbol{\Sigma}}_t(\widehat{\boldsymbol{\theta}}_n)\| \xi_{n,1-2\alpha}, \tag{3.1}$$

where $\xi_{n,1-2\alpha} = q_{1-2\alpha}(\{|\widehat{\eta}_{it}|, 1 \leq i \leq m, 1 \leq t \leq n\})$. The latter estimator takes advantage of the fact that the components of $\boldsymbol{\eta}_t$ are identically distributed under **A2**.

### 3.1 Asymptotic joint distribution of $\widehat{\boldsymbol{\theta}}_n$ and a quantile of absolute returns

We start by introducing the assumptions that are employed to establish the asymptotic distribution of $(\widehat{\boldsymbol{\theta}}_n', \xi_{n,1-2\alpha})$.

We now assume that the estimator $\widehat{\boldsymbol{\theta}}_n$ admits a Bahadur representation. Write $a \overset{c}{=} b$ for $a = b + c$.

**A3:** We have $\widehat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0$, a.s. Moreover, the following expansion holds

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \overset{o_P(1)}{=} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \boldsymbol{\Delta}_{t-1} \boldsymbol{V}(\boldsymbol{\eta}_t), \tag{3.2}$$

where $\boldsymbol{V}(\cdot)$ is a measurable function, $\boldsymbol{V} : \mathbb{R}^m \mapsto \mathbb{R}^K$ for some positive integer $K$, and $\boldsymbol{\Delta}_{t-1}$ is a $d \times K$ matrix, measurable with respect to the sigma-field generated by $\{\boldsymbol{\eta}_u,\ u < t\}$. The variables $\boldsymbol{\Delta}_t$ and $\boldsymbol{V}(\boldsymbol{\eta}_t)$ belong to $L^2$ with $E\boldsymbol{V}(\boldsymbol{\eta}_t) = 0$, $\mathrm{var}\{\boldsymbol{V}(\boldsymbol{\eta}_t)\} = \boldsymbol{\Upsilon}$ is nonsingular and $E\boldsymbol{\Delta}_t = \boldsymbol{\Lambda}$ is full row rank.

Assumption **A3** holds for a variety of MGARCH models and estimators[6] (see Appendix A for

[6] In the univariate setting, the asymptotic theory of estimation for GARCH parameters has been extensively

examples). The next assumption imposes smoothness of the functions $\boldsymbol{m}$ and $\boldsymbol{\Sigma}$ with respect to the parameter.

**A4:** For any sequence $(x_i)$, the functions $\boldsymbol{\theta} \mapsto \boldsymbol{m}(\boldsymbol{x}_1, \boldsymbol{x}_2, \dots; \boldsymbol{\theta})$ and $\boldsymbol{\theta} \mapsto \boldsymbol{\Sigma}(\boldsymbol{x}_1, \boldsymbol{x}_2, \dots; \boldsymbol{\theta})$ are continuously differentiable over $\Theta$.

The next theorem establishes the asymptotic normality of $(\widehat{\boldsymbol{\theta}}'_n, \xi_{n,1-2\alpha})$. Let

$$\boldsymbol{\Psi} = E(\boldsymbol{\Delta}_t \boldsymbol{\Upsilon} \boldsymbol{\Delta}'_t), \quad \boldsymbol{\Omega}' = E\left[\left\{\text{vec}\left(\boldsymbol{\Sigma}_t^{-1}\right)\right\}' \left\{\frac{\partial}{\partial \boldsymbol{\theta}'} \text{vec}\left(\boldsymbol{\Sigma}_t\right)\right\}\right], \quad \boldsymbol{W}_\alpha = \text{Cov}(\boldsymbol{V}(\boldsymbol{\eta}_t), N_t),$$

$\gamma_\alpha = \text{var}(N_t)$, with $N_t = \sum_{j=1}^m \left(\mathbf{1}_{\{|\eta_{jt}| < \xi_{1-2\alpha}\}} - 1 + 2\alpha\right)$, and, denoting by $f$ the density of $|\eta_{1t}|$, $\boldsymbol{\Xi}_{\boldsymbol{\theta}\xi} = \frac{-1}{m}\left\{\xi_{1-2\alpha}\boldsymbol{\Psi}\boldsymbol{\Omega} + \frac{1}{f(\xi_{1-2\alpha})}\boldsymbol{\Lambda}\boldsymbol{W}_\alpha\right\}$, $\zeta_{1-2\alpha} = \frac{1}{m^2}\left\{\xi_{1-2\alpha}^2 \boldsymbol{\Omega}'\boldsymbol{\Psi}\boldsymbol{\Omega} + \frac{2\xi_{1-2\alpha}}{f(\xi_{1-2\alpha})}\boldsymbol{\Omega}'\boldsymbol{\Lambda}\boldsymbol{W}_\alpha + \frac{\gamma_\alpha}{f^2(\xi_{1-2\alpha})}\right\}$.

**Theorem 3.1** *Assume that* **A2-A4** *hold. Let* $\alpha \in (0, 0.5)$*. Suppose that* $|\eta_{1t}|$ *admits a density* $f$ *which is continuous and strictly positive in a neighborhood of* $\xi_{1-2\alpha}$*. Then*

$$\sqrt{n}\begin{pmatrix} \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \\ \xi_{n,1-2\alpha} - \xi_{1-2\alpha} \end{pmatrix} \stackrel{\mathcal{L}}{\to} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Xi} := \begin{pmatrix} \boldsymbol{\Psi} & \boldsymbol{\Xi}_{\boldsymbol{\theta}\xi} \\ \boldsymbol{\Xi}'_{\boldsymbol{\theta}\xi} & \zeta_{1-2\alpha} \end{pmatrix}\right). \tag{3.3}$$

Details on how to estimate the asymptotic covariance matrix $\boldsymbol{\Xi}$ can be found in Appendix 3.4.

## 3.2 Conditional VaR parameter

The notion of VaR parameter, introduced for univariate GARCH models by Francq and Zakoïan (2015), allows to summarize the conditional risk, that is the joint effects of the volatility coefficients and the tails of the innovation process, in a single vector of coefficients. Its extension to the multivariate framework requires the following assumption.

**A5:** There exists a continuously differentiable function $\boldsymbol{G} : \mathbb{R}^d \mapsto \mathbb{R}^d$ such that for any $\boldsymbol{\theta} \in \Theta$, any $K > 0$, and any sequence $(\mathbf{x}_i)_i$ on $\mathbb{R}^m$,

$$m(\mathbf{x}_1, \mathbf{x}_2, \dots; \boldsymbol{\theta}) = m(\mathbf{x}_1, \mathbf{x}_2, \dots; \boldsymbol{\theta}^*), \quad \text{and}$$
$$K\boldsymbol{\Sigma}(\mathbf{x}_1, \mathbf{x}_2, \dots; \boldsymbol{\theta}) = \boldsymbol{\Sigma}(\mathbf{x}_1, \mathbf{x}_2, \dots; \boldsymbol{\theta}^*), \quad \text{where } \boldsymbol{\theta}^* = \boldsymbol{G}(\boldsymbol{\theta}, K).$$

studied, in particular for the QMLE by Berkes, Horváth and Kokoszka (2003) and for the LAD (Least Absolute Deviation) estimator by Ling (2005). In the multivariate setting, the asymptotic properties of the QMLE or alternative estimators were established, for particular classes, by Comte and Lieberman (2003), Boswijk and van der Weide (2011), Francq and Zakoian (2012), Pedersen and Rahbek (2014), Francq, Horváth and Zakoian (2015), Francq and Zakoian (2016) among others.

In other words, a change of the scale in the components of $\boldsymbol{\eta}$ can be compensated by a change of the parameter. This assumption is obviously satisfied for all commonly used parametric forms of $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$.[7] Under sphericity and the stability by scale assumption **A5**, the conditional VaR can be expressed in function of the expected returns vector and a reparameterized volatility matrix. Let $\alpha < 1/2$, so that $\mathrm{VaR}^{(\alpha)}(\eta) > 0$ under **A2**. It follows from **A5** that a formula of the form (2.9) holds, namely

$$\mathrm{VaR}_{t-1}^{(\alpha)}(r_t) = -\mathbf{a}_{t-1}'\boldsymbol{m}_t(\boldsymbol{\theta}_0^{(\alpha)}) + \|\mathbf{a}_{t-1}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0^{(\alpha)})\| \tag{3.4}$$

where

$$\boldsymbol{\theta}_0^{(\alpha)} = \boldsymbol{G}\left\{\boldsymbol{\theta}_0, \mathrm{VaR}^{(\alpha)}(\eta)\right\}. \tag{3.5}$$

The new parameter $\boldsymbol{\theta}_0^{(\alpha)}$ is referred to as the *conditional VaR parameter*, for a given risk level. It does not depend on the portfolio composition. An estimator of the conditional VaR parameter can be defined as

$$\widehat{\boldsymbol{\theta}}_n^{(\alpha)} = G\left\{\widehat{\boldsymbol{\theta}}_n, \widehat{\mathrm{VaR}}_n^{(\alpha)}(\eta)\right\}$$

with obvious notations. The asymptotic properties of $\widehat{\boldsymbol{\theta}}_n^{(\alpha)}$ are a direct consequence of Theorem 3.1.

**Corollary 3.1 (CAN of the VaR-parameter estimator)** *Under the assumptions of Theorem 3.1,* $\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n^{(\alpha)} - \boldsymbol{\theta}_0^{(\alpha)}\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Xi}^* := \dot{\boldsymbol{G}}\boldsymbol{\Xi}\dot{\boldsymbol{G}}'\right)$ *where* $\dot{\boldsymbol{G}} = \left[\frac{\partial G(\boldsymbol{\theta},\xi)}{\partial(\boldsymbol{\theta}',\xi)}\right]_{(\boldsymbol{\theta}_0,\xi_{1-2\alpha})}$.

**Remark 3.1 (Usefulness of the conditional VaR parameter)** Quantifying the estimation risk is in general a difficult task, due to the stochastic nature of the conditional risk. However, when the VaR takes the form (2.9), the asymptotic distribution of $\widehat{\boldsymbol{\theta}}_n^{(\alpha)}$ provides a quantification of the estimation risk. It can be used to compare the relative asymptotic efficiencies of estimators. Suppose, for instance, that estimators $\widehat{\boldsymbol{\theta}}_n^{(i)}, i = 1, \dots, m$, satisfying (3.2) are available and let $\boldsymbol{\Xi}^{(i)}$ denote the corresponding asymptotic covariance matrices in (3.3). Then, we can say that, as far as the estimation of the conditional VaR at level $\alpha$ is concerned, the $i$-th estimator is asymptotically more efficient than the $j$-th iff

$$\dot{\boldsymbol{G}}(\boldsymbol{\Xi}^{(j)} - \boldsymbol{\Xi}^{(i)})\dot{\boldsymbol{G}}' \quad \text{is a positive semidefinite matrix.}$$

[7] For instance, in the case of the VAR(1) model $\boldsymbol{y}_t = \boldsymbol{\phi}\boldsymbol{y}_{t-1} + \boldsymbol{\epsilon}_t$ with a BEKK-GARCH(1,1) model (3.7) for $\boldsymbol{\epsilon}_t$, and $\boldsymbol{\theta} = (\mathrm{vec}(\boldsymbol{\phi})', \mathrm{vec}(\boldsymbol{A})', \mathrm{vec}(\boldsymbol{B})', \mathrm{vec}(\boldsymbol{C})')'$, we find $\boldsymbol{\theta}^* = (\mathrm{vec}(\boldsymbol{\phi})', K\mathrm{vec}(\boldsymbol{A})', \mathrm{vec}(\boldsymbol{B})', K^2\mathrm{vec}(\boldsymbol{C})')'$.

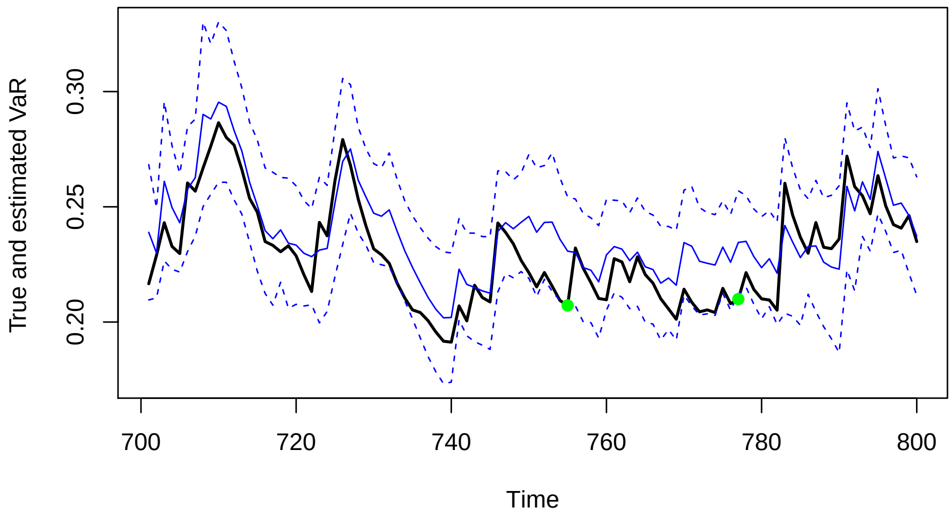


Figure 1: True 1%-VaR (full black line), estimated 1%-VaR (full blue line) and estimated 95%-confidence interval (dotted blue line), on a simulation of a fixed portfolio of a bivariate BEKK.

### 3.3 Asymptotic confidence intervals for the VaR's of portfolios

Let $\widehat{\boldsymbol{\Xi}}$ denote a consistent estimator of $\boldsymbol{\Xi}$. Let $\alpha_0 \in (0,1)$. In view of (3.1), by the delta method, an approximate $(1-\alpha_0)\%$ confidence interval (CI) for $\mathrm{VaR}_t(\alpha)$ has bounds given by

$$\widehat{\mathrm{VaR}}^{(\alpha)}_{S,t-1}(r_t) \pm \frac{1}{\sqrt{n}}\Phi^{-1}(1-\alpha_0/2)\left\{\boldsymbol{\delta}'_{t-1}\widehat{\boldsymbol{\Xi}}\boldsymbol{\delta}_{t-1}\right\}^{1/2}, \tag{3.6}$$

where $\Phi^{-1}(u)$ denotes the $u$-quantile of the standard Gaussian distribution, $u \in (0,1)$, and

$$\boldsymbol{\delta}'_{t-1} = \left[\mathbf{a}'_{t-1}\frac{\partial\widetilde{\boldsymbol{m}}(\widehat{\boldsymbol{\theta}}_n)}{\partial\boldsymbol{\theta}'} + \frac{(\mathbf{a}_{t-1}\otimes\mathbf{a}_{t-1})'}{2\|\mathbf{a}'_{t-1}\widetilde{\boldsymbol{\Sigma}}_t(\widehat{\boldsymbol{\theta}}_n)\|}\frac{\partial\mathrm{vec}\widetilde{H}_t(\widehat{\boldsymbol{\theta}}_n)}{\partial\boldsymbol{\theta}'} \qquad \|\mathbf{a}'_{t-1}\widetilde{\boldsymbol{\Sigma}}_t(\widehat{\boldsymbol{\theta}}_n)\|\right],$$

with $\widetilde{\boldsymbol{H}}_t(\cdot) = \widetilde{\boldsymbol{\Sigma}}_t(\cdot)\widetilde{\boldsymbol{\Sigma}}'_t(\cdot)$. Drawing such CIs allows to take into account the estimation risk inherent to the evaluation of the VaR of the portfolio. Note that the level $\alpha_0$ of risk estimation is independent from the market risk level $\alpha$.

An illustration is displayed in Figure 1, for the simulation of a bivariate BEKK model (see Appendix 3.4). The model parameters were estimated on 700 observations. The figure provides the true and estimated conditional 1%-VaRs, for $t > 700$, as well a CIs at 95% for the true conditional VaR, of a portfolio with fixed composition. This graph allows to visualize simultaneously the market risk (through the magnitude of the VaR) and the estimation risk (through the width of the CIs).

### 3.4 Estimating the asymptotic covariance matrix $\Xi$

In Theorem 3.1, most quantities involved in the asymptotic covariance matrix $\Xi$ can be estimated by empirical means, replacing $\boldsymbol{\theta}_0$ by the estimate $\widehat{\boldsymbol{\theta}}_n$ and the $\boldsymbol{\eta}_t$'s by the corresponding residuals. We focus on the estimation of $\boldsymbol{\Omega}$, which is the most delicate problem due to the presence of the derivatives of $\boldsymbol{\Sigma}_t$.

If a recursive linear relationship between $\boldsymbol{\Sigma}_t$ and its past-values existed, then the derivatives could be computed recursively (as the derivatives of the $\sigma_t$ or $\sigma_t^2$ in standard univariate GARCH models). Unfortunately, the standard multivariate volatility models do not allow to derive such a recursive relationship. Let us distinguish two general class of models, depending on the type of stochastic recursive equation (SRE) involved in the dynamics.

#### 3.4.1 Linear SRE on $\boldsymbol{H}_t$

A typical example is the BEKK model of Engle and Kroner (1995). As in Pedersen and Rahbek (2014), we focus on the BEKK-GARCH(1,1) model, in which $\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)$ is the symmetric square root of $\boldsymbol{H}_t$, given by

$$\boldsymbol{\epsilon}_t = \boldsymbol{H}_t^{1/2}\boldsymbol{\eta}_t, \qquad \boldsymbol{H}_t = \boldsymbol{C}_0 + \boldsymbol{A}_0\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}'\boldsymbol{A}_0' + \boldsymbol{B}_0\boldsymbol{H}_{t-1}\boldsymbol{B}_0' \tag{3.7}$$

where $\boldsymbol{A}_0, \boldsymbol{B}_0$ and $\boldsymbol{C}_0$ are real $m \times m$ matrices, with $\boldsymbol{C}_0$ positive definite, such that $\boldsymbol{H}_t$ is a positive definite matrix. For some $m \times m$ matrices $\boldsymbol{A}, \boldsymbol{B}$ and $\boldsymbol{C} > 0$, let $\boldsymbol{\theta} = (\mathrm{vec}(\boldsymbol{A})', \mathrm{vec}(\boldsymbol{B})', \mathrm{vec}(\boldsymbol{C})')'$. The derivatives of $\mathrm{vec}(\boldsymbol{H}_t)$ can be computed as follows, omitting $\boldsymbol{\theta}$ for ease of notation. From $\mathrm{vec}(\boldsymbol{H}_t) = \mathrm{vec}(\boldsymbol{C}) + (\boldsymbol{A}\otimes\boldsymbol{A})\mathrm{vec}(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t') + (\boldsymbol{B}\otimes\boldsymbol{B})\mathrm{vec}(\boldsymbol{H}_{t-1})$, it follows that, for $j = 1, \ldots, 3d$,

$$\begin{aligned}\frac{\partial \mathrm{vec}(\boldsymbol{H}_t)}{\partial\theta_j} &= \frac{\partial \mathrm{vec}(\boldsymbol{C})}{\partial\theta_j} + \frac{\partial(\boldsymbol{A}\otimes\boldsymbol{A})}{\partial\theta_j}\mathrm{vec}(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t') \\ &\quad + \frac{\partial(\boldsymbol{B}\otimes\boldsymbol{B})}{\partial\theta_j}\mathrm{vec}(\boldsymbol{H}_{t-1}) + (\boldsymbol{B}\otimes\boldsymbol{B})\frac{\partial \mathrm{vec}(\boldsymbol{H}_{t-1})}{\partial\theta_j}.\end{aligned}$$

For any $m \times n$ matrix $\boldsymbol{M}$, let the $dm \times n$ matrix $\partial\boldsymbol{M} = \left(\frac{\partial\boldsymbol{M}'}{\partial\theta_1}, \ldots, \frac{\partial\boldsymbol{M}'}{\partial\theta_d}\right)'$. Let $\boldsymbol{X}_t = (\mathrm{vec}'(\boldsymbol{H}_t), \{\partial\mathrm{vec}(\boldsymbol{H}_t)\}')'$. We have, in block matrix notation,

$$\boldsymbol{X}_t = \begin{pmatrix} \boldsymbol{B}\otimes\boldsymbol{B} & \boldsymbol{0} \\ \partial(\boldsymbol{B}\otimes\boldsymbol{B}) & \boldsymbol{I}_d\otimes(\boldsymbol{B}\otimes\boldsymbol{B}) \end{pmatrix}\boldsymbol{X}_{t-1} + \boldsymbol{e}_t, \tag{3.8}$$

where

$$\boldsymbol{e}_t = \begin{pmatrix} \mathrm{vec}(\boldsymbol{C}) \\ \partial\mathrm{vec}(\boldsymbol{C}) \end{pmatrix} + \begin{pmatrix} \boldsymbol{A}\otimes\boldsymbol{A} \\ \partial(\boldsymbol{A}\otimes\boldsymbol{A}) \end{pmatrix}\mathrm{vec}(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t').$$

Equation (3.8) allows to compute recursively the matrix $\boldsymbol{H}_t$ and its derivatives, provided that some initial values are chosen.

It remains to compute the derivatives of $\boldsymbol{\Sigma}_t = \boldsymbol{H}_t^{1/2}$. Without generality loss, this matrix can be assumed to be symmetric and positive definite. We note that $\boldsymbol{\Sigma}_t \frac{\partial \boldsymbol{\Sigma}_t}{\partial \theta_i} + \frac{\partial \boldsymbol{\Sigma}_t}{\partial \theta_i} \boldsymbol{\Sigma}_t = \frac{\partial \boldsymbol{H}_t}{\partial \theta_i}$. Thus

$$(\boldsymbol{I}_m \otimes \boldsymbol{\Sigma}_t + \boldsymbol{\Sigma}_t \otimes \boldsymbol{I}_m) \operatorname{vec}\left(\frac{\partial \boldsymbol{\Sigma}_t}{\partial \theta_i}\right) = \operatorname{vec}\left(\frac{\partial \boldsymbol{H}_t}{\partial \theta_i}\right), \tag{3.9}$$

which allows to compute the derivative of $\boldsymbol{\Sigma}_t$ provided $\boldsymbol{I}_m \otimes \boldsymbol{\Sigma}_t + \boldsymbol{\Sigma}_t \otimes \boldsymbol{I}_m$ is non-singular. In fact

$$\boldsymbol{I}_m \otimes \boldsymbol{\Sigma}_t + \boldsymbol{\Sigma}_t \otimes \boldsymbol{I}_m = (\boldsymbol{I}_m \otimes \boldsymbol{\Sigma}_t)(\boldsymbol{I}_{m^2} + \boldsymbol{\Sigma}_t \otimes \boldsymbol{\Sigma}_t^{-1}).$$

The eigenvalues of $\boldsymbol{\Sigma}_t^{-1}$ and $\boldsymbol{\Sigma}_t$ being positive, the eigenvalues of the latter parenthesis are larger than 1. The invertibility of $\boldsymbol{I}_m \otimes \boldsymbol{\Sigma}_t + \boldsymbol{\Sigma}_t \otimes \boldsymbol{I}_m$ follows and we have

$$\operatorname{vec}\left(\frac{\partial \boldsymbol{\Sigma}_t}{\partial \theta_i}\right) = (\boldsymbol{I}_m \otimes \boldsymbol{\Sigma}_t + \boldsymbol{\Sigma}_t \otimes \boldsymbol{I}_m)^{-1} \operatorname{vec}\left(\frac{\partial \boldsymbol{H}_t}{\partial \theta_i}\right).$$

### 3.4.2 Linear SRE's on the individual volatilities and the conditional correlation matrix

Consider parameterizations of the form $\boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \boldsymbol{D}_t(\boldsymbol{\theta})\boldsymbol{R}_t^{1/2}(\boldsymbol{\theta})$ where $\boldsymbol{D}_t(\boldsymbol{\theta})$ is the diagonal matrix of the individual volatilities (at $\boldsymbol{\theta}_0$), and $\boldsymbol{R}_t^{1/2}(\boldsymbol{\theta})$ denotes the symmetric positive definite square-root of the conditional correlation matrix $\boldsymbol{R}_t(\boldsymbol{\theta})$ (that is $\{\boldsymbol{R}_t^{1/2}(\boldsymbol{\theta})\}^2 = \boldsymbol{R}_t(\boldsymbol{\theta})$). For all commonly used models, the derivatives of the individual volatilities (or their squares) can be straightforwardly computed, using the SRE on the vector of individual volatilities. The matrix $\frac{\partial}{\partial \theta_i}\boldsymbol{D}_t(\boldsymbol{\theta})$ follows, for any component $\theta_i$ of $\boldsymbol{\theta}$. Turning to the derivatives of $\boldsymbol{R}_t^{1/2}(\boldsymbol{\theta})$, we note that, similar to (3.9),

$$\operatorname{vec}\left(\frac{\partial \boldsymbol{R}_t^{1/2}}{\partial \theta_i}\right) = \left(\boldsymbol{I}_m \otimes \boldsymbol{R}_t^{1/2} + \boldsymbol{R}_t^{1/2} \otimes \boldsymbol{I}_m\right)^{-1} \operatorname{vec}\left(\frac{\partial \boldsymbol{R}_t}{\partial \theta_i}\right).$$

Usual DCC models provide a SRE on the conditional correlation matrix $\boldsymbol{R}_t$, from which the derivatives of $\boldsymbol{R}_t^{1/2}$ can be computed using the previous equality. Consider the cDCC model (see Appendix C). We have $\boldsymbol{R}_t = \boldsymbol{Q}_t^{*-1/2}\boldsymbol{Q}_t\boldsymbol{Q}_t^{*-1/2}$, and

$$\boldsymbol{Q}_t = (1-\alpha-\beta)\boldsymbol{S} + \alpha \boldsymbol{Q}_{t-1}^{*1/2}\boldsymbol{D}_{t-1}^{-1}\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}'\boldsymbol{D}_{t-1}^{-1}\boldsymbol{Q}_{t-1}^{*1/2} + \beta \boldsymbol{Q}_{t-1},$$

where $\boldsymbol{S}$ is a correlation matrix. The diagonal terms of $\boldsymbol{Q}_t$ are given by

$$q_{ii,t} = (1-\alpha-\beta) + \left(\alpha \frac{\epsilon_{i,t-1}^2}{\sigma_{i,t-1}^2} + \beta\right) q_{ii,t-1},$$

from which the derivatives of $\boldsymbol{Q}_t^*$ can be recursively computed. The derivatives of $\boldsymbol{Q}_t^{*1/2}$ follow from (3.9), which in the diagonal case reduces to $\frac{\partial \boldsymbol{Q}_t^{*1/2}}{\partial \theta_i} = \frac{1}{2}\boldsymbol{Q}_t^{*-1/2}\frac{\partial \boldsymbol{Q}_t^*}{\partial \theta_i}$. Now we turn to the non diagonal terms. We have, for $i \neq j$,

$$q_{ij,t} = (1-\alpha-\beta)S_{ij} + \alpha\sqrt{q_{ii,t-1}}\frac{\epsilon_{i,t-1}}{\sigma_{i,t-1}}\sqrt{q_{jj,t-1}}\frac{\epsilon_{j,t-1}}{\sigma_{j,t-1}} + \beta q_{ij,t-1},$$

from which the derivatives of $q_{ij,t}$ follow recursively. The conclusion follows.

### 3.5 CI's based on a conditional resampling scheme

For certain estimation methods/models the asymptotic distribution of the estimator $\widehat{\boldsymbol{\theta}}_n$ may not be available. Even when it is, as shown in the previous section, the asymptotic variance $\boldsymbol{\Xi}$ may be difficult to compute. In particular, apart from the estimation of $\boldsymbol{\Omega}$, evaluating the density function $f$ of the innovations distribution at the desired quantile may be delicate. An alternative, which we will now illustrate, is a bootstrap procedure. We will use the well-known result that, under the sphericity assumption, $\|\boldsymbol{\eta}_t\|$ and $\boldsymbol{\eta}_t/\|\boldsymbol{\eta}_t\|$ are independent, the latter random variable being uniformly distributed over the unit sphere $\mathcal{S}^{m-1}$.

We consider the following resampling scheme, given observations $\boldsymbol{y}_1, \dots, \boldsymbol{y}_n$ and initial values:

1. Compute $\widehat{\boldsymbol{\theta}}_n = \widehat{\boldsymbol{\theta}}_n(\boldsymbol{y}_1, \dots, \boldsymbol{y}_n)$, the residuals $\widetilde{\boldsymbol{\eta}}_t$, and the estimator $\widehat{\mathrm{VaR}}^{(\alpha)}_{S,t-1}(r_t) =: \widehat{\mathrm{VaR}}(r_t)$.

2. Generate independent vectors $\boldsymbol{s}_u^*, u = 1, \dots, n$, that are uniformly distributed over $\mathcal{S}^{m-1}$. independently, generate vectors $\widetilde{\boldsymbol{U}}_u^*$, that are uniformly distributed on $(\widetilde{\boldsymbol{U}}_1, \dots, \widetilde{\boldsymbol{U}}_n)$ where $\widetilde{\boldsymbol{U}}_u = \boldsymbol{S}_u^{-1/2}(\widetilde{\boldsymbol{\eta}}_u - \widetilde{\boldsymbol{\eta}})$, $\boldsymbol{S}_u$ is the sample covariance matrix of the residuals $\widetilde{\boldsymbol{\eta}}_u$ and $\widetilde{\boldsymbol{\eta}}$ is their sample mean. Compute $\boldsymbol{\eta}_u^* = \|\widetilde{\boldsymbol{U}}_u^*\|\boldsymbol{s}_u^*$ and let $\boldsymbol{y}_u^* = \widetilde{\boldsymbol{m}}_u^*(\widehat{\boldsymbol{\theta}}_n) + \widetilde{\boldsymbol{\Sigma}}_u^*(\widehat{\boldsymbol{\theta}}_n)\boldsymbol{\eta}_u^*$, where $\widetilde{\boldsymbol{m}}_u^*(\widehat{\boldsymbol{\theta}}_n) = \boldsymbol{m}(\boldsymbol{y}_{t-1}^*, \dots, \boldsymbol{y}_1^*, \widetilde{\boldsymbol{y}}_0, \widetilde{\boldsymbol{y}}_{-1}, \dots, \widehat{\boldsymbol{\theta}}_n)$ and $\widetilde{\boldsymbol{\Sigma}}_u^*(\widehat{\boldsymbol{\theta}}_n) = \boldsymbol{\Sigma}(\boldsymbol{y}_{t-1}^*, \dots, \boldsymbol{y}_1^*, \widetilde{\boldsymbol{y}}_0, \widetilde{\boldsymbol{y}}_{-1}, \dots, \widehat{\boldsymbol{\theta}}_n)$.

3. Compute $\widehat{\boldsymbol{\theta}}_n^* = \widehat{\boldsymbol{\theta}}_n(\boldsymbol{y}_1^*, \dots, \boldsymbol{y}_n^*)$, the resampling residuals $\widetilde{\boldsymbol{\eta}}_u^* = \widetilde{\boldsymbol{\Sigma}}_u^{-1}(\widehat{\boldsymbol{\theta}}_n^*)\{\boldsymbol{y}_u^* - \widetilde{\boldsymbol{m}}_u(\widehat{\boldsymbol{\theta}}_n^*)\})$ where $\widetilde{\boldsymbol{m}}_u(\widehat{\boldsymbol{\theta}}_n^*) = \boldsymbol{m}(\boldsymbol{y}_{t-1}, \dots, \boldsymbol{y}_1, \widetilde{\boldsymbol{y}}_0, \widetilde{\boldsymbol{y}}_{-1}, \dots, \widehat{\boldsymbol{\theta}}_n^*)$ and $\widetilde{\boldsymbol{\Sigma}}_u(\widehat{\boldsymbol{\theta}}_n) = \boldsymbol{\Sigma}(\boldsymbol{y}_{t-1}, \dots, \boldsymbol{y}_1, \widetilde{\boldsymbol{y}}_0, \widetilde{\boldsymbol{y}}_{-1}, \dots, \widehat{\boldsymbol{\theta}}_n^*)$, and the estimator

$$\widehat{\mathrm{VaR}}^*(r_t) = -\mathbf{a}_{t-1}'\widetilde{\boldsymbol{m}}_t(\widehat{\boldsymbol{\theta}}_n^*) + \|\mathbf{a}_{t-1}'\widetilde{\boldsymbol{\Sigma}}_t(\widehat{\boldsymbol{\theta}}_n^*)\|\xi_{n,1-2\alpha}^*, \tag{3.10}$$

where $\xi_{n,1-2\alpha}^* = q_{1-2\alpha}\left(\{|\widehat{\eta}_{iu}^*|, 1 \le i \le m, 1 \le u \le n\}\right)$.

4. Repeat $B$ times Steps 1-3, resulting in $\widehat{\mathrm{VaR}}_1^*(r_t), \dots \widehat{\mathrm{VaR}}_B^*(r_t)$, say.

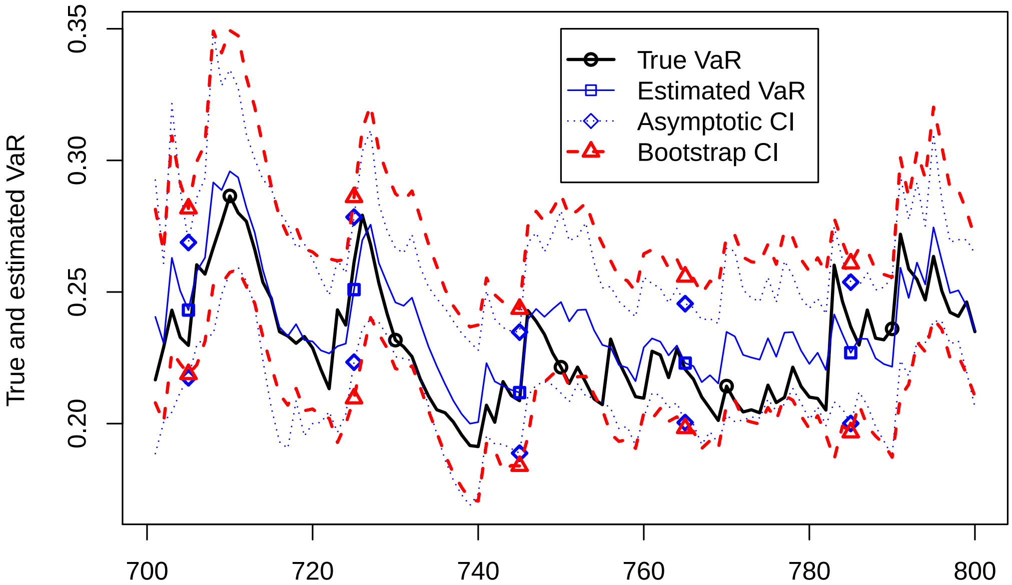

Figure 2: Confidence Intervals based on asymptotic results vs bootstrap.

Note that in Step 3, the conditional moments $\widetilde{\boldsymbol{\Sigma}}_t(\cdot)$ and $\widetilde{\boldsymbol{m}}_t(\cdot)$ are built using the real data (not the bootstraped ones), because we are estimating a conditional VaR (not a marginal one). Using the pivot method (see e.g. Davison and Hinkley (1997)), we get a CI for the conditional VaR at the confidence level $1-\alpha_0$ as

$$\left[\widehat{\text{VaR}}(r_t) - \left\{\widehat{\text{VaR}}^*_{(1-\alpha_0/2)}(r_t) - \widehat{\text{VaR}}(r_t)\right\}, \widehat{\text{VaR}}(r_t) - \left\{\widehat{\text{VaR}}^*_{(\alpha_0/2)}(r_t) - \widehat{\text{VaR}}(r_t)\right\}\right],$$

with standard notation for the order statistic. An illustration is displayed in Figure 2, for the same setting as for Figure 1. As can be seen, the CI's obtained by the bootstrap approach are similar to those obtained using the asymptotic results. For more complex models, or for estimators for which the asymptotic distribution is unknown or cannot be estimated, the latter CI's would be impossible to derive, while the bootstrap approach described above could be implemented without further difficulties. The validity of this procedure is however an open issue.

## 4 VaR estimation without the sphericity assumption

Rombouts and Verbeek (2009) proposed a semi-parametric method for evaluating the VaR of portfolios, which relies on: i) estimating $\boldsymbol{\theta}_0$, ii) using a Kernel estimator of the (multivariate) density of $\boldsymbol{\eta}_t$, iii) evaluating by numerical integration the conditional VaR of a portfolio. While this approach seems attractive from a practical point of view, its asymptotic properties are unknown. Deriving asymptotic confidence intervals for the VaR, which is the aim of this paper, would probably be extremely difficult with this method.

In this section, we study an alternative semi-parametric method which is amenable to asymptotic properties. This approach, called FHS, relies on

i) interpreting the conditional VaR at time $t$ as the $\alpha$-quantile of a linear combination (depending on $t$) of the components of the innovations;

ii) replacing the innovations by the GARCH residuals and computing the empirical $\alpha$-quantile of the estimated linear combination.

The conditional VaR of the portfolio return is

$$\mathrm{VaR}_{t-1}^{(\alpha)}(r_t) = \mathrm{VaR}_{t-1}^{(\alpha)}\left\{b_t(\boldsymbol{\theta}_0) + \boldsymbol{c}_t'(\boldsymbol{\theta}_0)\boldsymbol{\eta}_t\right\}$$

where $b_t(\boldsymbol{\theta}) = \mathbf{a}_{t-1}'\boldsymbol{m}_t(\boldsymbol{\theta})$ and $\boldsymbol{c}_t'(\boldsymbol{\theta}) = \mathbf{a}_{t-1}'\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$. The conditional VaR at time $t$ can thus be interpreted as the sum of the conditional mean and a quantile of a time-varying linear combination of the components of the iid noise. It can be estimated by

$$\widehat{\mathrm{VaR}}_{FHS,t-1}^{(\alpha)}(r_t) = -q_\alpha\left(\left\{b_t(\widehat{\boldsymbol{\theta}}_n) + \boldsymbol{c}_t'(\widehat{\boldsymbol{\theta}}_n)\widehat{\boldsymbol{\eta}}_s, \quad 1 \le s \le n\right\}\right). \tag{4.1}$$

**Remark 4.1 (On the name FHS)** In (4.1), all residuals are used to estimate the VaR. Alternatively, the conditional VaR could be estimated by randomly drawing $N$ residuals among the $\widehat{\boldsymbol{\eta}}_s$'s, for some specified number $N$ (hence the term "simulation" in FHS).

**Remark 4.2 (Higher horizons)** The approach can be extended to higher horizons. For $N$ independent draws of the $\widehat{\boldsymbol{\eta}}_s$'s, $N$ scenarios $\boldsymbol{y}_t^{(1)}, \dots, \boldsymbol{y}_t^{(N)}$ for $\boldsymbol{y}_t$ are obtained. For each value $\boldsymbol{y}_t^{(i)}$, another set of $N$ independent draws of the $\widehat{\boldsymbol{\eta}}_s$'s, produces $N$ scenarios $\boldsymbol{y}_{t+1}^{(i,1)}, \dots, \boldsymbol{y}_{t+1}^{(i,N)}$ for $\boldsymbol{y}_{t+1}$. Proceeding recursively, at horizon $H$ we get $N^H$ scenarios $\boldsymbol{y}_{t+H-1}^{(i_1,\dots,i_H)}$ for $\boldsymbol{y}_{t+H-1}$, where $i_j \in \{1, \dots, N\}$. Such scenarios allow to update the sequence of weights $\boldsymbol{a}_s$, for $s = t, \dots, t+H-1$. We deduce $N^H$ scenarios $r_{t+H-1}^{(i_1,\dots,i_H)}$ for $r_{t+H-1}$. The VaR of the portfolio at horizon $H$ conditional on the available information at time $t-1$ can thus be estimated by

$$\widehat{\mathrm{VaR}}_{FHS,t-1}^{(H,\alpha)}(r_{t+H-1}) = -q_\alpha\left(\left\{r_{t+H-1}^{(i_1,\dots,i_{H+1})}, \quad i_j \in \{1, \dots, N\}\right\}\right). \tag{4.2}$$

Let $\boldsymbol{c} : \boldsymbol{\Theta}_{\boldsymbol{\theta}} \mapsto \mathbb{R}^m$ and $b : \boldsymbol{\Theta}_{\boldsymbol{\theta}} \mapsto \mathbb{R}$ denote continuously differentiable vector-valued functions. Let $\xi_\alpha(\boldsymbol{\theta})$ denote the theoretical $\alpha$-quantile of $b(\boldsymbol{\theta}) + \boldsymbol{c}'(\boldsymbol{\theta})\boldsymbol{\eta}_t(\boldsymbol{\theta})$, where $\boldsymbol{\eta}_t(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\{\boldsymbol{y}_t - \boldsymbol{m}_t(\boldsymbol{\theta})\}$.

Let $\xi_{n,\alpha}(\boldsymbol{\theta}) = q_\alpha\left(\{b(\boldsymbol{\theta}) + \boldsymbol{c}'(\boldsymbol{\theta})\boldsymbol{\eta}_t(\boldsymbol{\theta}), 1 \le t \le n\}\right)$. Let $\boldsymbol{A}_\alpha = \mathrm{Cov}(\boldsymbol{V}(\boldsymbol{\eta}_t), \mathbf{1}_{\{b(\boldsymbol{\theta}_0)+\boldsymbol{c}'(\boldsymbol{\theta}_0)\boldsymbol{\eta}_t<\xi_\alpha(\boldsymbol{\theta}_0)\}})$,

$$\boldsymbol{\omega}' = \left[\boldsymbol{c}'(\boldsymbol{\theta}_0)E(\boldsymbol{C}_t) - \frac{\partial b}{\partial \boldsymbol{\theta}'}(\boldsymbol{\theta}_0) \qquad \boldsymbol{d}'_\alpha\left\{(\boldsymbol{c}'(\boldsymbol{\theta}_0)\otimes \boldsymbol{I}_m)E(\boldsymbol{\Omega}^*_t) - \frac{\partial \boldsymbol{c}}{\partial \boldsymbol{\theta}'}(\boldsymbol{\theta}_0)\right\}\right],$$

where $\boldsymbol{d}_\alpha = E(\boldsymbol{\eta}_t \mid b(\boldsymbol{\theta}_0) + \boldsymbol{c}'(\boldsymbol{\theta}_0)\boldsymbol{\eta}_t = \xi_\alpha(\boldsymbol{\theta}_0))$ and

$$\begin{aligned}
\boldsymbol{\Omega}^*_t &= \begin{pmatrix} I_m \otimes \boldsymbol{e}'_1 \\ \vdots \\ I_m \otimes \boldsymbol{e}'_m \end{pmatrix} (I_m \otimes \boldsymbol{\Sigma}_t^{-1}) \frac{\partial}{\partial \boldsymbol{\theta}'}\left\{\mathrm{vec}(\boldsymbol{\Sigma}_t)\right\}, \\
\boldsymbol{C}_t &= \left\{\boldsymbol{I}_m \otimes \mathrm{vec}'\left(\frac{\partial \boldsymbol{m}_t}{\partial \boldsymbol{\theta}'}\right)\right\} \begin{pmatrix} I_d \otimes \boldsymbol{\Sigma}_t^{-1}\boldsymbol{e}_1 \\ \vdots \\ I_d \otimes \boldsymbol{\Sigma}_t^{-1}\boldsymbol{e}_m \end{pmatrix}.
\end{aligned}$$

The following result establishes the asymptotic distribution of $\xi_{n,\alpha}(\widehat{\boldsymbol{\theta}}_n)$.

**Theorem 4.1** *Assume that* **A1, A3** *hold. Suppose that the variable* $\boldsymbol{c}'(\boldsymbol{\theta}_0)\boldsymbol{\eta}_t$ *admits a density* $f_{\boldsymbol{c}}$ *which is continuous and strictly positive in a neighborhood of* $x_0 = \xi_\alpha(\boldsymbol{\theta}_0) - b(\boldsymbol{\theta}_0)$. *Then*

$$\sqrt{n}\{\xi_{n,\alpha}(\widehat{\boldsymbol{\theta}}_n) - \xi_\alpha(\boldsymbol{\theta}_0)\} \overset{\mathcal{L}}{\to} \mathcal{N}\left(0, \sigma^2 := \boldsymbol{\omega}'\boldsymbol{\Psi}\boldsymbol{\omega} + 2\boldsymbol{\omega}'\boldsymbol{\Lambda}\boldsymbol{A}_\alpha + \frac{\alpha(1-\alpha)}{f^2_{\boldsymbol{c}}(x_0)}\right).$$

This theorem can be used to derive CIs for the conditional VaR at time $t$ of the portfolio return, with $b(\boldsymbol{\theta}) = \mathbf{a}'_{t-1}\boldsymbol{m}_t(\boldsymbol{\theta})$ and $\boldsymbol{c}'(\boldsymbol{\theta}) = \mathbf{a}'_{t-1}\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$. A Nadaraya-Watson estimator of $\boldsymbol{d}_\alpha$ is, with standard notation,

$$\widehat{\boldsymbol{d}}_{\alpha,t} = \frac{\sum_{s=1}^n \widehat{\boldsymbol{\eta}}_s K_h\left(b(\widehat{\boldsymbol{\theta}}_n) + \boldsymbol{c}'(\widehat{\boldsymbol{\theta}}_n)\widehat{\boldsymbol{\eta}}_s - \xi_{n,\alpha}(\widehat{\boldsymbol{\theta}}_n)\right)}{\sum_{s=1}^n K_h\left(b(\widehat{\boldsymbol{\theta}}_n) + \boldsymbol{c}'(\widehat{\boldsymbol{\theta}}_n)\widehat{\boldsymbol{\eta}}_s - \xi_{n,\alpha}(\widehat{\boldsymbol{\theta}}_n)\right)}.$$

A consistent estimator $\hat{\sigma}^2_{t-1}$ of $\sigma^2$ can be obtained by replacing the other theoretical quantities introduced before the theorem by their empirical counterparts, and by using the approach described in Appendix 3.4 to compute the derivatives of $\boldsymbol{\Sigma}_t$ and $\boldsymbol{m}_t$ for particular models. An approximate $(1-\alpha_0)\%$ CI for $\mathrm{VaR}^{(\alpha)}_{t-1}(r_t)$ is thus given by

$$\widehat{\mathrm{VaR}}^{(\alpha)}_{FHS,t-1}(r_t) \pm \frac{1}{\sqrt{n}}\Phi^{-1}(1-\alpha_0/2)\hat{\sigma}_{t-1}. \tag{4.3}$$

At higher horizons, deriving asymptotic CI's for the VaR in (4.2) seems a formidable task. Alternatively, the bootstrap procedure described in Section 3.5 could be extended to non-elliptical distributions and to higher horizons, but this is beyond the scope of this paper.

**Remark 4.3 (Data driven portfolio's composition)** For certain portfolios, the composition may depend on $\widehat{\boldsymbol{\theta}}_n$. We then write $\boldsymbol{a}_{t-1}(\widehat{\boldsymbol{\theta}}_n)$ instead of $\boldsymbol{a}_{t-1}$. Take the example of the Markowitz's Minimum Variance Portfolio (MVP). In a general multivariate model (2.3) with $m(\cdot) = 0$, the MVP is defined by

$$r_t^* = \boldsymbol{\epsilon}_t' \boldsymbol{a}_{t-1}^*, \quad \boldsymbol{a}_{t-1}^* = \frac{\boldsymbol{\Sigma}_t^{-2}(\boldsymbol{\theta}_0)\boldsymbol{e}}{\boldsymbol{e}'\boldsymbol{\Sigma}_t^{-2}(\boldsymbol{\theta}_0)\boldsymbol{e}}. \tag{4.4}$$

If Theorem 4.1 is applied with $b(\boldsymbol{\theta}) = 0$ and

$$\boldsymbol{c}'(\boldsymbol{\theta}) = \mathbf{a}_{t-1}'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t(\boldsymbol{\theta}) \quad \text{where} \quad \mathbf{a}_{t-1}'(\boldsymbol{\theta}) = \boldsymbol{e}'\boldsymbol{\Sigma}_t^{-2}(\boldsymbol{\theta})/\boldsymbol{e}'\boldsymbol{\Sigma}_t^{-2}(\boldsymbol{\theta})\boldsymbol{e}, \tag{4.5}$$

then $\xi_\alpha(\boldsymbol{\theta}_0)$ corresponds to the conditional VaR of the theoretical MVP. If Theorem 4.1 is applied with $b(\boldsymbol{\theta}) = 0$ and

$$\boldsymbol{c}'(\boldsymbol{\theta}) = \mathbf{a}_{t-1}'(\widehat{\boldsymbol{\theta}}_n)\boldsymbol{\Sigma}_t(\boldsymbol{\theta}) \quad \text{where} \quad \mathbf{a}_{t-1}'(\widehat{\boldsymbol{\theta}}_n) = \boldsymbol{e}'\boldsymbol{\Sigma}_t^{-2}(\widehat{\boldsymbol{\theta}}_n)/\boldsymbol{e}'\boldsymbol{\Sigma}_t^{-2}(\widehat{\boldsymbol{\theta}}_n)\boldsymbol{e}, \tag{4.6}$$

then $\xi_\alpha(\boldsymbol{\theta}_0)$ corresponds to the conditional VaR of the estimated MVP. Note that the asymptotic variance $\sigma^2$ depends on the derivatives of $\boldsymbol{c}$ (which are different in (4.5) and (4.6)) via the vector $\boldsymbol{\omega}$. It seems less interesting to evaluate the statistical risk of the theoretical MVP through (4.5) than that of the estimated MVP through (4.6), since this is the portfolio that is actually used in practice.

## 5 Numerical illustrations

The first part of the section presents a Monte-Carlo experiment that investigates the empirical coverage properties of the confidence intervals proposed in Sections 3.3 and 4. Real data examples are presented in the second part. [8]

### 5.1 Coverage probability of the VaR confidence intervals

We simulated $N = 1,000$ independent trajectories of length $n+1$ of a BEKK-GARCH(1,1) model (3.7), with $m = 2$ components and the volatility parameters $\text{vech}(\boldsymbol{\Omega}_0) = (0.001, 0, 0.001)'$, $\text{vec}(\boldsymbol{A}_0) = (0.1, 0.1, 0.1, 0.1)'$, $\text{diag}(\boldsymbol{B}_0) = (0.9, 0.95)'$. In Design S, $\boldsymbol{\eta}_t$ is distributed as a normalized Student with 9 degrees of freedom; in Design NS, $\boldsymbol{\eta}_t$ is distributed as the Asymmetric Exponential Power Distribution (AEPD) with parameters $\alpha = 0.5$, $p_1 = 1$ and $p_2 = 2$. The class of AEPD was

[8] The code and data used in the paper, as well as additional numerical illustrations, are available from the authors upon request, or on their web pages.

Table 1: Coverage properties of the $(1-\alpha_0)\%$ CI for the $\alpha\%$ VaR: relative frequencies (in %) of VaR in the estimated CI's over the 1,000 replications.

| | $n=1,000$ | | | | | | $n=2,000$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | 1% | | | 5% | | | 1% | | | 5% | | |
| $1-\alpha_0$ | 90% | 95% | 99% | 90% | 95% | 99% | 90% | 95% | 99% | 90% | 95% | 99% |
| Design S | 89.1 | 92.7 | 97.4 | 90.6 | 94.0 | 97.4 | 90.5 | 95.5 | 98.4 | 91.6 | 94.9 | 98.7 |
| Design NS | 8.6 | 10.5 | 15.2 | 20.9 | 25.4 | 34.2 | 6 | 7.4 | 9.7 | 12.8 | 16.6 | 23.5 |

Table 2: Coverage properties of the $(1-\alpha_0)\%$ CI for the $\alpha\%$ VaR: relative frequencies of VaR in the estimated CI's over the 1,000 replications.

| | $n=1,000$ | | | | | | $n=2,000$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | 1% | | | 5% | | | 1% | | | 5% | | |
| $1-\alpha_0$ | 90% | 95% | 99% | 90% | 95% | 99% | 90% | 95% | 99% | 90% | 95% | 99% |
| Design S | 89.5 | 94.1 | 97.7 | 89.7 | 94.3 | 98.2 | 89.7 | 95.4 | 99 | 90.9 | 95.3 | 98.3 |
| Design NS | 3.9 | 4.8 | 7.5 | 13.6 | 17.5 | 26.3 | 1.9 | 2.8 | 3.7 | 7.0 | 9.4 | 14.0 |

introduced by Zhu and Zinde-Walsh (2009) and allows for skewness with different decay rates of density in the left and right tails. On each simulation, the first $n$ observations were used for estimation of the theoretical value-at-risk $\text{VaR}_n^{(\alpha)}(r_{n+1})$ of an equally-weighted crystalized portfolio ($\mu_i = 1$ for $i = 1, 2$) at time $t = n+1$, under and without the sphericity assumption. We then checked if this theoretical VaR belonged to the confidence interval defined by (3.6) in the spherical case and by (4.3) without sphericity (with $t-1$ replaced by $n$ and $r_t$ replaced by $r_{n+1}$). For nominal coverage probabilities of 90%, 95% and 99%, respectively, the empirical coverage probabilities over the $N =$ 1,000 replications should belong to the intervals $[87.5\%, 92.4\%]$, $[93.1\%, 96.7\%]$ and $[98.1\%, 99.7\%]$, respectively, with probability 99%.

Table 3 shows that, in the spherical case, the empirical coverage probabilities are very close to

Table 3: Coverage properties for the FHS method.

| | $n=1,000$ | | | | | | $n=2,000$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | 1% | | | 5% | | | 1% | | | 5% | | |
| $1-\alpha_0$ | 90% | 95% | 99% | 90% | 95% | 99% | 90% | 95% | 99% | 90% | 95% | 99% |
| Design S | 74.5 | 79.8 | 88.6 | 81.3 | 87.2 | 93.2 | 78.1 | 84.4 | 91.2 | 78.2 | 85.3 | 93.0 |
| Design NS | 18.5 | 22.9 | 28.8 | 32.4 | 36.2 | 45.7 | 14.1 | 16.2 | 20.1 | 23.0 | 27.5 | 34.9 |

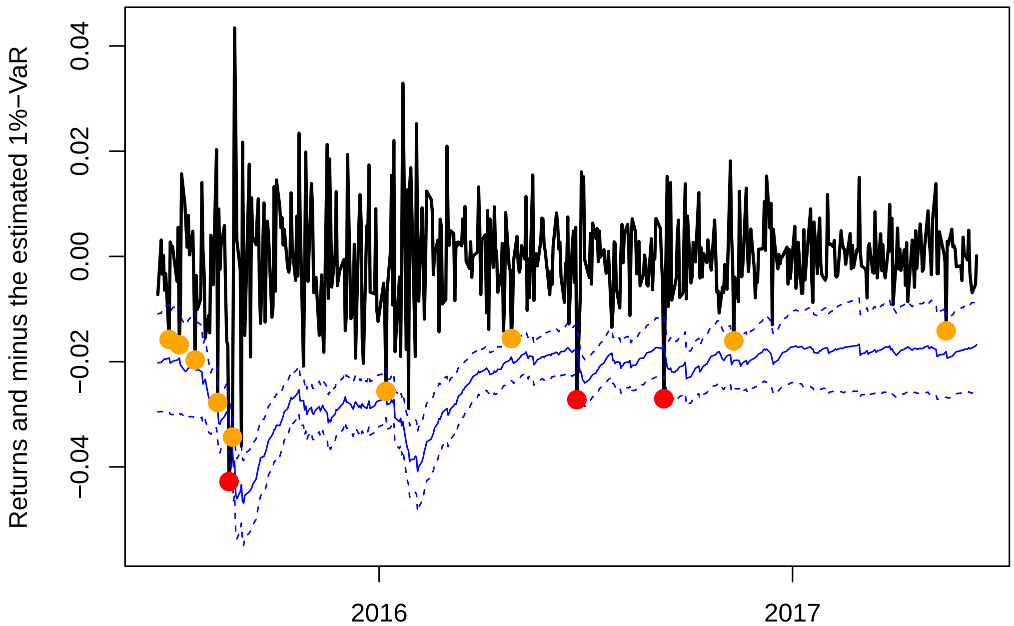


Figure 3: Returns of the portfolio (dark line) for the period 18/06/2015 to 13/06/2017, estimated 1%-VaR and 95%-confidence interval (full and dotted blue lines), based on the estimation of a BEKK model for the stocks.

their nominal values, at least when $n = 2,\ 000$. As expected, the empirical coverage probabilities provided by the spherical method are no more valid when the DGP has a non spherical innovation. This is due to the fact that the VaR's (and thus their CI's) are not consistently estimated when the innovations are asymmetrically distributed. Not however that our choice of parameters for the AEPD induces a more prominent asymmetry in the tails than what can be estimated from real financial data (see Table 2 in Zhu and Zinde-Walsh (2009)).

### 5.2 Portfolios of stocks

We considered the daily returns of 5 major NASDAQ companies : Apple, Coca-cola, Exxon Mobil, Intel and JPMorgan, from January 4, 2000 to June 13, 2017. The total number of observations is $n = 4389$. The data have been cleaned up to take into account stock splits.

We first estimated a BEKK model on the 5 stock returns over the whole sample except the last 500 returns. We considered an equally-weighted crystalized portfolio ($\mu_i = 1$ for $i = 1, \dots, 5$) and the VaR estimator based on the sphericity assumption. Figure 3, displaying the last 500 returns of the portfolio, shows that three returns are below the lower bound of the 95%-CI of the 1%-VaR. For such returns, there is strong evidence of violation of the theoretical VaR. For several other returns belonging to the CI, violation can be suspected.

A standard approach for evaluating VaR models is to use backtesting. Instead of the BEKK, we estimated the more popular DCC-GARCH(1,1) model on the first $n_1 = 3000$ observations and computed the residuals $\widehat{\boldsymbol{\eta}}_u, u = 1, \dots, n_1$. Instead of crystalized portfolios, we considered MVPs. In the case where the distribution of $\boldsymbol{\eta}_1$ is spherical, the theoretical conditional VaR of a MVP is obtained by computing the opposite of the $\alpha$-quantile of $\boldsymbol{a}_{t-1}^{*'}\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)\boldsymbol{\eta}_1$, which is simply given by

$$\mathrm{VaR}_{t-1}^{(\alpha)}(r_t^*) = \left\| \boldsymbol{a}_{t-1}^{*'}\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0) \right\| F_{|\eta_1|}^{-1}(1-2\alpha) = \frac{1}{\sqrt{\boldsymbol{e}'\boldsymbol{\Sigma}_t^{-2}(\boldsymbol{\theta}_0)\boldsymbol{e}}} F_{|\eta_1|}^{-1}(1-2\alpha) \tag{5.1}$$

with an invertible cumulative distribution function $F_{\eta_1}$, and $\alpha \in (0, 1/2)$.

Figure 4 displays the returns of the estimated Markowitz MVP

$$\widehat{r}_t^* = \frac{\boldsymbol{e}'\widetilde{\boldsymbol{\Sigma}}_t^{-2}(\widehat{\boldsymbol{\theta}}_{n_1})\boldsymbol{\epsilon}_t}{\boldsymbol{e}'\widetilde{\boldsymbol{\Sigma}}_t^{-2}(\widehat{\boldsymbol{\theta}}_{n_1})\boldsymbol{e}}, \qquad t = n_1 + 1, \dots, n$$

together with $\widehat{\mathrm{VaR}}_{S,t-1}^{(1\%)}(r_t^*)$ (left panel) and $\widehat{\mathrm{VaR}}_{FHS,t-1}^{(1\%)}(r_t^*)$ (right panel), as defined by

$$\widehat{\mathrm{VaR}}_{S,t-1}^{(\alpha)}(r_t^*) = \frac{\xi_{n_1,1-2\alpha}}{\sqrt{\boldsymbol{e}'\widetilde{\boldsymbol{\Sigma}}_t^{-2}(\widehat{\boldsymbol{\theta}}_{n_1})\boldsymbol{e}}},$$

$$\widehat{\mathrm{VaR}}_{FHS,t-1}^{(\alpha)}(r_t^*) = -q_\alpha\left(\left\{\frac{\boldsymbol{e}'\widetilde{\boldsymbol{\Sigma}}_t^{-1}(\widehat{\boldsymbol{\theta}}_{n_1})\widehat{\boldsymbol{\eta}}_u}{\boldsymbol{e}'\widetilde{\boldsymbol{\Sigma}}_t^{-2}(\widehat{\boldsymbol{\theta}}_{n_1})\boldsymbol{e}}, u = 1, \dots, n_1\right\}\right).$$

The most striking output is that the two methods give virtually indistinguishable estimated VaRs for the Markowitz portfolio. Applying on the multivariate residuals $\widehat{\boldsymbol{\eta}}_u$ a sphericity test recently proposed by Francq, Jimenez Gamero and Meintanis (2017), we found that the sphericity hypothesis cannot be rejected at any reasonable level. [9] Table 4 provides the $p$-values of three backtests (see Christoffersen (2003) for details) on the last $n - n_1 = 1389$ observations: the unconditional coverage (UC) test that the probability of violation is equal to the nominal level $\alpha$, the independence (IND) test that the violations are independent, and the conditional coverage (CC) test. The VaR estimation procedures clearly pass the backtests, except in one or two cases. In view of the sphericity test and these backtests, the spherical and FHS approaches are equivalent on these data. [10]

[9] This test is of Kolmogorov-Smirnov-type and relies on the characteristic function. In the aforementioned paper, it is shown on simulations that this test has good power properties for conditional distributions which are sufficiently far from sphericity. Applying the $\mathrm{KS}^{(2)}$ test of Section 6 with $L = 8$, and $B = 100$ bootstrap replications, we obtained an empirical $p$-value equal to 0.57.

[10] The aforementioned backtests do not account for the impact of the estimation errors. In a fully parametric dynamic framework, Pei (2010) studied the effect of estimation on backtests. Developing similar tests in our semi-parametric framework is beyond the scope of the present paper.

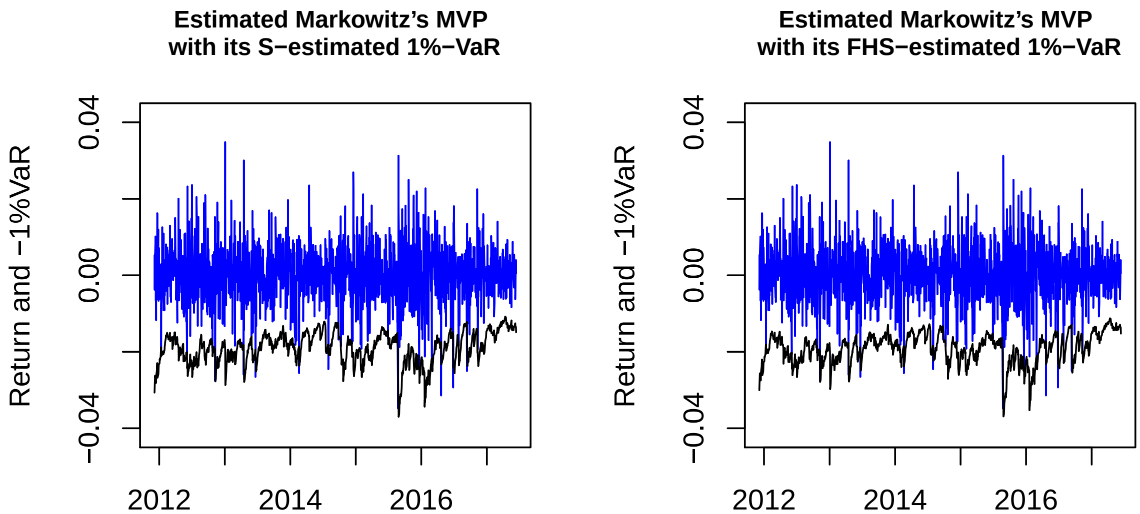


Figure 4: Returns of estimated optimal portfolios of 5 stocks and their estimated VaR's.

Table 4: $p$-values of three backtests for MVP and minimal VaR portfolios

| Method | Portfolio | $\alpha$ | % of Viol | UC | IND | CC |
|---|---|---|---|---|---|---|
| Spherical | MVP | 1% | $\frac{17}{1389} = 1.22\%$ | 0.418 | 0.202 | 0.319 |
| FHS | MVP | 1% | $\frac{16}{1389} = 1.15\%$ | 0.579 | 0.175 | 0.342 |
| FHS | Minimal 1%-VaR | 1% | $\frac{13}{1389} = 0.94\%$ | 0.808 | 0.108 | 0.267 |
| Spherical | MVP | 5% | $\frac{67}{1389} = 4.82\%$ | 0.762 | 0.053 | 0.147 |
| FHS | MVP | 5% | $\frac{79}{1389} = 5.69\%$ | 0.249 | 0.044 | 0.068 |
| FHS | Minimal 5%-VaR | 5% | $\frac{68}{1389} = 4.90\%$ | 0.858 | 0.020 | 0.067 |

# 6 Conclusion

This paper develops a unified theory for the inference of conditional VaRs of dynamic portfolios. The dynamics of the underlying vector process of returns is governed by a quite general stationary multivariate GARCH-type model. The portfolio is based on a combination of individual returns which can be time-varying. We showed that the sphericity assumption on the innovations distribution allows i) to define the concept of VaR parameter for which we provided an asymptotically Gaussian estimator; ii) to quantify the estimation risk via asymptotic CI's on the VaR parameter. Without the sphericity assumption, asymptotic results were also derived for the FHS estimator. For both approaches, with or without the sphericity assumption, we showed how to build asymptotic CIs for the conditional VaR and thus to visualize on the same graph both market and estimation risks. As far as the comparison between the two approaches is concerned, our results and experiments not reported here allow us to draw the following lessons, by distinguishing two different problems:

i) **Estimating the conditional VaR** by the spherical method is simpler and more accurate when sphericity holds. On the other hand, it may yield inconsistent VaR estimators when sphericity is in failure. The FHS method performs well in both cases and outperforms the first approach in the absence of sphericity.

ii) **Evaluating the asymptotic accuracy** of the conditional VaR estimators can be achieved using Theorems 3.1 and 4.1. Implementation of the latter asymptotic results is more involved but is worthwhile when sphericity is doubtful. An alternative bootstrap procedure can also be used when the asymptotic distribution is not available or is untractable. Conclusions drawn from our experiments are that the asymptotic and bootstrap approaches give similar results when both are available, the latter being obviously much more time consuming.

The practical implications of our results concern the derivation of reserves for financial positions. By neglecting the estimation risk, practitioners may erroneously believe that the risk is controlled at a given level. The problem is even more important in highly volatile periods, for which the accuracy of risk estimators tends to lower. Our results could clearly be extended to other risk measures, but we leave these extensions for future research.

# Appendix: Complementary results

## A Illustrations of the Bahadur representation A3

### A.1 For the Gaussian QML

Let us illustrate (3.2) in Assumption **A3** when $\boldsymbol{m}(\cdot) = 0$ and the criterion used to estimate $\boldsymbol{\theta}_0$ is the Gaussian QML. We have

$$\widehat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} n^{-1} \sum_{t=1}^{n} \widetilde{\ell}_t(\boldsymbol{\theta}) \tag{A.1}$$

where

$$\widetilde{\ell}_t(\boldsymbol{\theta}) = \boldsymbol{\epsilon}_t' \widetilde{\boldsymbol{H}}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\epsilon}_t + \log|\widetilde{\boldsymbol{H}}_t(\boldsymbol{\theta})|, \qquad \widetilde{\boldsymbol{H}}_t(\boldsymbol{\theta}) = \widetilde{\boldsymbol{\Sigma}}_t(\boldsymbol{\theta})\widetilde{\boldsymbol{\Sigma}}_t'(\boldsymbol{\theta})$$

and

$$\widetilde{\boldsymbol{\Sigma}}_t(\boldsymbol{\theta}_0) = \boldsymbol{\Sigma}(\boldsymbol{\epsilon}_{t-1}, \dots, \boldsymbol{\epsilon}_1, \widetilde{\boldsymbol{\epsilon}}_0, \widetilde{\boldsymbol{\epsilon}}_{-1}, \dots, \boldsymbol{\theta}_0),$$

where $\widetilde{\boldsymbol{\epsilon}}_{-i}$, for $i \geq 0$, denote arbitrary initial values. Under appropriate assumptions not discussed here, we have the following expansion

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \overset{o_P(1)}{=} \boldsymbol{J}^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial \ell_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}},$$

where

$$\boldsymbol{J} = E\left(-\frac{\partial^2 \ell_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right) \quad \text{and} \quad \ell_t(\boldsymbol{\theta}) = \boldsymbol{\epsilon}_t'\boldsymbol{H}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\epsilon}_t + \log|\boldsymbol{H}_t(\boldsymbol{\theta})|,$$

with

$$\boldsymbol{H}_t(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_t(\boldsymbol{\theta})\boldsymbol{\Sigma}_t'(\boldsymbol{\theta}), \quad \boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0) = \boldsymbol{\Sigma}(\boldsymbol{\epsilon}_{t-1}, \dots, ).$$

Moreover, for $j = 1, \dots, d$, we have, using the equality $\mathrm{Tr}(\boldsymbol{A}'\boldsymbol{B}) = \mathrm{vec}'(\boldsymbol{A})\mathrm{vec}(\boldsymbol{B})$,

$$\begin{aligned}
\frac{\partial \ell_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_j} &= \mathrm{Tr}\left\{(\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0))'(\boldsymbol{I}_m - \boldsymbol{\eta}_t\boldsymbol{\eta}_t')\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0)\frac{\partial \boldsymbol{H}_t(\boldsymbol{\theta}_0)}{\partial \theta_j}\right\} \\
&= \mathrm{vec}'\left\{\frac{\partial \boldsymbol{H}_t(\boldsymbol{\theta}_0)}{\partial \theta_j}\right\}\mathrm{vec}\left\{(\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0))'(\boldsymbol{I}_m - \boldsymbol{\eta}_t\boldsymbol{\eta}_t')\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0)\right\} \\
&= \mathrm{vec}'\left\{\frac{\partial \boldsymbol{H}_t(\boldsymbol{\theta}_0)}{\partial \theta_j}\right\}\left\{\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0) \otimes \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0)\right\}'\mathrm{vec}\left\{\boldsymbol{I}_m - \boldsymbol{\eta}_t\boldsymbol{\eta}_t'\right\}.
\end{aligned}$$

It follows that

$$\frac{\partial \ell_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} = \frac{\partial \mathrm{vec}'\boldsymbol{H}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\left\{\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0) \otimes \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0)\right\}'\mathrm{vec}\left\{\boldsymbol{I}_m - \boldsymbol{\eta}_t\boldsymbol{\eta}_t'\right\}.$$

Hence (3.2) holds with

$$\boldsymbol{\Delta}_{t-1} = -\boldsymbol{J}^{-1}\frac{\partial \mathrm{vec}'\boldsymbol{H}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\left\{\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0) \otimes \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0)\right\}'$$

and

$$\boldsymbol{V}(\boldsymbol{\eta}_t) = \mathrm{vec}\left\{\boldsymbol{I}_m - \boldsymbol{\eta}_t\boldsymbol{\eta}_t'\right\}.$$

### A.2 For the EbE estimator of generalized CCC models

Francq and Zakoian (2016) studied the asymptotic properties of the so-called Equation-by-Equation (EbE) estimation method. In this approach, instead of estimating a $m$-multivariate volatility model, $m$ univariate GARCH-type models are estimated EbE in the first step, and a correlation matrix is estimated in the second step. Let $\boldsymbol{m}(\cdot) = 0$, and assume

$$\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0) = \boldsymbol{D}_t\boldsymbol{R}^{1/2}$$

where $\boldsymbol{D}_t = \mathrm{diag}(\sigma_{1t}, \ldots, \sigma_{mt})$ and $\boldsymbol{R} = (R_{ij})$ is a constant correlation matrix. Suppose that that $\sigma_{kt}^2$ is parameterized by some parameter $\boldsymbol{\zeta}_0^{(k)}$, so that

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

where $\sigma_k$ is a positive function and $\eta_{kt}^*$ is the $k$-th component of $\boldsymbol{R}^{1/2}\boldsymbol{\eta}_t$ (see Francq and Zakoian (2016) for precise assumptions). Each volatility being allowed to depend on the past of all components of $\boldsymbol{\epsilon}_t$, the model can be called generalized CCC. The parameter $\boldsymbol{\theta} = \boldsymbol{\theta} := (\boldsymbol{\zeta}', \boldsymbol{\rho}')'$ here consists in the volatility parameters $\boldsymbol{\zeta} = (\zeta^{(1)'}, \ldots, \zeta^{(m)'})'$ and the correlation parameters

$$\boldsymbol{\rho} = (R_{21}, \ldots, R_{m1}, R_{32}, \ldots, R_{m2}, \ldots, R_{m,m-1})'.$$

The components of $\boldsymbol{\zeta}$ are estimated in a first step by the QML method applied to each volatility equation, while the correlation matrix is estimated by the sample autocorrelation. Equation (B.2) in Francq and Zakoian (2016) shows that (3.2) in Assumption **A3** holds for the EbE estimator of the generalized CCC model.

### A.3 For the VTE of the CCC model

Consider the CCC-GARCH$(p, q)$ model

$$\begin{cases} \boldsymbol{\epsilon}_t &= \boldsymbol{H}_t^{1/2}\boldsymbol{\eta}_t, \\ \boldsymbol{H}_t &= \boldsymbol{D}_t\boldsymbol{R}_0\boldsymbol{D}_t, \quad \boldsymbol{D}_t^2 = \mathrm{diag}(\underline{\boldsymbol{h}}_t), \\ \underline{\boldsymbol{h}}_t - \boldsymbol{h}_0 &= \sum_{i=1}^q \boldsymbol{A}_{0i}\left(\underline{\boldsymbol{\epsilon}}_{t-i} - \boldsymbol{h}_0\right) + \sum_{j=1}^p \boldsymbol{B}_{0j}\left(\underline{\boldsymbol{h}}_{t-j} - \boldsymbol{h}_0\right), \end{cases} \tag{A.3}$$

where $\underline{\boldsymbol{\epsilon}}_t = \left(\epsilon_{1t}^2, \cdots, \epsilon_{mt}^2\right)'$ and $\boldsymbol{R}_0$ is a correlation matrix. The matrices $\boldsymbol{A}_{0i}$ and $\boldsymbol{B}_{0j}$ are matrices of size $m \times m$ with positive coefficients and $\boldsymbol{h}_0$ is a vector of dimension $m$ such that

$$\left\{\boldsymbol{I}_m - \sum_{i=1}^r (\boldsymbol{A}_{0i} + \boldsymbol{B}_{0i})\right\}\boldsymbol{h}_0$$

has strictly positive coefficients (with $r = \max\{p, q\}$). The parameter vector is denoted $\boldsymbol{\theta} = (\boldsymbol{h}', \boldsymbol{\gamma}')'$, with

$$\boldsymbol{\gamma} = (\boldsymbol{\alpha}_1', \ldots, \boldsymbol{\alpha}_q', \boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_p', \boldsymbol{\rho}')',$$

where

$$\begin{aligned} \boldsymbol{\rho}' &= (\rho_{21}, \ldots, \rho_{m1}, \rho_{32}, \ldots, \rho_{m2}, \ldots, \rho_{m,m-1}) \in \mathbb{R}^{m(m-1)/2} \\ \boldsymbol{\alpha}_i &= \mathrm{vec}\,\boldsymbol{A}_i \in \mathbb{R}^{m^2}, \quad i = 1, \ldots, q, \end{aligned}$$

and

$$\boldsymbol{\beta}_j = \operatorname{vec} \boldsymbol{B}_j \in \mathbb{R}^{m^2}, \qquad j = 1, \dots, p.$$

Using initial values, for any $\boldsymbol{\gamma}$ belonging to some compact set $\Theta_{\boldsymbol{\gamma}}$, the $\widetilde{\boldsymbol{H}}_t$'s are recursively defined, for $t \geq 1$, by

$$\begin{cases} \widetilde{\boldsymbol{H}}_t &= \widetilde{\boldsymbol{D}}_t \boldsymbol{R} \widetilde{\boldsymbol{D}}_t, \quad \widetilde{\boldsymbol{D}}_t = \{\operatorname{diag}(\underline{\widetilde{\boldsymbol{h}}}_t)\}^{1/2}, \\ \underline{\widetilde{\boldsymbol{h}}}_t &= \underline{\widetilde{\boldsymbol{h}}}_t(\boldsymbol{\theta}) = \boldsymbol{h} + \sum_{i=1}^q \boldsymbol{A}_i \left(\underline{\boldsymbol{\epsilon}}_{t-i} - \boldsymbol{h}\right) + \sum_{j=1}^p \boldsymbol{B}_j \left(\underline{\widetilde{\boldsymbol{h}}}_{t-j} - \boldsymbol{h}\right). \end{cases}$$

The VTE of the parameter $\boldsymbol{h}_0$ is defined by the empirical mean

$$\widehat{\boldsymbol{h}}_n = \frac{1}{n} \sum_{t=1}^n \underline{\boldsymbol{\epsilon}}_t.$$

The VTE of the parameter $\boldsymbol{\gamma}_0$ is then defined by $\widehat{\boldsymbol{\gamma}}_n = \arg\min_{\boldsymbol{\gamma} \in \Theta_{\boldsymbol{\gamma}}} \widetilde{\mathcal{L}}_n(\boldsymbol{\gamma})$, where

$$\widetilde{\mathcal{L}}_n(\boldsymbol{\gamma}) = n^{-1} \sum_{t=1}^n \widetilde{\ell}_{t,n}$$

and

$$\widetilde{\ell}_{t,n} = \widetilde{\ell}_t(\widehat{\boldsymbol{h}}_n, \boldsymbol{\gamma}), \qquad \widetilde{\ell}_t = \widetilde{\ell}_t(\boldsymbol{h}, \boldsymbol{\gamma}) = \boldsymbol{\epsilon}_t' \widetilde{\boldsymbol{H}}_t^{-1} \boldsymbol{\epsilon}_t + \log |\widetilde{\boldsymbol{H}}_t|.$$

Letting $\widehat{\boldsymbol{\theta}}_n = (\widehat{\boldsymbol{h}}_n', \widehat{\boldsymbol{\gamma}}_n')'$, the VTE of $\boldsymbol{\theta}_0$, Francq, Horváth and Zakoïan (2015) showed that

$$\sqrt{n} \left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) = \boldsymbol{L}_n \boldsymbol{X}_n \tag{A.4}$$

where $\boldsymbol{L}_n$ converges in probability to some positive-definite matrix $\boldsymbol{L}$,

$$\boldsymbol{X}_n := \begin{pmatrix} \sqrt{n}\left(\widehat{\boldsymbol{h}}_n - \boldsymbol{h}_0\right) \\ \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\partial}{\partial \boldsymbol{\gamma}} \widetilde{\ell}_t(\boldsymbol{\theta}_0) \end{pmatrix} = \begin{pmatrix} \frac{\boldsymbol{C}}{\sqrt{n}} \sum_{t=1}^n (\boldsymbol{U}_t^2 - \boldsymbol{I}_m) \underline{\boldsymbol{h}}_t \\ \frac{1}{\sqrt{n}} \sum_{t=1}^n \boldsymbol{\Phi}_{t-1} \boldsymbol{V}_t \end{pmatrix} + o_P(1),$$

where $\boldsymbol{C}$ is a non-random matrix, $\boldsymbol{\Phi}_{t-1}$ is a matrix which is measurable with respect to the past, and

$$\boldsymbol{U}_t = \operatorname{diag}(\boldsymbol{R}_0^{1/2} \boldsymbol{\eta}_t), \quad \boldsymbol{V}_t = \operatorname{vec}(\boldsymbol{I}_m - \boldsymbol{R}_0^{-1/2} \boldsymbol{\eta}_t \boldsymbol{\eta}_t' \boldsymbol{R}_0^{1/2}).$$

It can be noted that

$$(\boldsymbol{U}_t^2 - \boldsymbol{I}_m) \underline{\boldsymbol{h}}_t = \boldsymbol{D}_t^2 \underline{\boldsymbol{\eta}}_t^*,$$

where

$$\underline{\boldsymbol{\eta}}_t^* = \left(\eta_{1t}^{*2} - 1, \cdots, \eta_{mt}^{*2} - 1\right)'$$

and

$$\boldsymbol{\eta}_t^* = (\eta_{1t}^*, \cdots, \eta_{mt}^*)' = \boldsymbol{R}_0^{1/2}\boldsymbol{\eta}_t.$$

Note that $E\underline{\boldsymbol{\eta}}_t^* = \mathbf{0}$.

Thus, (3.2) in Assumption **A3** holds for the VTE of the CCC model with, in particular,

$$\boldsymbol{V}(\boldsymbol{\eta}_t) = \left(\underline{\boldsymbol{\eta}}_t^{*\prime}, \boldsymbol{V}_t'\right)'.$$

# B Proofs

## B.1 Proof of Theorem 3.1

Note that

$$\xi_{n,1-2\alpha} \quad = \quad \arg\min_{z\in\mathbb{R}} \frac{1}{n}\sum_{t=1}^{n}\sum_{k=1}^{m}\rho_{1-2\alpha}(|\widehat{\eta}_{kt}| - z),$$

where $\rho_{1-2\alpha}(u) = u(1-2\alpha - \mathbf{1}_{\{u\le 0\}})$. Thus

$$\sqrt{n}(\xi_{n,1-2\alpha} - \xi_{1-2\alpha}) = \arg\min_{z\in\mathbb{R}} Q_n(z)$$

where

$$Q_n(z) \quad = \quad \sum_{k=1}^{m}\sum_{t=1}^{n}\left\{\rho_{1-2\alpha}\left(|\widehat{\eta}_{kt}| - \xi_{1-2\alpha} - \frac{z}{\sqrt{n}}\right) - \rho_{1-2\alpha}(|\eta_{kt}| - \xi_{1-2\alpha})\right\}.$$

Let $\boldsymbol{e}_k$ denote the $k$-th column of the $m\times m$ identity matrix $\boldsymbol{I}_m$. Let $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)$. Let $\boldsymbol{\eta}_t(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\{\boldsymbol{y}_t - \boldsymbol{m}_t(\boldsymbol{\theta})\} = (\eta_{1t}(\boldsymbol{\theta}), \ldots, \eta_{mt}(\boldsymbol{\theta}))'$. We have, for $j = 1, \ldots, d$,

$$\begin{aligned}
\frac{\partial\eta_{kt}}{\partial\theta_j}(\boldsymbol{\theta}_0) &= -\boldsymbol{e}_k'\boldsymbol{\Sigma}_t^{-1}\frac{\partial\boldsymbol{m}_t}{\partial\theta_j} - \boldsymbol{e}_k'\boldsymbol{\Sigma}_t^{-1}\frac{\partial\boldsymbol{\Sigma}_t}{\partial\theta_j}\boldsymbol{\Sigma}_t^{-1}\{\boldsymbol{y}_t - \boldsymbol{m}_t(\boldsymbol{\theta}_0)\} \\
&= -\boldsymbol{e}_k'\boldsymbol{\Sigma}_t^{-1}\frac{\partial\boldsymbol{m}_t}{\partial\theta_j} + \mathrm{Tr}\left\{-\boldsymbol{\eta}_t\boldsymbol{e}_k'\boldsymbol{\Sigma}_t^{-1}\frac{\partial\boldsymbol{\Sigma}_t}{\partial\theta_j}\right\} \\
&= -\boldsymbol{e}_k'\boldsymbol{\Sigma}_t^{-1}\frac{\partial\boldsymbol{m}_t}{\partial\theta_j} - \sum_{\ell=1}^{m}\eta_{\ell t}\boldsymbol{e}_k'\boldsymbol{\Sigma}_t^{-1}\left\{\frac{\partial}{\partial\theta_j}\boldsymbol{\Sigma}_{\cdot\ell,t}\right\},
\end{aligned}$$

where $\boldsymbol{\Sigma}_{\cdot\ell,t}$ is the $\ell$-th column of $\boldsymbol{\Sigma}_t$. Let

$$\boldsymbol{\Omega}_{kt}^* = (I_m\otimes\boldsymbol{e}_k'\boldsymbol{\Sigma}_t^{-1})\frac{\partial}{\partial\boldsymbol{\theta}'}\{\mathrm{vec}(\boldsymbol{\Sigma}_t)\}, \quad \boldsymbol{C}_{kt} = \mathrm{vec}\left\{\boldsymbol{e}_k'\boldsymbol{\Sigma}_t^{-1}\frac{\partial\boldsymbol{m}_t}{\partial\boldsymbol{\theta}'}\right\}, \quad \boldsymbol{M}_{kt}' = \boldsymbol{C}_{kt}' + \boldsymbol{\eta}_t'\boldsymbol{\Omega}_{kt}^*.$$

A Taylor expansion of $\eta_{kt}(\boldsymbol{\theta})$ around $\boldsymbol{\theta}_0$ thus yields,

$$\begin{aligned}
\widehat{\eta}_{kt} &= \eta_{kt} - \sum_{j=1}^{d}\left(\boldsymbol{e}_k'\boldsymbol{\Sigma}_t^{-1}\frac{\partial\boldsymbol{m}_t}{\partial\theta_j} + \sum_{\ell=1}^{m}\eta_{\ell t}\boldsymbol{e}_k'\boldsymbol{\Sigma}_t^{-1}\left\{\frac{\partial}{\partial\theta_j}\boldsymbol{\Sigma}_{\cdot\ell,t}\right\}\right)(\widehat{\theta}_{nj} - \theta_{0j}) + o_P(n^{-1/2}) \\
&= \eta_{kt} - \boldsymbol{M}_{kt}'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o_P(n^{-1/2}). \qquad \text{(B.1)}
\end{aligned}$$

Note that for any sequence $(b_n)$ tending to zero and any real number $a$, we have, for $n$ large enough, $|a - b_n| = |a| - ub_n$ where $u = 1$ if $a > 0$ or if $a = 0$ and $b_n < 0$, and $u = -1$ otherwise. Thus

$$|\widehat{\eta}_{kt}| = \left|\eta_{kt} - \boldsymbol{M}'_{kt}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\right| + o_P(n^{-1/2}) = |\eta_{kt}| - u_{kt}\boldsymbol{M}'_{kt}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o_P(n^{-1/2}),$$

where $u_{kt} = \pm 1$, the sign of $u_{kt}$ being equal to that of $\eta_{kt}$ when $\eta_{kt} \neq 0$, and to the sign of $-\boldsymbol{M}'_{kt}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ when $\eta_{kt} = 0$. Using the identity

$$\rho_{1-2\alpha}(u - v) - \rho_{1-2\alpha}(u) = -v(1 - 2\alpha - \mathbf{1}_{\{u<0\}}) + \int_0^v \left\{\mathbf{1}_{\{u\leq s\}} - \mathbf{1}_{\{u<0\}}\right\} ds$$

for $u \neq 0$ (see Equation (A.3) in Koenker and Xiao, 2006), we thus have

$$Q_n(z) = \sum_{k=1}^{m} zX_{n,k} + Y_{n,k} + I_{n,k}(z) + J_{n,k}(z),$$

where

$$X_{n,k} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\mathbf{1}_{\{|\eta_{kt}|<\xi_{1-2\alpha}\}} - 1 + 2\alpha),$$

$$Y_{n,k} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} R_{t,n,k}(\mathbf{1}_{\{|\eta_{kt}|<\xi_{1-2\alpha}\}} - 1 + 2\alpha),$$

$$I_{n,k}(z) = \sum_{t=1}^{n}\int_0^{z/\sqrt{n}}(\mathbf{1}_{\{|\eta_{kt}|\leq\xi_{1-2\alpha}+s\}} - \mathbf{1}_{\{|\eta_{kt}|<\xi_{1-2\alpha}\}})ds,$$

$$J_{n,k}(z) = \sum_{t=1}^{n}\int_{z/\sqrt{n}}^{(z+R_{t,n,k})/\sqrt{n}}(\mathbf{1}_{\{|\eta_{kt}|\leq\xi_{1-2\alpha}+s\}} - \mathbf{1}_{\{|\eta_{kt}|<\xi_{1-2\alpha}\}})ds,$$

with $R_{t,n,k} \stackrel{o_P(1)}{=} u_{kt}\boldsymbol{M}'_{kt}\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$. We have $I_{n,k}(z) \to \frac{z^2}{2}f(\xi_{1-2\alpha})$ in probability as $n \to \infty$ (see Appendix B.2). Moreover, by the change of variable $u = s - z/\sqrt{n}$, we have $J_{n,k}(z) = J^{(1)}_{n,k}(z) + J^{(2)}_{n,k}(z)$ where

$$J^{(1)}_{n,k}(z) = \sum_{t=1}^{n}\int_0^{R_{t,n,k}/\sqrt{n}}\left(\mathbf{1}_{\{|\eta_{kt}|-\xi_{1-2\alpha}-z/\sqrt{n}\leq u\}} - \mathbf{1}_{\{|\eta_{kt}|-\xi_{1-2\alpha}-z/\sqrt{n}<0\}}\right) du,$$

$$J^{(2)}_{n,k}(z) = \sum_{t=1}^{n}\int_0^{R_{t,n,k}/\sqrt{n}}\left(\mathbf{1}_{\{|\eta_{kt}|-\xi_{1-2\alpha}-z/\sqrt{n}<0\}} - \mathbf{1}_{\{|\eta_{kt}|-\xi_{1-2\alpha}<0\}}\right) du.$$

Let $\mathbf{1}^*_{\{X\in(a,b)\}} = \mathbf{1}_{\{X<b\}} - \mathbf{1}_{\{X<a\}}$ for any real numbers $a, b$ and any real random variable $X$. We have

$$J^{(2)}_{n,k}(z) = \sum_{t=1}^{n}\left\{u_{kt}\boldsymbol{M}'_{kt}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o_P(n^{-1/2})\right\}\mathbf{1}^*_{\{|\eta_{kt}|-\xi_{1-2\alpha}\in(0,z/\sqrt{n})\}}$$

$$\stackrel{o_P(1)}{=} \left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} u_{kt}\mathbf{1}^*_{\{|\eta_{kt}|-\xi_{1-2\alpha}\in(0,z/\sqrt{n})\}}\boldsymbol{M}'_{kt}\right)\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0).$$

Note that, for $z>0$,

$$E(u_{kt}\mathbf{1}^*_{\{|\eta_{kt}|-\xi_{1-2\alpha}\in(0,z/\sqrt{n})\}}) \quad = \quad E(\mathbf{1}_{\{\eta_{kt}-\xi_{1-2\alpha}\in(0,z/\sqrt{n})\}}) - E(\mathbf{1}_{\{-\eta_{kt}-\xi_{1-2\alpha}\in(0,z/\sqrt{n})\}}) = 0,$$

in view of the symmetry of the distribution of $\eta_{kt}$ under the sphericity assumption **A2**. The same equality holds for $z \leq 0$. Now, for $z>0$ and $\ell \neq k$,

$$E(u_{kt}\eta_{\ell t}\mathbf{1}^*_{\{|\eta_{kt}|-\xi_{1-2\alpha}\in(0,z/\sqrt{n})\}}) \quad = \quad E(\eta_{\ell t}\mathbf{1}_{\{\eta_{kt}-\xi_{1-2\alpha}\in(0,z/\sqrt{n})\}}) - E(\eta_{\ell t}\mathbf{1}_{\{-\eta_{kt}-\xi_{1-2\alpha}\in(0,z/\sqrt{n})\}}) = 0,$$

because $(\eta_{\ell t},\eta_{kt})$ and $(\eta_{\ell t},-\eta_{kt})$ have the same distribution under **A2**. For $k=\ell$ we have

$$E(|\eta_{kt}|\mathbf{1}^*_{\{|\eta_{kt}|-\xi_{1-2\alpha}\in(0,z/\sqrt{n})\}}) \quad = \quad \xi_{1-2\alpha}f(\xi_{1-2\alpha})\frac{z}{\sqrt{n}} + o(1/\sqrt{n}).$$

The same equalities hold for $z \leq 0$. Thus, we have

$$E\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} u_{kt}\mathbf{1}^*_{\{|\eta_{kt}|-\xi_{1-2\alpha}\in(0,z/\sqrt{n})\}}\boldsymbol{M}'_{kt}\right) \quad \overset{o_P(1)}{=} \quad z\xi_{1-2\alpha}f(\xi_{1-2\alpha})\boldsymbol{e}'_k E\left(\boldsymbol{\Sigma}_t^{-1}\left\{\frac{\partial}{\partial\boldsymbol{\theta}'}\boldsymbol{\Sigma}_{\cdot k,t}\right\}\right).$$

Similar arguments show that

$$\begin{aligned}
&\operatorname{Var}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} u_{kt}\mathbf{1}^*_{\{|\eta_{kt}|-\xi_{1-2\alpha}\in(0,z/\sqrt{n})\}}\boldsymbol{C}'_{kt}\right)\\
= \;& \frac{1}{n}\sum_{t=1}^{n} E(\mathbf{1}^*_{\{|\eta_{kt}|-\xi_{1-2\alpha}\in(0,z/\sqrt{n})\}})E(\boldsymbol{C}'_{kt}\boldsymbol{C}_{kt}) = o(1),\\
&\operatorname{Var}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} u_{kt}\mathbf{1}^*_{\{|\eta_{kt}|-\xi_{1-2\alpha}\in(0,z/\sqrt{n})\}}\boldsymbol{\eta}'_t\boldsymbol{\Omega}^*_{kt}\right)\\
= \;& \frac{1}{n}\sum_{t=1}^{n} \operatorname{Var}\left(u_{kt}\mathbf{1}^*_{\{|\eta_{kt}|-\xi_{1-2\alpha}\in(0,z/\sqrt{n})\}}\boldsymbol{\eta}'_t\boldsymbol{\Omega}^*_{kt}\right) = o(1).
\end{aligned}$$

It follows that

$$J^{(2)}_{n,k}(z) \quad \overset{o_P(1)}{=} \quad z\xi_{1-2\alpha}f(\xi_{1-2\alpha})\boldsymbol{e}'_k E\left(\boldsymbol{\Sigma}_t^{-1}\left\{\frac{\partial}{\partial\boldsymbol{\theta}'}\boldsymbol{\Sigma}_{\cdot k,t}\right\}\right)\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0),$$

and

$$\sum_{k=1}^{m} J^{(2)}_{n,k}(z) \overset{o_P(1)}{=} z\xi_{1-2\alpha}f(\xi_{1-2\alpha})\sum_{k=1}^{m}\boldsymbol{e}'_k E\left(\boldsymbol{\Sigma}_t^{-1}\left\{\frac{\partial}{\partial\boldsymbol{\theta}'}\boldsymbol{\Sigma}_{\cdot k,t}\right\}\right)\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0).$$

Moreover,

$$\begin{aligned}
\sum_{k=1}^{m}\boldsymbol{e}'_k E\left(\boldsymbol{\Sigma}_t^{-1}\left\{\frac{\partial}{\partial\boldsymbol{\theta}'}\boldsymbol{\Sigma}_{\cdot k,t}\right\}\right) &= \sum_{k=1}^{m} E\left[\left(\boldsymbol{e}_k\otimes\left\{\frac{\partial}{\partial\boldsymbol{\theta}'}\boldsymbol{\Sigma}_{\cdot k,t}\right\}\right)'\operatorname{vec}\left(\boldsymbol{\Sigma}_t^{-1}\right)\right]'\\
&= E\left[\left\{\operatorname{vec}\left(\boldsymbol{\Sigma}_t^{-1}\right)\right\}'\sum_{k=1}^{m}\left(\boldsymbol{e}_k\otimes\left\{\frac{\partial}{\partial\boldsymbol{\theta}'}\boldsymbol{\Sigma}_{\cdot k,t}\right\}\right)\right]\\
&= E\left[\left\{\operatorname{vec}\left(\boldsymbol{\Sigma}_t^{-1}\right)\right\}'\left\{\frac{\partial}{\partial\boldsymbol{\theta}'}\operatorname{vec}\left(\boldsymbol{\Sigma}_t\right)\right\}\right] = \boldsymbol{\Omega}'.
\end{aligned}$$

As in Francq and Zakoian (2015), it can be shown that $\sum_{k=1}^m J_{n,k}^{(1)}(z)$ converges in distribution to a variable which does not depend on $z$. Therefore,

$$\sum_{k=1}^m J_{n,k}(z) \overset{o_P(1)}{=} z\xi_{1-2\alpha} f(\xi_{1-2\alpha}) \boldsymbol{\Omega}' \sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + A$$

where $A$ is a random variable which is independent of $z$. By the arguments given in Francq and Zakoïan (2015), we can conclude that

$$\sqrt{n}(\xi_{n,1-2\alpha} - \xi_{1-2\alpha}) \quad \overset{o_P(1)}{=} \quad -\frac{\xi_{1-2\alpha}}{m} \boldsymbol{\Omega}' \sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) - \frac{1}{f(\xi_{1-2\alpha})} \frac{1}{m\sqrt{n}} \sum_{t=1}^n N_t. \qquad \text{(B.2)}$$

In view of **A3** we have

$$\mathrm{Cov}_{as}\left(\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0), \frac{1}{m\sqrt{n}} \sum_{t=1}^n N_t\right) \quad = \quad \frac{1}{m} \boldsymbol{\Lambda} \boldsymbol{W}_\alpha,$$

and thus,

$$\mathrm{Var}_{as}\{\sqrt{n}(\xi_{n,1-2\alpha} - \xi_{1-2\alpha})\} = \frac{1}{m^2}\left\{\xi_{1-2\alpha}^2 \boldsymbol{\Omega}' \boldsymbol{\Psi} \boldsymbol{\Omega} + \frac{2\xi_{1-2\alpha}}{f(\xi_{1-2\alpha})} \boldsymbol{\Omega}' \boldsymbol{\Lambda} \boldsymbol{W}_\alpha + \frac{\gamma_\alpha}{f^2(\xi_{1-2\alpha})}\right\},$$

$$\mathrm{Cov}_{as}\left(\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0), \sqrt{n}(\xi_{n,1-2\alpha} - \xi_{1-2\alpha})\right) = \frac{-1}{m}\left\{\xi_{1-2\alpha} \boldsymbol{\Psi} \boldsymbol{\Omega} + \frac{1}{f(\xi_{1-2\alpha})} \boldsymbol{\Lambda} \boldsymbol{W}_\alpha\right\}.$$

The convergence in distribution (3.3) follows by the Central Limit Theorem of Billingsley (1961) for ergodic, stationary and square integrable martingale differences, applied to the sequence $\begin{pmatrix} \boldsymbol{\Delta}_{t-1} \boldsymbol{V}(\boldsymbol{\eta}_t) \\ N_t \end{pmatrix}$. □

## B.2 Proof that $I_{n,k}(z) \to \frac{z^2}{2} f(\xi_{1-2\alpha})$ in probability as $n \to \infty$

For ease of notation, we omit the index $k$. Write $\eta_t$ instead of $\eta_{kt}$ and $I_n(z)$ instead of $I_{n,k}(z)$. Note that

$$\begin{aligned} I_n(z) &= \sum_{t=1}^n \mathbf{1}_{\{|\eta_t| > \xi_{1-2\alpha}\}} \int_0^{z/\sqrt{n}} \mathbf{1}_{\{|\eta_t| \le \xi_{1-2\alpha} + s\}} ds \\ &= \sum_{t=1}^n \mathbf{1}_{\{|\eta_t| > \xi_{1-2\alpha}\}} \mathbf{1}_{\{|\eta_t| - \xi_{1-2\alpha} \le z/\sqrt{n}\}} \int_{|\eta_t| - \xi_{1-2\alpha}}^{z/\sqrt{n}} ds \\ &= \sum_{t=1}^n \left(\frac{z}{\sqrt{n}} - X_t\right) \mathbf{1}_{0 < X_t < z/\sqrt{n}}, \qquad X_t = |\eta_t| - \xi_{1-2\alpha}. \end{aligned}$$

Let

$$W_{n,t} = \left(\frac{z}{\sqrt{n}} - X_t\right) \mathbf{1}_{0<X_t<z/\sqrt{n}} - E\left\{\left(\frac{z}{\sqrt{n}} - X_t\right) \mathbf{1}_{0<X_t<z/\sqrt{n}}\right\}.$$

We have, for any integer $p > 0$,

$$\begin{aligned} E\left\{\left(\frac{z}{\sqrt{n}} - X_t\right)^p \mathbf{1}_{0<X_t<z/\sqrt{n}}\right\} &= \int_0^{z/\sqrt{n}} \left(\frac{z}{\sqrt{n}} - x\right)^p f(x + \xi_{1-2\alpha})dx \\ &= n^{-(p+1)/2} \int_0^z (z-u)^p f\{(u + \xi_{1-2\alpha})/\sqrt{n}\}du \\ &\sim \frac{z^{p+1}}{p+1} f(\xi_{1-2\alpha}) n^{-(p+1)/2}, \quad \text{as } n \to \infty. \end{aligned}$$

Thus, by Markov's inequality, for any $\epsilon > 0$,

$$\begin{aligned} P\left(\left|\sum_{t=1}^n W_{n,t}\right| > \epsilon\right) &\leq \frac{E\left(\sum_{t=1}^n W_{n,t}\right)^2}{\epsilon^2} \\ &= \frac{\sum_{t=1}^n EW_{n,t}^2}{\epsilon^2} \sim \frac{z^3}{3\epsilon^2} f(\xi_{1-2\alpha}) n^{-1/2} = o(1), \quad \text{as } n \to \infty. \end{aligned}$$

It follows that $\sum_{t=1}^n W_{n,t} \to 0$, in probability as $n \to \infty$. Thus,

$$I_n(z) \sim nE\left\{\left(\frac{z}{\sqrt{n}} - X_t\right) \mathbf{1}_{0<X_t<z/\sqrt{n}}\right\} \sim \frac{z^2}{2} f(\xi_{1-2\alpha}),$$

in probability as $n \to \infty$. □

## B.3 Proof of Corollary 3.1

The asymptotic normality follows from Theorem 3.1 and the following Taylor expansion of $G$ around $(\boldsymbol{\theta}_0, \xi_{1-2\alpha})$

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n^* - \boldsymbol{\theta}_0^*\right) = \left[\frac{\partial G(\boldsymbol{\theta}, \xi)}{\partial(\boldsymbol{\theta}', \xi)}\right]_{(\boldsymbol{\theta}_0, \xi_{1-2\alpha})} \begin{pmatrix} \sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \\ \sqrt{n}(\xi_{n,1-2\alpha} - \xi_{1-2\alpha}) \end{pmatrix} + o_P(1).$$

□

## B.4 Proof of Theorem 4.1

Noting that $\xi_{n,\alpha}(\widehat{\boldsymbol{\theta}}_n) = \arg\min_{z\in\mathbb{R}} \frac{1}{n}\sum_{t=1}^n \rho_\alpha\{b(\widehat{\boldsymbol{\theta}}_n) + \boldsymbol{c}'(\widehat{\boldsymbol{\theta}}_n)\widehat{\boldsymbol{\eta}}_t - z\}$, we have

$$\sqrt{n}\{\xi_{n,\alpha}(\widehat{\boldsymbol{\theta}}_n) - \xi_\alpha(\boldsymbol{\theta}_0)\} = \arg\min_{z\in\mathbb{R}} \mathcal{O}_n(z)$$

where

$$\mathcal{O}_n(z) = \sum_{t=1}^{n} \left\{ \rho_\alpha \left( b(\widehat{\boldsymbol{\theta}}_n) + \boldsymbol{c}'(\widehat{\boldsymbol{\theta}}_n)\widehat{\boldsymbol{\eta}}_t - \xi_\alpha(\boldsymbol{\theta}_0) - \frac{z}{\sqrt{n}} \right) - \rho_\alpha\{b(\boldsymbol{\theta}_0) + \boldsymbol{c}'(\boldsymbol{\theta}_0)\boldsymbol{\eta}_t - \xi_\alpha(\boldsymbol{\theta}_0)\} \right\}.$$

It follows from (B.1) that

$$\widehat{\boldsymbol{\eta}}_t = \boldsymbol{\eta}_t - \boldsymbol{C}_t(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) - (I_m \otimes \boldsymbol{\eta}_t')\boldsymbol{\Omega}_t^*(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o_P(n^{-1/2}).$$

Noting that $\boldsymbol{c}(\boldsymbol{\theta}_0)'(\boldsymbol{I}_m \otimes \boldsymbol{\eta}_t)'\boldsymbol{\Omega}_t^* = \sum_{j=1}^m c_j(\boldsymbol{\theta}_0)\boldsymbol{\eta}_t'\boldsymbol{\Omega}_{jt}^* = \boldsymbol{\eta}_t'\{\boldsymbol{c}'(\boldsymbol{\theta}_0) \otimes \boldsymbol{I}_m\}\boldsymbol{\Omega}_t^*$, a Taylor expansion around $\boldsymbol{\theta}_0$ thus yields

$$\begin{aligned}
& b(\widehat{\boldsymbol{\theta}}_n) + \boldsymbol{c}'(\widehat{\boldsymbol{\theta}}_n)\widehat{\boldsymbol{\eta}}_t - \{b(\boldsymbol{\theta}_0) + \boldsymbol{c}'(\boldsymbol{\theta}_0)\boldsymbol{\eta}_t\} \\
=& \left\{ \frac{\partial b}{\partial \boldsymbol{\theta}'}(\boldsymbol{\theta}_0) - \boldsymbol{c}'(\boldsymbol{\theta}_0)\boldsymbol{C}_t \right\} (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + \boldsymbol{\eta}_t' \left\{ \frac{\partial \boldsymbol{c}}{\partial \boldsymbol{\theta}'}(\boldsymbol{\theta}_0) - (\boldsymbol{c}'(\boldsymbol{\theta}_0) \otimes \boldsymbol{I}_m)\boldsymbol{\Omega}_t^* \right\} (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \\
=& \boldsymbol{n}_t'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o_P(n^{-1/2}),
\end{aligned}$$

where $\boldsymbol{n}_t'$ is the row vector

$$\boldsymbol{n}_t' = \left[ \frac{\partial b}{\partial \boldsymbol{\theta}'}(\boldsymbol{\theta}_0) - \boldsymbol{c}'(\boldsymbol{\theta}_0)\boldsymbol{C}_t \quad \boldsymbol{\eta}_t' \left\{ \frac{\partial \boldsymbol{c}}{\partial \boldsymbol{\theta}'}(\boldsymbol{\theta}_0) - (\boldsymbol{c}'(\boldsymbol{\theta}_0) \otimes \boldsymbol{I}_m)\boldsymbol{\Omega}_t^* \right\} \right] := \left[ \boldsymbol{c}_t' \;\; \boldsymbol{\eta}_t'\boldsymbol{F}_t \right].$$

Proceeding as in the proof of Theorem 3.1, we find that

$$\mathcal{O}_n(z) = zX_n + Y_n + I_n(z) + J_n(z), \quad \text{where}$$

$$\begin{aligned}
X_n &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\mathbf{1}_{\{b(\boldsymbol{\theta}_0) + \boldsymbol{c}'(\boldsymbol{\theta}_0)\boldsymbol{\eta}_t < \xi_\alpha(\boldsymbol{\theta}_0)\}} - \alpha), \\
Y_n &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} S_{t,n}(\mathbf{1}_{\{b(\boldsymbol{\theta}_0) + \boldsymbol{c}'(\boldsymbol{\theta}_0)\boldsymbol{\eta}_t < \xi_\alpha(\boldsymbol{\theta}_0)\}} - \alpha), \\
I_n(z) &= \sum_{t=1}^{n} \int_0^{z/\sqrt{n}} (\mathbf{1}_{\{b(\boldsymbol{\theta}_0) + \boldsymbol{c}'(\boldsymbol{\theta}_0)\boldsymbol{\eta}_t \le \xi_\alpha(\boldsymbol{\theta}_0) + s\}} - \mathbf{1}_{\{b(\boldsymbol{\theta}_0) + \boldsymbol{c}'(\boldsymbol{\theta}_0)\boldsymbol{\eta}_t < \xi_\alpha(\boldsymbol{\theta}_0)\}})ds, \\
J_n(z) &= \sum_{t=1}^{n} \int_{z/\sqrt{n}}^{(z+S_{t,n})/\sqrt{n}} (\mathbf{1}_{\{b(\boldsymbol{\theta}_0) + \boldsymbol{c}'(\boldsymbol{\theta}_0)\boldsymbol{\eta}_t \le \xi_\alpha(\boldsymbol{\theta}_0) + s\}} - \mathbf{1}_{\{b(\boldsymbol{\theta}_0) + \boldsymbol{c}'(\boldsymbol{\theta}_0)\boldsymbol{\eta}_t < \xi_\alpha(\boldsymbol{\theta}_0)\}})ds,
\end{aligned}$$

with $S_{t,n} \overset{o_P(1)}{=} -\boldsymbol{n}_t'\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$. By arguments already used, we have $I_n(z) \to \frac{z^2}{2} f_{\boldsymbol{c}}\{x_0\}$ in probability as $n \to \infty$, and $J_n(z) = J_n^{(1)}(z) + J_n^{(2)}(z)$ where $J_n^{(1)}(z)$ converges in distribution to a variable which does not depend on $z$ and

$$\begin{aligned}
J_n^{(2)}(z) &= \sum_{t=1}^{n} \int_0^{S_{t,n}/\sqrt{n}} \left( \mathbf{1}_{\{-x_0+\boldsymbol{c}'(\boldsymbol{\theta}_0)\boldsymbol{\eta}_t - z/\sqrt{n}<0\}} - \mathbf{1}_{\{-x_0+\boldsymbol{c}'(\boldsymbol{\theta}_0)\boldsymbol{\eta}_t<0\}} \right) du \\
&= \sum_{t=1}^{n} \left\{ -\boldsymbol{n}_t'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o_P(n^{-1/2}) \right\} \mathbf{1}^*_{\{-x_0+\boldsymbol{c}'(\boldsymbol{\theta}_0)\boldsymbol{\eta}_t \in (0, z/\sqrt{n})\}} \\
&\overset{o_P(1)}{=} \left( \frac{-1}{\sqrt{n}} \sum_{t=1}^{n} \mathbf{1}^*_{\{-x_0+\boldsymbol{c}'(\boldsymbol{\theta}_0)\boldsymbol{\eta}_t \in (0, z/\sqrt{n})\}} \boldsymbol{n}_t' \right) \sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0).
\end{aligned}$$

First suppose for $z > 0$. We have, Now, in view of the independence between $\boldsymbol{\eta}_t$ and $\boldsymbol{F}_t$, we have, for $z > 0$,

$$\begin{aligned}
& E\left( \boldsymbol{\eta}_t' \mathbf{1}^*_{\{-x_0+\boldsymbol{c}'(\boldsymbol{\theta}_0)\boldsymbol{\eta}_t \in (0, z/\sqrt{n})\}} \boldsymbol{F}_t \right) \\
= \; & E\left\{ \boldsymbol{\eta}_t' \boldsymbol{F}_t \mid -x_0 + \boldsymbol{c}'(\boldsymbol{\theta}_0)\boldsymbol{\eta}_t \in \left(0, \frac{z}{\sqrt{n}}\right) \right\} \left\{ \frac{z}{\sqrt{n}} f_{\boldsymbol{c}}(x_0) + o\left(\frac{1}{\sqrt{n}}\right) \right\} \\
= \; & \frac{z}{\sqrt{n}} f_{\boldsymbol{c}}(x_0) \boldsymbol{d}_\alpha' E(\boldsymbol{F}_t) + o\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

Similar computations show that the last equality continues to hold for $z < 0$. Similarly,

$$E\left( \mathbf{1}^*_{\{-x_0+\boldsymbol{c}'(\boldsymbol{\theta}_0)\boldsymbol{\eta}_t \in (0, z/\sqrt{n})\}} \boldsymbol{c}_t' \right) = \frac{z}{\sqrt{n}} f_{\boldsymbol{c}}(x_0) E(\boldsymbol{c}_t') + o\left(\frac{1}{\sqrt{n}}\right).$$

By arguments already used, it follows that

$$J_n^{(2)}(z) \overset{o_P(1)}{=} z f_{\boldsymbol{c}}(x_0) \left[ -E(\boldsymbol{c}_t') \quad - \boldsymbol{d}_\alpha' E(\boldsymbol{F}_t) \right] \sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = z f_{\boldsymbol{c}}(x_0) \boldsymbol{w}' \sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0).$$

Finally,

$$\mathcal{O}_n(z) = \frac{z^2}{2} f_{\boldsymbol{c}}(x_0) + z \left\{ X_n + f_{\boldsymbol{c}}(x_0) \boldsymbol{w}' \sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \right\} + O_P(1).$$

We conclude that, similarly to (B.2),

$$\begin{aligned}
\sqrt{n}\{\xi_{n,\alpha}(\widehat{\boldsymbol{\theta}}_n) - \xi_\alpha(\boldsymbol{\theta}_0)\} \overset{o_P(1)}{=} \; & -\boldsymbol{w}' \sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \\
& - \frac{1}{f_{\boldsymbol{c}}(x_0)} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\mathbf{1}_{\{b(\boldsymbol{\theta}_0)+\boldsymbol{c}'(\boldsymbol{\theta}_0)\boldsymbol{\eta}_t < \xi_\alpha(\boldsymbol{\theta}_0)\}} - \alpha).
\end{aligned}$$

The convergence in distribution follows. □

# C DCC-GARCH dynamic portfolios

In this appendix, we consider the case where the return vector $\boldsymbol{\epsilon}_t$ follows a DCC GARCH model of the form $\boldsymbol{\epsilon}_t = \boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)\boldsymbol{\eta}_t$ with $\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0) = \boldsymbol{D}_t \boldsymbol{R}_t^{1/2}$. The diagonal matrix $\boldsymbol{D}_t = \mathrm{diag}(\sigma_{1t}, \ldots, \sigma_{mt})$ is assumed to satisfy the GARCH(1,1) equation

$$\boldsymbol{h}_t = \boldsymbol{\omega}_0 + \boldsymbol{A}_0 \underline{\boldsymbol{\epsilon}}_{t-1} + \boldsymbol{B}_0 \boldsymbol{h}_{t-1} \tag{C.1}$$

where $\boldsymbol{h}_t = \left(\sigma_{1t}^2, \cdots, \sigma_{mt}^2\right)'$, $\underline{\boldsymbol{\epsilon}}_t = \left(\epsilon_{1t}^2, \cdots, \epsilon_{mt}^2\right)'$, $\boldsymbol{A}_0$ and $\boldsymbol{B}_0$ are $m \times m$ matrices with positive coefficients, $\boldsymbol{\omega}_0$ is a vector of strictly positive coefficients, and $\boldsymbol{B}_0$ is assumed to be diagonal. Assume also that the correlation matrix $\boldsymbol{R}_t$ satisfies the cDCC version of Aielli (2013), which is a modification of the original DCC formulation introduced by Engle (2002). The cDCC model is defined by

$$\boldsymbol{R}_t = \boldsymbol{Q}_t^{*-1/2} \boldsymbol{Q}_t \boldsymbol{Q}_t^{*-1/2}, \quad \boldsymbol{Q}_t = (1 - \alpha_0 - \beta_0)\boldsymbol{S}_0 + \alpha_0 \boldsymbol{Q}_{t-1}^{*1/2} \boldsymbol{\eta}_{t-1}^* \boldsymbol{\eta}_{t-1}^{*'} \boldsymbol{Q}_{t-1}^{*1/2} + \beta_0 \boldsymbol{Q}_{t-1},$$

where $\alpha_0, \beta_0 \geq 0, \alpha_0 + \beta_0 < 1$, $\boldsymbol{S}_0$ is a correlation matrix, $\boldsymbol{Q}_t^*$ is the diagonal matrix with the same diagonal elements as $\boldsymbol{Q}_t$, and $\boldsymbol{\eta}_t^* = \boldsymbol{D}_t^{-1} \boldsymbol{\epsilon}_t$. The unknown parameter $\boldsymbol{\theta}_0$ contains the volatility parameters $\boldsymbol{\omega}_0$, $\boldsymbol{A}_0$ and $\mathrm{diag}(\boldsymbol{B}_0)$, and the conditional correlation parameters $\alpha_0$, $\beta_0$ and the sub-diagonal elements of $\boldsymbol{S}_0$.

To estimate $\boldsymbol{\theta}_0$, we used a three-step estimation procedure similar to that employed by Aielli (2013). The individual volatility parameters $\boldsymbol{\omega}_0$, $\boldsymbol{A}_0$ and $\boldsymbol{B}_0$ are estimated equation-by-equation, from the $m$ augmented univariate GARCH models followed by the components of $\boldsymbol{\epsilon}_t$ (see Appendix A.2). This step is slightly different from Step 1 in Definition 3.2 of Aielli (2013) because we do not assume that $\boldsymbol{A}_0$ is diagonal in (C.1), which allows for possible volatility spillovers. The two other steps are unchanged: $\alpha_0$ and $\beta_0$ are estimated by maximizing a QML of the EbE residuals $\widehat{\boldsymbol{\eta}}_t^* = \widehat{\boldsymbol{D}}_t^{-1} \boldsymbol{\epsilon}_t$, and the last parameter $\boldsymbol{S}_0$ is estimated empirically. More precisely, let $\widehat{\boldsymbol{R}}_t = \widehat{\boldsymbol{R}}_t(\alpha, \beta)$ with

$$\begin{aligned}
\widehat{\boldsymbol{R}}_t &= \widehat{\boldsymbol{Q}}_t^{*-1/2} \widehat{\boldsymbol{Q}}_t \widehat{\boldsymbol{Q}}_t^{*-1/2}, \quad \widehat{\boldsymbol{Q}}_t = (1 - \alpha - \beta)\boldsymbol{S}_n + \alpha \widehat{\boldsymbol{Q}}_{t-1}^{*1/2} \widehat{\boldsymbol{\eta}}_{t-1}^* \widehat{\boldsymbol{\eta}}_{t-1}^{*'} \widehat{\boldsymbol{Q}}_{t-1}^{*1/2} + \beta \widehat{\boldsymbol{Q}}_{t-1}, \\
\boldsymbol{S}_n &= \boldsymbol{S}_n(\alpha, \beta) = \frac{1}{n} \sum_{t=1}^{n} \widehat{\boldsymbol{Q}}_t^{*1/2} \widehat{\boldsymbol{\eta}}_t^* \widehat{\boldsymbol{\eta}}_t^{*'} \widehat{\boldsymbol{Q}}_t^{*1/2}, \quad \widehat{\boldsymbol{Q}}_t^* = \mathrm{diag}(\widehat{q}_{11,t}, \ldots, \widehat{q}_{mm,t})
\end{aligned}$$

and $\widehat{q}_{ii,t} = (1 - \alpha - \beta) + (\alpha \widehat{\eta}_{i,t-1}^{*2} + \beta)\widehat{q}_{ii,t-1}$ for $i = 1, \ldots, m$. The estimators of the DCC parameters are then defined by

$$\begin{aligned}
(\widehat{\alpha}_n, \widehat{\beta}_n) &= \arg \min_{(\alpha, \beta)} \sum_{t=1}^{n} \widehat{\boldsymbol{\eta}}_{t-1}^{*'} \widehat{\boldsymbol{R}}_t^{-1} \widehat{\boldsymbol{\eta}}_{t-1}^* + \log \left| \widehat{\boldsymbol{R}}_t \right|, \\
\widehat{\boldsymbol{S}}_n &= \boldsymbol{S}_n^{*-1/2}(\widehat{\alpha}_n, \widehat{\beta}_n) \boldsymbol{S}_n(\widehat{\alpha}_n, \widehat{\beta}_n) \boldsymbol{S}_n^{*-1/2}(\widehat{\alpha}_n, \widehat{\beta}_n),
\end{aligned}$$

with $\boldsymbol{S}_n^*(\widehat{\alpha}_n, \widehat{\beta}_n) = \mathrm{diag}\boldsymbol{S}_n(\widehat{\alpha}_n, \widehat{\beta}_n)$ and usual notations.

# References


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

**Barone-Adesi, G., Giannopoulos, K., and L. Vosper** (1999) VaR without correlations for nonlinear portfolios. *Journal of Futures Markets* 19, 583–602.

**Bauwens, L., Hafner, C.M. 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., 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.

**Boswijk H.P., and R. van der Weide** (2011) Method of moments estimation of GO-GARCH models. *Journal of Econometrics* 163, 118–126.

**Bradley B.O. and M. S. Taqqu** (2002) Financial Risk and Heavy Tails. In *Heavy-tailed distributions in Finance*, Ed. by S.T. Rachev, North Holland.

**Christoffersen, P.F.** (2003) *Elements of financial risk management.* Academic Press, London.

**Christoffersen, P.F. and S. Gonçalves** (2005) Estimation Risk in Financial Risk Management. *Journal of Risk* 7, 1–28.

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

**Davison, A. C. and D.V. Hinkley** (1997). *Bootstrap methods and their application.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

**Embrechts, P. and G. Puccetti** (2010) Risk Aggregation. *In P.Jaworski et al. (eds.), Copula Theory and Its Applications, Lecture Notes in Statistics* 198, Springer, Berlin/Heidelberg.

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

**Escanciano, J.C., and J. Olmo** (2010) Backtesting parametric VaR with estimation risk. *Journal of Business and Economic Statistics* 28, 36–51.

**Escanciano, J.C. and J. Olmo** (2011) Robust backtesting tests for value-at-risk models. *Journal of Financial Econometrics* 9, 132–161.

**Farkas, W., Fringuellotti, F. and R. Tunaru** (2016) Model Risk Adjustments for Regulatory Capital Calculations. Working paper EFMA.

**Fiorentini, G., and E., Sentana** (2016) Consistent non-Gaussian Pseudo Maximum Likelihood Estimators. DP CEMFI.

**Francq, C., Horváth, L. and J.M. Zakoïan** (2015) Variance targeting estimation of multivariate GARCH models. Forthcoming in the *Journal of Financial Econometrics.*

**Francq, C., Jiménez Gamero, M.D. and S. Meintanis** (2017) Tests for conditional ellipticity in multivariate GARCH models *Journal of Econometrics* 196, 305–319.

**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.

**Francq, C. and J.M. Zakoïan** (2015) Risk-parameter estimation in volatility models. *Journal of Econometrics* 184, 158–173.

**Francq, C. and J.M. Zakoïan** (2016) Estimating multivariate GARCH models equation by equation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 78, 613–635.

**Gouriéroux, C. and J.M. Zakoïan** (2013) Estimation adjusted VaR. *Econometric Theory* 29, 735–770.

**Koenker, R. and Z. Xiao** (2006) Quantile autoregression. *Journal of the American Statistical Association* 101, 980–990.

**Kuester, K., Mittnik, S. and M. Paolella** (2006) Value-at-Risk prediction: a comparison of alternative strategies. *Journal of Financial Econometrics* 4, 53–89.

**Ling, S.** (2004) Estimation and testing stationarity for double-autoregressive models. *Journal of the Royal Statistical Society B* 66, 63–78.

**Ling, S.** (2005) Self-weighted least absolute deviation estimation for infinite variance autoregressive models. *Journal of the Royal Statistical Society B* 67, 381–393.

**Mancini, L. and F. Trojani** (2011) Robust Value-at-Risk prediction. *Journal of Financial Econometrics* 9, 281–313.

**McNeil, A., Frey, R. and P. Embrechts** (2005) *Quantitative risk management.* Princeton Series in Finance, Princeton.

**Pedersen, R.S. and A. Rahbek** (2014) Multivariate variance targeting in the BEKK-GARCH model. *The Econometrics Journal* 17, 24–55.

**Pei, P.** (2010) Backtesting portfolio Value-at-Risk with estimated portfolio weights. CAEPR Working Paper #010-2010.

**Rombouts, J.V.K. and M. Verbeek** (2009) Evaluating portfolio Value-at-Risk using semi-parametric GARCH models. *Quantitative Finance* 17, 737–745.

**Sentana, E.** (2003) Mean-variance portfolio allocation with a value at risk constraint. *Revista de Economia Financiera* 1, 4–14.

**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.

**Spierdijk, L.** (2014) Confidence Intervals for ARMA-GARCH Value-at-Risk: The Case of Heavy Tails and Skewness. *Computational Statistics and Data Analysis*, forthcoming.

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

**Zhu, D. and V. Zinde-Walsh** (2009) Properties and estimation of asymmetric exponential power distribution. *Journal of Econometrics* 148, 86–99.