Joint inference on market and estimation risks in dynamic portfolios

Francq, Christian and Zakoian, Jean-Michel

November 2015

Joint inference on market and estimation risks in dynamic portfolios

Christian Francq* and Jean-Michel Zakoïan†

Abstract

We study the estimation risk induced by univariate and multivariate methods for evaluating the conditional Value-at-Risk (VaR) of a portfolio of assets. The composition of the portfolio can be time-varying and the individual returns are assumed to follow a general multivariate dynamic model. Under sphericity of the innovations distribution, we introduce in the multivariate framework a concept of VaR parameter, and we establish the asymptotic distribution of its estimator. A multivariate Filtered Historical Simulation method, which does not rely on sphericity, is also studied. We derive asymptotic confidence intervals for the conditional VaR, which allow to quantify simultaneously the market and estimation risks. The particular case of minimal variance and minimal VaR portfolios is considered. Potential usefulness, feasibility and drawbacks of the different approaches are illustrated via Monte-Carlo experiments and an empirical study based on stock returns.

Keywords: Confidence Intervals for VaR, DCC GARCH model, Estimation risk, Filtered Historical Simulation, Optimal Dynamic Portfolio.

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 $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_m)'$, risk aggregation concerns the

*CREST and University Lille 3, BP 60149, 59653 Villeneuve d'Ascq cedex, France. E-Mail: christian.francq@univ-lille3.fr

†Corresponding author: Jean-Michel Zakoïan, University Lille 3 and CREST, 15 boulevard Gabriel Péri, 92245 Malakoff Cedex, France. E-mail: zakoian@ensae.fr, Phone number: 33.1.41.17.77.25. The second author gratefully thanks the IDR ACP "Régulation et risques systémiques" for financial support.

risk implied by an aggregate financial position defined as a real-valued function of $\boldsymbol{\epsilon}$. For instance, under the terms of Basel II, banks often measure the risk of a vector $\boldsymbol{\epsilon}$ of financial positions by the Value-at-Risk (VaR) of $a_1\epsilon_1 + \cdots + a_m\epsilon_m$ 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 $\boldsymbol{\epsilon}$.

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. In this article, we will focus on the VaR, arguably the most popular risk measure in finance and insurance due to its importance within the Basel II capital adequacy framework.

1.1 Conditional VaR of a dynamic portfolio

Let $\boldsymbol{p}_t = (p_{1t}, \ldots, p_{mt})'$ denote the vector of prices of m assets at time t. Let $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \ldots, \epsilon_{mt})'$ denote the corresponding vector of log-returns, with $\epsilon_{it} = \log(p_{it}/p_{i,t-1})$ for $i = 1, \ldots, m$. Let V_t denote the value at time t of a portfolio composed of $\mu_{i,t-1}$ units of asset i, for $i = 1, \ldots, m$:

$$V_0 = \sum_{i=1}^{m} \mu_i p_{i0}, \quad V_t = \sum_{i=1}^{m} \mu_{i,t-1} p_{it}, \quad \text{for } t \geq 1 \tag{1.1}$$

where the $\mu_{i,t-1}$ are measurable functions of the prices up to time $t-1$, and the μ_i are constants. The return of the portfolio over the period $[t-1, t]$ is, for $t \geq 1$, assuming that $V_{t-1} \neq 0$,

$$\frac{V_t}{V_{t-1}} - 1 = \sum_{i=1}^{m} a_{i,t-1} e^{\epsilon_{it}} - 1 \approx \sum_{i=1}^{m} a_{i,t-1}\epsilon_{it} + a_{0,t-1}$$

where

$$a_{i,t-1} = \frac{\mu_{i,t-1} p_{i,t-1}}{\sum_{j=1}^{m} \mu_{j,t-2} p_{j,t-1}}, \quad i = 1, \ldots, m \quad \text{and} \quad a_{0,t-1} = -1 + \sum_{i=1}^{m} a_{i,t-1}.$$

We assume that, at date t, the investor may rebalance his portfolio under a "self-financing" constraint.

SF: The portfolio is rebalanced in such a way that $\sum_{i=1}^{m} \mu_{i,t-1} p_{it} = \sum_{i=1}^{m} \mu_{i,t} p_{it}$.

In other words, the value at time t of the portfolio bought at time $t-1$ equals the value at time t of the portfolio bought at time t. An obvious consequence of the self-financing assumption **SF**, is that the change of value of the portfolio between $t-1$ and t is only due to the change of value of the underlying assets:

$$V_t - V_{t-1} = \sum_{i=1}^{m} \mu_{i,t-1}(p_{i,t} - p_{i,t-1}).$$

Another consequence is that the weights $a_{i,t-1}$ sum up to 1, that is $a_{0,t-1} = 0$. Thus, under **SF** we have $\frac{V_t}{V_{t-1}} - 1 \approx \epsilon_t^{(P)}$, where

$$\epsilon_t^{(P)} = \sum_{i=1}^{m} a_{i,t-1}\epsilon_{it} = \mathbf{a}'_{t-1}\boldsymbol{\epsilon}_t, \qquad a_{i,t-1} = \frac{\mu_{i,t-1}p_{i,t-1}}{\sum_{j=1}^{m}\mu_{j,t-1}p_{j,t-1}}, \tag{1.2}$$

for $i = 1, \ldots, m$, and $\mathbf{a}_{t-1} = (a_{1,t-1}, \ldots, a_{m,t-1})'$. The *conditional* VaR of the portfolio's return process $(\epsilon_t^{(P)})$ at risk level $\alpha \in (0,1)$, denoted by $\mathrm{VaR}_{t-1}^{(\alpha)}(\epsilon^{(P)})$, is defined by

$$P_{t-1}\left[\epsilon_t^{(P)} < -\mathrm{VaR}_{t-1}^{(\alpha)}(\epsilon^{(P)})\right] = \alpha, \tag{1.3}$$

where P_{t-1} denotes the historical distribution conditional on $\{\boldsymbol{p}_u, u < t\}$.

1.2 Univariate vs multivariate modeling of the portfolio's dynamic

In order to estimate the conditional risk of the portfolio's return $\epsilon_t^{(P)}$ from observations $\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n$, two strategies can be advocated. A multivariate strategy requires a dynamic model for the vector of risk factors $\boldsymbol{\epsilon}_t$, while a univariate approach will be based on a dynamic model for the portfolio's return $(\epsilon_t^{(P)})$. According to Bauwens, Laurent and Rombouts (2006), "it is probably simpler to use the univariate framework if there are many assets, but we conjecture that using a multivariate specification may become a feasible alternative. Whether the univariate "repeated" approach is more adequate than the multivariate one is an open question." These issues were tackled, by means of Monte-carlo experiments and real data analysis, by McAleer and da Veiga (2008), and Santos, Nogales and Ruiz (2013).

In fact, deriving a univariate model for the portfolio's return may raise several difficulties.

i) Without further constraints on the past-dependent weights $a_{i,t-1}$, the resulting process $(\epsilon_t^{(P)})$ might not be stationary (details will be given below). Needless to say that developing statistical inference procedures in this situation can be cumbersome.

ii) By embedding the weights into the stochastic process, the univariate approach does not facilitate portfolio comparison. For instance the determination of an optimal portfolio in the mean-variance sense requires knowledge of the first two conditional moments of the vector process.

iii) More importantly, the univariate approach provides a VaR defined by

$$P_{t-1}^{*}\left[\epsilon_t^{(P)} < -\mathrm{VaR}_{t-1}^{(\alpha)*}(\epsilon^{(P)})\right] = \alpha, \tag{1.4}$$

where P_{t-1}^{*} denotes the distribution conditional on $\{\epsilon_u^{(P)}, u < t\}$, which is different from the VaR defined in (1.3). The latter takes into account the full information brought by the past prices.

We now describe more thoroughly the multivariate approach.

1.3 Multivariate modeling of the risk factors

The multivariate approach is based on a model which is independent of the weight sequence. Consider a general multivariate model of the form

$$\boldsymbol{\epsilon}_t = \boldsymbol{m}_t(\boldsymbol{\varphi}_0) + \boldsymbol{\Sigma}_t(\boldsymbol{\vartheta}_0)\boldsymbol{\eta}_t, \tag{1.5}$$

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 $\boldsymbol{\Sigma}_t(\boldsymbol{\vartheta}_0)$ and the $m \times 1$ vector $\boldsymbol{m}_t(\boldsymbol{\varphi}_0)$ are specified as functions parameterized by a d-dimensional parameter $\boldsymbol{\theta}_0 = (\boldsymbol{\varphi}_0', \boldsymbol{\vartheta}_0') \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ of the past values of $\boldsymbol{\epsilon}_t$:

$$\boldsymbol{m}_t(\boldsymbol{\varphi}_0) = \boldsymbol{m}(\boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \dots, \boldsymbol{\varphi}_0), \qquad \boldsymbol{\Sigma}_t(\boldsymbol{\vartheta}_0) = \boldsymbol{\Sigma}(\boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \dots, \boldsymbol{\vartheta}_0). \tag{1.6}$$

For the sake of generality, we do not consider a particular specification of the conditional mean $\boldsymbol{m}_t$, or the conditional variance $\boldsymbol{\Sigma}_t$.[1]

In view of (1.2)-(1.5), the portfolio's return satisfies

$$\epsilon_t^{(P)} = \mathbf{a}_{t-1}'\boldsymbol{m}_t(\boldsymbol{\varphi}_0) + \mathbf{a}_{t-1}'\boldsymbol{\Sigma}_t(\boldsymbol{\vartheta}_0)\boldsymbol{\eta}_t, \tag{1.7}$$

from which it follows that its conditional VaR at level α is given by

$$\mathrm{VaR}_{t-1}^{(\alpha)}(\epsilon^{(P)}) = -\mathbf{a}_{t-1}'\boldsymbol{m}_t(\boldsymbol{\varphi}_0) + \mathrm{VaR}_{t-1}^{(\alpha)}\left(\mathbf{a}_{t-1}'\boldsymbol{\Sigma}_t(\boldsymbol{\vartheta}_0)\boldsymbol{\eta}_t\right). \tag{1.8}$$

The VaR formula can be simplified if we assume that 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 $\boldsymbol{P}$. This is equivalent to assuming that

[1] The most widely used specifications of multivariate GARCH 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 (1.6) also includes multivariate extensions of the double-autoregressive models studied by Ling (2004).

A1: for any non-random vector $\boldsymbol{\lambda} \in \mathbb{R}^m$, $\boldsymbol{\lambda}'\boldsymbol{\eta}_t \stackrel{d}{=} \|\boldsymbol{\lambda}\|\eta_{1t}$,

where $\|\cdot\|$ denotes the euclidian norm on $\mathbb{R}^m$, η_{it} denotes the i-th component of $\boldsymbol{\eta}_t$, and $\stackrel{d}{=}$ stands for the equality in distribution. [2] Under the sphericity assumption **A1** we have

$$\mathrm{VaR}_{t-1}^{(\alpha)}(\epsilon^{(P)}) = -\mathbf{a}_{t-1}'\boldsymbol{m}_t(\boldsymbol{\varphi}_0) + \left\|\mathbf{a}_{t-1}'\boldsymbol{\Sigma}_t(\boldsymbol{\vartheta}_0)\right\| \mathrm{VaR}^{(\alpha)}(\eta)\,, \tag{1.9}$$

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

1.4 Estimation risk

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. As far as the VaR of a portfolio is concerned, as defined in (1.9), it is clear that uncertainty results from the estimation of the model parameter $\boldsymbol{\theta}_0$, as well as from the estimation of the VaR of η_{1t}. The econometric literature devoted to the estimation risk in dynamic models is scant. 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 (2015a) introduced the notion of risk parameter (to be discussed below) and derived asymptotic confidence intervals for the conditional VaR of univariate returns.

1.5 Aims of the paper

The first aim of this paper is to study the asymptotic properties of different multivariate 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. Based on formula (1.9), it consists in estimating parameter $\boldsymbol{\theta}_0$ in the first step, and replacing the VaR of η_t by an empirical quantile of the residuals. 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 a method for constructing confidence intervals for the conditional VaR of

[2] Note that the choice of any other norm in this assumption would not be compatible with the assumed unit covariance matrix for $\boldsymbol{\eta}_t$. Indeed, under **A1** we have $\mathrm{Var}(\boldsymbol{\lambda}'\boldsymbol{\eta}_t) = \boldsymbol{\lambda}'\boldsymbol{\lambda} = \|\boldsymbol{\lambda}\|^2\mathrm{Var}(\eta_{1t}) = \|\boldsymbol{\lambda}\|^2$.

portfolios, that is, a way to visualize the estimation risk. The third aim is to provide a framework for selecting portfolios, on the basis of their estimated conditional risks. The goal is to estimate the composition, as well as the risk, of dynamic "optimal portfolios" (in the sense of minimal conditional variance or minimal conditional VaR). The last aim is to compare, from a practical point of view, the univariate and multivariate approaches. Despite the previously underlined difficulties, the univariate approach is popular among practitioners because of its simplicity, and may provide good results in certain situations.

The rest of this paper is organized as follows. Section 2 is devoted to the asymptotic properties of the estimators of the conditional VaR under the sphericity assumption. This assumption is relaxed in Section 3. Comparisons of the different approaches are proposed in Section 4. Proofs and complementary results are collected in the Appendix.

2 Conditional VaR estimation under sphericity

Under the sphericity assumption **A1**, a natural strategy for estimating the conditional VaR of a portfolio is to estimate $\boldsymbol{\theta}_0$ by some consistent estimator $\widehat{\boldsymbol{\theta}}_n = (\widehat{\boldsymbol{\varphi}}_n', \widehat{\boldsymbol{\vartheta}}_n')'$ 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 some 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 components of the first-step residuals.

Let $\Theta = \Theta_\varphi \times \Theta_\vartheta$ denote the parameter space, and assume $\boldsymbol{\theta}_0 \in \Theta$. Let $\widehat{\boldsymbol{\theta}}_n = (\widehat{\boldsymbol{\varphi}}_n', \widehat{\boldsymbol{\vartheta}}_n')'$ denote an estimator of parameter $\boldsymbol{\theta}_0$, obtained from observations $\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_n$ and initial values $\widetilde{\boldsymbol{\epsilon}}_0, \widetilde{\boldsymbol{\epsilon}}_{-1}, \dots$. The vector of residuals is defined by $\widehat{\boldsymbol{\eta}}_t = \widetilde{\boldsymbol{\Sigma}}_t^{-1}(\widehat{\boldsymbol{\vartheta}}_n)\{\boldsymbol{\epsilon}_t - \widetilde{\boldsymbol{m}}_t(\widehat{\boldsymbol{\varphi}}_n)\}) = (\widehat{\eta}_{1t}, \dots, \widehat{\eta}_{mt})'$. Let $\widetilde{\boldsymbol{m}}_t(\boldsymbol{\varphi}) = \boldsymbol{m}(\boldsymbol{\epsilon}_{t-1}, \dots, \boldsymbol{\epsilon}_1, \widetilde{\boldsymbol{\epsilon}}_0, \widetilde{\boldsymbol{\epsilon}}_{-1}, \dots, \boldsymbol{\varphi})$, $\widetilde{\boldsymbol{\Sigma}}_t(\boldsymbol{\vartheta}) = \boldsymbol{\Sigma}(\boldsymbol{\epsilon}_{t-1}, \dots, \boldsymbol{\epsilon}_1, \widetilde{\boldsymbol{\epsilon}}_0, \widetilde{\boldsymbol{\epsilon}}_{-1}, \dots, \boldsymbol{\vartheta})$, for $t \geq 1$ and $(\boldsymbol{\varphi}', \boldsymbol{\vartheta}')' \in \Theta$. For $\alpha \in (0,1)$, let $q_\alpha(S)$ denote the α-quantile of a set $S \subset \mathbb{R}$. In view of (1.9), an estimator *based on the spherical assumption* of the conditional VaR at level α is

$$\widehat{\mathrm{VaR}}^{(\alpha)}_{S,t-1}(\epsilon^{(P)}) = -\mathbf{a}_{t-1}'\widetilde{\boldsymbol{m}}_t(\widehat{\boldsymbol{\varphi}}_n) + \|\mathbf{a}_{t-1}'\widetilde{\boldsymbol{\Sigma}}_t(\widehat{\boldsymbol{\vartheta}}_n)\|\xi_{n,1-2\alpha}, \tag{2.1}$$

where $\xi_{n,1-2\alpha} = q_{1-2\alpha}\left(\{|\widehat{\eta}_{it}|, 1 \leq i \leq m, 1 \leq t \leq n\}\right)$. The latter estimator takes advantage of the fact that the components of $\boldsymbol{\eta}_t$ are identically distributed under **A1**.

2.1 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 $(\widehat{\boldsymbol{\theta}}_n', \xi_{n,1-2\alpha})$.

A2: $(\boldsymbol{\epsilon}_t)$ is a strictly stationary and nonanticipative[3] solution of Model (1.5)-(1.6).

This assumption can be made explicit for particular classes of MGARCH models satisfying Model (1.5)-(1.6). We now assume that the estimator $\widehat{\boldsymbol{\theta}}_n$ admits a Bahadur representation.

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

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \overset{o_P(1)}{=} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \boldsymbol{\Delta}_{t-1} \boldsymbol{V}(\boldsymbol{\eta}_t), \tag{2.2}$$

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 $E\boldsymbol{\Delta}_t = \boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_\varphi \\ \boldsymbol{\Lambda}_\vartheta \end{pmatrix}$ is full row rank.

Assumption **A3** holds for a variety of MGARCH models and estimators[4] (see Appendix A for examples).

A4: For all $\boldsymbol{x} \in \mathbb{R}^K, \boldsymbol{y} \in \mathbb{R}^m$,

$$\boldsymbol{x}'\boldsymbol{V}(\boldsymbol{\eta}_t) + \boldsymbol{y}'\boldsymbol{\nu}_\alpha(\boldsymbol{\eta}_t) = 0, \ \ a.s. \quad \Longrightarrow \quad \boldsymbol{x} = \boldsymbol{0}_K, \ \ \boldsymbol{y} = \boldsymbol{0}_m,$$

where $\boldsymbol{\nu}_\alpha(\boldsymbol{\eta}_t) = (\mathbf{1}_{\{|\eta_{1t}|<\xi_{1-2\alpha}\}} - 1 + 2\alpha, \dots, \mathbf{1}_{\{|\eta_{mt}|<\xi_{1-2\alpha}\}} - 1 + 2\alpha)'$.

Assumption **A4** will be used to ensure the invertibility of the asymptotic covariance matrix of $(\widehat{\boldsymbol{\theta}}_n', \xi_{n,1-2\alpha})$. It is, in particular, satisfied if the random vectors $\boldsymbol{\eta}_t$ and $\boldsymbol{V}(\boldsymbol{\eta}_t)$ have a positive density over $\mathbb{R}^m$ and $\mathbb{R}^K$, respectively. The next assumption imposes smoothness of the functions $\boldsymbol{m}$ and $\boldsymbol{\Sigma}$ with respect to the parameter.

[3] In the sense that $\boldsymbol{\epsilon}_t$ is a measurable function of the variables $\boldsymbol{\eta}_u$ with $u \le t$.

[4] In the univariate setting, the asymptotic theory of estimation for GARCH parameters has been extensively 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) among others.

A5: The functions $\boldsymbol{\varphi} \mapsto \boldsymbol{m}(x_1, x_2, \dots; \boldsymbol{\varphi})$ and $\boldsymbol{\vartheta} \mapsto \boldsymbol{\Sigma}(x_1, x_2, \dots; \boldsymbol{\vartheta})$ are continuously differentiable over Θ_φ and Θ_ϑ respectively.

The next theorem establishes the asymptotic normality of $(\widehat{\boldsymbol{\theta}}_n', \xi_{n,1-2\alpha})$. Let $\boldsymbol{\Psi} = E(\boldsymbol{\Delta}_t \boldsymbol{\Upsilon} \boldsymbol{\Delta}_t') = \begin{pmatrix} \boldsymbol{\Psi}_{\varphi\varphi} & \boldsymbol{\Psi}_{\varphi\vartheta} \\ \boldsymbol{\Psi}_{\vartheta\varphi} & \boldsymbol{\Psi}_{\vartheta\vartheta} \end{pmatrix} = (\boldsymbol{\Psi}_{\cdot\varphi} \quad \boldsymbol{\Psi}_{\cdot\vartheta})$, $\boldsymbol{\Omega} = E\left[\left\{\mathrm{vec}\left(\boldsymbol{\Sigma}_t^{-1}\right)\right\}' \left\{\frac{\partial}{\partial \boldsymbol{\vartheta}'}\mathrm{vec}\left(\boldsymbol{\Sigma}_t\right)\right\}\right]$, $\boldsymbol{W}_\alpha = \mathrm{Cov}(\boldsymbol{V}(\boldsymbol{\eta}_t), N_t)$, $\gamma_\alpha = \mathrm{var}(N_t)$, 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 $|\eta_{1t}|$, $\boldsymbol{\Xi}_{\boldsymbol{\theta}\xi} = \frac{-1}{m}\left\{\xi_{1-2\alpha}\boldsymbol{\Psi}_{\cdot\vartheta}\boldsymbol{\Omega}' + \frac{1}{f(\xi_{1-2\alpha})}\boldsymbol{\Lambda}\boldsymbol{W}_\alpha\right\}$, $\zeta_{1-2\alpha} = \frac{1}{m^2}\left\{\xi_{1-2\alpha}^2\boldsymbol{\Omega}\boldsymbol{\Psi}_{\vartheta\vartheta}\boldsymbol{\Omega}' + \frac{2\xi_{1-2\alpha}}{f(\xi_{1-2\alpha})}\boldsymbol{\Omega}\boldsymbol{\Lambda}_\vartheta\boldsymbol{W}_\alpha + \frac{\gamma_\alpha}{f^2(\xi_{1-2\alpha})}\right\}$.

Theorem 2.1. *Assume that* **A1-A5** *hold. Let $\alpha \in (0, 0.5)$. Suppose that $|\eta_{1t}|$ admit a density f which is continuous and strictly positive in a neighborhood of $\xi_{1-2\alpha}$. Then*

$$\sqrt{n}\begin{pmatrix} \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \\ \xi_{n,1-2\alpha} - \xi_{1-2\alpha} \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Xi} := \begin{pmatrix} \boldsymbol{\Psi} & \boldsymbol{\Xi}_{\boldsymbol{\theta}\xi} \\ \boldsymbol{\Xi}_{\boldsymbol{\theta}\xi}' & \zeta_{1-2\alpha} \end{pmatrix}\right). \tag{2.3}$$

Moreover, $\boldsymbol{\Xi}$ is nonsingular.

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

2.2 Conditional VaR parameter

The notion of VaR parameter, introduced for univariate GARCH models by Francq and Zakoïan (2015a), 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.

A6: There exists a continuously differentiable function $\boldsymbol{G} : \mathbb{R}^{d_2+1} \mapsto \mathbb{R}^{d_2}$ such that for any $\boldsymbol{\vartheta} \in \Theta_\vartheta$, any $K > 0$, and any sequence $(\mathbf{x}_i)_i$ on $\mathbb{R}^m$

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

In other words, a change of the scale in the components of $\boldsymbol{\eta}$ can be compensated by a change of the variance parameter. This assumption is obviously satisfied for all commonly used parametric forms of $\boldsymbol{\Sigma}_t(\boldsymbol{\vartheta})$.[5] Under sphericity and the stability-by scale assumption **A6** on the volatility function $\boldsymbol{\Sigma}_t(\cdot)$, the conditional VaR can be expressed in function of the expected returns vector and

[5]For instance, in the case of the BEKK-GARCH(1,1) model (C.1), with $\boldsymbol{\vartheta} = (\mathrm{vec}(\boldsymbol{A})', \mathrm{vec}(\boldsymbol{B})', \mathrm{vec}(\boldsymbol{C})')'$, we find $\boldsymbol{\vartheta}^* = (K\mathrm{vec}(\boldsymbol{A})', \mathrm{vec}(\boldsymbol{B})', K^2\mathrm{vec}(\boldsymbol{C})')'$.

a reparameterized volatility matrix. Let $\alpha < 1/2$, so that $\mathrm{VaR}^{(\alpha)}(\eta) > 0$ under **A1**. It follows from **A6** that

$$\mathrm{VaR}_{t-1}^{(\alpha)}(\epsilon^{(P)}) = -\mathbf{a}_{t-1}'\boldsymbol{m}_t(\boldsymbol{\varphi}_0) + \|\mathbf{a}_{t-1}'\boldsymbol{\Sigma}_t(\boldsymbol{\vartheta}_0^*)\| \tag{2.4}$$

where

$$\boldsymbol{\vartheta}_0^* = \boldsymbol{G}\left\{\boldsymbol{\vartheta}_0, \mathrm{VaR}^{(\alpha)}(\eta)\right\}. \tag{2.5}$$

The new parameter $\boldsymbol{\theta}_0^* = (\boldsymbol{\varphi}_0', \boldsymbol{\vartheta}_0^{*'})'$ 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

$$\widehat{\boldsymbol{\theta}}_n^* = (\widehat{\boldsymbol{\varphi}}_n', \widehat{\boldsymbol{\vartheta}}_n^{*'})' \quad \text{where} \quad \widehat{\boldsymbol{\vartheta}}_n^* = G\left\{\widehat{\boldsymbol{\vartheta}}_n, \widehat{\mathrm{VaR}}_n^{(\alpha)}(\eta)\right\}$$

with obvious notations. The asymptotic properties of $\widehat{\boldsymbol{\theta}}_n^*$ are a direct consequence of Theorem 2.1.

Corollary 2.1 (CAN of the VaR-parameter estimator)**.** *Under the assumptions of Theorem 2.1,* $\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n^* - \boldsymbol{\theta}_0^*\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Xi}^* := \dot{\boldsymbol{G}}\boldsymbol{\Xi}\dot{\boldsymbol{G}}'\right)$ *where*

$$\dot{\boldsymbol{G}} = \begin{pmatrix} \boldsymbol{I}_{d_1} & \mathbf{0}_{d_1\times(d+1)} \\ \mathbf{0}_{(d+1)\times d_1} & \left[\frac{\partial G(\boldsymbol{\vartheta},\xi)}{\partial(\boldsymbol{\vartheta}',\xi)}\right]_{(\boldsymbol{\vartheta}_0,\xi_{1-2\alpha})} \end{pmatrix}.$$

The asymptotic distribution of $\widehat{\boldsymbol{\theta}}_n^*$ provides a quantification of the estimation risk.

2.3 Asymptotic confidence intervals for the VaR's of portfolios

In view of (2.4), the estimator in (2.1) of the conditional VaR of the portfolio at level α writes

$$\widehat{\mathrm{VaR}}_{S,t-1}^{(\alpha)}(\epsilon^{(P)}) = -\mathbf{a}_{t-1}'\boldsymbol{m}_t(\widehat{\boldsymbol{\varphi}}_n) + \|\mathbf{a}_{t-1}'\boldsymbol{\Sigma}_t(\widehat{\boldsymbol{\vartheta}}_n^*)\|. \tag{2.6}$$

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

$$\widehat{\mathrm{VaR}}_{S,t-1}^{(\alpha)}(\epsilon^{(P)}) \pm \frac{1}{\sqrt{n}}\Phi^{-1}(1-\alpha_0/2)\left\{\boldsymbol{\delta}_{t-1}'\widehat{\boldsymbol{\Xi}}^*\boldsymbol{\delta}_{t-1}\right\}^{1/2}, \tag{2.7}$$

where

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

$\widetilde{\boldsymbol{H}}_t(\cdot) = \widetilde{\boldsymbol{\Sigma}}_t(\cdot)\widetilde{\boldsymbol{\Sigma}}_t'(\cdot)$, and $\Phi^{-1}(u)$ denotes the u-quantile of the standard Gaussian distribution, $u \in (0,1)$. Drawing such CIs allows to take into account the estimation risk inherent to the

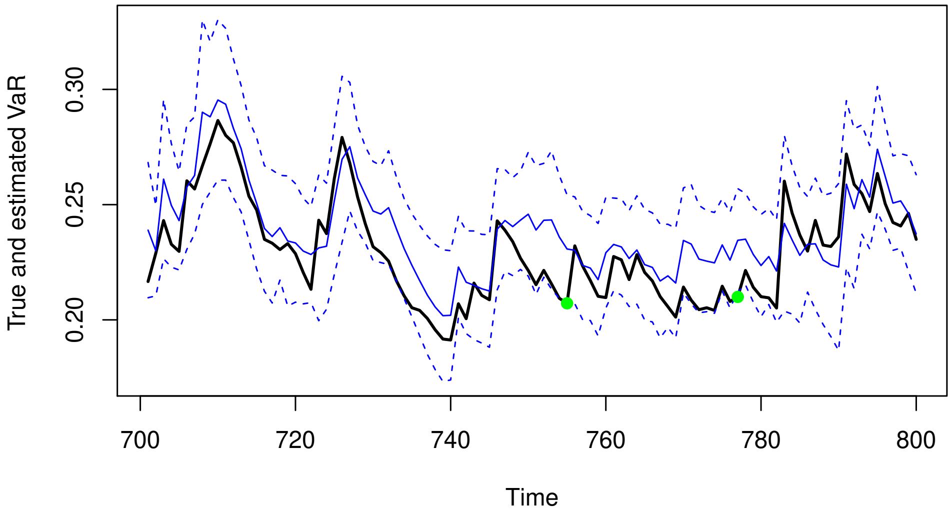

Figure 1: True 1%-VaR (full black line) and estimated 95%-confidence interval (dotted blue line) for the 1%-VaR, on a simulation of a fixed portfolio of a bivariate BEKK.

evaluation of the VaR of the portfolio. Note that the level α_0 of risk estimation is independent from the market risk level α.

An illustration is displayed in Figure 1, for the simulation of a bivariate BEKK model. 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).

2.4 Optimal dynamic portfolios

The portfolio with the smallest variance (the mean-variance efficient portfolio, that we call hereafter Markowitz's portfolio) is

$$\epsilon_t^{(P)*} = \boldsymbol{\epsilon}_t' \boldsymbol{a}_{0,t-1}^*, \quad \boldsymbol{a}_{0,t-1}^* = \frac{\boldsymbol{\Sigma}_t^{-2}(\boldsymbol{\vartheta}_0)\boldsymbol{e}}{\boldsymbol{e}'\boldsymbol{\Sigma}_t^{-2}(\boldsymbol{\vartheta}_0)\boldsymbol{e}}. \tag{2.8}$$

The theoretical conditional VaR of this portfolio is obtained by computing the opposite of the α-quantile of $\boldsymbol{a}_{0,t-1}^{*'}\boldsymbol{\Sigma}_t(\boldsymbol{\vartheta}_0)\boldsymbol{\eta}_1$, which is simply given by

$$\mathrm{VaR}_{t-1}^{(\alpha)}\left(\epsilon_t^{(P)*}\right) = \left\| \boldsymbol{a}_{0,t-1}^{*'}\boldsymbol{\Sigma}_t(\boldsymbol{\vartheta}_0) \right\| \xi_{1-2\alpha} = \frac{1}{\sqrt{\boldsymbol{e}'\boldsymbol{\Sigma}_t^{-2}(\boldsymbol{\vartheta}_0)\boldsymbol{e}}}\xi_{1-2\alpha} \tag{2.9}$$

under the sphericity assumption. Different alternative types of optimal portfolios have been introduced in the finance literature. In particular, several papers developed portfolio selection based on

VaR (see for instance Alexander and Baptista (2002), Campbell, Huisman and Koedijk (2001)). In the following, we derive the optimal dynamic composition of a portfolio that minimizes the VaR at level α. Such a portfolio can be called *optimal-VaR portfolio* at level α.

Under the sphericity assumption **A1**, the conditional VaR of the portfolio's return process $(\epsilon_t^{(P)})$ at risk level α is given by (1.9) which, omitting the parameter, writes

$$\mathrm{VaR}_{t-1}^{(\alpha)}(\epsilon^{(P)}) = -\mathbf{a}_{t-1}'\boldsymbol{m}_t + \left\|\mathbf{a}_{t-1}'\boldsymbol{\Sigma}_t\right\| \xi_{1-2\alpha} := q_{t-1}(\boldsymbol{a}_{t-1}),$$

where $\boldsymbol{a}_{t-1}$ satisfies $\boldsymbol{e}'\boldsymbol{a}_{t-1} = 1$. Let $\boldsymbol{a}_{\alpha,t-1}^* := \arg\min_{\{\boldsymbol{a}|\boldsymbol{e}'\boldsymbol{a}=1\}} q_{t-1}(\boldsymbol{a})$, the composition of the optimal-VaR portfolio in the spherical case. Let

$$\Delta_{t-1}^{(\alpha)} = \left(\boldsymbol{e}'\boldsymbol{\Sigma}_t^{-2}\boldsymbol{m}_t\right)^2 - \left(\boldsymbol{e}'\boldsymbol{\Sigma}_t^{-2}\boldsymbol{e}\right)\left(\boldsymbol{m}_t'\boldsymbol{\Sigma}_t^{-2}\boldsymbol{m}_t\right) + \left(\boldsymbol{e}'\boldsymbol{\Sigma}_t^{-2}\boldsymbol{e}\right)\xi_{1-2\alpha}^2. \tag{2.10}$$

Proposition 2.1. *Under the sphericity assumption* **A1***, the optimal-VaR portfolio at time t exists and is unique if and only if $\Delta_{t-1}^{(\alpha)} > 0$. The optimal composition is given by*

$$\boldsymbol{a}_{\alpha,t-1}^* = \frac{\boldsymbol{\Sigma}_t^{-2}(\boldsymbol{m}_t + \lambda\boldsymbol{e})}{\boldsymbol{e}'\boldsymbol{\Sigma}_t^{-2}(\boldsymbol{m}_t + \lambda\boldsymbol{e})} \quad \textit{where} \quad \lambda = \frac{-\boldsymbol{e}'\boldsymbol{\Sigma}_t^{-2}\boldsymbol{m}_t + \sqrt{\Delta_{t-1}^{(\alpha)}}}{\boldsymbol{e}'\boldsymbol{\Sigma}_t^{-2}\boldsymbol{e}} \tag{2.11}$$

and the optimal VaR is $q_{t-1}(\boldsymbol{a}_{\alpha,t-1}^*) = \lambda$.

In the particular case where $\boldsymbol{m}_t$ and $\boldsymbol{e}$ are colinear, that is $\boldsymbol{m}_t = m_t\boldsymbol{e}$ where $m_t \in \mathbb{R}$, we find that $\boldsymbol{a}_{\alpha,t-1}^*$ reduces to $\frac{\boldsymbol{\Sigma}_t^{-2}\boldsymbol{e}}{\boldsymbol{e}'\boldsymbol{\Sigma}_t^{-2}\boldsymbol{e}} := \boldsymbol{a}_{0,t-1}^*$, which is the optimal composition in the mean-variance sense. Note that $\boldsymbol{a}_{0,t-1}^* = \lim_{\alpha\to 0}\boldsymbol{a}_{\alpha,t-1}^*$. In this case, the optimal-VaR portfolio coincides with the Markowitz portfolio and this portfolio does not depend on α. Interestingly, this property no longer holds when $\boldsymbol{m}_t \neq m_t\boldsymbol{e}$: the optimal portfolio in (2.11) clearly depends on the risk level α. More precisely, the difference between the VaRs of the optimal-VaR and the Markowitz portfolios is

$$q_{t-1}(\boldsymbol{a}_{0,t-1}^*) - q_{t-1}(\boldsymbol{a}_{t-1}^*) = \frac{\left(\boldsymbol{e}'\boldsymbol{\Sigma}_t^{-2}\boldsymbol{e}\right)\left(\boldsymbol{m}_t'\boldsymbol{\Sigma}_t^{-2}\boldsymbol{m}_t\right) - \left(\boldsymbol{e}'\boldsymbol{\Sigma}_t^{-2}\boldsymbol{m}_t\right)^2}{\left(\boldsymbol{e}'\boldsymbol{\Sigma}_t^{-2}\boldsymbol{e}\right)\tau_{t-1}} \geq 0,$$

where $\tau_{t-1} = \left(\boldsymbol{e}'\boldsymbol{\Sigma}_t^{-2}\boldsymbol{e}\right)^{1/2}\xi_{1-2\alpha} + \sqrt{\Delta_{t-1}^{(\alpha)}}$. The nonnegativity of the numerator follows from the Cauchy-Schwarz inequality. This inequality is strict unless if $\boldsymbol{m}_t$ and $\boldsymbol{e}$ are colinear. Notice that the difference between the two VaRs increases with the non colinearity of theses two vectors. On the other hand, when α tends to 0, the difference vanishes.

3 Conditional VaR estimation without the sphericity assumption

In this section, we develop a method which does not require symmetries of the conditional distribution, inherent to the sphericity assumption.

3.1 FHS estimator and asymptotic CIs

To derive asymptotic results, we slightly modify the statistical framework by assuming that the estimator $\widehat{\boldsymbol{\theta}}_n$ is based on past observations $\boldsymbol{\epsilon}_{t-n}, \dots, \boldsymbol{\epsilon}_{t-1}$. We will use the FHS approach which relies on

i) interpreting the conditional VaR at time t as the α-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 α-quantile of the estimated linear combination.

The conditional VaR of the portfolio return is $\mathrm{VaR}_{t-1}^{(\alpha)}(\epsilon^{(P)}) = \mathrm{VaR}_{t-1}^{(\alpha)}\left\{\mathbf{a}_{t-1}'\boldsymbol{m}_t(\boldsymbol{\varphi}_0) + \mathbf{a}_{t-1}'\boldsymbol{\Sigma}_t(\boldsymbol{\vartheta}_0)\boldsymbol{\eta}_t\right\} = -\mathbf{a}_{t-1}'\boldsymbol{m}_t(\boldsymbol{\varphi}_0) - q_\alpha(t;\boldsymbol{\vartheta}_0)$ where $q_\alpha(t;\boldsymbol{\vartheta})$ denotes the theoretical α-quantile of $\boldsymbol{c}_t'(\boldsymbol{\vartheta})\boldsymbol{\eta}_1$, with the (considered as) non random vector $\boldsymbol{c}_t'(\boldsymbol{\vartheta}) = \mathbf{a}_{t-1}'\boldsymbol{\Sigma}_t(\boldsymbol{\vartheta})$. 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

$$\widehat{\mathrm{VaR}}_{FHS,t-1}^{(\alpha)}(\epsilon^{(P)}) = -\mathbf{a}_{t-1}'\boldsymbol{m}_t(\widehat{\boldsymbol{\varphi}}_n) - q_{n,\alpha}(t;\widehat{\boldsymbol{\theta}}_n),$$

where $q_{n,\alpha}(t;\widehat{\boldsymbol{\theta}}_n) = q_\alpha\left(\{\boldsymbol{c}_t'(\widehat{\boldsymbol{\vartheta}}_n)\widehat{\boldsymbol{\eta}}_s, \quad t-n \leq s \leq t-1\}\right)$.

Let $\boldsymbol{c} : \boldsymbol{\Theta}_{\boldsymbol{\vartheta}} \mapsto \mathbb{R}^m$ and $b : \boldsymbol{\Theta}_{\boldsymbol{\varphi}} \mapsto \mathbb{R}$ denote continuously differentiable vector-valued functions. Let $\xi_\alpha(\boldsymbol{\theta})$ denote the theoretical α-quantile of $b(\boldsymbol{\varphi})+\boldsymbol{c}'(\boldsymbol{\vartheta})\boldsymbol{\eta}_t(\boldsymbol{\theta})$, where $\boldsymbol{\eta}_t(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\vartheta})\{\boldsymbol{\epsilon}_t - \boldsymbol{m}_t(\boldsymbol{\varphi})\}$. Let $\xi_{n,\alpha}(\boldsymbol{\theta}) = q_\alpha\left(\{b(\boldsymbol{\varphi}) + \boldsymbol{c}'(\boldsymbol{\vartheta})\boldsymbol{\eta}_t(\boldsymbol{\theta}), 1 \leq t \leq n\}\right)$. We need to introduce the following identifiability assumption.

A7: For all $\boldsymbol{x} \in \mathbb{R}^K, y \in \mathbb{R}$,

$$\boldsymbol{x}'\boldsymbol{V}(\boldsymbol{\eta}_t) + y(\mathbf{1}_{b(\boldsymbol{\varphi}_0)+\boldsymbol{c}'(\boldsymbol{\vartheta}_0)\boldsymbol{\eta}_t < \xi_\alpha(\boldsymbol{\theta}_0)} - \alpha) = 0, \ a.s. \implies \boldsymbol{x} = 0_K, \ y = 0.$$

Let $\boldsymbol{A}_\alpha = \mathrm{Cov}(\boldsymbol{V}(\boldsymbol{\eta}_t), \mathbf{1}_{\{b(\boldsymbol{\varphi}_0)+\boldsymbol{c}'(\boldsymbol{\vartheta}_0)\boldsymbol{\eta}_t < \xi_\alpha(\boldsymbol{\theta}_0)\}})$,

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

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

$$\begin{aligned} \boldsymbol{\Omega}_t^* &= \begin{pmatrix} I_m \otimes \boldsymbol{e}_1' \\ \vdots \\ I_m \otimes \boldsymbol{e}_m' \end{pmatrix} (I_m \otimes \boldsymbol{\Sigma}_t^{-1}) \frac{\partial}{\partial \boldsymbol{\vartheta}'} \{\mathrm{vec}(\boldsymbol{\Sigma}_t)\}, \\ \boldsymbol{C}_t &= \left\{ \boldsymbol{I}_m \otimes \mathrm{vec}' \left(\frac{\partial \boldsymbol{m}_t}{\partial \boldsymbol{\varphi}'} \right) \right\} \begin{pmatrix} I_{d_1} \otimes \boldsymbol{\Sigma}_t^{-1} \boldsymbol{e}_1 \\ \vdots \\ I_{d_1} \otimes \boldsymbol{\Sigma}_t^{-1} \boldsymbol{e}_m \end{pmatrix}. \end{aligned}$$

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

Theorem 3.1. *Assume that* **A2, A3, A7** *hold. Suppose that the variable* $\boldsymbol{c}'(\boldsymbol{\vartheta}_0)\boldsymbol{\eta}_t$ *admits a density* $f_{\boldsymbol{c}}$ *which is continuous and strictly positive in a neighborhood of* $x_0 = \xi_\alpha(\boldsymbol{\theta}_0) - b(\boldsymbol{\varphi}_0)$*. Then*

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

Moreover $\sigma^2 > 0$.

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

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

A consistent estimator $\hat{\sigma}_{t-1}^2$ of σ^2 (based on the n observations anterior to time $t-1$) 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 C to compute the derivatives of $\boldsymbol{\Sigma}_t$ and $\boldsymbol{m}_t$ for particular models. An approximate $(1-\alpha_0)\%$ CI for $\mathrm{VaR}_{t-1}^{(\alpha)}(\epsilon^{(P)})$ is thus given by

$$\widehat{\mathrm{VaR}}_{FHS,t-1}^{(\alpha)}(\epsilon^{(P)}) \pm \frac{1}{\sqrt{n}} \Phi^{-1}(1-\alpha_0/2)\hat{\sigma}_{t-1}. \tag{3.1}$$

3.2 Efficiency comparisons in the static case

In this section, we compare the efficiencies of the multivariate and univariate approaches for estimating the VaR of a simplistic portfolio. We consider a static framework in which, in (1.5), $\boldsymbol{m}(\cdot) = 0$ and the matrix $\boldsymbol{\Sigma}_t(\boldsymbol{\vartheta}_0)$ is constant and diagonal, $\boldsymbol{\Sigma}_t(\boldsymbol{\vartheta}_0) = \boldsymbol{\Sigma}(\boldsymbol{\vartheta}_0) = \mathrm{diag}(\sigma_{01}, \ldots, \sigma_{0m})$, with $\boldsymbol{\vartheta}_0 = (\sigma_{01}^2, \ldots, \sigma_{0m}^2)'$. Moreover, the portfolio satisfies

$$\mathbf{a}_{t-1} = \mathbf{a} = (a_1, \ldots, a_m)', \quad \text{where } a_1, \ldots, a_m \geq 0, \text{ and } \sum_{j=1}^m a_j = 1.$$

Such a portfolio can be called *static* and it is obtained by taking in (1.1) the dynamic weights $\mu_{i,t-1} = V_{t-1}a_i/p_{i,t-1}$.[6] The return's portfolio $\mathbf{a}'\boldsymbol{\epsilon}_t$ is thus iid and its conditional VaR is constant. Under the sphericity assumption **A1**, we have

$$\mathrm{VaR}_{t-1}^{(\alpha)}(\epsilon^{(P)}) = \|\mathbf{a}'\boldsymbol{\Sigma}(\boldsymbol{\vartheta}_0)\|\xi_{1-2\alpha} = \{\widetilde{\mathbf{a}}'\boldsymbol{\vartheta}_0\}^{1/2}\xi_{1-2\alpha},$$

where $\widetilde{\mathbf{a}} = (a_1^2,\dots,a_m^2)'$. Let $\widehat{\boldsymbol{\vartheta}}_n = (\widehat{\sigma}_{n1}^2,\dots,\widehat{\sigma}_{nm}^2)'$ the Gaussian QMLE of $\boldsymbol{\vartheta}_0$, with $\widehat{\sigma}_{ni}^2 = \frac{1}{n}\sum_{t=1}^n \epsilon_{it}^2$. Under the sphericity assumption, the (constant) conditional VaR can be estimated by

$$\widehat{\mathrm{VaR}}_{S,t-1}^{(\alpha)}(\epsilon^{(P)}) = \|\mathbf{a}'\boldsymbol{\Sigma}(\widehat{\boldsymbol{\vartheta}}_n)\|\xi_{n,1-2\alpha} = \{\widetilde{\mathbf{a}}'\widehat{\boldsymbol{\vartheta}}_n\}^{1/2}\xi_{n,1-2\alpha}.$$

On the other hand, the FHS method, without the sphericity assumption, reduces to a univariate method in this setting. Indeed,

$$\boldsymbol{c}_t'(\widehat{\boldsymbol{\vartheta}}_n)\widehat{\boldsymbol{\eta}}_s = \boldsymbol{a}'\boldsymbol{\Sigma}(\widehat{\boldsymbol{\vartheta}}_n)\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\vartheta}}_n)\boldsymbol{\epsilon}_s = \boldsymbol{a}'\boldsymbol{\epsilon}_s,$$

and the estimator $\widehat{\mathrm{VaR}}_{FHS,t-1}^{(\alpha)}(\epsilon^{(P)})$ is simply $-q_\alpha(\{\boldsymbol{a}'\boldsymbol{\epsilon}_1,\dots,\boldsymbol{a}'\boldsymbol{\epsilon}_n\})$. An alternative univariate method exploits the symmetry of the distribution of $\mathbf{a}'\boldsymbol{\epsilon}_t$: let $\widehat{\mathrm{VaR}}_{U,t-1}^{(\alpha)}(\epsilon^{(P)}) = q_{1-2\alpha}\left(\{|\boldsymbol{a}'\boldsymbol{\epsilon}_1|,\dots,|\boldsymbol{a}'\boldsymbol{\epsilon}_n|\}\right)$.

The following result compares the asymptotic distributions of those three estimators of $\mathrm{VaR}_{t-1}^{(\alpha)}(\mathbf{a}'\boldsymbol{\epsilon}_t)$, when the distribution of $\boldsymbol{\eta}_t$ is Gaussian. Let ϕ denote the probability density function of the standard normal law.

Corollary 3.1. *For the static model $\boldsymbol{\epsilon}_t = \boldsymbol{\Sigma}(\boldsymbol{\vartheta}_0)\boldsymbol{\eta}_t$, where $\boldsymbol{\Sigma}(\boldsymbol{\vartheta}_0) = diag(\sigma_{01},\dots,\sigma_{0m})$ and $\boldsymbol{\eta}_t \rightsquigarrow \mathcal{N}(\mathbf{0},\boldsymbol{I}_m)$ we have*

$$\sqrt{n}\left\{\widehat{VaR}_{S,t-1}^{(\alpha)}(\epsilon^{(P)}) - VaR_{t-1}^{(\alpha)}(\mathbf{a}'\boldsymbol{\epsilon}_t)\right\} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0,\sigma_S^2(\alpha,\mathbf{a})\right),$$
$$\sqrt{n}\left\{\widehat{VaR}_{FHS,t-1}^{(\alpha)}(\epsilon^{(P)}) - VaR_{t-1}^{(\alpha)}(\epsilon^{(P)})\right\} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0,\sigma_{FHS}^2(\alpha,\mathbf{a})\right),$$
$$\sqrt{n}\left\{\widehat{VaR}_{U,t-1}^{(\alpha)}(\epsilon^{(P)}) - VaR_{t-1}^{(\alpha)}(\epsilon^{(P)})\right\} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0,\sigma_U^2(\alpha,\mathbf{a})\right),$$

where

$$\begin{aligned}\sigma_S^2(\alpha,\mathbf{a}) &= \frac{\xi_{1-2\alpha}^2}{2}\left\{\frac{\sum_{i=1}^m a_i^4\sigma_{0i}^4}{\sum_{i=1}^m a_i^2\sigma_{0i}^2} - \frac{\sum_{i=1}^m a_i^2\sigma_{0i}^2}{m}\right\} + \frac{2\alpha(1-2\alpha)}{4\phi^2(\xi_{1-2\alpha})}\frac{\sum_{i=1}^m a_i^2\sigma_{0i}^2}{m},\\ \sigma_{FHS}^2(\alpha,\mathbf{a}) &= \frac{\alpha(1-\alpha)}{\phi^2(\xi_{1-2\alpha})}\sum_{i=1}^m a_i^2\sigma_{0i}^2, \qquad \sigma_U^2(\alpha,\mathbf{a}) = \frac{2\alpha(1-2\alpha)}{4\phi^2(\xi_{1-2\alpha})}\sum_{i=1}^m a_i^2\sigma_{0i}^2.\end{aligned}$$

Moreover, $\sigma_S^2(\alpha,\mathbf{a}) < \sigma_U^2(\alpha,\mathbf{a}) < \sigma_{FHS}^2(\alpha,\mathbf{a})$ when $m \geq 2$.

[6] Symmetrically, it is possible to take fixed units of each asset in the composition of the portfolio. A portfolio will be called *crystallized* if, for each $i = 1,\dots m$, we have $\mu_{i,t-1} = \mu_i$ for all t.

Remark 3.1. For the static Gaussian model with $m \geq 2$, the multivariate estimator is thus asymptotically strictly more efficient than the univariate estimator, and the efficiency ratio is given by

$$\frac{\sigma_S^2(\alpha, \mathbf{a})}{\sigma_U^2(\alpha, \mathbf{a})} = \frac{1}{m}\left[1 + \frac{\xi_{1-2\alpha}^2 \phi^2(\xi_{1-2\alpha})}{\alpha(1-2\alpha)}\left\{\frac{\frac{1}{m}\sum_{i=1}^m a_i^4 \sigma_{0i}^4}{\left(\frac{1}{m}\sum_{i=1}^m a_i^2 \sigma_{0i}^2\right)^2} - 1\right\}\right].$$

It follows that

$$\frac{1}{m} \leq \frac{\sigma_S^2(\alpha, \mathbf{a})}{\sigma_U^2(\alpha, \mathbf{a})} \leq \frac{1}{m}\left[1 + (m-1)\frac{\xi_{1-2\alpha}^2 \phi^2(\xi_{1-2\alpha})}{\alpha(1-2\alpha)}\right]. \tag{3.2}$$

The lower bound is reached when the weight of each asset i in the portfolio is proportional to $1/\sigma_i$. The upper bound is reached when the portfolio reduces to one asset ($a_{oi} = 0$ for all except one i). It is maximized, for $\alpha = 0.069...$, by $0.652 + \frac{0.348}{m}$.

Remark 3.2. The computations required to obtain the asymptotic variance $\sigma_S^2(\alpha, \mathbf{a})$ are hardly extendable to the case where $\boldsymbol{\eta}_t$ follows another spherical distribution than the Gaussian. Simulation experiments reported in Appendix E show that for some fat tailed distributions the univariate method may be more accurate than the multivariate method.

Remark 3.3 (Estimation effect on the asymptotic accuracies)**.** In the multivariate estimation of the VaR, the estimation of $\boldsymbol{\vartheta}_0$ occurs in two places: in the estimation of $\{\widetilde{\mathbf{a}}'\boldsymbol{\vartheta}_0\}^{1/2}$ and in the estimator $\xi_{n,1-2\alpha}$ of the residuals quantile. To separate the two effects, let us introduce the infeasible estimator of the VaR

$$\widetilde{\mathrm{VaR}}_{t-1}^{(\alpha)}(\mathbf{a}'\boldsymbol{\epsilon}_t) = \{\widetilde{\mathbf{a}}'\boldsymbol{\vartheta}_0\}^{1/2}\xi_{n,1-2\alpha}.$$

The asymptotic variance $\widetilde{\sigma}_S^2(\alpha, \mathbf{a})$ of $\sqrt{n}\left(\widetilde{\mathrm{VaR}}_{t-1}^{(\alpha)}(\mathbf{a}'\boldsymbol{\epsilon}_t) - \mathrm{VaR}_{t-1}^{(\alpha)}(\mathbf{a}'\boldsymbol{\epsilon}_t)\right)$ is given by

$$\widetilde{\sigma}_S^2(\alpha, \mathbf{a}) = \frac{1}{m}\sum_{i=1}^m a_i^2 \sigma_{0i}^2 \left(-\frac{\xi_{1-2\alpha}^2}{2} + \frac{2\alpha(1-2\alpha)}{4\phi^2(\xi_{1-2\alpha})}\right)$$

and the ratio of asymptotic efficiency of the univariate estimator with respect to this theoretical estimator is independent of the portfolio,

$$\frac{\widetilde{\sigma}_S^2(\alpha, \mathbf{a})}{\sigma_U^2(\alpha, \mathbf{a})} = \frac{1}{m}\left\{1 - \frac{\xi_{1-2\alpha}^2 \phi^2(\xi_{1-2\alpha})}{\alpha(1-2\alpha)}\right\}.$$

Unsurprisingly, this ratio varies in $1/m$, the quantile $\xi_{n,1-2\alpha}$ being based on m times more observations than the univariate estimator of the VaR. The negative second term in the bracket comes from the fact that, in the Gaussian framework, quantiles based on residuals are more accurate than quantiles based on observations of the i.i.d. process (see for instance Francq and Zakoïan (2015a)). It follows that

$$\frac{\sigma_S^2(\alpha,\mathbf{a})}{\sigma_U^2(\alpha,\mathbf{a})}=\frac{\widetilde{\sigma}_S^2(\alpha,\mathbf{a})}{\sigma_U^2(\alpha,\mathbf{a})}+\frac{1}{m}\frac{\xi_{1-2\alpha}^2\phi^2(\xi_{1-2\alpha})}{\alpha(1-2\alpha)}\frac{\frac{1}{m}\sum_{i=1}^m a_i^4\sigma_{0i}^4}{\left(\frac{1}{m}\sum_{i=1}^m a_i^2\sigma_{0i}^2\right)^2},$$

where the first term in the right-hand side represents the effect of the estimation of $\boldsymbol{\vartheta}_0$ on the quantile of the iid process. The second term represents the price paid, in the multivariate method, for the estimation of $\boldsymbol{\vartheta}_0$ in $\{\widetilde{\mathbf{a}}'\boldsymbol{\vartheta}_0\}^{1/2}$.

3.3 Optimal-VaR portfolios

In the spherical case, the optimal-VaR portfolio is obtained in closed form, by Proposition 2.1, and it coincides with the Markowitz portfolio in the absence of conditional mean ($\boldsymbol{m}_t(\boldsymbol{\vartheta}_0)=\mathbf{0}$). None of these properties continues to hold in the non-spherical case. The portfolio with the smallest VaR, at a given level α, is defined by

$$\epsilon_t^{(\alpha)}=\boldsymbol{\epsilon}_t'\boldsymbol{a}_{t-1}^{(\alpha)},\quad \boldsymbol{a}_{t-1}^{(\alpha)}=\arg\min_{\boldsymbol{a}:\,\boldsymbol{a}'\boldsymbol{e}=1}\mathrm{VaR}_{t-1}^{(\alpha)}\left\{\boldsymbol{a}'\boldsymbol{m}_t(\boldsymbol{\vartheta}_0)+\boldsymbol{a}'\boldsymbol{\Sigma}_t(\boldsymbol{\vartheta}_0)\boldsymbol{\eta}_t\right\}. \tag{3.3}$$

In practical situation, $\boldsymbol{\theta}_0$ is unknown but the optimal-VaR portfolio can be estimated by $\widehat{\epsilon}_t^{(\alpha)}=\boldsymbol{\epsilon}_t'\widehat{\boldsymbol{a}}_{t-1}^{(\alpha)}$ where

$$\widehat{\boldsymbol{a}}_{t-1}^{(\alpha)}=\arg\min_{\boldsymbol{a}:\,\boldsymbol{a}'\boldsymbol{e}=1}-q_\alpha\left\{\boldsymbol{a}'\boldsymbol{m}_t(\widehat{\boldsymbol{\vartheta}}_n)+\boldsymbol{a}'\widetilde{\boldsymbol{\Sigma}}_t(\widehat{\boldsymbol{\vartheta}}_n)\widehat{\boldsymbol{\eta}}_u,u=1,\dots,n\right\}.$$

4 Numerical illustrations

The first two parts of the section presents a selection of Monte-Carlo experiments aiming at studying the performance of the previous approaches in finite sample. Real data examples will be presented in the third part.[7]

4.1 Invalidity of the univariate approach when the composition is time varying

For simplicity, we consider a crystallized equally weighted portfolio of 3 assets (of initial price $p_{i0}=1000$) $V_t=\sum_{i=1}^3 p_{it}$. Thus, the return portfolio composition is time varying, with coefficients

[7] The code and data used in the paper are available on the web site
http://perso.univ-lille3.fr/~cfrancq/Christian-Francq/VaRPortfolio.html

$\mathbf{a}_{t-1} = (a_{1,t-1}, a_{2,t-1}, a_{3,t-1})'$ and $a_{i,t-1} = p_{i,t-1}/\sum_{j=1}^{3} p_{j,t-1}$. Assume that the vector of the log-returns is iid, centered, with variance $\mathrm{Var}(\boldsymbol{\epsilon}_t) = \boldsymbol{\Sigma}^2 = \boldsymbol{DRD}$, with

$$\boldsymbol{D} = \begin{pmatrix} 0.01 & 0 & 0 \\ 0 & 0.02 & 0 \\ 0 & 0 & 0.04 \end{pmatrix}, \quad \boldsymbol{R} = \begin{pmatrix} 1 & -0.855 & 0.855 \\ -0.855 & 1 & -0.810 \\ 0.855 & -0.810 & 1 \end{pmatrix}.$$

The composition $\boldsymbol{a}_{t-1}$ of the portfolio is plotted in Figure 2. It can be shown that this vector is non stationary[8]. More precisely, by the Chung-Fuks theorem, more and more frequently the composition $\boldsymbol{a}_{t-1}$ of the portfolio approaches one of the three single-asset portfolios $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$.

It is thus non surprising to see that the univariate return series $\epsilon_t^{(P)}$ (not reported here) exhibits some nonstationarity features, in particular marginal heteroscedasticity. However, because the series also presents conditional heteroscedasticity, we fitted a GARCH(1,1) model which corresponds to common practice. The parameters of this model are estimated online, starting from $t = 200$. As in Section 3.2, $\widehat{\mathrm{VaR}}^{(\alpha)}_{FHS,t-1}(\epsilon^{(P)}) = -q_\alpha\left(\{\boldsymbol{a}'_{t-1}\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{a}'_{t-1}\boldsymbol{\epsilon}_{t-1}\}\right)$. These empirical quantiles were computed starting from $t = 150$. The spherical method, based on the estimation of $\boldsymbol{\Sigma}$, was computed on the same range of observations. Figure 3 displays the sample paths of the true conditional VaR as well as the 3 estimated VaRs. It can be seen that the spherical method converges faster to the true value than the FHS method. On the other hand, the univariate method fails to converge to the theoretical conditional VaR. This can be explained by the difference between the information sets (point iii) in Section 1.2), and also by the non stationarity of the univariate series of portfolio returns. appropriate for this non stationary series.

The results of this section are in agreement with Santos et al. (2013) who found that, on real and simulated series, multivariate models outperform univariate models. Therefore, we shall not consider the univariate approach in the subsequent illustrations.

4.2 Comparison of the multivariate approaches on DCC models

In this section, we consider more involved/realistic models, namely the Dynamic Conditional Correlation (DCC) GARCH model (see Appendix D for a presentation).

We simulated N independent trajectories of length n for the corrected DCC (cDCC) GARCH(1,1) model of Aielli (2013). On each simulation, the first n_1 observations are used to

[8]Indeed, the ratio $\log(a_{1,t}/a_{2,t}) = \sum_{k=1}^{t}(\epsilon_{1,k} - \epsilon_{2,k})$ is non stationary: the non-singularity of $\boldsymbol{\Sigma}$ entails that the variance of $\epsilon_{1,k} - \epsilon_{2,k}$ is non degenerated. This property holds under more general assumptions, for instance if the sequence $(\epsilon_{1,k} - \epsilon_{2,k})$ is mixing and nondegenerated.

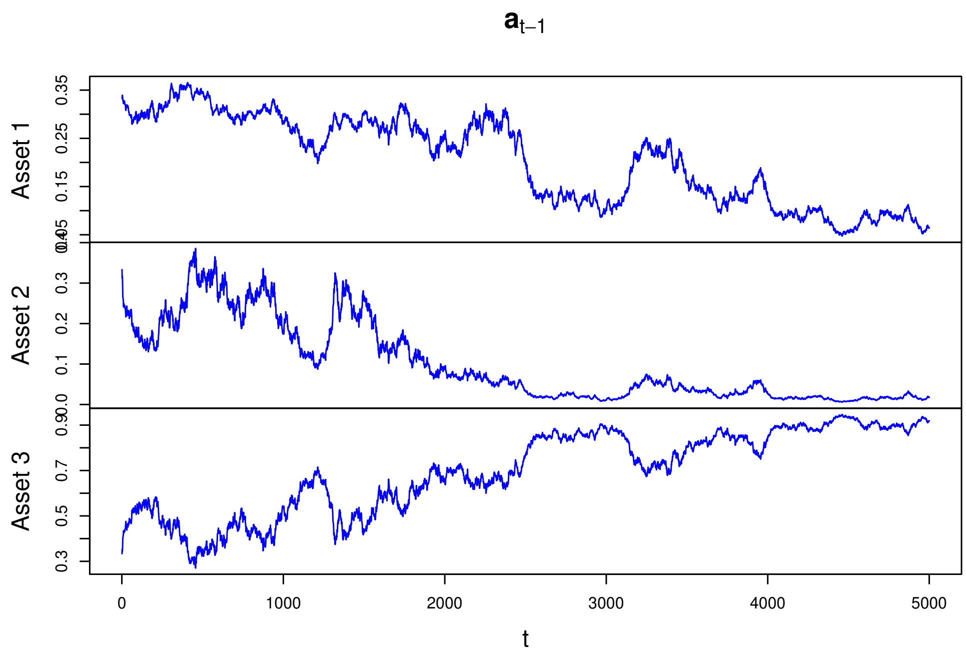

Figure 2: Time-varying composition of the crystallized portfolio.

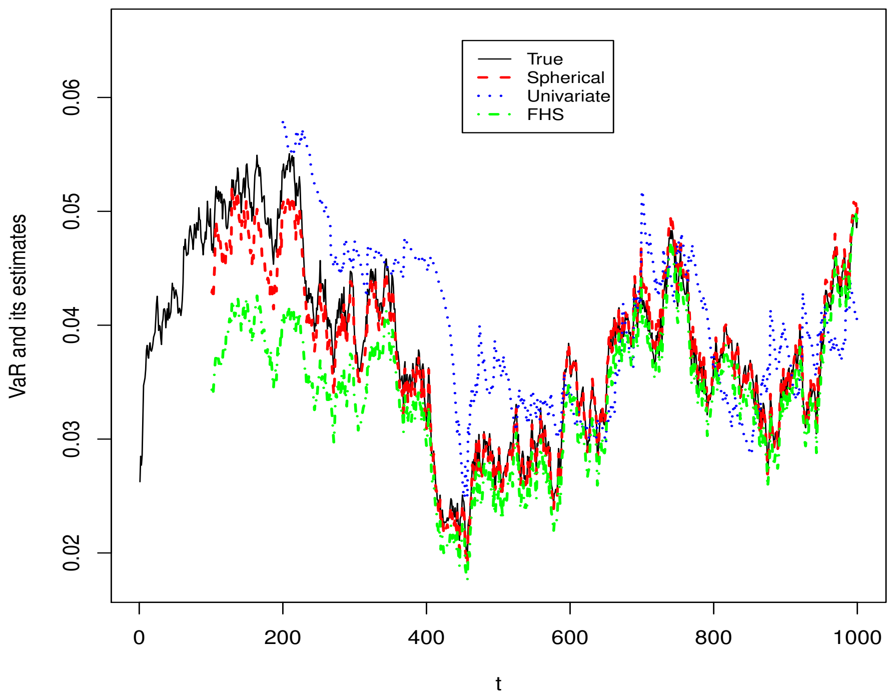

Figure 3: True and estimated VaRs of the crystallized portfolio.

obtain an estimator $\widehat{\boldsymbol{\vartheta}}_{n_1}$ of $\boldsymbol{\vartheta}_0$ by the three-step estimator defined in Appendix D, and to compute $\xi_{n_1,1-2\alpha} = q_{1-2\alpha}\{|\widehat{\eta}_{it}|, i=1,\dots,m, t=1,\dots,n_1\}$. On the last $n-n_1$ simulations, *i.e.* for $t=n_1+1,\dots,n$, we compared the theoretical $\mathrm{VaR}^{(\alpha)}_{t-1}\left(\epsilon_t^{(P)*}\right)$ of the Markowitz portfolio (2.8) with the two estimates obtained from the spherical and FHS methods, given respectively by

$$\widehat{\mathrm{VaR}}^{(\alpha)}_{S,t-1}(\epsilon^{(P)*}) = \frac{\xi_{n_1,1-2\alpha}}{\sqrt{\boldsymbol{e}'\widetilde{\boldsymbol{\Sigma}}_t^{-2}(\widehat{\boldsymbol{\vartheta}}_{n_1})\boldsymbol{e}}} \tag{4.1}$$

and

$$\widehat{\mathrm{VaR}}^{(\alpha)}_{FHS,t-1}(\epsilon^{(P)*}) = -q_\alpha\left(\left\{\frac{\boldsymbol{e}'\widetilde{\boldsymbol{\Sigma}}_t^{-1}(\widehat{\boldsymbol{\vartheta}}_{n_1})\widehat{\boldsymbol{\eta}}_u}{\boldsymbol{e}'\widetilde{\boldsymbol{\Sigma}}_t^{-2}(\widehat{\boldsymbol{\vartheta}}_{n_1})\boldsymbol{e}}, u=1,\dots,n_1\right\}\right). \tag{4.2}$$

We considered portfolios of $m=2$ assets. The different designs, displayed in Appendix D, correspond to spherical (designs A-H) or non spherical (designs A*-H*) distributions.

We took $N=100$ independent replications, and $n-n_1=1000$ out-of-sample predictions for each simulation. In each design, we then compared the corresponding $10,000$ theoretical values of the VaR defined by (2.9) with their estimates (4.1)-(4.2) obtained by the spherical and FHS methods. Denote by MSE_S and MSE_{FHS} the mean square errors (MSE) of prediction of the two methods. Table 1 displays the relative efficiency (RE) of the spherical method with respect to the FHS method, as measured by the ratio $\mathrm{MSE}_{FHS}/\mathrm{MSE}_S$. In Designs A-H, the spherical method is generally more efficient than the FHS method (for Designs C and D, the spherical method can be two times more efficient than the other method). This is not surprising because the distribution of the innovations is spherical in each of the designs A-H. The bottom panel of Table 1 shows that, when the density is strongly asymmetric, the FHS method can be much more efficient than the spherical method. It can be seen that the empirical REs decrease when the sample size increases, reflecting the inconsistency of the spherical method.

From Table 1 and other simulations experiments conducted with crystallized and minimal-VaR portfolios (see Appendix E), the two multivariate methods appear comparable when the conditional distribution is spherical. Both are quite satisfactory and in agreement with the theoretical results. In the non spherical case, the spherical approach is no longer reliable contrary to the FHS method.

4.3 Optimal portfolios of exchange rates

We considered the daily returns of 5 exchange rates against the Euro: the Canadian Dollar (CAD), the Chinese Yuan (CNY), the British Pound (GBP), the Japanese Yen (JPY) and the United States

Table 1: Relative efficiency of the Spherical method with respect to the FHS method for the Markowitz portfolio.

n_1	α	A	B	C	D	E	F	G	H
500	1%	1.181	1.109	2.567	2.350	1.076	1.174	1.232	1.424
	5%	1.209	1.029	1.813	1.403	1.181	1.115	1.122	1.186
1000	1%	1.301	1.105	2.354	1.623	1.533	1.511	1.572	1.549
	5%	1.144	1.025	2.070	0.999	1.249	1.077	1.332	1.011
		A*	B*	C*	D*	E*	F*	G*	H*
500	1%	0.077	0.061	0.027	0.037	0.115	0.104	0.102	0.122
	5%	0.325	0.321	0.133	0.250	0.447	0.518	0.462	0.526
1000	1%	0.079	0.032	0.016	0.018	0.059	0.030	0.029	0.036
	5%	0.344	0.188	0.094	0.108	0.304	0.239	0.211	0.290

Dollar (USD). The data come from the European Central Bank and cover the period from January 14, 2000 to May 5, 2015. The total number of observations is $n = 2582$.

We first estimated a BEKK model on the 5 exchange rates over the whole sample except the last 100 returns. We consider an equally-weighted crystalized portfolio ($\mu_i = 1$ for $i = 1, \dots, 5$). Figure 4, displaying the last 100 returns of the portfolio, shows that one return (23/01/2015[9]) is clearly below the lower bound of the 95%-CI of the 1%-VaR. For such a return, there is strong evidence of violation of the theoretical VaR. For three other returns belonging to the CI (18/12/2014, 15/01/2015 and 25/02/2015), 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 $n_1 = 2000$ observations and computed the residuals $\widehat{\boldsymbol{\eta}}_u, u = 1, \dots, n_1$. Instead of crystalized portfolios, we considered optimal portfolios. The top panels of Figure 5 display the returns of the estimated Markowitz portfolio

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

together with $\widehat{\mathrm{VaR}}_{S,t-1}^{(1\%)}(\epsilon^{(P)*})$ (left panel) and $\widehat{\mathrm{VaR}}_{FHS,t-1}^{(1\%)}(\epsilon^{(P)*})$ (right panel), as defined by (4.1)-(4.2). The most striking output is that the two methods give virtually indistinguishable estimated VaRs for the Markowitz portfolio. The bottom panels present the sample paths of the portfolios $\widehat{\epsilon}_t^{(\alpha)}$ of minimal VaR (at levels 1% and 5%) together with their VaRs defined by

[9] The European Central Bank announced a large-scale bond-buying program to address the risks of deflation in Eurozone which entailed large exchange rates variations.

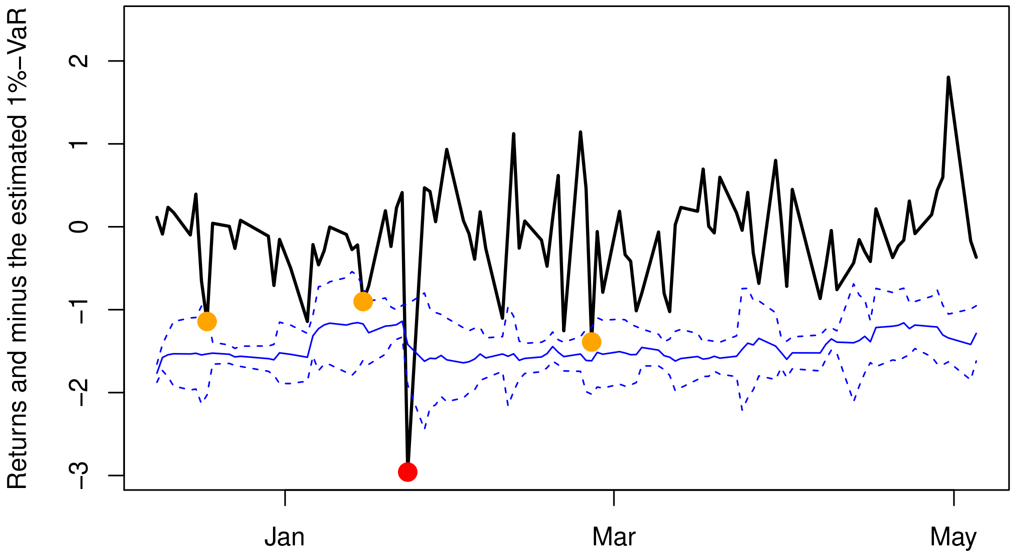

Figure 4: Returns of the portfolio (dark line) for the period 09/12/2014 to 05/05/2015, estimated 1%-VaR and 95%-confidence interval (full and dotted blue lines), based on the estimation of a BEKK model for the exchange rates.

$-q_\alpha\left(\{\widehat{\boldsymbol{\eta}}_u'\widetilde{\boldsymbol{\Sigma}}_t(\widehat{\boldsymbol{\vartheta}}_{n_1})\widehat{\boldsymbol{a}}_{t-1}^{(\alpha)}, u=1,\dots,n_1\}\right)$. The global shapes of these portfolios paths are similar to Markowitz portfolio's path, but they differ at some points. In the spherical case the optimal portfolios (VaR and Markowitz) should coincide, but the difference could be due to estimation or optimization. Applying the sphericity test recently proposed by Francq, Jimenez Gamero and Meintanis (2015), we found that the sphericity hypothesis cannot be rejected at any reasonable level.[10] Table 2 provides the p-values of three backtests (see Christoffersen (2003) for details): the unconditional coverage (UC) test that the probability of violation is equal to the nominal level α, 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 two cases. For the Markowitz portfolio and, to a lesser extent for the 5% minimal VaR portfolio, with both VaRs estimated by FHS, the numbers of violations are below the 5% level. In view of the sphericity test and these backtests, the spherical approach seems more reliable than the FHS on these data.

[10] 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 0.73.

5 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, by circumventing the non stationarity of the resulting portfolio, multivariate approaches are more reliable than the univariate approach based on the sole univariate modeling of the portfolio's returns. Moreover, they account for richer information than aggregate information conveyed by the past portfolio returns. Beyond these intuitive arguments, we established, both theoretically and empirically, the invalidity of the univariate approach. We also 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 IC's on the VaR parameter; iii) to obtain the minimal-VaR portfolios in closed form and estimate their conditional VaRs. 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 allow us to draw the following lessons, by distinguishing three 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) **Determining optimal-VaR portfolios** is greatly facilitated under the sphericity assumption. Without this assumption, the composition of an optimal-VaR portfolio has to be determined numerically, which can be cumbersome in high dimension.

iii) **Evaluating the asymptotic accuracy** of the conditional VaR estimators can be achieved using Theorems 2.1 and 3.1. Implementation of the latter asymptotic results is more involved but is worthwhile when sphericity is doubtful.

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

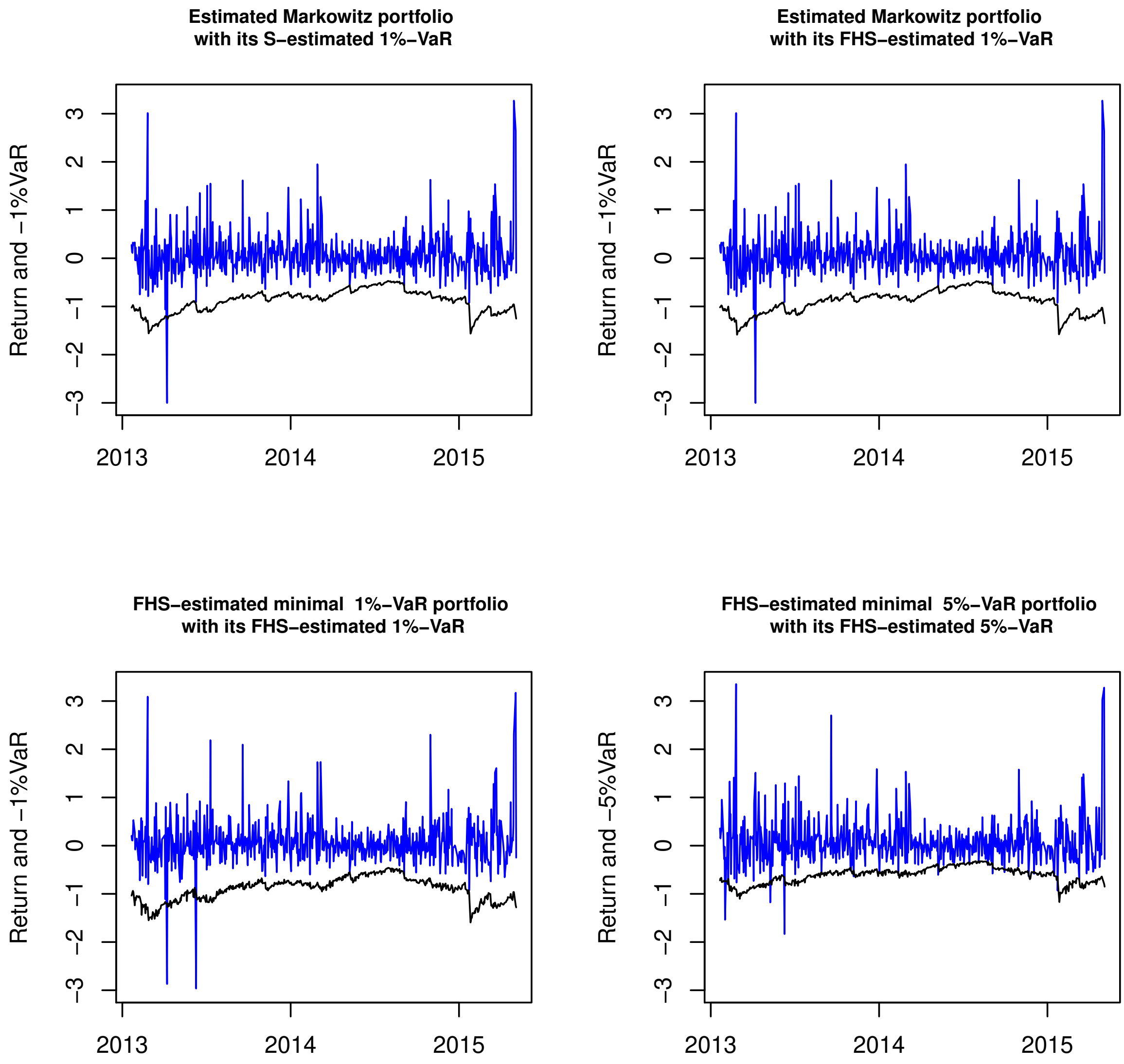

Figure 5: Returns of estimated optimal portfolios of 5 exchange rates and their estimated VaR's.

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.

Table 2: p-values of three backtests

Portfolio	Method	α	% of Viol	UC	IND	CC
Markowitz	Spherical	1%	2/582	0.065	0.906	0.182
Markowitz	FHS	1%	2/582	0.065	0.906	0.182
Minimal 1%-VaR	FHS	1%	3/582	0.195	0.860	0.426
Markowitz	Spherical	5%	20/582	0.067	0.232	0.092
Markowitz	FHS	5%	18/582	0.023	0.283	0.043
Minimal 5%-VaR	FHS	5%	19/582	0.041	0.257	0.065

Appendices

A Illustrations of the Bahadur representation A3

A.1 For the Gaussian QML

Let us illustrate (2.2) in Assumption **A3** when $\boldsymbol{m}(\cdot) = 0$ and the criterion used to estimate $\boldsymbol{\theta}_0 = \boldsymbol{\vartheta}_0$ 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 (2.2) 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 (2015b) 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

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

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

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

where σ_k is a positive function and η_{kt}^* is the k-th component of $\boldsymbol{R}^{1/2}\boldsymbol{\eta}_t$ (see Francq and Zakoian (2015b) 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 $\boldsymbol{\vartheta} = \boldsymbol{\theta} := (\boldsymbol{\zeta}', \boldsymbol{\rho}')'$ here consists in the volatility parameters $\boldsymbol{\zeta} = (\zeta^{(1)'}, \dots, \zeta^{(m)'})'$ and the correlation parameters

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

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 (2015b) shows that (2.2) 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 $\boldsymbol{\vartheta} = (\boldsymbol{h}', \boldsymbol{\gamma}')'$, with

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

where

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

and

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

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

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

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

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

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

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

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

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

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

Thus, (2.2) in Assumption **A3** holds for the VTE of the CCC model with, in particular,

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

B Proofs

B.1 Proof of Theorem 2.1

Note that

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

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

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

where

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

Let $\boldsymbol{e}_k$ denote the k-th column of the $m \times m$ identity matrix $\boldsymbol{I}_m$. Let $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_t(\boldsymbol{\vartheta}_0)$. Let $\boldsymbol{\eta}_t(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\vartheta})\{\boldsymbol{\epsilon}_t - \boldsymbol{m}_t(\boldsymbol{\varphi})\} = (\eta_{1t}(\boldsymbol{\theta}), \dots, \eta_{mt}(\boldsymbol{\theta}))'$. We have, for $j = 1, \dots, d_1$,

$$\frac{\partial \eta_{kt}}{\partial \theta_j}(\boldsymbol{\theta}_0) = -\boldsymbol{e}_k' \boldsymbol{\Sigma}_t^{-1} \frac{\partial \boldsymbol{m}_t}{\partial \theta_j}$$

and for $j = d_1 + 1, \dots, d$,

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

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

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

A Taylor expansion of $\eta_{kt}(\boldsymbol{\theta})$ around $\boldsymbol{\theta}_0$ thus yields, with obvious notations for the components of $\boldsymbol{\varphi}$ and $\boldsymbol{\vartheta}$,

$$\begin{aligned}
\widehat{\eta}_{kt} &= \eta_{kt} - \sum_{j=1}^{d_1} \boldsymbol{e}_k' \boldsymbol{\Sigma}_t^{-1} \frac{\partial \boldsymbol{m}_t}{\partial \varphi_j} (\widehat{\varphi}_{nj} - \varphi_{0j}) \\
&\quad - \sum_{j=1}^{d_2} \sum_{\ell=1}^{m} \eta_{\ell t} \boldsymbol{e}_k' \boldsymbol{\Sigma}_t^{-1} \left\{\frac{\partial}{\partial \vartheta_j} \boldsymbol{\Sigma}_{\cdot \ell, t}\right\} (\widehat{\vartheta}_{nj} - \vartheta_{0j}) + o_P(n^{-1/2}) \\
&= \eta_{kt} - \boldsymbol{C}_{kt}'(\widehat{\boldsymbol{\varphi}}_n - \boldsymbol{\varphi}_0) - \boldsymbol{\eta}_t' \boldsymbol{\Omega}_{kt}^* (\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) + o_P(n^{-1/2}) \\
&= \eta_{kt} - \boldsymbol{M}_{kt}'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o_P(n^{-1/2}).
\end{aligned} \tag{B.1}$$

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

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

where $u_{kt} = \pm 1$, the sign of u_{kt} being equal to that of η_{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

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

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

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

where

$$\begin{aligned}
X_{n,k} &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\mathbf{1}_{\{|\eta_{kt}|<\xi_{1-2\alpha}\}}-1+2\alpha),\\
Y_{n,k} &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}R_{t,n,k}(\mathbf{1}_{\{|\eta_{kt}|<\xi_{1-2\alpha}\}}-1+2\alpha),\\
I_{n,k}(z) &= \sum_{t=1}^{n}\int_0^{z/\sqrt{n}}(\mathbf{1}_{\{|\eta_{kt}|\le\xi_{1-2\alpha}+s\}}-\mathbf{1}_{\{|\eta_{kt}|<\xi_{1-2\alpha}\}})ds,\\
J_{n,k}(z) &= \sum_{t=1}^{n}\int_{z/\sqrt{n}}^{(z+R_{t,n,k})/\sqrt{n}}(\mathbf{1}_{\{|\eta_{kt}|\le\xi_{1-2\alpha}+s\}}-\mathbf{1}_{\{|\eta_{kt}|<\xi_{1-2\alpha}\}})ds,
\end{aligned}$$

with $R_{t,n,k}\overset{o_P(1)}{=}u_{kt}\boldsymbol{M}'_{kt}\sqrt{n}(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)$. 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

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

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

$$\begin{aligned}
J^{(2)}_{n,k}(z) &= \sum_{t=1}^{n}\left\{u_{kt}\boldsymbol{M}'_{kt}(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)+o_P(n^{-1/2})\right\}\mathbf{1}^*_{\{|\eta_{kt}|-\xi_{1-2\alpha}\in(0,z/\sqrt{n})\}}\\
&\overset{o_P(1)}{=} \left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}u_{kt}\mathbf{1}^*_{\{|\eta_{kt}|-\xi_{1-2\alpha}\in(0,z/\sqrt{n})\}}\boldsymbol{M}'_{kt}\right)\sqrt{n}(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)\\
&= \left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}u_{kt}\mathbf{1}^*_{\{|\eta_{kt}|-\xi_{1-2\alpha}\in(0,z/\sqrt{n})\}}\boldsymbol{C}'_{kt}\right)\sqrt{n}(\widehat{\boldsymbol{\varphi}}_n-\boldsymbol{\varphi}_0)\\
&\quad+\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}u_{kt}\mathbf{1}^*_{\{|\eta_{kt}|-\xi_{1-2\alpha}\in(0,z/\sqrt{n})\}}\boldsymbol{\eta}'_t\boldsymbol{\Omega}^*_{kt}\right)\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n-\boldsymbol{\vartheta}_0).
\end{aligned}$$

Note that, for $z>0$,

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

in view of the symmetry of the distribution of η_{kt} under the sphericity assumption **A1**. The same equality holds for $z\le 0$.

Now, for $z>0$ and $\ell \neq k$,

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

because $(\eta_{\ell t},\eta_{kt})$ and $(\eta_{\ell t},-\eta_{kt})$ have the same distribution under **A1**. For $k=\ell$ we have

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

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

$$\begin{aligned}
& E\left(\frac{1}{\sqrt{n}}\sum_{t=1}^n u_{kt}\mathbf{1}^*_{\{|\eta_{kt}|-\xi_{1-2\alpha}\in(0,z/\sqrt{n})\}}\boldsymbol{M}'_{kt}\right) \\
\stackrel{o_P(1)}{=} \;& z\xi_{1-2\alpha}f(\xi_{1-2\alpha})\begin{bmatrix}\mathbf{0}_{1\times d_1} & \boldsymbol{e}'_k E\left(\boldsymbol{\Sigma}_t^{-1}\left\{\frac{\partial}{\partial\boldsymbol{\vartheta}'}\boldsymbol{\Sigma}_{\cdot k,t}\right\}\right)\end{bmatrix}.
\end{aligned}$$

Similar arguments show that

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

It follows that

$$\begin{aligned}
J^{(2)}_{n,k}(z) \;&\stackrel{o_P(1)}{=}\; z\xi_{1-2\alpha}f(\xi_{1-2\alpha})\begin{bmatrix}\mathbf{0}_{1\times d_1} & \boldsymbol{e}'_k E\left(\boldsymbol{\Sigma}_t^{-1}\left\{\frac{\partial}{\partial\boldsymbol{\vartheta}'}\boldsymbol{\Sigma}_{\cdot k,t}\right\}\right)\end{bmatrix}\sqrt{n}(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0) \\
&\stackrel{o_P(1)}{=}\; z\xi_{1-2\alpha}f(\xi_{1-2\alpha})\boldsymbol{e}'_k E\left(\boldsymbol{\Sigma}_t^{-1}\left\{\frac{\partial}{\partial\boldsymbol{\vartheta}'}\boldsymbol{\Sigma}_{\cdot k,t}\right\}\right)\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n-\boldsymbol{\vartheta}_0),
\end{aligned}$$

and

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

Moreover,

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

As in Francq and Zakoian (2015a), 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,

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

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

$$\begin{aligned} &\sqrt{n}(\xi_{n,1-2\alpha} - \xi_{1-2\alpha}) \\ \overset{o_P(1)}{=} \ & -\frac{\xi_{1-2\alpha}}{m}\boldsymbol{\Omega}\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) - \frac{1}{f(\xi_{1-2\alpha})}\frac{1}{m\sqrt{n}}\sum_{t=1}^n \boldsymbol{e}'\boldsymbol{\nu}_\alpha(\boldsymbol{\eta}_t). \end{aligned} \tag{B.2}$$

In view of **A3** we have

$$\mathrm{Cov}_{as}\left(\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0), \frac{1}{m\sqrt{n}}\sum_{t=1}^n \boldsymbol{e}'\boldsymbol{\nu}_\alpha(\boldsymbol{\eta}_t)\right) = \frac{1}{m}\boldsymbol{\Lambda}\boldsymbol{W}_\alpha,$$

and thus,

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

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

The convergence in distribution (2.3) follows by the Central Limit Theorem of Billingsley (1961) for ergodic, stationary and square integrable martingale differences, applied to the sequence $\begin{pmatrix} \boldsymbol{\Delta}_{t-1}\boldsymbol{V}(\boldsymbol{\eta}_t) \\ \boldsymbol{e}'\boldsymbol{\nu}_\alpha(\boldsymbol{\eta}_t) \end{pmatrix}$.

To conclude, we prove the nonsingularity of matrix $\boldsymbol{\Xi}$. Suppose that $(\boldsymbol{x}', y)\boldsymbol{\Xi}(\boldsymbol{x}', y)' = 0$ where $\boldsymbol{x} \in \mathbb{R}^d, y \in \mathbb{R}$. In view of the expansion

$$\begin{aligned} &\sqrt{n}\begin{pmatrix} \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \\ \xi_{n,1-2\alpha} - \xi_{1-2\alpha} \end{pmatrix} \\ \overset{o_P(1)}{=} &\begin{pmatrix} \boldsymbol{I}_d & \boldsymbol{0} \\ -\frac{\xi_{1-2\alpha}}{m}[\boldsymbol{0}_{1\times d_1} \quad \boldsymbol{\Omega}] & \frac{-1}{f(\xi_{1-2\alpha})} \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{t=1}^n \boldsymbol{\Delta}_{t-1}\boldsymbol{V}(\boldsymbol{\eta}_t) \\ \frac{1}{m\sqrt{n}}\sum_{t=1}^n \boldsymbol{e}'\boldsymbol{\nu}_\alpha(\boldsymbol{\eta}_t) \end{pmatrix} \end{aligned}$$

we must have

$$\boldsymbol{x}'\boldsymbol{\Delta}_{t-1}\boldsymbol{V}(\boldsymbol{\eta}_t) + y\left\{-\frac{\xi_{1-2\alpha}}{m}[\boldsymbol{0}_{1\times d_1} \quad \boldsymbol{\Omega}]\boldsymbol{\Delta}_{t-1}\boldsymbol{V}(\boldsymbol{\eta}_t) - \frac{1}{mf(\xi_{1-2\alpha})}\boldsymbol{e}'\boldsymbol{\nu}_\alpha(\boldsymbol{\eta}_t)\right\} = c, \quad a.s.$$

for some constant c. Because $V(\boldsymbol{\eta}_t)$ and $\boldsymbol{\nu}_\alpha(\boldsymbol{\eta}_t)$ are centered, we must have $c = 0$. By **A4**, it follows that $\boldsymbol{x}'\boldsymbol{\Delta}_{t-1} - y\frac{\xi_{1-2\alpha}}{m}[\boldsymbol{0}_{1\times d_1} \quad \boldsymbol{\Omega}]\boldsymbol{\Delta}_{t-1} = 0$ and $y\frac{1}{mf(\xi_{1-2\alpha})}\boldsymbol{e} = \boldsymbol{0}$. The last equality entails $y = 0$ from which it follows that $\boldsymbol{x}'\boldsymbol{\Delta}_{t-1} = \boldsymbol{0} = \boldsymbol{x}'\boldsymbol{\Lambda}$. Because $\boldsymbol{\Lambda}$ is full row rank, this entails $\boldsymbol{x} = \boldsymbol{0}$ and the proof is complete. $\square$

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 η_t instead of η_{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 $\epsilon > 0$,

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

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

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

in probability as $n \to \infty$. □

B.3 Proof of Corollary 2.1

The asymptotic normality follows from Theorem 2.1 and the following Taylor expansion of G around $(\boldsymbol{\vartheta}_0, \xi_{1-2\alpha})$

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

□

B.4 Proof of Proposition 2.1

Using the Lagrangian method, define $\mathcal{L}(\boldsymbol{a}, \lambda) = q_{t-1}(\boldsymbol{a}) - \lambda(\boldsymbol{e}'\boldsymbol{a} - 1)$ for $\lambda \in \mathbb{R}$. The first order conditions write

$$\frac{\partial \mathcal{L}(\boldsymbol{a}, \lambda)}{\partial \boldsymbol{a}} = -\boldsymbol{m}_t + \frac{\xi_{1-2\alpha}}{\|\mathbf{a}'\boldsymbol{\Sigma}_t\|}\boldsymbol{\Sigma}_t^2\boldsymbol{a} - \lambda\boldsymbol{e} = 0, \qquad \boldsymbol{e}'\boldsymbol{a} = 1.$$

The optimum is thus of the form $\boldsymbol{a} = K\boldsymbol{\Sigma}_t^{-2}(\boldsymbol{m}_t + \lambda\boldsymbol{e})$, for some constant K. Provided that α is small enough so that $\xi_{1-2\alpha} > 0$, the first order conditions entail

$$K > 0, \quad \left\{(\boldsymbol{m}_t + \lambda\boldsymbol{e})'\boldsymbol{\Sigma}_t^{-2}(\boldsymbol{m}_t + \lambda\boldsymbol{e})\right\}^{1/2} = \xi_{1-2\alpha}, \quad K\boldsymbol{e}'\boldsymbol{\Sigma}_t^{-2}(\boldsymbol{m}_t + \lambda\boldsymbol{e}) = 1. \tag{B.3}$$

The first equality has two solutions in λ, provided that

$$\left(\boldsymbol{e}'\boldsymbol{\Sigma}_t^{-2}\boldsymbol{m}_t\right)^2 - \left(\boldsymbol{e}'\boldsymbol{\Sigma}_t^{-2}\boldsymbol{e}\right)\left(\boldsymbol{m}_t'\boldsymbol{\Sigma}_t^{-2}\boldsymbol{m}_t\right) + \left(\boldsymbol{e}'\boldsymbol{\Sigma}_t^{-2}\boldsymbol{e}\right)\xi_{1-2\alpha}^2 > 0. \tag{B.4}$$

This condition is satisfied for α small enough. Taking into account the first and third conditions of (B.3), there is a unique solution for the Lagrangian multiplicator λ. Finally, the optimal composition is given by (2.11) and the optimal VaR is $q_{t-1}(\boldsymbol{a}^*_{\alpha,t-1}) = \lambda$. □

B.5 Proof of Theorem 3.1

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

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

It follows from (B.1) that

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

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

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

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

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

Proceeding as in the proof of Theorem 2.1, we find that

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

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

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

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

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

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

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

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

By arguments already used, it follows that

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

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),

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

The convergence in distribution follows. The positivity of σ^2 is established using **A4*** and the arguments given to prove the non singularity of $\boldsymbol{\Xi}$ in Theorem 2.1. $\square$

B.6 Proof of Corollary 3.1

The first convergence in distribution is obtained by applying Theorem 2.1. Note that $f(x) = 2\phi(x)1_{x>0}$ and $\xi_{1-2\alpha} = \Phi^{-1}(1-\alpha)$. We thus have $\boldsymbol{\Psi} = 2\boldsymbol{\Sigma}^4(\boldsymbol{\vartheta}_0)$. The other terms involved in Theorem 2.1 are as follows. We have

$$\begin{aligned} \boldsymbol{\Omega} &= \frac{1}{2}\left(\sigma_{01}^{-2}, \ldots, \sigma_{0m}^{-2}\right)', \quad \boldsymbol{V}(\boldsymbol{\eta}_t) = (\eta_{1t}^2 - 1, \ldots, \eta_{mt}^2 - 1)', \\ \boldsymbol{\Delta}_{t-1} &= \boldsymbol{\Lambda} = \boldsymbol{\Sigma}^2(\boldsymbol{\vartheta}_0), \\ \boldsymbol{W}_\alpha &= 2\phi'(\xi_{1-2\alpha})\boldsymbol{e}, \quad \gamma_\alpha = 2m\alpha(1-2\alpha), \\ \boldsymbol{\Xi}_{\boldsymbol{\vartheta}\xi} &= \mathbf{0}_m, \quad \zeta_{1-2\alpha} = \frac{1}{2m}\left(-\xi_{1-2\alpha}^2 + \frac{\alpha(1-2\alpha)}{\phi^2(\xi_{1-2\alpha})}\right). \end{aligned}$$

Thus, we have

$$\begin{aligned}
& \sqrt{n}\left(\widehat{\mathrm{VaR}}^{(\alpha)}_{S,t-1}(\mathbf{a}'\boldsymbol{\epsilon}_t) - \mathrm{VaR}^{(\alpha)}_{t-1}(\mathbf{a}'\boldsymbol{\epsilon}_t)\right) \\
= & \sqrt{n}\left(\{\widetilde{\mathbf{a}}'\widehat{\boldsymbol{\vartheta}}_n\}^{1/2} - \{\widetilde{\mathbf{a}}'\boldsymbol{\vartheta}_0\}^{1/2}\right)\xi_{n,1-2\alpha} + \sqrt{n}\{\widetilde{\mathbf{a}}'\boldsymbol{\vartheta}_0\}^{1/2}(\xi_{n,1-2\alpha} - \xi_{1-2\alpha}) \\
\stackrel{d}{=} & \left(\frac{\xi_{1-2\alpha}\widetilde{\mathbf{a}}'}{2\{\widetilde{\mathbf{a}}'\boldsymbol{\vartheta}_0\}^{1/2}}, \{\widetilde{\mathbf{a}}'\boldsymbol{\vartheta}_0\}^{1/2}\right)\sqrt{n}\begin{pmatrix} \widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0 \\ \xi_{n,1-2\alpha} - \xi_{1-2\alpha}\end{pmatrix} \\
\xrightarrow{\mathcal{L}} & \mathcal{N}\left(\mathbf{0}, \left(\frac{\xi_{1-2\alpha}\widetilde{\mathbf{a}}'}{2\{\widetilde{\mathbf{a}}'\boldsymbol{\vartheta}_0\}^{1/2}}, \{\widetilde{\mathbf{a}}'\boldsymbol{\vartheta}_0\}^{1/2}\right)\begin{pmatrix} \boldsymbol{\Psi} & \boldsymbol{\Xi}_{\boldsymbol{\vartheta}\xi} \\ \boldsymbol{\Xi}'_{\boldsymbol{\vartheta}\xi} & \zeta_{1-2\alpha}\end{pmatrix}\begin{pmatrix} \frac{\widetilde{\mathbf{a}}\xi_{1-2\alpha}}{2\{\widetilde{\mathbf{a}}'\boldsymbol{\vartheta}_0\}^{1/2}} \\ \{\widetilde{\mathbf{a}}'\boldsymbol{\vartheta}_0\}^{1/2}\end{pmatrix}\right).
\end{aligned}$$

The first convergence in distribution follows. The two other convergences are standard results for the empirical quantiles of iid variables. The inequality $\sigma^2_S(\alpha, \mathbf{a}) < \sigma^2_U(\alpha, \mathbf{a})$ follows from (3.2) and the fact that $\xi^2_{1-2\alpha}\phi^2(\xi_{1-2\alpha})/2\alpha(1-2\alpha) < 0.326$.

Note that the asymptotic variance of the FHS estimator can be retrieved by applying Theorem 3.1: we find that $\boldsymbol{\omega} = \mathbf{0}$ and

$$x_0 = -\left(\sum_{i=1}^m a_i^2\sigma_{0i}^2\right)^{1/2}\phi^{-1}(1-\alpha).$$

□

C Estimating the asymptotic covariance matrix

In Theorem 2.1, most quantities involved in the asymptotic covariance matrix $\boldsymbol{\Xi}$ 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 σ_t or σ_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.

C.1 Linear SRE on $\boldsymbol{H}_t$

A typical example is the BEKK model of Engle and Kroner (1995). As in Pedersen and Rahbek (2013), we focus on the BEKK-GARCH(1,1) model, in which $\boldsymbol{\Sigma}_t(\boldsymbol{\vartheta}_0)$ is the symmetric square root of $\boldsymbol{H}_t$, given by

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

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 $\boldsymbol{\vartheta} = (\mathrm{vec}(\boldsymbol{A})', \mathrm{vec}(\boldsymbol{B})', \mathrm{vec}(\boldsymbol{C})')'$. The derivatives of $\mathrm{vec}(\boldsymbol{H}_t)$ can be computed as follows, omitting $\boldsymbol{\vartheta}$ 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 $j = 1, \dots, 3d$,

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

For any $m \times n$ matrix $\boldsymbol{M}$, let the $dm \times n$ matrix $\partial \boldsymbol{M} = \left(\frac{\partial \boldsymbol{M}'}{\partial \vartheta_1}, \dots, \frac{\partial \boldsymbol{M}'}{\partial \vartheta_d}\right)'$. Let $\boldsymbol{X}_t = (\mathrm{vec}'(\boldsymbol{H}_t), \{\partial \mathrm{vec}(\boldsymbol{H}_t)\}')'$. We have, in block matrix notation,

$$\boldsymbol{X}_t = \begin{pmatrix} \boldsymbol{B} \otimes \boldsymbol{B} & \boldsymbol{0} \\ \partial(\boldsymbol{B} \otimes \boldsymbol{B}) & \boldsymbol{I}_d \otimes (\boldsymbol{B} \otimes \boldsymbol{B}) \end{pmatrix} \boldsymbol{X}_{t-1} + \boldsymbol{e}_t, \tag{C.2}$$

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 (C.2) 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

$$\boldsymbol{\Sigma}_t \frac{\partial \boldsymbol{\Sigma}_t}{\partial \vartheta_i} + \frac{\partial \boldsymbol{\Sigma}_t}{\partial \vartheta_i}\boldsymbol{\Sigma}_t = \frac{\partial \boldsymbol{H}_t}{\partial \vartheta_i}.$$

Thus

$$(\boldsymbol{I}_m \otimes \boldsymbol{\Sigma}_t + \boldsymbol{\Sigma}_t \otimes \boldsymbol{I}_m)\,\mathrm{vec}\left(\frac{\partial \boldsymbol{\Sigma}_t}{\partial \vartheta_i}\right) = \mathrm{vec}\left(\frac{\partial \boldsymbol{H}_t}{\partial \vartheta_i}\right), \tag{C.3}$$

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

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

C.2 Linear SRE's on the individual volatilities and the conditional correlation matrix

Consider parameterizations of the form $\boldsymbol{\Sigma}_t(\boldsymbol{\vartheta}) = \boldsymbol{D}_t(\boldsymbol{\vartheta})\boldsymbol{R}_t^{1/2}(\boldsymbol{\vartheta})$ where $\boldsymbol{D}_t(\boldsymbol{\vartheta})$ is the diagonal matrix of the individual volatilities (at $\boldsymbol{\vartheta}_0$), and $\boldsymbol{R}_t^{1/2}(\boldsymbol{\vartheta})$ denotes the symmetric positive definite square-root of the conditional correlation matrix $\boldsymbol{R}_t(\boldsymbol{\vartheta})$ (that is $\{\boldsymbol{R}_t^{1/2}(\boldsymbol{\vartheta})\}^2 = \boldsymbol{R}_t(\boldsymbol{\vartheta})$). 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 $\frac{\partial}{\partial \vartheta_i}\boldsymbol{D}_t(\boldsymbol{\vartheta})$ follows, for any component ϑ_i of $\boldsymbol{\vartheta}$. Turning to the derivatives of $\boldsymbol{R}_t^{1/2}(\boldsymbol{\vartheta})$, we note that, similar to (C.3),

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

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 D). 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 (C.3), which in the diagonal case reduces to $\frac{\partial \boldsymbol{Q}_t^{*1/2}}{\partial \vartheta_i} = \frac{1}{2}\boldsymbol{Q}_t^{*-1/2}\frac{\partial \boldsymbol{Q}_t^*}{\partial \vartheta_i}$. 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.

D 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 $\boldsymbol{\epsilon}_t = \boldsymbol{\Sigma}_t(\boldsymbol{\vartheta}_0)\boldsymbol{\eta}_t$ with $\boldsymbol{\Sigma}_t(\boldsymbol{\vartheta}_0) = \boldsymbol{D}_t\boldsymbol{R}_t^{1/2}$. The diagonal matrix $\boldsymbol{D}_t = \operatorname{diag}(\sigma_{1t}, \ldots, \sigma_{mt})$ is assumed to satisfy the GARCH(1,1) equation

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

where $\boldsymbol{h}_t = (\sigma_{1t}^2, \cdots, \sigma_{mt}^2)'$, $\underline{\boldsymbol{\epsilon}}_t = (\epsilon_{1t}^2, \cdots, \epsilon_{mt}^2)'$, $\boldsymbol{A}_0$ 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

$$\boldsymbol{R}_t = \boldsymbol{Q}_t^{*-1/2}\boldsymbol{Q}_t\boldsymbol{Q}_t^{*-1/2}, \quad \boldsymbol{Q}_t = (1-\alpha_0-\beta_0)\boldsymbol{S}_0 + \alpha_0\boldsymbol{Q}_{t-1}^{*1/2}\boldsymbol{\eta}_{t-1}^*\boldsymbol{\eta}_{t-1}^{*\prime}\boldsymbol{Q}_{t-1}^{*1/2} + \beta_0\boldsymbol{Q}_{t-1},$$

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 $\boldsymbol{\vartheta}_0$ contains the volatility parameters $\boldsymbol{\omega}_0$, $\boldsymbol{A}_0$ and $\mathrm{diag}(\boldsymbol{B}_0)$, and the conditional correlation parameters α_0, β_0 and the sub-diagonal elements of $\boldsymbol{S}_0$.

To estimate $\boldsymbol{\vartheta}_0$, 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 (D.1), which allows for possible volatility spillovers. The two other steps are unchanged: α_0 and β_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

$$\begin{aligned}
\widehat{\boldsymbol{R}}_t &= \widehat{\boldsymbol{Q}}_t^{*-1/2}\widehat{\boldsymbol{Q}}_t\widehat{\boldsymbol{Q}}_t^{*-1/2}, \quad \widehat{\boldsymbol{Q}}_t = (1-\alpha-\beta)\boldsymbol{S}_n + \alpha\widehat{\boldsymbol{Q}}_{t-1}^{*1/2}\widehat{\boldsymbol{\eta}}_{t-1}^*\widehat{\boldsymbol{\eta}}_{t-1}^{*\prime}\widehat{\boldsymbol{Q}}_{t-1}^{*1/2} + \beta\widehat{\boldsymbol{Q}}_{t-1}, \\
\boldsymbol{S}_n &= \boldsymbol{S}_n(\alpha,\beta) = \frac{1}{n}\sum_{t=1}^n \widehat{\boldsymbol{Q}}_t^{*1/2}\widehat{\boldsymbol{\eta}}_t^*\widehat{\boldsymbol{\eta}}_t^{*\prime}\widehat{\boldsymbol{Q}}_t^{*1/2}, \quad \widehat{\boldsymbol{Q}}_t^* = \mathrm{diag}(\widehat{q}_{11,t}, \dots, \widehat{q}_{mm,t})
\end{aligned}$$

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

$$\begin{aligned}
(\widehat{\alpha}_n, \widehat{\beta}_n) &= \arg\min_{(\alpha,\beta)} \sum_{t=1}^n \widehat{\boldsymbol{\eta}}_{t-1}^{*\prime}\widehat{\boldsymbol{R}}_t^{-1}\widehat{\boldsymbol{\eta}}_{t-1}^* + \log\left|\widehat{\boldsymbol{R}}_t\right|, \\
\widehat{\boldsymbol{S}}_n &= \boldsymbol{S}_n^{*-1/2}(\widehat{\alpha}_n, \widehat{\beta}_n)\boldsymbol{S}_n(\widehat{\alpha}_n, \widehat{\beta}_n)\boldsymbol{S}_n^{*-1/2}(\widehat{\alpha}_n, \widehat{\beta}_n),
\end{aligned}$$

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

The parameters used in the Monte-Carlo experiments of Section 4.2 are displayed in Table 3. In Designs A-D the first return is less volatile and less conditionally heteroscedastic than the second return, whereas the two returns have the same dynamic in Designs E-H. Two sets of designs are also distinguished by strong dynamic correlations ($\alpha_0 + \beta_0 = 0.99$) with a strong correlation between the returns ($\boldsymbol{S}_0(1,2) = 0.7$) or constant conditional correlations with null cross-correlation

($\alpha_0 = \beta_0 = 0$ and $\boldsymbol{S}_0(1,2) = 0$). Finally, the designs are distinguished by the distribution of the innovations, which can be the standard normal or the Student distribution with 7 degrees of freedom $\mathcal{S}t_7$ (standardized to obtain unit variance). For generating non spherical distributions, we simulated vectors $\boldsymbol{\eta}_t$ with independent components, distributed according to the Asymmetric Exponential Power Distribution (AEPD) introduced by Zhu and Zinde-Walsh (2009). This class of distributions allows for skewness with different decay rates of density in the left and right tails. This led to the new Designs A*-H*, in which the $\mathcal{N}(0, \boldsymbol{I}_2)$ is replaced by the AEPD with parameters $\alpha = 0.4$, $p_1 = 1.182$ and $p_2 = 1.802$ (which are the values estimated by Zhu and Zinde-Walsh on the S&P500), and the Student distribution $\mathcal{S}t_7$ is replaced by the AEPD with parameters $\alpha = 0.5$, $p_1 = 1$ and $p_2 = 2$ (which gives a strongly asymmetric density). The AEPD densities have also been standardized to obtain zero mean and unit variance.

Table 3: Design of Monte Carlo experiments.

	$\boldsymbol{\omega}_0'$	$(\mathrm{vec}\boldsymbol{A}_0)'$	$\mathrm{diag}\boldsymbol{B}_0$	$\boldsymbol{S}_0(1,2)$	α	β	$P_{\boldsymbol{\eta}}$
A	$(10^{-6}, 4\times 10^{-6})$	(0.01, 0.01, 0.01, 0.07)	(0, 0.92)	0.7	0.04	0.95	$\mathcal{N}(0, \boldsymbol{I}_2)$
B	$(10^{-6}, 4\times 10^{-6})$	(0.01, 0.01, 0.01, 0.07)	(0, 0.92)	0.7	0.04	0.95	$\mathcal{S}t_7$
C	$(10^{-6}, 4\times 10^{-6})$	(0.01, 0.01, 0.01, 0.07)	(0, 0.92)	0	0	0	$\mathcal{N}(0, \boldsymbol{I}_2)$
D	$(10^{-6}, 4\times 10^{-6})$	(0.01, 0.01, 0.01, 0.07)	(0, 0.92)	0	0	0	$\mathcal{S}t_7$
E	$(10^{-5}, 10^{-5})$	(0.07, 0.00, 0.00, 0.07)	(0.92, 0.92)	0.7	0.04	0.95	$\mathcal{N}(0, \boldsymbol{I}_2)$
F	$(10^{-5}, 10^{-5})$	(0.07, 0.00, 0.00, 0.07)	(0.92, 0.92)	0.7	0.04	0.95	$\mathcal{S}t_7$
G	$(10^{-5}, 10^{-5})$	(0.07, 0.00, 0.00, 0.07)	(0.92, 0.92)	0	0	0	$\mathcal{N}(0, \boldsymbol{I}_2)$
H	$(10^{-5}, 10^{-5})$	(0.07, 0.00, 0.00, 0.07)	(0.92, 0.92)	0	0	0	$\mathcal{S}t_7$

Designs A*-H* are the same as Designs A-H, except that $P_{\boldsymbol{\eta}}$ follows an asymmetric AEPD.

E Additional numerical illustrations

We will first complete the asymptotic results of Corollary 3.1 by some finite sample experiments, also allowing for non-Gaussian errors distributions. Then, we will illustrate the nonstationarity of the portfolio's returns. The last part of the section is devoted to dynamic portfolios generated by the DCC GARCH.

E.1 Relative efficiency comparisons in the static case

The computations required to obtain the asymptotic variance $\sigma_S^2(\alpha, \mathbf{a})$ in Corollary 3.1 are elementary but tedious, and they are hardly extendable to the case where $\boldsymbol{\eta}_t$ follows another spherical distribution than the Gaussian. We will compare the asymptotic distributions of the two VaR estimators $\widehat{\mathrm{VaR}}_{S,t-1}^{(\alpha)}(\epsilon^{(P)})$ and $\widehat{\mathrm{VaR}}_{U,t-1}^{(\alpha)}(\epsilon^{(P)})$ with their empirical distributions, on simulations of $\boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \boldsymbol{I}_m)$ with $m = 6$ individual returns; we will also compare the two estimators when $\boldsymbol{\eta}_t$ follows a bivariate Student spherical distribution, standardized so that $\mathrm{var}(\boldsymbol{\eta}_t) = \boldsymbol{I}_2$. The latter distribution is obtained by setting $\boldsymbol{\eta}_t = w_t \boldsymbol{Z}_t$, where (w_t) and $(\boldsymbol{Z}_t)$ and two independent iid sequences such that $(\nu - 2)/w_t^2 \sim \chi_\nu^2$ and $\boldsymbol{Z}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_2)$. Figure 6 displays the boxplots of the estimation errors for the two methods, over 10,000 independent replications of samples of length $n = 500$. As expected from the theory, the multivariate method is more efficient than the univariate method in the normal case (top panels), especially when the portfolio is equally weighted (diversified portfolio). In agreement with Remark 3.1, the effect is less pronounced when only one asset is present (undiversified portfolio). The ratio of the empirical MSE's of the univariate over the multivariate methods is 6.08 in the diversified case, and 1.40 in the undiversified case, which closely corresponds to the values provided by the asymptotic theory (respectively 6 and 1.408). The two bottom panels correspond to the Student spherical distribution of parameter $\nu = 5$. In that case (and for the undiversified (single-asset) portfolio with $\alpha = 0.069$), the univariate method can be more accurate than the multivariate method. The intuition behind this result is that the multivariate method requires empirical moments of order two, for which the variances are very large when $\nu = 5$. Figure 7 compares the three methods on Gaussian innovations. Recall that the FHS method coincides with the univariate method without the symmetry assumption (hence the label Asym). The ranking of the three methods on finite sample ($n = 500$) coincides with the asymptotic ranking.

E.2 Sample path of returns of the crystallized portfolio in the static model

The nonstationarity of the univariate return series $\epsilon_t^{(P)}$ was shown in Section 4.1. Figure 8 illustrates this feature. The increased variance in the second part of the sample reflects the fact that the portfolio tends to be less and less diversified (see Figure 2).

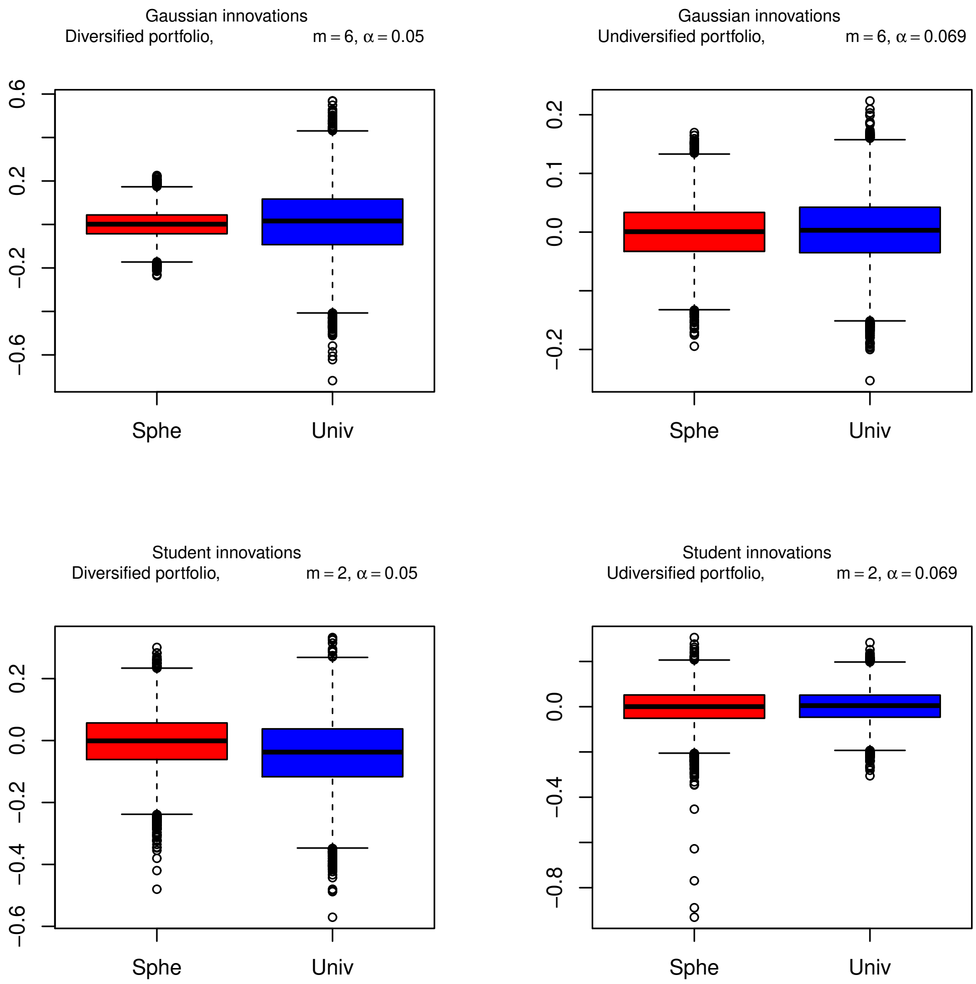

Figure 6: Distribution of the estimation errors for the multivariate and univariate methods.

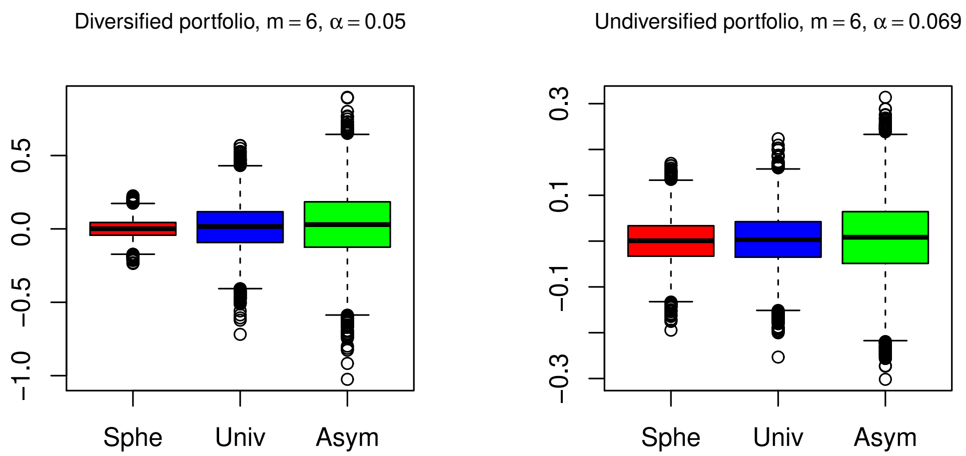

Figure 7: Distribution of the estimation errors for the multivariate and univariate methods.

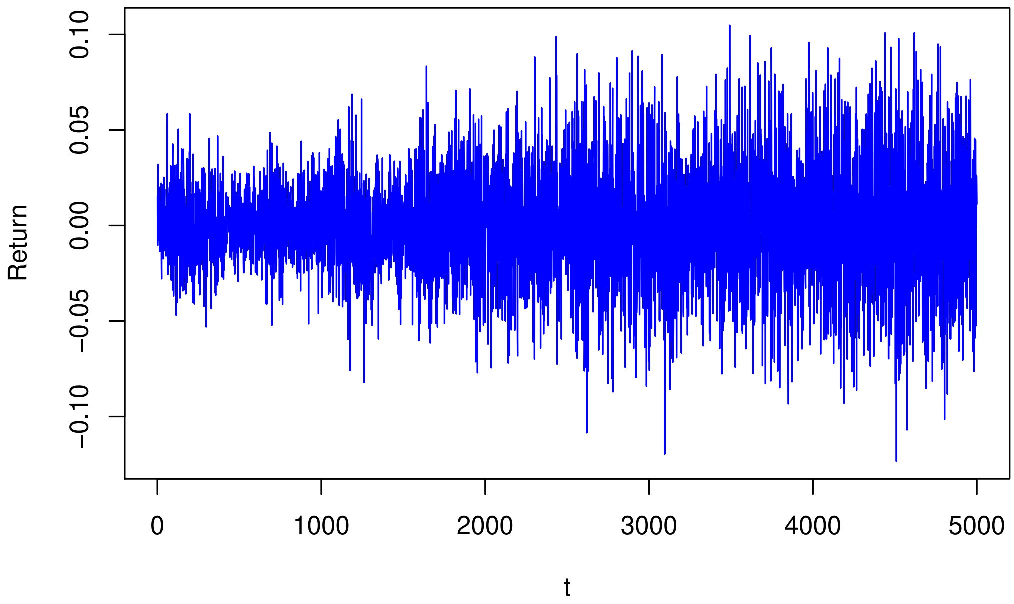

Figure 8: Returns of the crystallized portfolio.

E.3 Sample paths of returns and VaRs of portfolios for DCC models

As a complement to Section 4.2, simulations experiments were conducted with crystallized and minimal-VaR portfolios. With the spherical method, as already seen, the minimal-VaR portfolio coincides with the Markowitz portfolio. Using the FHS method, the portfolio with the smallest α-level conditional VaR can be estimated by

$$\widehat{\epsilon}_t^{(\alpha)} = \boldsymbol{\epsilon}_t' \widehat{\boldsymbol{a}}_{t-1}^{(\alpha)}, \quad \widehat{\boldsymbol{a}}_{t-1}^{(\alpha)} = \arg\min_{\boldsymbol{a}:\, \boldsymbol{a}'\boldsymbol{e}=1} -q_\alpha \left\{ \boldsymbol{a}' \widetilde{\boldsymbol{\Sigma}}_t(\widehat{\boldsymbol{\vartheta}}_{n_1}) \widehat{\boldsymbol{\eta}}_u, u = 1, \dots, n_1 \right\}.$$

where $q_\alpha(S)$ denotes the α-quantile of a set S of real values. In Figure 9, we visualize a typical result obtained for Design D with $n_1 = 1000$ and $n - n_1 = 1000$. This figure displays the returns of the crystallized portfolio obtained by taking an identical proportion of the two components of the portfolio (*i.e.* $\mu_{1,t} = \mu_{2,t}$ for all t), and also the same initial values for the components (*i.e.* $p_{1,0} = p_{2,0}$). As can be seen, the variability of this portfolio is much higher than that of the minimal variance portfolio $\epsilon_t^{(P)*}$ defined by (2.8). The bottom panels display the estimated optimal portfolio $\widehat{\epsilon}_t^{(P)*}$ obtained by replacing $\boldsymbol{\vartheta}_0$ with $\widehat{\boldsymbol{\vartheta}}_n$ in $\epsilon_t^{(P)*}$. In can be seen that $\epsilon_t^{(P)*}$ and $\widehat{\epsilon}_t^{(P)*}$ are very close. Similarly $\mathrm{VaR}_{t-1}^{(\alpha)}(\epsilon^{(P)})$ at level $\alpha = 1\%$ (top-right panel) and its estimates $\widehat{\mathrm{VaR}}_{S,t-1}^{(\alpha)}(\epsilon^{(P)})$ and $\widehat{\mathrm{VaR}}_{FHS,t-1}^{(\alpha)}(\epsilon^{(P)})$ are virtually indistinguishable. On the contrary, Figure 10 shows that $\widehat{\mathrm{VaR}}_{S,t-1}^{(\alpha)}(\epsilon^{(P)})$ may have a much more important bias than $\widehat{\mathrm{VaR}}_{FHS,t-1}^{(\alpha)}(\epsilon^{(P)})$ when the distribution of $\boldsymbol{\eta}_t$ is not spherical. The minimal variance (Markowitz) portfolio and its 1% conditional VaR are displayed in the top right panel. The FHS-estimates given in the bottom-right panel are very accurate, whereas the estimate VaR given by the spherical method (bottom-left panel) is clearly too small. The top panels of Figure 10 represent the returns of the Markowitz and minimal 1%-VaR portfolios, together with their 1%-VaR. With the spherical method, the estimated minimal 1%-VaR (bottom-left panel) is actually the estimated Markowitz portfolio (because under the sphericity assumption these two portfolios coincide). The estimation provided by the FHS method (bottom-right panel) is more satisfactory because it resembles more the top-right panel. From these figures and Table 1, the FHS method seems to be more attractive than the method based on the sphericity assumption.

References

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

Alexander, G.J. and A. M. Baptista (2002) Economic implications of using a mean-VaR model for portfolio selection: a comparison with mean-variance analysis. *Journal of Economic Dynamics and Control* 26, 1159–1193.

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.

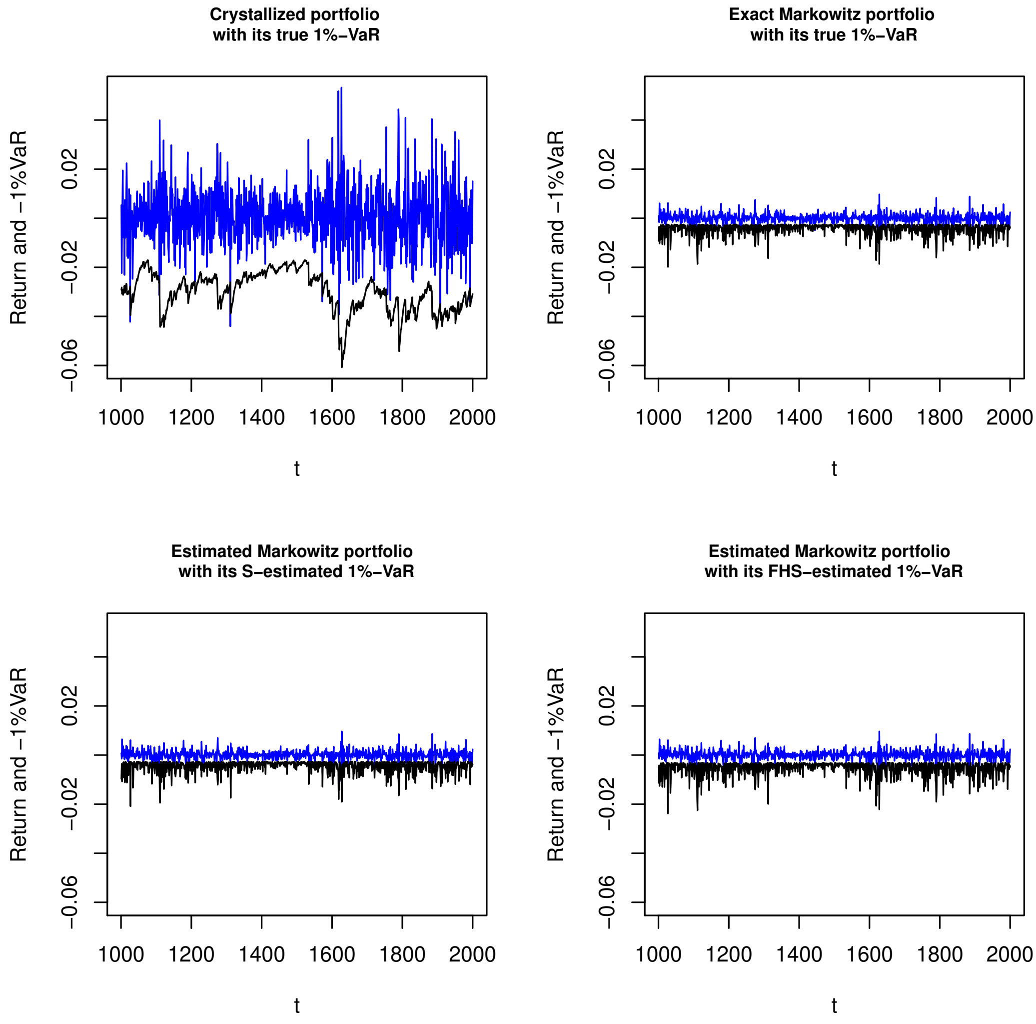

Figure 9: In design D, returns and VaR for a crystallized portfolio and for the optimal (minimal variance) portfolio: true values for the top panels and estimated values of the optimal portfolio and its VaR (black line) at level 1% for the bottom panels, by the Spherical and FHS methods.

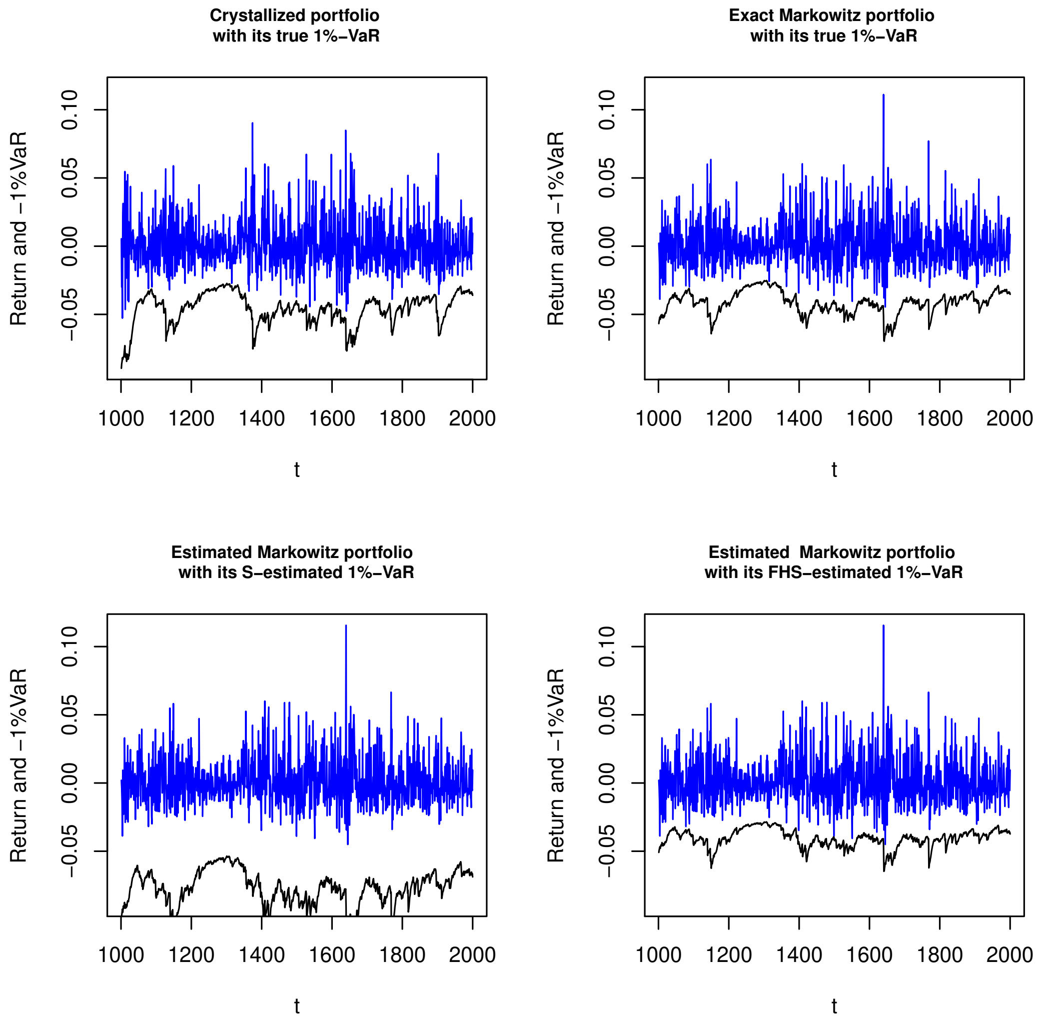

Figure 10: As Figure 9, but for design H* in which the innovations are not spherically distributed.

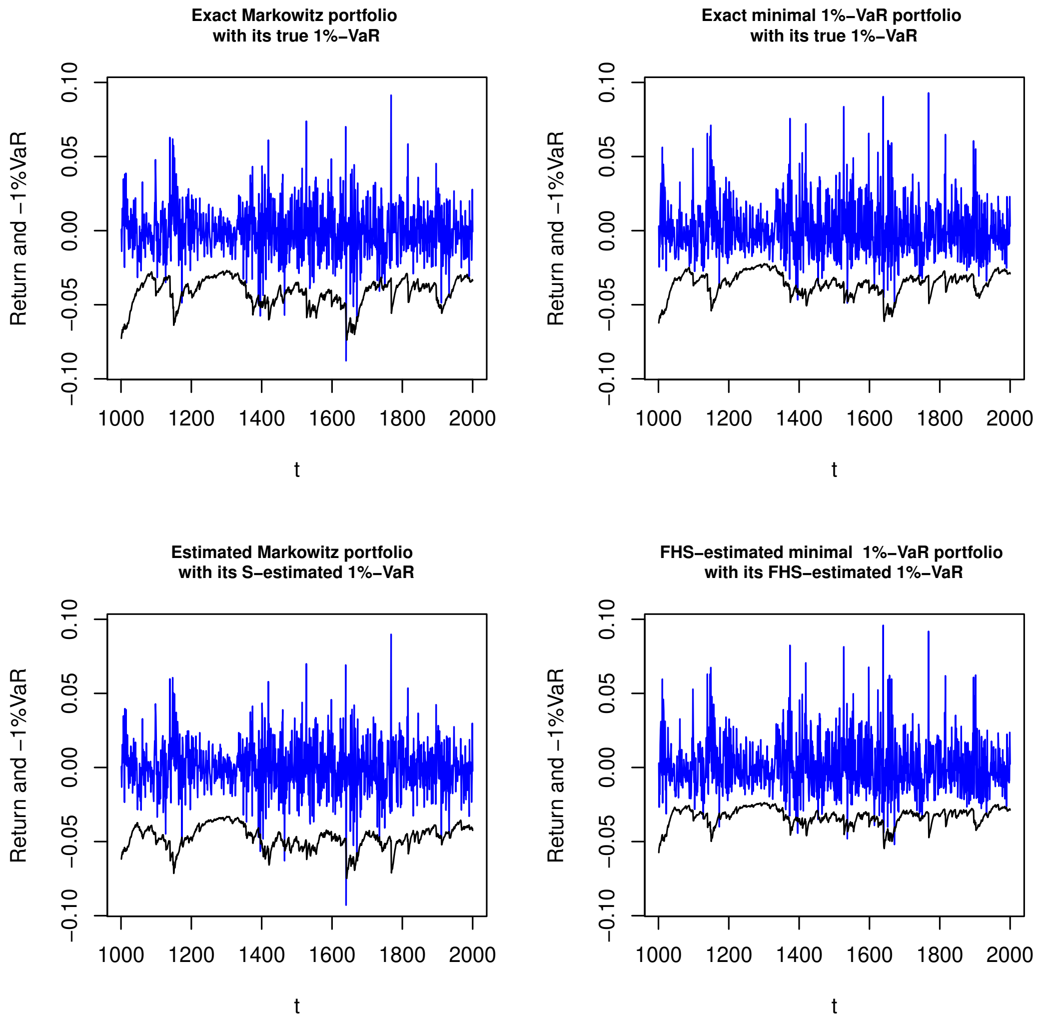

Figure 11: As Figure 10, but for the minimum VaR portfolio.

Campbell, R., Huisman, R. and K. Koedijk Optimal portfolio selection in a value-at-Risk framework. *Journal of Banking & Finance* 25, 1789–1804.

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.

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.

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 (2015) Tests for sphericity in multivariate GARCH models. MPRA Paper No. 67411.

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 (2015a) Risk-parameter estimation in volatility models. *Journal of Econometrics* 184, 158–173.

Francq, C. and J.M. Zakoïan (2015b) Estimating multivariate GARCH models equation by equation. Forthcoming in the *Journal of the Royal Statistical Society: Series B (Statistical Methodology).*

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.

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.

McAleer, M. and B. da Veiga (2008) Single-index and portfolio models for forecasting Value-at-Risk. *Journal of Forecasting* 27, 217–235.

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

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.

Santos, A.A.P., Nogales F.J. and E. Ruiz (2013) Comparing univariate and multivariate models to forecast portfolio Value-at-Risk. *Journal of Financial Econometrics* 11, 400–441.

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.