Virtual Historical Simulation for estimating the conditional VaR of large portfolios

Christian Francq [a,b], Jean-Michel Zakoïan [a,b,*]

[a] *ENSAE-CREST, 5 Avenue Henri Le Chatelier, 91120 Palaiseau, France*
[b] *University of Lille, BP 60149, 59653 Villeneuve d'Ascq cedex, France*

ARTICLE INFO

ABSTRACT

In order to estimate the conditional risk of a portfolio's return, two strategies can be advocated. A multivariate strategy requires estimating a dynamic model for the vector of risk factors, which is often challenging, when at all possible, for large portfolios. A univariate approach based on a dynamic model for the portfolio's return seems more attractive. However, when the combination of the individual returns is time varying, the portfolio's return series is typically non stationary which may invalidate statistical inference. An alternative approach consists in reconstituting a "virtual portfolio", whose returns are built using the current composition of the portfolio and for which a stationary dynamic model can be estimated. This paper establishes the asymptotic properties of this method, that we call Virtual Historical Simulation. Numerical illustrations on simulated and real data are provided.

© 2019 Published by Elsevier B.V.

1. Introduction

The quantitative standards laid down under Basel Accord II and III allow banks to develop internal models for setting aside capital. Methods that incorporate time dependence to quantify market risks are able to use knowledge of the conditional distribution. In particular, the conditional Value-at-Risk (VaR) of financial returns, with a given risk level α (typically, $\alpha = 1\%$ or 5%) is nothing else, from a statistical point of view, than the negated α-quantile of the conditional distribution of the portfolio returns. Estimating conditional quantiles, or more generally conditional risk measures, of a time series of financial returns is thus crucial for risk management.

It is also essential, for risk management purposes, to be able to evaluate the accuracy of such estimators of conditional risks. Uncertainty implied by statistical procedures in the implementation of risk measures may lead to false security in financial markets (see e.g. Farkas et al. (2016) and the references therein). Estimation risk thus needs to be accounted for, in addition to market risk. However, evaluating the estimation risk for the conditional Value-at-Risk (VaR) is generally challenging for two main reasons. Firstly, because the stochastic nature of the conditional VaR does not allow in general to reduce the problem to the estimation of a parameter. Making inference on a stochastic process is obviously more intricate than on a parameter. Secondly, quantiles being obtained as the solutions of optimization problems based on non-smooth functions, establishing asymptotic properties of conditional VaR estimators may become a difficult task.

Increasing attention has been directed in the recent econometric literature to the inference of risk measures in dynamic risk models. Francq and Zakoïan (2015) derived asymptotic confidence intervals (CI) for the conditional VaR of a series of financial returns driven by a parametric dynamic model. Robust backtesting procedures were developed by Escanciano and

---

Olmo (2010) and Escanciano and Olmo (2011), and Gouriéroux and Zakoïan (2013) studied the effect of estimation on the coverage probabilities. Several articles proposed resampling methods: among others, Christoffersen and Gonçalves (2005) and Spierdijk (2016) considered using bootstrap procedures for constructing CIs for VaR; Hurlin et al. (2017) proposed bootstrap-based comparison tests of two conditional risk measures. Beutner et al. (2018) established the validity of a fixed-design residual bootstrap method for the two-step conditional VaR estimator of Francq and Zakoïan (2015). See Nieto and Ruiz (2016) for an extensive survey of the methods for constructing and evaluating VaR forecasts that have been proposed in the literature.

Most existing studies on risk measure inference focus on the risk of a single financial asset. The aim of the present article is to estimate conditional VaR's for *portfolios* of financial assets. From a statistical point of view, the extension is far from trivial. First, because evaluating the quantile of a linear combination of variables may require knowledge of the complete joint distribution of such variables. When the object of interest is a *conditional* quantile, this approach requires specifying a dynamic model for the vector of returns of the assets involved in the portfolio. Second, portfolios compositions are generally time-varying, in particular if the agents adopt a mean–variance approach which, in a dynamic framework, requires specifying the first two conditional moments. This typically entails non-stationarity of the portfolio's return time series, as we shall see in more detail.

A natural approach for obtaining the conditional VaR of a portfolio relies on specifying a multivariate GARCH model for the vector of underlying asset returns. Rombouts and Verbeek (2009) proposed a semi-parametric approach relying on estimating the conditional density of the innovations vector and evaluating numerically the conditional VaR of a portfolio. The asymptotic properties of similar multivariate methods – with or without the assumption of sphericity of the innovations vector – were investigated by Francq and Zakoïan (2018). As noted by Rombouts and Verbeek the advantage of multivariate approaches is to "take into account the dynamic interrelationships between the portfolio components, while the model underlying the VaR calculations is independent of the portfolio composition". On the other hand, for large portfolios multivariate approaches often become untractable due to the well-known dimensionality curse.

In this paper, we consider *univariate* procedures aiming at handling portfolios constructed with a large number of assets. We first consider a "naive" approach in which a standard volatility model is directly fitted to the portfolio returns. Despite its empirical relevance, we will see that the naive approach is not amenable to asymptotic statistical inference (due to the inherent non stationarity of the observed portfolio's time series). We study the asymptotic properties of an alternative procedure relying on a "virtual portfolio" constructed with the *current composition* of the portfolio, on which a univariate model is fitted. This procedure – which we call Virtual Historical Simulation (VHS) – is amenable to asymptotic statistical inference. From a numerical point of view, it allows to avoid difficulties caused by the dimensionality curse in estimation of multivariate volatility models for vectors of asset returns.

The VHS method is related to other approaches introduced in Finance. The Basel Committee and European Union directives (UCITS) recommend that banks backtest their VaR measures against both "clean" and "dirty" P&L's of their trading portfolios (see Holton (2014)). Dirty P&L's are the actual P&L's reported at the end of the time horizon. They can be impacted by changes in the composition of the portfolio that occur during the VaR horizon. Since such position changes may have exogenous causes that cannot be anticipated, it is relevant to backtest the VaR with the so-called clean P&L, which is the hypothetical P&L that would occur if the composition of the portfolio remained unchanged and if market moves were the only source of P&L change (see Pérignon et al. (2008)). Clean P&L thus excludes P&L arising from intra-Day trading, new trades, changes in reserves, fees and commissions. Clean P&L's are often used in the backtesting process, but they can also be used for VaR estimation. The VHS method exploits the idea of cleaning the P&L's for computing the VaR. At each past date t, one can compute a virtual return – the opposite of a clean (or hypothetical) P&L – that would occur if day t positions were exactly those of the current date. Even if each bank uses its own internal VaR model, most financial institutions compute VaR through filtered or simple historical simulations on plain or hypothetical (virtual) returns (see Laurent and Omidi Firouzi (2017). This is the aim of the present paper to study the asymptotic properties of such VaR evaluation methods.

The paper is organized as follows. Section 2 defines the conditional VaR of a portfolio whose composition at the current date may depend on the historical prices, and presents the naive and VHS estimation methods. In Section 3 we derive the asymptotic properties of the VHS procedure based on the Gaussian Quasi-Maximum Likelihood (QML) criterion, under general assumptions on the volatility model. Section 4 presents some numerical illustrations based on Monte Carlo experiments and real financial data. Proofs are collected in Appendix.

2. Estimating the conditional VaR

2.1. 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 $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{mt})'$ denote the corresponding vector of log-returns, with $y_{it} = \log(p_{it}/p_{i,t-1})$ 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 $i = 1, \ldots, m$:

$$V_0 = \sum_{i=1}^{m} \mu_i p_{i0}, \quad V_t = \sum_{i=1}^{m} \mu_{i,t-1} p_{it}, \quad \text{for } t \geq 1 \tag{2.1}$$

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 $V_{t-1} \neq 0$,

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

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 $\frac{V_t}{V_{t-1}} - 1 \approx r_t$, where

$$r_t = \sum_{i=1}^{m} a_{i,t-1} y_{it} = \mathbf{a}'_{t-1} \mathbf{y}_t, \qquad a_{i,t-1} = \frac{\mu_{i,t-1} p_{i,t-1}}{\sum_{j=1}^{m} \mu_{j,t-1} p_{j,t-1}}, \tag{2.2}$$

for $i = 1, \ldots, m$, and $\mathbf{a}_{t-1} = (a_{1,t-1}, \ldots, a_{m,t-1})'$. A portfolio is usually called *crystallized* when the number of units of each asset is time independent, that is $\mu_{i,t-1} = \mu_i$ for each $i = 1, \ldots, m$ and for all t. We will call *static* a portfolio with fixed proportion in value of each return, that is when $a_{i,t-1} = a_i$ for each $i = 1, \ldots, m$ and for all t.

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

$$\mathrm{VaR}_{t-1}^{(\alpha)}(r_t) = \inf\{x \ : \ P_{t-1}(-r_t \leq x) \geq 1 - \alpha\} \tag{2.3}$$

where P_{t-1} denotes the historical distribution conditional on the information I_{t-1} available at time $t - 1$. The specification of I_{t-1} will depend on the approach used. Multivariate approaches use full information, that is all past prices of all assets. In the next approach we describe a univariate approach which only uses the past returns of the portfolio.

2.2. The naive approach

A natural approach for evaluating the conditional VaR in (2.3) when $I_{t-1} = \sigma(r_s, s < t)$ is to estimate a univariate GARCH model, or any time series model, on the series of portfolio returns. We will see that this approach, which can be called "naive", may be misleading due to the fact that the return's portfolio is a time-varying combination of the individual returns.

For simplicity, we consider a crystallized portfolio, with weight μ_i and initial price p_{i0} for the asset $i \in \{1, \ldots, m\}$. The composition $\mathbf{a}_{t-1}$ of such a portfolio is non stationary in general. Indeed, we have

$$\log\left(\frac{a_{i,t}}{a_{j,t}}\right) = \log\left(\frac{\mu_i p_{i,0}}{\mu_j p_{j,0}}\right) + \sum_{k=1}^{t} \Delta_{i,j,k}, \qquad \Delta_{i,j,k} = y_{i,k} - y_{j,k},$$

and $(\sum_{k=1}^{t} \Delta_{i,j,k})_{t \geq 1}$ is a non stationary integrated process of order 1 under general assumptions.[1] More precisely, a consequence of the following lemma is that, with probability tending to one, the composition $\mathbf{a}_{t-1}$ of the portfolio converges to the set of the vectors $\boldsymbol{e}_i$ of the canonical basis (corresponding to single-asset portfolios): $P(\mathbf{a}_{t-1} \in \{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m\}) \to 1$ as $t \to \infty$.

Lemma 2.1. *Consider a process $(\Delta_k)_{k \geq 1}$. Assume that there exist real sequences $a_n > 0$ and b_n, both tending to zero, such that*

$$Z_n := a_n \sum_{k=1}^{n} \Delta_k + b_n \xrightarrow{\mathcal{L}} Z \text{ as } n \to \infty, \tag{2.4}$$

---

[1] By the Chung–Fuks theorem, this is the case when $\boldsymbol{y}_t$ is iid with zero mean and a non-singular covariance matrix Σ. The non stationarity of the process also holds, for instance, if the sequence $(\Delta_{i,j,k})_k$ is mixing and nondegenerated.

for some random variable Z, whose cdf is continuous at 0 and such that $p = P(Z > 0) \in (0, 1)$. Then, for any $c > 0$, we have $P(\sum_{k=1}^{n} \Delta_k > c) \to p$ and $P(\sum_{k=1}^{n} \Delta_k < -c) \to 1 - p$ as $n \to \infty$.

Note that a generalized central limit theorem of the form (2.4) holds for any iid sequence (Δ_k) whenever the distribution of Δ_k belongs to the domain of attraction of Z, which then follows a stable distribution. If the assumptions of Lemma 2.1 hold with $\Delta_k = \Delta_{i,j,k}$ for any pair (i, j), with $i \neq j$, then all the ratios $a_{i,t}/a_{j,t}$ are arbitrarily close to either 1 or 0 with probability tending to 1 as $t \to \infty$. In that case, the composition $\mathbf{a}_{t-1}$ tends to be totally undiversified, but is not always close to the same single-asset composition $\mathbf{e}_i$. If the dynamics of the individual returns y_{it} are not identical, the dynamics of the return r_t will be time-varying, and the naive method based on a fixed stationary GARCH model is likely to produce poor results.

Simulation experiments reported in Section 4 confirm that for crystallized portfolios, the naive approach may behave badly due to the non stationarity of the univariate returns r_t. Of course, for static portfolios the non stationarity issue vanishes, but such portfolios may be considered as artificial.[2] The next section studies a remedy to the non stationarity issue, while keeping the univariate framework.

2.3. The VHS approach

An alternative to the naive approach consists in reconstituting a "virtual portfolio", whose returns are built using the current composition of the portfolio. Given the portfolio composition $\mathbf{a}_{t_0-1} = \mathbf{x}$, say, at time t_0, we construct a process of virtual returns

$$r_t^*(\mathbf{x}) = \mathbf{x}' \mathbf{y}_t, \qquad t \in \mathbb{Z}$$

and we consider the information set $I_{t_0-1} = \sigma(r_s^*(\mathbf{x}), s < t_0)$. Note that, in general, $r_t^*(\mathbf{x}) \neq r_t$ because the composition of the (non virtual) portfolio is time varying ($\mathbf{a}_{t-1} \neq \mathbf{x}$, in general, for $t \neq t_0$). Given the stationarity of $(\mathbf{y}_t)$, it is clear that the series of virtual returns $\{r_t^*(\mathbf{x})\}$ is also stationary, with conditional moments

$$E_{t-1}[r_t^*(\mathbf{x})] =: \mu_t(\mathbf{x}), \qquad \mathrm{var}_{t-1}[r_t^*(\mathbf{x})] =: \sigma_t^2(\mathbf{x}),$$

where $E_{t-1}(X) = E(X \mid r_s^*(\mathbf{x}), s < t)$ for any variable X, and the variance is defined accordingly. Thus, $r_t^*(\mathbf{x})$ follows a model of the form

$$r_t^*(\mathbf{x}) = \mu_t(\mathbf{x}) + \sigma_t(\mathbf{x})u_t, \quad \text{where } E_{t-1}(u_t) = 0 \text{ and } \mathrm{var}_{t-1}(u_t) = 1. \tag{2.5}$$

Noting that $r_{t_0} = r_{t_0}^*(\mathbf{a}_{t_0-1})$, the conditional VaR at time t_0 thus satisfies

$$\mathrm{VaR}_{t_0-1}^{*(\alpha)}(r_{t_0}) = -\mu_{t_0}(\mathbf{a}_{t_0-1}) + \sigma_{t_0}(\mathbf{a}_{t_0-1})\mathrm{VaR}_{t_0-1}^{*(\alpha)}(u_{t_0}) \tag{2.6}$$

where $\mathrm{VaR}_{t-1}^{*(\alpha)}(X)$ is the VaR of X at level α conditional on I_{t-1}.

Note that the martingale difference (u_t) may not be iid, as the following example illustrates.

Example 2.1. Consider the bivariate ARCH(1) process, defined as the stationary non anticipative solution of the model

$$\mathbf{y}_t = \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \Sigma_t \eta_t, \quad \Sigma_t = \begin{pmatrix} \sigma_{1t}^2 := \omega_1 + \alpha_{11}y_{1,t-1}^2 + \alpha_{12}y_{2,t-1}^2 & 0 \\ 0 & \sigma_{2t}^2 := \omega_2 + \alpha_{21}y_{1,t-1}^2 \end{pmatrix},$$

where η_t is iid $(\mathbf{0}, \mathbf{I})$, and assuming that the components η_{1t} and η_{2t} are independent. Let a portfolio which is fully invested in the first asset, hence $r_t = (1, 0)\mathbf{y}_t = y_{1t}$ for all t. Denote by $\mathcal{F}_{1t}$ the σ-field generated by $\{y_{1u}, u \leq t\}$. We have $E(r_t|\mathcal{F}_{1,t-1}) = 0$ and

$$\begin{aligned} E(r_t^2|\mathcal{F}_{1,t-1}) &= E(\sigma_{1t}^2|\mathcal{F}_{1,t-1}) = \omega_1 + \alpha_{11}y_{1,t-1}^2 + \alpha_{12}E(y_{2,t-1}^2|\mathcal{F}_{1,t-1}) \\ &= \omega_1 + \alpha_{11}y_{1,t-1}^2 + \alpha_{12}\sigma_{2,t-1}^2 E(\eta_{2,t-1}^2|\mathcal{F}_{1,t-1}) \\ &= \omega_1 + \alpha_{11}y_{1,t-1}^2 + \alpha_{12}\sigma_{2,t-1}^2 := \sigma_t^2. \end{aligned}$$

It follows that (r_t) satisfies the model $r_t = \sigma_t u_t$, where

$$u_t = \frac{\sigma_{1t}}{\sigma_t}\eta_{1t} = \left(1 + \frac{\alpha_{12}\sigma_{2,t-1}^2(\eta_{2,t-1}^2 - 1)}{\omega_1 + \alpha_{11}y_{1,t-1}^2 + \alpha_{12}\sigma_{2,t-1}^2} \right)^{1/2} \eta_{1t}.$$

It is then clear that $(u_t, \mathcal{F}_{1,t})$ is a martingale difference but (u_t) is generally not iid (except when $\alpha_{12} = 0$ or $\eta_{2,t}^2$ is degenerated).

---

[2] As it would require rebalancing every day in order to maintain a fixed composition in percentages. In general, portfolios are rebalanced to minimize risk, leading to a non-static composition.

Even in the simple previous example, the conditional quantile $\text{VaR}^{*(\alpha)}_{t_0-1}(u_{t_0})$ popping up in (2.6) cannot be explicitly computed. Whether or not this quantity could be estimated nonparametrically is beyond the scope of this paper. Instead, we consider a "hybrid" VaR defined by

$$\text{VaR}^{(\alpha)}_{H,t_0-1}(r_{t_0}) = -\mu_{t_0}(\mathbf{a}_{t_0-1}) + \sigma_{t_0}(\mathbf{a}_{t_0-1})\text{VaR}^{(\alpha)}(u) \tag{2.7}$$

where $\text{VaR}^{(\alpha)}(u)$ is the marginal VaR of u_t at level α. An estimator of $\text{VaR}^{(\alpha)}_{H,t_0-1}(r_{t_0})$ is obtained as follows: given $\mathbf{a}_{t_0-1} = \mathbf{x}$,

STEP 1: Compute the virtual historical returns $r_t^*(\mathbf{x})$ for $t = 1, \ldots, n$.

STEP 2: Estimate $\mu_t(\mathbf{x})$ and $\sigma_t(\mathbf{x})$. Denote by $\hat{\mu}_t(\mathbf{x})$ and $\hat{\sigma}_t(\mathbf{x})$ the resulting estimators, and by $\hat{u}_t = \{r_t^*(\mathbf{x}) - \hat{\mu}_t(\mathbf{x})\}/\hat{\sigma}_t(\mathbf{x})$ the residuals.

STEP 3: Compute the α-quantile $\xi_{n,\alpha}$ of $\{\hat{u}_s, 1 \leq s \leq n\}$ and define an estimator of $\text{VaR}^{(\alpha)}_{H,t_0-1}(r_{t_0})$ as

$$\widehat{\text{VaR}}^{(\alpha)}_{VHS,t_0-1}(r_{t_0}) = -\hat{\mu}_{t_0}(\mathbf{x}) - \hat{\sigma}_{t_0}(\mathbf{x})\xi_{n,\alpha}. \tag{2.8}$$

Step 2 can be implemented by estimating a parametric model. This approach will be developed in Section 3.

This procedure is particularly appropriate for large portfolios, when the large dimension of the vector of underlying assets precludes – or at least formidably complicates – estimation of multivariate volatility models. Moreover, the following example shows that for large portfolios a univariate GARCH model is a reasonable assumption for the virtual returns.

Example 2.2. Suppose that m is large and that the vector of log-returns is driven by a vector $\boldsymbol{f}_t$ of K factors (with $K \ll m$) as

$$\boldsymbol{y}_t = \boldsymbol{\beta}\boldsymbol{f}_t + \boldsymbol{u}_t$$

where $\boldsymbol{\beta}$ is a $m \times K$ matrix, $\text{var}_{t-1}(\boldsymbol{f}_t) = \boldsymbol{F}_t$ is a full-rank matrix and $\text{var}_{t-1}(\boldsymbol{u}_t) = \Sigma_t$. With a composition fixed to $\mathbf{x}$, the virtual portfolio's returns thus satisfy

$$r_t^* = \mathbf{x}'\boldsymbol{\beta}\boldsymbol{f}_t + \mathbf{x}'\boldsymbol{u}_t.$$

Suppose that the portfolio is well-diversified, so that the components x_i of $\mathbf{x}$ satisfy $x_i = O(1/m)$ for $i = 1, \ldots, m$ as $m \to \infty$. Under appropriate assumptions $\text{var}_{t-1}(\mathbf{x}'\boldsymbol{u}_t)$ converges to 0 as $m \to \infty$. On the other hand, $\text{var}_{t-1}(\mathbf{x}'\boldsymbol{\beta}\boldsymbol{f}_t) = \mathbf{x}'\boldsymbol{\beta}\boldsymbol{F}_t\boldsymbol{\beta}'\mathbf{x} = O_P(1)$ and does not vanish as m increases under appropriate assumptions.[3] It follows that $r_t^* \approx \mathbf{x}'\boldsymbol{\beta}\boldsymbol{f}_t$. If now $K = 1$ and the (real-valued) factor f_t is the solution of a GARCH model, the process $\mathbf{x}'\boldsymbol{\beta}_t$ will follow the same model up to a change of scale. It is therefore natural to fit a GARCH model for the virtual returns under these assumptions when m is large.

3. Asymptotic properties of the VHS approach

To obtain asymptotic properties of the VHS procedure, we make the following parametric assumptions on Model (2.5). For simplicity, we consider the model without conditional mean, that is $\mu_t(\mathbf{x}) = 0$. For some (known) function $\sigma : \mathbb{R}^\infty \times \Theta \to (0, \infty)$, let

$$\sigma_t(\mathbf{x}; \boldsymbol{\theta}) = \sigma(r_{t-1}^*(\mathbf{x}), r_{t-2}^*(\mathbf{x}), \ldots; \boldsymbol{\theta}), \tag{3.1}$$

where $\boldsymbol{\theta}_0 = \boldsymbol{\theta}_0(\mathbf{x})$ is the true value of the finite dimensional parameter $\boldsymbol{\theta}$, belonging to some compact subset Θ of $\mathbb{R}^d$. To alleviate notations, we will denote the virtual returns by $\epsilon_t := r_t^*(\mathbf{x})$ and replace $\sigma_t(\mathbf{x}; \boldsymbol{\theta})$ by $\sigma_t(\boldsymbol{\theta})$. Model (2.5) thus writes

$$\epsilon_t = \sigma_t(\boldsymbol{\theta}_0)u_t, \quad \text{where for all } \boldsymbol{\theta} \in \Theta, \quad \sigma_t(\boldsymbol{\theta}) = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \ldots; \boldsymbol{\theta}). \tag{3.2}$$

Given (virtual) observations $\epsilon_1, \ldots, \epsilon_n$, and arbitrary initial values $\tilde{\epsilon}_i$ for $i \leq 0$, we define

$$\tilde{\sigma}_t(\boldsymbol{\theta}) = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \ldots, \epsilon_1, \tilde{\epsilon}_0, \tilde{\epsilon}_{-1}, \ldots; \boldsymbol{\theta}).$$

The Gaussian QML criterion is defined by

$$\tilde{Q}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^n \tilde{\ell}_t, \qquad \tilde{\ell}_t = \tilde{\ell}_t(\boldsymbol{\theta}) = \frac{\epsilon_t^2}{\tilde{\sigma}_t^2(\boldsymbol{\theta})} + \log \tilde{\sigma}_t^2(\boldsymbol{\theta}). \tag{3.3}$$

---

[3] For instance if the matrix $\boldsymbol{\beta}$ does not contain too many zeros or, more precisely, if at least one column $\boldsymbol{\beta}_j$ of $\boldsymbol{\beta}$ is such that $\liminf |\mathbf{x}'\boldsymbol{\beta}_j| > 0$ as $m \to \infty$.

Let the QML estimator of $\boldsymbol{\theta}_0$,

$$\hat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}\in\Theta} \tilde{Q}_n(\boldsymbol{\theta}), \tag{3.4}$$

which has the particularity of being based on virtual returns rather than on observations.

To study the asymptotic properties of the VHS estimator,

$$\widehat{\mathrm{VaR}}_{VHS,t_0-1}^{(\alpha)}(r_{t_0}) = -\tilde{\sigma}_{t_0}(\hat{\boldsymbol{\theta}}_n)\xi_{n,\alpha}, \tag{3.5}$$

we introduce the following additional assumptions. Recall that iidness of the sequence (u_t) is not a natural assumption in our framework (see Example 2.1). Let $\boldsymbol{D}_t(\boldsymbol{\theta}) = \sigma_t^{-1}(\boldsymbol{\theta})\partial\sigma_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$, $\boldsymbol{D}_t = \boldsymbol{D}_t(\boldsymbol{\theta}_0)$. Let also $\mathcal{F}_t$ the sigma-field generated by $\{u_k, k \leq t\}$, and $\mathcal{F}_{t:t-s}$ the sigma-field generated by $\{u_k, t-s \leq k \leq t\}$ with $s > 0$.

A1: The sequence (u_t) is stationary, with $E|u_t|^{4+\nu} < \infty$ for some $\nu > 0$, and mixing coefficients $\{\alpha(h)\}_{h\geq 0}$ satisfying, for $\epsilon \in (0, \nu)$,

$$\sum_{h=1}^{\infty} h^{r^*}\{\alpha(h)\}^{\frac{\nu-\epsilon}{4+\nu-\epsilon}} < \infty \text{ for some } r^* > \frac{2\kappa\{2+\nu-\epsilon\}}{\nu-\epsilon-2\kappa\{2+\nu-\epsilon\}} \text{ and } \kappa \in \left(0, \frac{\nu-\epsilon}{4(2+\nu-\epsilon)}\right).$$

Suppose that $E(u_t \mid \mathcal{F}_{t-1}) = 0$ and $E(u_t^2 \mid \mathcal{F}_{t-1}) = 1$. Let ξ_α the α-quantile of u_t. Assume that the conditional distribution of u_t given $\mathcal{F}_{t-1}$ has a density f_{t-1} such that $f_{t-1}(\xi_\alpha) > 0$ a.s. and $E\sup_{\xi\in V(\xi_\alpha)} f_{t-1}^4(\xi) < \infty$ for some neighborhood $V(\xi_\alpha)$ of ξ_α. Assume also that this density is continuous at ξ_α uniformly in $\mathcal{F}_{t-1}$, in the sense that for sufficiently small $\varepsilon > 0$, there exists a stationary and ergodic sequence (K_t) such that $K_{t-1} \in \mathcal{F}_{t-1}$ and

$$\sup_{x\in[\xi_\alpha-\varepsilon, \xi_\alpha+\varepsilon]} |f_{t-1}(x) - f_{t-1}(\xi_\alpha)| \leq K_{t-1}\varepsilon$$

with $EK_t^4 < \infty$ a.s.

A2: (ϵ_t) is a strictly stationary and ergodic solution of (3.2), and there exists $s_0 > 0$ such that $E|\epsilon_1|^{s_0} < \infty$.

A3: There exists a sequence $\boldsymbol{D}_{t,T_n}$ such that $\boldsymbol{D}_t = \boldsymbol{D}_{t,T_n} + o_P(1)$ as $n \to \infty$, where $T_n \to \infty$ and $T_n = O(n^\kappa)$ (with κ defined in **A1**), $\boldsymbol{D}_{t,T_n} \in \mathcal{F}_{t-1:t-T_n}$ and for any $r \geq 0$

$$E\|\boldsymbol{D}_t\|^r < \infty, \qquad \sup_{n\geq 1} E\|\boldsymbol{D}_{t,T_n}\|^r < \infty.$$

A4: For some $\underline{\omega} > 0$, almost surely, $\sigma_t(\boldsymbol{\theta})$ and $\tilde{\sigma}_t(\boldsymbol{\theta})$ belong to $(\underline{\omega}, \infty]$ for any $\boldsymbol{\theta} \in \Theta$ and any $t \geq 1$. For $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$, we have $\sigma_t(\boldsymbol{\theta}_1) = \sigma_t(\boldsymbol{\theta}_2)$ a.s. if and only if $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$. Moreover, for any $\boldsymbol{x} \in \mathbb{R}^m$, $\boldsymbol{x}'\boldsymbol{D}_t(\boldsymbol{\theta}_0) = 0$ a.s. entails $\boldsymbol{x} = 0$.

A5: There exist a random variable C which is measurable with respect to $\{\epsilon_t, t < 0\}$ and a constant $\rho \in (0, 1)$ such that $\sup_{\boldsymbol{\theta}\in\Theta} |\sigma_t(\boldsymbol{\theta}) - \tilde{\sigma}_t(\boldsymbol{\theta})| \leq C\rho^t$.

A6: The function $\boldsymbol{\theta} \mapsto \sigma(x_1, x_2, \ldots; \boldsymbol{\theta})$ has continuous second-order derivatives, and

$$\sup_{\boldsymbol{\theta}\in\Theta} \left\| \frac{\partial\sigma_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} - \frac{\partial\tilde{\sigma}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} \right\| + \left\| \frac{\partial^2\sigma_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} - \frac{\partial^2\tilde{\sigma}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} \right\| \leq C\rho^t,$$

where C and ρ are as in **A5**.

A7: There exists a neighborhood $V(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ and $\tau > 0$ such that

$$\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)} \left\| \frac{1}{\sigma_t(\boldsymbol{\theta})}\frac{\partial\sigma_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} \right\|^4, \quad \sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)} \left\| \frac{1}{\sigma_t(\boldsymbol{\theta})}\frac{\partial^2\sigma_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} \right\|^2, \quad \sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)} \left| \frac{\sigma_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta})} \right|^{\frac{2(4+\nu)(1+\tau)}{2+\nu}},$$

have finite expectations, where ν is as in **A1**.

Assumptions **A2**, **A4** and **A5**, together with $E(u_t^2 \mid \mathcal{F}_{t-1}) = 1$, are sufficient to prove the strong consistency of the QML estimator $\hat{\boldsymbol{\theta}}_n$. Assumption **A6** allows to show that the initial values $\tilde{\epsilon}_i$ do not matter for the asymptotic normality of $\hat{\boldsymbol{\theta}}_n$. Assumptions **A1**, **A3** and **A7** are used to apply a CLT for a mixing triangular array based on the approximation of $\boldsymbol{D}_t$ by $\boldsymbol{D}_{t,T_n}$.

For particular volatility models, some of the assumptions can be simplified as the following lemma shows.

Lemma 3.1. *For the standard GARCH(1,1) model*

$$\epsilon_t = \sigma_t u_t, \quad \sigma_t^2 = \omega_0 + \alpha_0\epsilon_{t-1}^2 + \beta_0\sigma_{t-1}^2, \qquad \omega_0 > 0, \alpha_0 > 0, \beta_0 > 0, \tag{3.6}$$

*where (u_t) satisfies **A1**, Assumptions **A2–A7** reduce to: (i) $E\log(\alpha_0 u_t^2 + \beta_0) < 0$; (ii) u_t^2 has a non-degenerate distribution; (iii) $\Theta = \{(\omega, \alpha, \beta)\}$ is a compact subset of $(0, \infty)^3$ such that, for all $\theta \in \Theta$, $\omega > \underline{\omega}$ for some $\underline{\omega} > 0$ and $\beta < 1$. (iv) $E|\epsilon_t|^{s_0} < \infty$ for $s_0 > 0$, where (ϵ_t) is the strictly stationary solution implied by (i).*

We are now in a position to state our main result.

Theorem 3.1. *Assume $\xi_\alpha < 0$. Let **A1–A7** hold. Then $\hat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0$ a.s. as $n \to \infty$, and*

$$\begin{pmatrix} \sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \\ \sqrt{n}(\xi_\alpha - \xi_{n,\alpha}) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \boldsymbol{\Sigma}_\alpha), \qquad \boldsymbol{\Sigma}_\alpha = \begin{pmatrix} \boldsymbol{J}^{-1}\boldsymbol{S}^{11}\boldsymbol{J}^{-1} & \boldsymbol{\Lambda}_\alpha \\ \boldsymbol{\Lambda}'_\alpha & \zeta_\alpha \end{pmatrix},$$

where $\boldsymbol{J} = E(\boldsymbol{D}_t\boldsymbol{D}'_t)$ with $\boldsymbol{D}_t = \boldsymbol{D}_t(\boldsymbol{\theta}_0)$, and

$$\boldsymbol{\Lambda}_\alpha = \frac{\xi_\alpha}{4E[f_{t-1}(\xi_\alpha)]}\boldsymbol{J}^{-1}\boldsymbol{S}^{11}\boldsymbol{J}^{-1}\boldsymbol{\Psi}_\alpha + \frac{1}{2E[f_{t-1}(\xi_\alpha)]}\boldsymbol{J}^{-1}\boldsymbol{S}^{12}_\alpha, \qquad \boldsymbol{\Psi}_\alpha = E[f_{t-1}(\xi_\alpha)\boldsymbol{D}_t],$$

$$\zeta_\alpha = \frac{1}{\{E[f_{t-1}(\xi_\alpha)]\}^2}\left(\frac{1}{4}\xi_\alpha^2\boldsymbol{\Psi}'_\alpha\boldsymbol{J}^{-1}\boldsymbol{S}^{11}\boldsymbol{J}^{-1}\boldsymbol{\Psi}_\alpha + \xi_\alpha\boldsymbol{\Psi}'_\alpha\boldsymbol{J}^{-1}\boldsymbol{S}^{12}_\alpha + \boldsymbol{S}^{22}_\alpha\right),$$

$$\boldsymbol{S}^{11} = E\left[\left(u_t^2 - 1\right)^2 \boldsymbol{D}_t\boldsymbol{D}'_t\right], \qquad \boldsymbol{S}^{22}_\alpha = \sum_{h=-\infty}^{\infty} cov\left(\mathbf{1}_{\{u_t<\xi_\alpha\}}, \mathbf{1}_{\{u_{t-h}<\xi_\alpha\}}\right),$$

$$\boldsymbol{S}^{12}_\alpha = \boldsymbol{S}^{21'}_\alpha = \sum_{h=0}^{\infty} cov[(u_t^2 - 1)\boldsymbol{D}_t, \mathbf{1}_{\{u_{t+h}<\xi_\alpha\}}].$$

Remark 3.1. When the errors u_t are iid, the expression of the asymptotic covariance matrix simplifies considerably. More precisely, we have

$$\boldsymbol{S}^{11} = (\kappa_4 - 1)\boldsymbol{J}, \qquad \boldsymbol{S}^{22}_\alpha = \alpha(1-\alpha), \quad \boldsymbol{S}^{12}_\alpha = \boldsymbol{S}^{21'}_\alpha = [E\left(u_t^2\mathbf{1}_{\{u_t<\xi_\alpha\}}\right) - \alpha]E(\boldsymbol{D}_t) := p_\alpha E(\boldsymbol{D}_t),$$

$$\boldsymbol{\Lambda}_\alpha = \left(\frac{\xi_\alpha(\kappa_4 - 1)}{4} + \frac{p_\alpha}{2f(\xi_\alpha)}\right)\boldsymbol{J}^{-1}\boldsymbol{\Omega}, \qquad \boldsymbol{\Omega} = E(\boldsymbol{D}_t),$$

$$\zeta_\alpha = \frac{\kappa_4 - 1}{4}\xi_\alpha^2\boldsymbol{\Omega}'\boldsymbol{J}^{-1}\boldsymbol{\Omega} + \frac{\xi_\alpha p_\alpha}{f(\xi_\alpha)}\boldsymbol{\Omega}'\boldsymbol{J}^{-1}\boldsymbol{\Omega} + \frac{\alpha(1-\alpha)}{f^2(\xi_\alpha)} = \frac{\kappa_4 - 1}{4}\xi_\alpha^2 + \frac{\xi_\alpha p_\alpha}{f(\xi_\alpha)} + \frac{\alpha(1-\alpha)}{f^2(\xi_\alpha)}.$$

where f denotes the density of u_t, and $\kappa_4 = Eu_t^4$. For the last equality, we used the relation $\boldsymbol{\Omega}'\boldsymbol{J}^{-1}\boldsymbol{\Omega} = 1$ (see (8) in Francq and Zakoïan (2013)). Hence, when (u_t) is iid we have

$$\boldsymbol{\Sigma}_\alpha = \begin{pmatrix} (\kappa_4 - 1)\boldsymbol{J}^{-1} & \left(\frac{\xi_\alpha(\kappa_4-1)}{4} + \frac{p_\alpha}{2f(\xi_\alpha)}\right)\boldsymbol{J}^{-1}\boldsymbol{\Omega} \\ \left(\frac{\xi_\alpha(\kappa_4-1)}{4} + \frac{p_\alpha}{2f(\xi_\alpha)}\right)\boldsymbol{\Omega}'\boldsymbol{J}^{-1} & \frac{\kappa_4-1}{4}\xi_\alpha^2 + \frac{\xi_\alpha p_\alpha}{f(\xi_\alpha)} + \frac{\alpha(1-\alpha)}{f^2(\xi_\alpha)} \end{pmatrix}.$$

In particular, the asymptotic variance of $\sqrt{n}(\xi_\alpha - \xi_{n,\alpha})$ only depends on the errors distribution, not on the volatility parameter $\boldsymbol{\theta}_0$. Nevertheless, the estimation of $\boldsymbol{\theta}_0$ affects the asymptotic accuracy, since the asymptotic variance would be equal to $\frac{\alpha(1-\alpha)}{f^2(\xi_\alpha)}$ if the u_t's were observed.

Remark 3.2. Consistent estimation of $\boldsymbol{\Sigma}_\alpha$ is crucial for evaluating the estimation risk. Some matrices involved in the asymptotic covariance matrix have the form of expectations, and can therefore be straightforwardly estimated by their empirical counterpart. This is the case of $\boldsymbol{J}$ and $\boldsymbol{S}^{11}$. To estimate $\boldsymbol{S}^{22}_\alpha$ and $\boldsymbol{S}^{12}_\alpha$, a classical HAC (heteroskedasticity and autocorrelation consistent) estimator can be used (see for instance Andrews (1991), Newey and West (1987)). Estimation of the matrix $\boldsymbol{\Psi}_\alpha$ is more tricky. We propose two approaches: (i) noting that

$$\boldsymbol{\Psi}_\alpha = \lim_{h\to 0}\frac{1}{h}E\left[\{P_{t-1}(u_t < \xi_\alpha + h) - P_{t-1}(u_t < \xi_\alpha)\}\boldsymbol{D}_t\right]$$

$$= \lim_{h\to 0}\frac{1}{h}E\left[E_{t-1}\{\mathbf{1}_{\{u_t<\xi_\alpha+h\}} - \mathbf{1}_{\{u_t<\xi_\alpha\}}\}\boldsymbol{D}_t\right] = \lim_{h\to 0}\frac{1}{h}E\left[(\mathbf{1}_{\{u_t<\xi_\alpha+h\}} - \mathbf{1}_{\{u_t<\xi_\alpha\}})\boldsymbol{D}_t\right],$$

a natural estimator for $\boldsymbol{\Psi}_\alpha$ is

$$\widehat{\boldsymbol{\Psi}}_\alpha = \frac{1}{nh_n}\sum_{t=1}^{n}(\mathbf{1}_{\{\hat{u}_t<\xi_{n,\alpha}+h_n\}} - \mathbf{1}_{\{\hat{u}_t<\xi_{n,\alpha}\}})\hat{\boldsymbol{D}}_t,$$

where $\hat{\boldsymbol{D}}_t = \tilde{\sigma}_t^{-1}(\hat{\boldsymbol{\theta}}_n)\partial\tilde{\sigma}_t(\hat{\boldsymbol{\theta}}_n)/\partial\boldsymbol{\theta}$ and h_n is a bandwidth parameter; (ii) alternatively, one may approximate f_{t-1} by the density of u_t conditional on u_{t-1} (instead of the infinite past). A standard kernel estimator for this density at point ξ_α is

$$\hat{f}_{t-1}(\xi_\alpha) = \frac{\sum_{s=1}^{n} K_{h_2}(\hat{u}_{t-1} - \hat{u}_{s-1})K_{h_1}(\xi_\alpha - \hat{u}_s)}{\sum_{s=1}^{n} K_{h_2}(\hat{u}_{t-1} - \hat{u}_{s-1})},$$

where h_1, h_2 are bandwidths, $K_h(x) = h^{-1}K(x/h)$ for any bandwidth h, and K is a bounded symmetric kernel function (see e.g. Hansen (2004)). Matrix Ψ_α can thus be estimated by

$$\widehat{\Psi}_\alpha^\circ = \frac{1}{n}\sum_{t=1}^{n}\hat{f}_{t-1}(\xi_\alpha)\hat{D}_t.$$

The asymptotic properties under Assumptions **A1–A7** of both estimators, $\widehat{\Psi}_\alpha$ and $\widehat{\Psi}_\alpha^\circ$, are left for further research.

We have the Taylor expansion

$$F_{t_0}(\hat{\boldsymbol{\theta}}_n, \xi_{n,\alpha}) := \tilde{\sigma}_{t_0}(\hat{\boldsymbol{\theta}}_n)\xi_{n,\alpha} = F_{t_0}(\boldsymbol{\theta}_0, \xi_\alpha) + \frac{\partial}{\partial(\boldsymbol{\theta}', \xi)}F_{t_0}(\boldsymbol{\theta}_n^*, \xi_n^*)\sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \\ \xi_\alpha - \xi_{n,\alpha} \end{pmatrix},$$

where $(\boldsymbol{\theta}_n^*, \xi_n^*)$ is between $(\boldsymbol{\theta}_n, \xi_{n,\alpha})$ and $(\boldsymbol{\theta}_0, \xi_\alpha)$. Given a fixed information set I_{t_0-1}, it can be shown using **A5–A6** that the observations available at time $t_0 - 1$ have no effect on the asymptotic distribution of Theorem 3.1. Thus, an approximate *conditional* $(1 - \alpha_0)\%$ CI for the hybrid VaR (2.7) has bounds given by

$$\widehat{\mathrm{VaR}}_{VHS,t_0-1}^{(\alpha)}(r_{t_0}) \pm \frac{1}{\sqrt{n}}\Phi^{-1}(1 - \alpha_0/2)\left\{\boldsymbol{\delta}_{t_0-1}'\widehat{\Sigma}_\alpha\boldsymbol{\delta}_{t_0-1}\right\}^{1/2}, \tag{3.7}$$

where $\widehat{\mathrm{VaR}}_{VHS,t_0-1}^{(\alpha)}(r_{t_0})$ is defined in (3.5), $\widehat{\Sigma}_\alpha$ is a consistent estimator of Σ_α and

$$\boldsymbol{\delta}_{t_0-1}' = \left(\frac{\partial\tilde{\sigma}_{t_0}(\hat{\boldsymbol{\theta}}_n)}{\partial\boldsymbol{\theta}}\xi_{n,\alpha} \qquad \tilde{\sigma}_{t_0}(\hat{\boldsymbol{\theta}}_n)\right).$$

In other words,

$$\lim_{n\to\infty}P_{t_0-1}\left[\left|\mathrm{VaR}_{H,t_0-1}^{(\alpha)}(r_{t_0}) - \widehat{\mathrm{VaR}}_{VHS,t_0-1}^{(\alpha)}(r_{t_0})\right| \le \frac{1}{\sqrt{n}}\Phi^{-1}(1 - \alpha_0/2)\left\{\boldsymbol{\delta}_{t_0-1}'\widehat{\Sigma}_\alpha\boldsymbol{\delta}_{t_0-1}\right\}^{1/2}\right] = 1 - \alpha_0.$$

Remark 3.3. It is worth noting that the previous CI has to be understood conditionally on I_{t_0-1} for fixed t_0. In particular, it cannot be directly applied when $t_0 - 1 = n$. Indeed, the bounds in (3.7) are no longer random in this case.[4] See Beutner et al. (2019) for an asymptotic justification of this type of CI for conditional objects.

4. Numerical illustrations

An alternative to the univariate approaches is a multivariate strategy, requiring a dynamic model for the vector of risk factors $\boldsymbol{y}_t$. We describe in Appendix A two multivariate approaches — the spherical method and the Filtered Historical Simulation (FHS) method. Such methods might perform better than univariate ones because they incorporate information stemming from individual returns rather than an aggregate information stemming from portfolio returns. However, at least for large portfolios, multivariate approaches can be very challenging, when at all possible.

The first section is devoted to numerical comparisons of the different methods. We first study the reliability of the naive approach in a static framework where the returns are iid and the portfolio is crystallized. In a dynamic framework, we then compare the performance of the univariate and multivariate approaches when the dynamic multivariate model is well specified and the dimension is small. Next, we consider misspecified GARCH models and, possibly, higher dimensions. The second section concerns real data examples based on large sets of US stocks.[5]

4.1. Monte-Carlo experiments

The theoretical conditional VaR in (2.3) depends on the information set I_{t-1}. In our first two sets of experiments, the vector of individual returns will be simulated using multivariate GARCH and the different estimators will be compared to the same target VaR obtained by taking the full information set (i.e. including the returns of the individual assets, not only the returns of the portfolio). In the third set of experiments, we will estimate a misspecified multivariate model for which the true conditional VaR is no longer available.

---

[4] Except if one assumes, as in Gong et al. (2010), that the number of conditioning values is finite as in the ARCH(q) case.

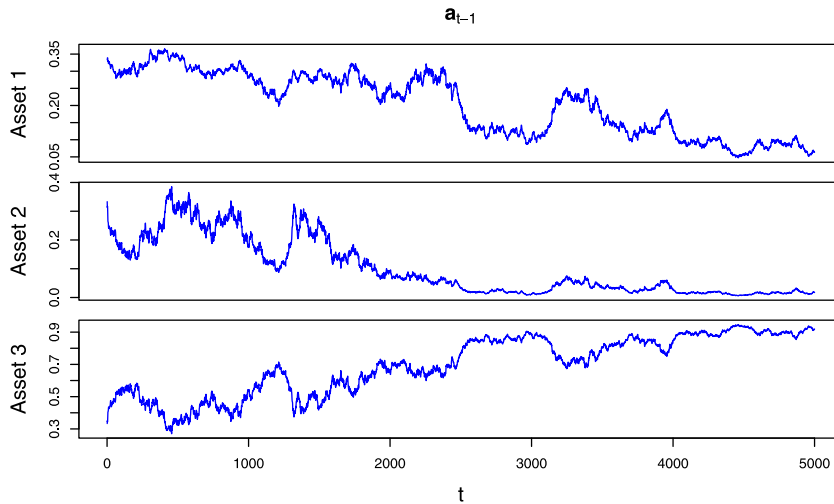[5] The code and data used in the paper are available on the web site http://perso.univ-lille3.fr/~cfrancq/Christian-Francq/VaRPortfolio.html.

Fig. 1. Time-varying composition of the simulated crystallized portfolio.

4.1.1. Static model and crystallized portfolio

For a simple illustration of Lemma 2.1, we consider a crystallized equally weighted portfolio of 3 assets (of initial price $p_{i0} = 1000$): $V_t = \sum_{i=1}^{3} p_{it}$ and $\mu_{i,t} = 1$ for $i = 1, 2, 3$. Thus, the return portfolio composition is time varying, with coefficients $\mathbf{a}_{t-1} = (a_{1,t-1}, a_{2,t-1}, a_{3,t-1})'$ 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, Gaussian, centered, with variance $\mathrm{Var}(\mathbf{y}_t) = \mathbf{\Sigma}^2 = \boldsymbol{DRD}$, 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 $\mathbf{a}_{t-1}$ of the portfolio is plotted in Fig. 1. As we have seen in Section 2.2, this vector is non stationary. More precisely, as discussed in Section 2.2, with increasing probability the composition $\mathbf{a}_{t-1}$ of the portfolio is arbitrarily close to one of the three single-asset portfolios $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$. Experiments conducted with different parameters (in particular different correlation matrices) led to the same conclusion.

It is thus non surprising to see that the univariate return series r_t plotted in Fig. 2 exhibits some nonstationarity features, in particular unconditional heteroscedasticity. The increased variance in the second part of the sample reflects the fact that the portfolio tends to be less and less diversified (see Fig. 1).

When $\mathbf{\Sigma}_t$ and $\boldsymbol{m}_t$ are constant, the FHS estimator in (A.7) reduces to the opposite of the quantile of the portfolio's virtual returns: $\widehat{\mathrm{VaR}}_{FHS,t-1}^{(\alpha)}(r_t) := -q_\alpha\left(\{\mathbf{a}_{t-1}'\mathbf{y}_1, \ldots, \mathbf{a}_{t-1}'\mathbf{y}_{t-1}\}\right)$ and this estimator coincides with the VHS estimator. These empirical quantiles were computed starting from $t = 100$. The naive estimator is the opposite of the quantile of the portfolio's returns: $\widehat{\mathrm{VaR}}_{N,t-1}^{(\alpha)}(r_t) = -q_\alpha\left(\{\mathbf{a}_1'\mathbf{y}_1, \ldots, \mathbf{a}_{t-1}'\mathbf{y}_{t-1}\}\right)$. The spherical method, based on the estimation of $\mathbf{\Sigma}$, was computed on the same range of observations. In view of (A.6), we have $\widehat{\mathrm{VaR}}_{S,t-1}^{(\alpha)}(r_t) = \sqrt{\mathbf{a}_{t-1}'\widehat{\mathbf{\Sigma}}_{t-1}\mathbf{a}_{t-1}}q_{1-2\alpha}\left(\{\widehat{\mathbf{\Sigma}}_{t-1}^{-1}\mathbf{y}_1, \ldots, \widehat{\mathbf{\Sigma}}_{t-1}^{-1}\mathbf{y}_{t-1}\}\right)$, where $\widehat{\mathbf{\Sigma}}_{t-1}$ is the empirical covariance matrix of the observations $\mathbf{y}_1, \ldots, \mathbf{y}_{t-1}$. Fig. 3 displays the sample paths of the true conditional VaR as well as the estimated VaRs. It can be seen that the spherical method converges faster to the true value than the FHS = VHS method. Unsurprisingly, the univariate naive method fails to converge to the theoretical conditional VaR based on the full information set. Now we turn to a less artificial setting.

4.1.2. Well-specified multivariate GARCH models

In this section, we simulate the process of log-returns $\mathbf{y}_t = \mathbf{\Sigma}_t \boldsymbol{\eta}_t$ from the corrected Dynamic Conditional Correlation (cDCC) GARCH model of Aielli (2013). For the multivariate approaches, we estimate the same cDCC-GARCH(1,1) model. For the univariate approaches we estimate GARCH(1,1) models, which are generally misspecified (see Example 2.1).

We consider the minimum variance portfolio variance given by

$$
r_t = \mathbf{y}_t'\mathbf{a}_{t-1}, \quad \mathbf{a}_{t-1} = \frac{\mathbf{\Sigma}_t^{-2}\boldsymbol{e}}{\boldsymbol{e}'\mathbf{\Sigma}_t^{-2}\boldsymbol{e}}, \quad \text{where} \quad \boldsymbol{e} = (1, \ldots, 1)' \in \mathbb{R}^m. \tag{4.1}
$$

We simulated N independent trajectories of length n for the cDCC-GARCH(1,1) model. On each simulation, the first n_1 observations are used (i) to obtain an estimator $\widehat{\boldsymbol{\vartheta}}_{n_1}$ of the parameters involved in $\mathbf{\Sigma}_t$ by the three-step estimator defined in Appendix C of Francq and Zakoïan (2018), and (ii) to estimate the quantiles required for the VaR estimator. On the last
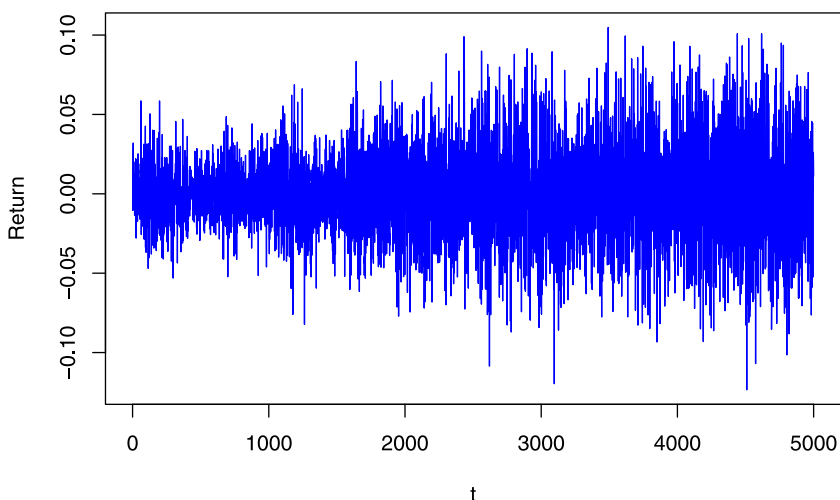
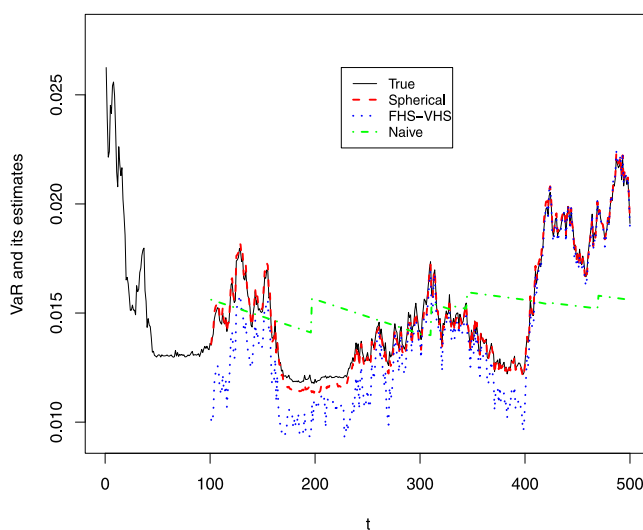**Fig. 2.** Returns of the simulated crystallized portfolio.

Fig. 3. True and estimated VaRs of the crystallized portfolio.

$n - n_1$ simulations, *i.e.* for $t = n_1 + 1, \ldots, n$, we compared the theoretical VaR of the portfolio with the three estimates obtained from the spherical, FHS and VHS methods. We considered portfolios of $m = 2$ assets. The different designs, displayed in Appendix C, 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 100,000 theoretical values of the conditional VaR with their estimates obtained by the spherical, FHS and VHS methods. Denote by MSE_S, MSE_{FHS} and MSE_{VHS} the mean square errors (MSE) of prediction of the three methods. Table 1 displays the relative efficiency (RE) of the spherical method with respect to the FHS and the VHS methods, as measured by the ratios $\mathrm{MSE}_{FHS}/\mathrm{MSE}_S$ and $\mathrm{MSE}_{VHS}/\mathrm{MSE}_S$. It should be underlined that all MSE's are computed with respect to the full information VaR's, which a priori favors the multivariate methods. Let us first briefly compare the two multivariate approaches: in designs A–H (with spherical distributions) the spherical method is as expected more efficient than the FHS method (for Designs C and D, the spherical method can be twice more efficient than the other multivariate method). On the contrary, the bottom panel of Table 1 reveals that, when the density is strongly asymmetric, the FHS method can be much more efficient than the spherical method.

One question of interest is whether the VHS estimator (targeted to estimate the hybrid VaR) can be used to approximate this full-information VaR. Let us first compare the VHS method to the multivariate approaches in the spherical case. The univariate VHS method is apparently dominated by the multivariate methods in designs A–H but one has to recall that the reference VaR to which the methods are compared is designed for the multivariate framework. Better results are obtained for the VHS method in designs E–H (identically distributed returns) as opposed to A–D, and for designs

Table 1

Relative efficiency of the Spherical method with respect to the FHS method ($\text{MSE}_{FHS}/\text{MSE}_S$) and with respect to the VHS method ($\text{MSE}_{VHS}/\text{MSE}_S$) for predicting the full-information VaR. Designs A–H correspond to spherical distributions, Designs A*-H* correspond to non-spherical distributions (see Appendix C). The number of independent replications of each design is 100 and, for each simulation, n_1 observations were used for estimation and 1000 out-of-sample predictions were computed.

n_1	α		A	B	C	D	E	F	G	H
500	1%	$\text{MSE}_{FHS}/\text{MSE}_S$	1.12	1.11	2.57	2.35	1.08	1.17	1.23	1.42
		$\text{MSE}_{VHS}/\text{MSE}_S$	53.9	16.9	188.	82.0	1.63	1.53	2.42	2.18
	5%	$\text{MSE}_{FHS}/\text{MSE}_S$	1.21	1.03	1.81	1.40	1.18	1.12	1.12	1.19
		$\text{MSE}_{VHS}/\text{MSE}_S$	28.7	9.76	130.	74.6	1.69	1.56	2.31	1.99
1000	1%	$\text{MSE}_{FHS}/\text{MSE}_S$	1.30	1.11	2.35	1.62	1.53	1.51	1.57	1.36
		$\text{MSE}_{VHS}/\text{MSE}_S$	91.6	23.4	303.	79.8	1.93	2.53	4.43	2.23
	5%	$\text{MSE}_{FHS}/\text{MSE}_S$	1.14	1.03	2.07	1.00	1.25	1.08	1.33	1.01
		$\text{MSE}_{VHS}/\text{MSE}_S$	55.4	15.7	267.	82.5	1.75	2.44	4.14	2.01
			A*	B*	C*	D*	E*	F*	G*	H*
500	1%	$\text{MSE}_{FHS}/\text{MSE}_S$	0.08	0.06	0.03	0.04	0.12	0.10	0.10	0.12
		$\text{MSE}_{VHS}/\text{MSE}_S$	2.70	4.45	2.78	3.32	0.18	0.14	0.16	0.18
	5%	$\text{MSE}_{FHS}/\text{MSE}_S$	0.32	0.32	0.13	0.25	0.45	0.52	0.46	0.53
		$\text{MSE}_{VHS}/\text{MSE}_S$	5.3	11.0	10.2	14.9	0.65	0.66	0.71	0.73
1000	1%	$\text{MSE}_{FHS}/\text{MSE}_S$	0.08	0.03	0.02	0.02	0.06	0.03	0.03	0.04
		$\text{MSE}_{VHS}/\text{MSE}_S$	2.20	2.43	2.31	1.67	0.05	0.04	0.07	0.06
	5%	$\text{MSE}_{FHS}/\text{MSE}_S$	0.34	0.19	0.09	0.11	0.30	0.24	0.21	0.29
		$\text{MSE}_{VHS}/\text{MSE}_S$	3.78	6.68	10.2	8.72	0.26	0.35	0.59	0.44

{A,B,E,F} as opposed to {C,D,G,H} (independent returns). This can be intuitively explained as follows. Univariate methods are expected to behave better when the trajectories of the underlying returns are close, which is the case when the two assets have similar dynamic models and are strongly dependent. It is thus non surprising to note that the worst results for the VHS (by comparison with the multivariate methods) occur for designs C-D (independent returns with very different dynamics), and the best results occur with designs E-F (dependent identically distributed returns).

In the case of non-spherical distributions (designs A*-H*) the same conclusions hold: the more dependent the assets and the closer their dynamics, the better results for the VHS. This method clearly outperforms the spherical approach in designs E*-H*. Unexpectedly, the results reveal that the univariate VHS method can even outperform the FHS approach (design E* with $n_1 = 1000$).

From these experiments, it appears that the accuracy of the approximation provided by the VHS approach is very dependent from the model parameters. However, these first two examples clearly favor the multivariate methods by assuming that the dynamic model is well specified. In the next example, we consider a data generating process that does not belong to the GARCH class, and we will compare the different methods using backtests.

4.1.3. Misspecified GARCH models

We simulated m-multivariate factor models, with two GARCH factors of the form

$$f_{1t} = \sigma_{1t}\eta_{1t}, \quad f_{2t} = \sigma_{2t}\eta_{2t},$$

where $(\eta_{1t})_t$ and $(\eta_{2t})_t$ are two independent sequences of iid $\mathcal{N}(0, 1)$-distributed random variables. The volatilities follow standard GARCH(1, 1) equations of the form

$$\sigma_{it}^2 = \omega_i + \alpha_i f_{i,t-1}^2 + \beta_i \sigma_{i-1,t}^2.$$

We took $(\omega_1, \alpha_1, \beta_1) = (1, 0.09, 0.87)$ and $(\omega_2, \alpha_2, \beta_2) = (0.1, 0.7, 0.01)$, so that the dynamics of the two factors be quite distinct. The even and odd components of our simulated factor model are respectively of the form

$$y_{2k,t} = f_{2t} + e_{2k,t}, \qquad y_{2k+1,t} = f_{1t} + e_{2k+1,t},$$

where $(e_{k,t})_t$, for $k = 1, \ldots, m$, are idiosyncratic independent iid noises with law $\mathcal{N}(0, 0.1^2)$. To obtain a graphical comparison of the VaR estimates, we first simulated a trajectory of size 1100 of the factor model with $m = 4$. A crystallized portfolio of composition $(1/m, \ldots, 1/m)$ at time $t = 1$ has been considered. For the multivariate approaches we estimated cDCC-GARCH(1,1) models, while for the univariate approaches we estimated GARCH(1,1) models. The four competing estimators of the 5% VaR_{t-1} at time $t = 1001$ were estimated on the basis on the first 1000 simulated values $y_1, \ldots, y_{t-1}$. Then, VaR at time $t = 1002$ was estimated based on the past 1000 simulations $y_2, \ldots, y_{t-1}$. We continued this way until we obtained the last VaR estimations at time $t = 1100$. Fig. 4 shows that the estimates obtained by the Spherical, FHS and VHS methods are very close (actually, they are not distinguishable on the figure), whereas the estimates obtained by the naive method behave differently. This can be explained by the fact that the portfolio is crystallized but not static. In other words, even if the portfolio is constituted of an equal quantity of the m simulated assets, the return r_t is not a fixed average of the individuals returns y_{kt} (see Fig. 5).

To compare the methods by using formal backtests, we considered the same framework of GARCH estimations on rolling windows of length 1000, but the methods have been backtested on a longer period of length 2000. Moreover, in
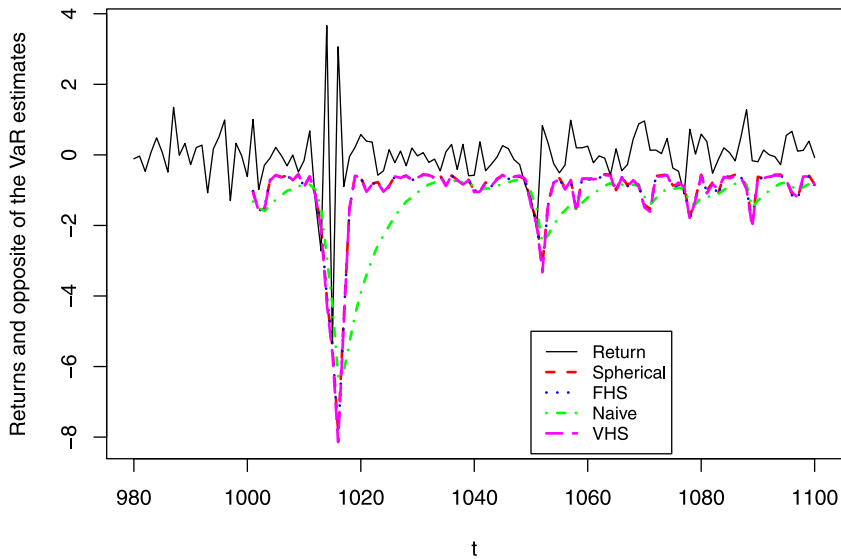
Fig. 4. Comparison of 4 VaRs at the horizon 1 for a crystallized portfolio of $m = 4$ simulated assets.
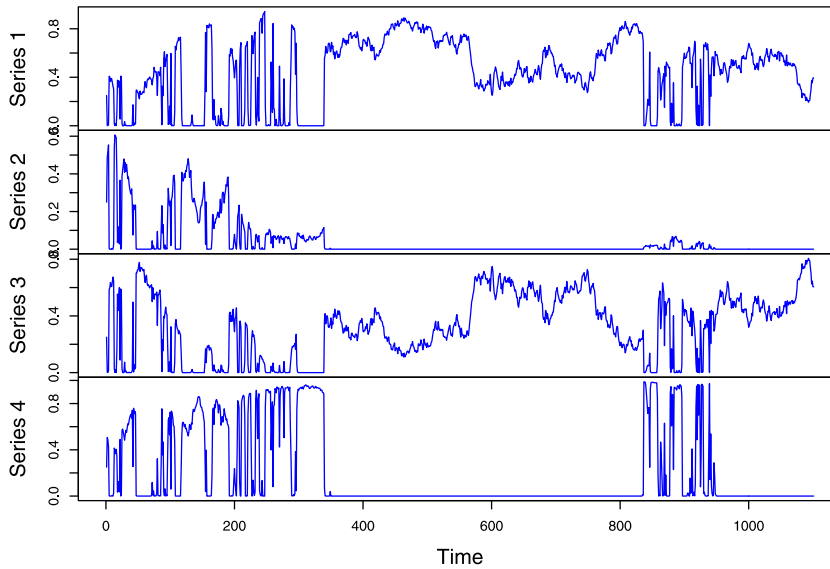

Fig. 5. Time-varying composition of the return of the crystallized portfolio as function of the returns of the individual assets.

order to obtain a clearcut comparison between the naive method and the VHS method, the composition of the portfolio has to be highly time-varying. We thus simulated portfolios whose composition alternates as follows: we take an equal proportion of the returns of the even assets $\epsilon_{2k,t}$ during a period of length 100, and then we switch to an equal proportion of the odd assets $\epsilon_{2k+1,t}$ during another period of length 100. Table 2 summarizes the results of the 4 VaR estimation methods for $m = 2, 4, 8, 100$. This simulation exercise is intensive since 2000 DCC-GARCH models were estimated for each of the two multivariate methods, and 2000 univariate GARCH(1,1) models were estimated for each of the univariate methods. The spherical and FHS methods become rapidly too time consuming when the number m of returns increases, because multivariate m-GARCH models have to be estimated. Interestingly, the numerical complexity of the univariate methods does not increase much with m, so that Table 2 reports results on portfolios of $m = 8$ and $m = 100$ assets for the univariate methods only.

In this table, the column Viol gives the relative frequency of violations (in %), while the columns LRuc, LRind and LRcc give respectively the p-values of the unconditional coverage test that the probability of violation is equal to the nominal 5% level, the independence test that the violations are independent and the conditional coverage test of Christoffersen (2003). Conclusions drawn from those backtests, which solely focus on the violations, are that all methods are validated

Table 2
Backtests of the 5%-VaR estimates for the factor model. The composition of the portfolio changes every 100 observations. For the multivariate (resp. univariate) approaches DCC (resp. GARCH(1,1)) models are estimated on 1000 observations. All statistics are computed on 2000 observations.

		Viol[a]	LRuc[b]	LRind[b]	LRcc[b]	$\overline{\text{VaR}}$[c]	AV[d]	ES[d]	Loss[d]	DM[e]
$m = 2$	Naive	5.20	68.34	79.24	88.89	4.64	1.87	6.56	0.33	–
	VHS	5.55	26.71	72.63	50.81	4.26	1.07	5.30	0.27	5.e−10
	Spherical	5.30	54.19	86.71	81.87	4.28	1.10	5.49	0.27	1.e−09
	FHS	5.60	22.67	90.67	47.82	4.25	1.10	5.33	0.27	3.e−09
$m = 4$	Naive	5.55	26.71	12.95	17.12	4.51	1.90	6.08	0.33	–
	VHS	4.35	17.30	90.93	39.26	4.60	1.18	5.45	0.28	1.e−07
	Spherical	5.20	68.34	5.95	15.59	4.36	1.19	5.61	0.28	2.e−08
	FHS	4.30	14.15	87.19	33.50	4.60	1.19	5.55	0.28	2.e−07
$m = 8$	Naive	5.10	83.79	56.34	82.87	4.87	1.61	6.10	0.33	–
	VHS	5.50	31.23	69.02	55.44	4.44	1.05	5.38	0.28	1.e−08
$m = 100$	Naive	4.90	83.69	6.93	18.8	4.53	2.16	7.65	0.34	–
	VHS	5.25	61.07	81.42	85.46	4.56	1.09	5.61	0.29	6.e−09

[a]% of violations of the estimated 5%-VaRs.
[b]p-values of backtests of the violations.
[c]Average VaR.
[d]Average amount of violation criteria.
[e]p-values of DM tests that the naive method has the same loss.

on these experiments. In particular, it is interesting to notice that the naive method does not behave so poorly in terms of backtests. It is necessary to introduce alternative statistics to differentiate the different approaches. From a portfolio's manager point of view, it is interesting to minimize the average VaR (denoted $\overline{\text{VaR}}$ in the table) in order to minimize the reserves. For both regulator and manager, it is also important to minimize the amount of violation. The column AV displays the average amount of violation, and the column ES gives the expected shortfall, that is the average loss when the VaR is violated: for each estimator $\widehat{VaR}_t$ of the conditional VaR, let

$$\text{AV} = \frac{\sum_{t=1}^{n} -(r_t + \widehat{VaR}_t)\mathbf{1}_{\{r_t < -\widehat{VaR}_t\}}}{\sum_{t=1}^{n} \mathbf{1}_{\{r_t < -\widehat{VaR}_t\}}}, \qquad \text{ES} = \frac{\sum_{t=1}^{n} -r_t \mathbf{1}_{\{r_t < -\widehat{VaR}_t\}}}{\sum_{t=1}^{n} \mathbf{1}_{\{r_t < -\widehat{VaR}_t\}}}.$$

These statistics clearly show that the naive approach is inefficient compared to its competitors. With this method, the amount of violation tends to be higher whatever the size m of the portfolio. For these statistics AV and ES, the VHS approach appears comparable to the multivariate methods when comparison is possible, that is when m is not too large. Alternative comparisons are provided by introducing the loss function

$$\text{Loss} = \frac{1}{n}\sum_{t=1}^{n} -(r_t + \widehat{VaR}_t)(\alpha - \mathbf{1}_{\{r_t < -\widehat{VaR}_t\}}),$$

also considered by Giacomini and Komunjer (2005), and Gneiting (2011) in the context of forecast evaluation. The last column of Table 2 reports, for each of the three non-naive methods, p-values of the Diebold and Mariano (1995) test for the null that the naive method produces the same loss against the alternative that it induces higher loss. The null is rejected in each situation, leading to the same conclusion as before: the naive method is outperformed by its three competitors when m is small, and by the VHS method when m is large.

4.2. Real data

We start by plotting the returns of an actual crystallized portfolio, showing the same kind of behavior as the simulated portfolio of Fig. 2. The portfolio is obtained by equally weighting 3 stocks, ADM (Advanced Micro Devices), BFB (Brown-Forman Corporation) and AMZ (Amazon). on the period 1997-05-15 to 2018-09-05 (5363 observations). The composition $\mathbf{a}_{t-1}$ of the portfolio is plotted in Fig. 6. It is seen that the third asset becomes preponderant at the end of the period. The plot of the portfolio's returns in Fig. 7 shows that the changes in the portfolio composition induce apparent non-stationarities. Contrary to the simulated portfolio of Fig. 2, the volatility decreases when the portfolio becomes more concentrated, which is explained by the fact the third asset is also the less volatile. Next, we will compare the univariate methods on portfolios whose composition is strongly time varying. Then, we will consider portfolios that are regularly rebalanced so that their compositions do not vary too much.

4.2.1. Estimating the conditional VaR of portfolios of US stocks

We now consider portfolios built from a set of $m = 49$ US stocks covering 2489 trading days, from January 4, 1999 to December 31, 2008. The data have been kindly provided to us by Sébastien Laurent, and are described in Laurent et al. (2016). The top panel of Fig. 8 displays the returns of a crystallized portfolio which was fully diversified at the beginning of the period, *i.e* with composition $\mathbf{a}_0 = (1/m, \ldots, 1/m)$ at time $t = 1$, and for which the number of units of each asset
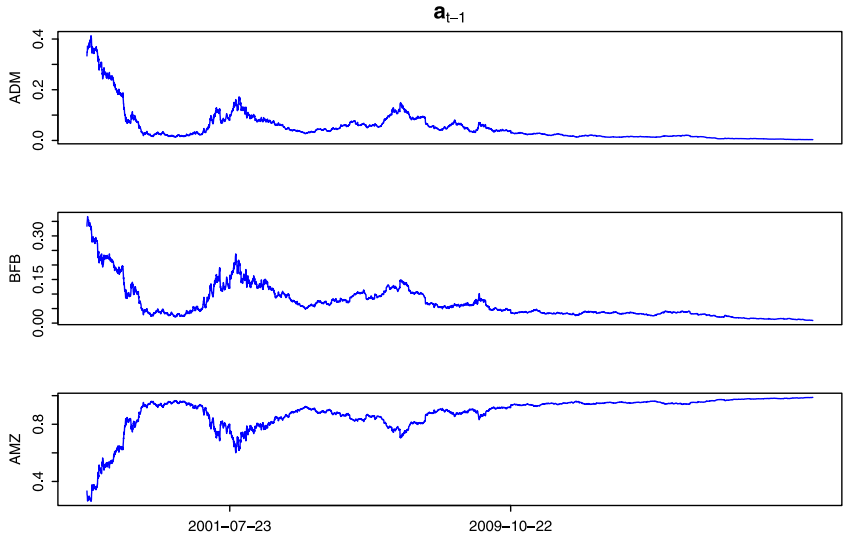
Fig. 6. Time-varying composition of the crystallized portfolio based on the three US stocks.
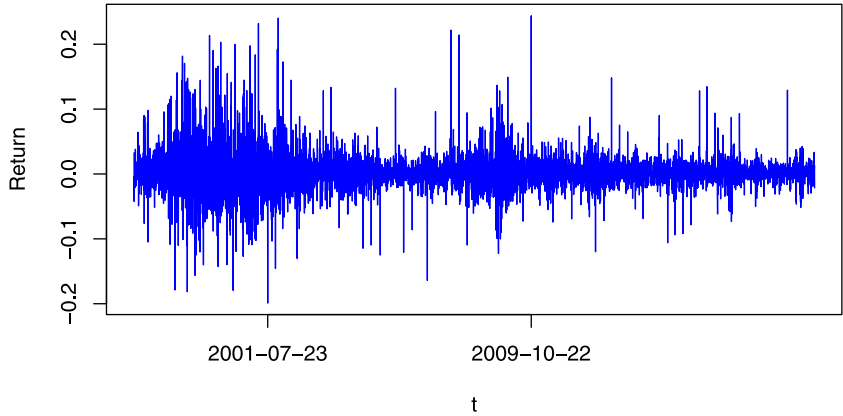

Fig. 7. Returns of the crystallized portfolio based on the three US stocks.

$\mu_{i,t} = 1$ is time-invariant. The bottom panel of this figure displays $\max_{i \in \{1,...,49\}} a_{i,t}$ as function of t. This figure shows that the composition of the portfolio is time-varying, and that the portfolio tends to become more and more concentrated. At the beginning of the period, the return of the crystallized portfolio is an equi-weighted average of the individual returns, but at the end of the period, one of the individual returns tends to have a prominent weight (this individual return is the UNILEVER stock from February 20, 2003 onwards).

Given the large number of assets, we did not implement multivariate approaches to estimate the VaR of this portfolio. Estimating a multivariate GARCH(1,1) model by QML in this setting would require inverting very large correlation/covariance matrices at each step of the optimization algorithm. There exists multivariate approaches – either based on constrained models or using alternatives to the full QML (e.g. the composite likelihood as in Engle et al. (2017), or the Equation-by-Equation method of Francq and Zakoïan (2016)), or using intraday data (e.g. Boudt et al. (2017)) – which do not fit into our semiparametric GARCH framework.

Fig. 9 displays the estimates of the 5%-VaR obtained from the Naive and VHS methods. Starting from $t = 1001$, the estimates are computed from all the previous observations $r_1, \ldots, r_{t-1}$. As can be seen from the figure, the naive and VHS methods provide very similar results and the backtests used in the previous section are not able to distinguish them. At first sight, this result is quite surprising, but it can be explained by the fact that the composition of the portfolio varies relatively slowly and, even if the composition is changing, the dynamics does not change drastically because most of the individual returns follow similar GARCH models. The interesting conclusion is that, even if the naive method is not supported by rigorous theoretical results, it may work surprisingly well in practice.

In a second experiment, we considered a portfolio whose composition changes every week, between an equi-weighted average of the stocks MS, F, GM and an equi-weighted average of the stocks CVX, XOM, ECX. At the beginning of the week,
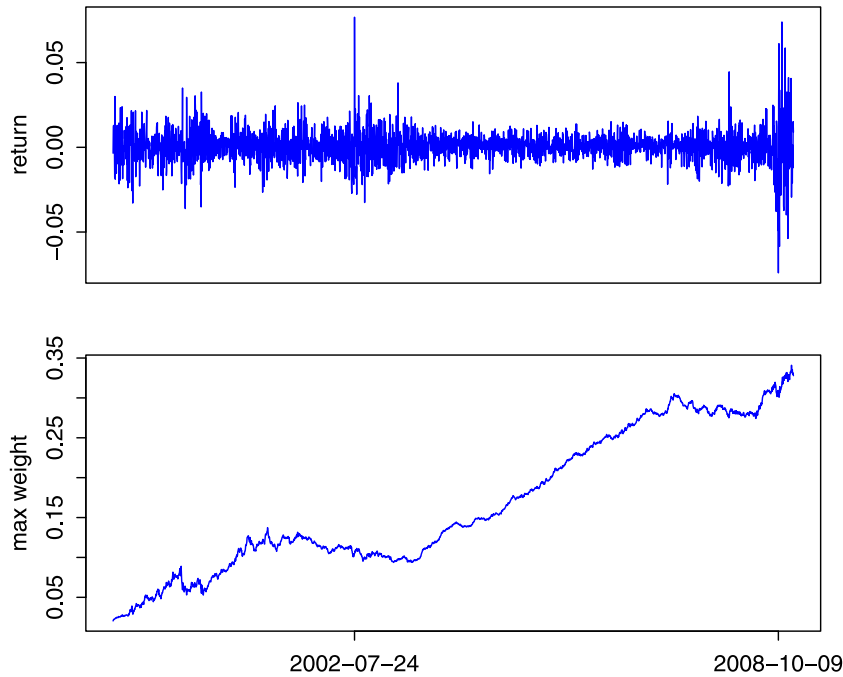
Fig. 8.   Returns and maximum weight of a crystallized portfolio of 49 US stocks.
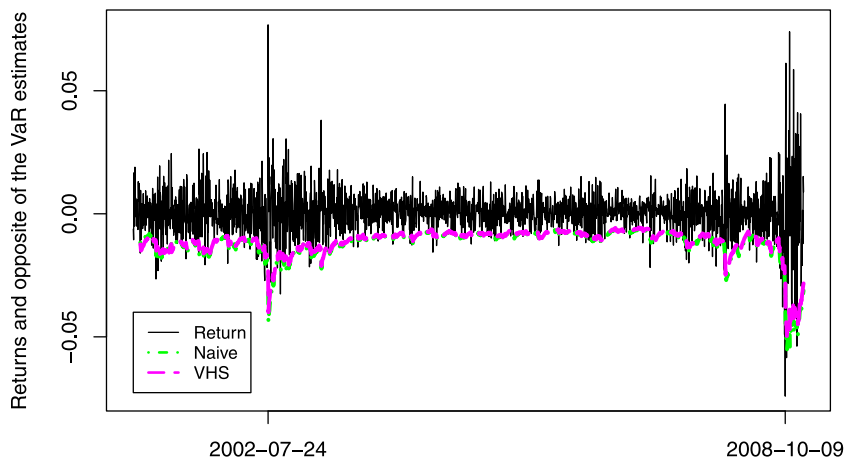

Fig. 9.   Naive and VHS estimations of the 5%-VaR at horizon 1 for a portfolio of US stocks.

the portfolio is thus composed of the same amount of the three assets, then the portfolio is crystallized until the end of the week. The two sets of stocks have been chosen because of their different dynamic behaviors. Results are displayed in Table 3. Once again, the naive and VHS methods are not much different, but there are some discrepancies in favor of the VHS method (as indicated by the DM test).

4.2.2. Comparing naive and VHS methods on rebalanced portfolios

In practice, portfolios are often periodically re-balanced. Intuitively, the naive and VHS approaches should behave similarly in this situation. To verify this intuition, we now build portfolios with the $m = 19$ stocks that illustrate Chapter 17 of Boyd and Vandenberghe (2018). The data set covers the period from 2004-01-02 to 2013-12-31 (2517 values). We consider a portfolio which is equally diversified at the beginning of the period (*i.e.* we take $\mu_{i,t} = \mu_i = V_0/(mp_{i,0})$ at $t = 0$, corresponding to the beginning date 2004-01-02), and we re-balance the portfolio every T periods (such that $\mu_{i,t} = V_t/(mp_{i,t})$ at $t = kT$ for all $k \in \mathbb{N}$). If the portfolio is re-balanced at any time, *i.e.* $T = 1$, then $a_{i,t} = 1/m$ for all t, and thus the naive and VHS methods coincide. To limit transaction costs, the investor can maintain the same asset

Table 3
Backtests of estimates of the VaR at the level α obtained by univariate methods, for a portfolio of US stocks whose composition changes every week over the period 1999–2008.

		Viol	LRuc	LRind	LRcc	$\overline{\text{VaR}}$	AV	ES	Loss	DM
$\alpha = 5\%$	Naive	6.234	1.484	91.914	11.426	0.024	0.012	0.036	1.9e−03	–
	VHS	5.782	11.812	78.518	47.234	0.024	0.010	0.035	1.7e−03	4e−04
$\alpha = 1\%$	Naive	1.307	18.858	40.646	49.032	0.038	0.021	0.061	6e−04	–
	VHS	1.207	36.968	29.420	59.235	0.036	0.016	0.057	5e−04	7e−03
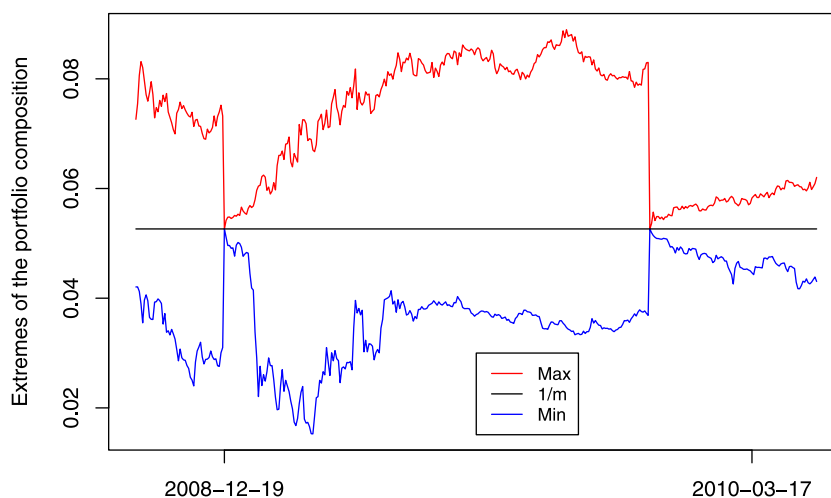
Fig. 10. Maximal (in red) and minimal (in blue) value of the composition $\mathbf{a}_{t-1}$ of a portfolio that is re-balanced every $T = 250$ days.
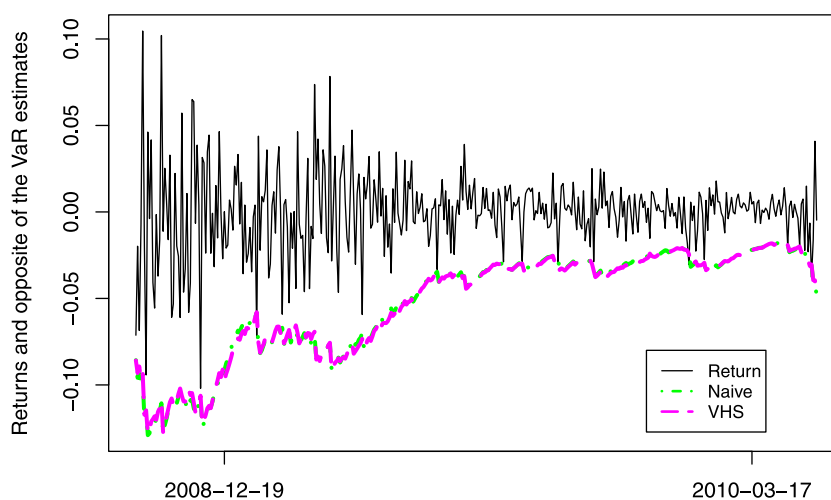
Fig. 11. Estimated 1%-VaR by the naive (dotted green line) and VHS (magenta line) methods.

allocation for an extended period of time. The so-called lazy portfolios or permanent portfolios are re-balanced every year or at any time its asset allocation strays too far from its initial state. Fig. 10 shows that, when the portfolio is re-balanced every year, the composition of the portfolio can deviate much from the fully diversified portfolio (for which $a_{i,t} = 1/m$ for all $i \in \{1, \dots, m\}$). This does not necessarily entail a huge difference between the VaR's estimated by the naive and VHS methods. Actually, for the nominal risk level $\alpha = 1\%$, the maximum difference between the estimated VaR has been observed on 2008-12-19, with a naive VaR of 0.0916 and a VHS VaR of 0.0978. As illustrated by Fig. 11 the two estimated VaR are hardly distinguishable. This is interesting because it entails that the asymptotic theory built for the VHS method should also apply for the naive method. In other words, the naive method is not so naive if the portfolio is re-balanced from time to time, as recommended by finance professionals.

5. Conclusion

This paper developed a method for estimating the conditional VaR of a portfolio of asset returns, without relying on a joint dynamic modeling of the vector of returns. For large portfolios, using optimization routines in multivariate approaches often entails formidable numerical difficulties. By circumventing the dimensionality curse, univariate methods provide an operational alternative.

The naive method discussed in this paper has no theoretical grounds because it implicitly, and erroneously when the composition is time varying, relies on the stationarity of the returns process. In many cases, however, it behaves satisfactorily as our numerical experiments revealed. For the VHS method, we developed an asymptotic theory for a general class of dynamic models, which are not directly estimated on observations but rather on reconstituted returns. The obtained asymptotic results allow to quantify the estimation risk that should be taken into account in risk management. From our numerical experiments, the multivariate methods can be recommended when the size of the portfolio is small and the estimated multivariate GARCH model is likely to be well-specified, which is in line with the existing literature, in particular Santos et al. (2013). When the number of underlying assets is large, or when finding an appropriate multivariate specification is difficult, the univariate methods offer a valuable alternative. The VHS method requires computing virtual returns, which is negligible computational burden. Thus, the two univariate methods display similar numerical complexity. However, when the naive and VHS estimators differ substantially, the former should not be considered as reliable. When they provide similar results, the asymptotic results obtained for the VHS method could also be used to evaluate the accuracy of the naive method.

Acknowledgments

Appendix A. Multivariate approaches

We assume in this appendix that the vector of log-returns follows a general multivariate model of the form

$$\boldsymbol{y}_t = \boldsymbol{m}_t(\boldsymbol{\vartheta}_0) + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t = \Sigma_t(\boldsymbol{\vartheta}_0)\boldsymbol{\eta}_t, \tag{A.1}$$

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; $\boldsymbol{\eta}_t$ is independent from the $\boldsymbol{y}_{t-i}$ for $i > 0$; the $m \times m$ non-singular matrix $\Sigma_t(\boldsymbol{\vartheta}_0)$ and the $m \times 1$ vector $\boldsymbol{m}_t(\boldsymbol{\vartheta}_0)$ are functions of the past values of $\boldsymbol{y}_t$ which are parameterized by a d-dimensional parameter $\boldsymbol{\vartheta}_0$:

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

Multivariate approaches require specifying the first two conditional moments in (A.2) of the vector of individual returns. While the conditional mean is generally modeled using a small-order AR process, there are plenty of GARCH-type specifications for the conditional variance. See for instance Bauwens et al. (2006), Francq and Zakoïan (2019, Chapter 10) or Bauwens et al. (2012) for presentations of the most commonly used specifications.

In view of (A.1) and (2.2), the portfolio's return satisfies

$$r_t = \mathbf{a}_{t-1}' \boldsymbol{m}_t(\boldsymbol{\vartheta}_0) + \mathbf{a}_{t-1}' \Sigma_t(\boldsymbol{\vartheta}_0)\boldsymbol{\eta}_t, \tag{A.3}$$

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

$$\mathrm{VaR}_{t-1}^{(\alpha)}(r_t) = -\mathbf{a}_{t-1}' \boldsymbol{m}_t(\boldsymbol{\vartheta}_0) + \mathrm{VaR}_{t-1}^{(\alpha)}\left(\mathbf{a}_{t-1}' \Sigma_t(\boldsymbol{\vartheta}_0)\boldsymbol{\eta}_t\right). \tag{A.4}$$

A.1. Conditional VaR estimation under conditional ellipticity

The VaR formula can be simplified if we assume that the errors $\boldsymbol{\eta}_t$ have a spherical distribution, that is, for any non-random vector $\boldsymbol{\lambda} \in \mathbb{R}^m$, $\boldsymbol{\lambda}'\boldsymbol{\eta}_t \overset{d}{=} \|\boldsymbol{\lambda}\|\eta_{1t}$, where $\|\cdot\|$ denotes the euclidian norm on $\mathbb{R}^m$, η_{it} denotes the ith component of $\boldsymbol{\eta}_t$, and $\overset{d}{=}$ stands for the equality in distribution. Note that assuming sphericity of the distribution of $\boldsymbol{\eta}_t$ amounts to assuming ellipticity of the conditional distribution of $\boldsymbol{y}_t$ satisfying (A.1)–(A.2). Under the sphericity assumption we have

$$\mathrm{VaR}_{t-1}^{(\alpha)}(r_t) = -\mathbf{a}_{t-1}' \boldsymbol{m}_t(\boldsymbol{\vartheta}_0) + \left\|\mathbf{a}_{t-1}' \Sigma_t(\boldsymbol{\vartheta}_0)\right\| \mathrm{VaR}^{(\alpha)}(\eta), \tag{A.5}$$

where $\mathrm{VaR}^{(\alpha)}(\eta)$ is the (marginal) VaR of η_{1t}. Under the sphericity assumption, by (A.5) a natural strategy for estimating the conditional VaR of a portfolio is to estimate $\boldsymbol{\vartheta}_0$ by some consistent estimator $\widehat{\boldsymbol{\vartheta}}_n$ in a first step, to extract the residuals and to estimate $\mathrm{VaR}^{(\alpha)}(\eta)$ in a second step.

An estimator of the conditional VaR at level α accounting for the conditional ellipticity is thus

$$\widehat{\text{VaR}}_{S,t-1}^{(\alpha)}(r_t) = -\mathbf{a}_{t-1}'\widetilde{\boldsymbol{m}}_t(\widehat{\boldsymbol{\vartheta}}_n) + \|\mathbf{a}_{t-1}'\widetilde{\Sigma}_t(\widehat{\boldsymbol{\vartheta}}_n)\|\xi_{n,1-2\alpha}, \tag{A.6}$$

where $\xi_{n,1-2\alpha}$ is the empirical $(1-2\alpha)$-quantile of all components of the residuals $\widehat{\boldsymbol{\eta}}_t = \widetilde{\Sigma}_t^{-1}(\widehat{\boldsymbol{\vartheta}}_n)\{\boldsymbol{y}_t - \widetilde{\boldsymbol{m}}_t(\widehat{\boldsymbol{\vartheta}}_n)\}$.[6] Here $\widetilde{\boldsymbol{m}}_t(\widehat{\boldsymbol{\vartheta}}_n)$ and $\widetilde{\Sigma}_t(\widehat{\boldsymbol{\vartheta}}_n)$ denote the estimated conditional mean and variance of $\boldsymbol{y}_t$ based on initial values $\widetilde{\boldsymbol{y}}_i$ for $i \leq 0$. Francq and Zakoïan (2018) derived, under appropriate assumptions, the asymptotic joint distribution of $\widehat{\boldsymbol{\vartheta}}_n$ and $\xi_{n,1-2\alpha}$.

A.2. Conditional VaR estimation without the sphericity assumption

The FHS approach (see Barone-Adesi et al. (1999), Mancini and Trojani (2011) and the references therein) does not require any symmetry assumption. It relies on estimating the conditional quantile of a linear combination of the components of the innovation, where the coefficients depend on both the model parameter and the past returns. Indeed, the conditional VaR of the portfolio return is

$$\text{VaR}_{t-1}^{(\alpha)}(r_t) = \text{VaR}_{t-1}^{(\alpha)}\left\{\mathbf{a}_{t-1}'\boldsymbol{m}_t(\boldsymbol{\vartheta}_0) + \mathbf{a}_{t-1}'\Sigma_t(\boldsymbol{\vartheta}_0)\boldsymbol{\eta}_t\right\}.$$

A natural estimator is thus

$$\widehat{\text{VaR}}_{FHS,t-1}^{(\alpha)}(r_t) = -q_\alpha\left(\left\{\mathbf{a}_{t-1}'\widetilde{\boldsymbol{m}}_t(\widehat{\boldsymbol{\vartheta}}_n) + \mathbf{a}_{t-1}'\widetilde{\Sigma}_t(\widehat{\boldsymbol{\vartheta}}_n)\widehat{\boldsymbol{\eta}}_s, \quad 1 \leq s \leq n\right\}\right), \tag{A.7}$$

where $q_\alpha(S)$ denotes the α-quantile of the elements of any finite set $S \subset \mathbb{R}$.

Appendix B. Proofs

B.1. Proof of Lemma 2.1

We have

$$P\left(\sum_{k=1}^n \Delta_k > c\right) = P(Z_n > c_n) = P(Z_n > 0) - P(Z_n \in (0, c_n])\mathbf{1}_{\{c_n>0\}} + P(Z_n \in (c_n, 0])\mathbf{1}c_n \leq 0$$

with $c_n = a_n c + b_n$. We have $P(Z_n > 0) \to p$ and, for any $\varepsilon > 0$, there exists $c_\varepsilon > 0$ such that $\lim_{n\to\infty} P(Z_n \in (0, c_n]) \leq \lim_{n\to\infty} P(Z_n \in [-c_\varepsilon, c_\varepsilon]) \leq \varepsilon$. The conclusion follows. $\square$

B.2. Proof of Lemma 3.1

We start by showing that Assumption **A2** is satisfied. Let $a(z) = \alpha_0 z^2 + \beta_0$ and let

$$\epsilon_t = \sqrt{\omega_0}\left\{1 + \sum_{i=1}^\infty a(u_{t-1})\ldots a(u_{t-i})\right\}^{1/2} u_t,$$

which is well defined under the condition in (i). The process (ϵ_t) is strictly stationary and ergodic by **A1**. The second part of **A2** follows from (iv). Next, we have

$$\sigma_t^2(\boldsymbol{\theta}_0) = \sigma_{t,T_n}^2 + \widetilde{\sigma}_{t,T_n}^2, \qquad \sigma_{t,T_n}^2 = \omega_0\left\{1 + \sum_{k=1}^{T_n}\prod_{i=1}^k a(u_{t-i})\right\}.$$

Note that, under the strict stationarity condition $E\log a(u_1) < 0$, we have

$$\widetilde{\sigma}_{t,T_n}^2 = \omega_0\sum_{k=T_n+1}^\infty \prod_{i=1}^k a(u_{t-i}) \to 0 \text{ a.s. when } T_n \to \infty. \tag{B.1}$$

We also set $\epsilon_t^2 = \epsilon_{t,T_n}^2 + \widetilde{\epsilon}_{t,T_n}^2$, where $\epsilon_{t,T_n} = u_t\sigma_{t,T_n} \in \mathcal{F}_{t:t-T_n}$. We thus have

$$\sigma_{t,T_n}^2 = \sum_{i=0}^{T_n-2}\beta_0^i\left(\omega_0 + \alpha_0\epsilon_{t-i-1,T_n-i-1}^2\right) + \beta_0^{T_n-1}\sigma_{t-T_n+1,1}^2.$$

---

[6] By assumption, the components of $\boldsymbol{\eta}_t$ have the same symmetric distribution.

Now, we show that **A3** holds true for the GARCH(1,1) model. The first and second components of $\boldsymbol{D}_t$ are bounded, and thus can be handled easily. The last component of $\boldsymbol{D}_t$ has the form

$$_\beta\sigma_t^2 := \frac{1}{2\sigma_t^2}\frac{\partial\sigma_t^2(\boldsymbol{\theta}_0)}{\partial\beta} = {}_\beta\sigma_{t,T_n}^2 + {}_\beta\widetilde{\sigma}_{t,T_n}^2, \qquad {}_\beta\sigma_{t,T_n}^2 = \frac{\sum_{i=1}^{T_n-2}i\beta_0^{i-1}\left(\omega_0 + \alpha_0\epsilon_{t-i,T_n-i}^2\right)}{\sigma_{t,T_n}^2}.$$

Note that $_\beta\sigma_{t,T_n}^2 \in \mathcal{F}_{t-1:t-T_n}$ and, using the inequality $x/(1+x) \le x^s$ for any $x \ge 0$ and any $s \in (0, 1)$, we have

$$_\beta\sigma_{t,T_n}^2 \le \frac{1}{(1-\beta_0)^2} + \sum_{i=1}^{T_n-2}\frac{i\beta_0^{i-1}\alpha_0\epsilon_{t-i,T_n-i}^2}{\omega_0 + \beta_0^i\alpha_0\epsilon_{t-i,T_n-i}^2} \le \frac{1}{(1-\beta_0)^2} + \frac{\alpha_0^s}{\beta_0\omega_0^s}\sum_{i=1}^{T_n-2}i\left\{\beta_0^s\right\}^i\epsilon_{t-i,T_n-i}^{2s}.$$

Therefore, under **A2**, for any $r \ge 1$, choosing $s > 0$ such that $2sr \le s_0$, the Hölder inequality shows that

$$\sup_n\left\|{}_\beta\sigma_{t,T_n}^2\right\|_r \le \frac{1}{(1-\beta_0)^2} + \frac{\alpha_0^s}{\beta_0\omega_0^s}\left\|\epsilon_1^{2s}\right\|_r\sum_{i=1}^{\infty}i\left\{\beta_0^s\right\}^i < \infty, \quad \left\|{}_\beta\sigma_t^2\right\|_r < \infty,$$

where $\|\cdot\|_r$ denotes the L_r norm. Now note that

$$
\begin{aligned}
_\beta\widetilde{\sigma}_{t,T_n}^2 &= \frac{1}{2\sigma_{t,T_n}^2}\frac{\partial\sigma_t^2(\boldsymbol{\theta}_0)}{\partial\beta} + \frac{1}{2}\frac{\partial\sigma_t^2(\boldsymbol{\theta}_0)}{\partial\beta}\left(\frac{1}{\sigma_t^2(\boldsymbol{\theta}_0)} - \frac{1}{\sigma_{t,T_n}^2}\right) - {}_\beta\sigma_{t,T_n}^2 \\
&= \frac{1}{2\sigma_{t,T_n}^2}\sum_{i=1}^{T_n-2}i\beta_0^{i-1}\left(\omega_0 + \alpha_0\epsilon_{t-i,T_n-i}^2\right) + \frac{1}{2\sigma_{t,T_n}^2}\sum_{i=1}^{T_n-2}i\beta_0^{i-1}\alpha_0(\epsilon_{t-i}^2 - \epsilon_{t-i,T_n-i}^2) - {}_\beta\sigma_{t,T_n}^2 \\
&\quad + \frac{1}{2\sigma_{t,T_n}^2}\sum_{i=T_n-1}^{\infty}i\beta_0^{i-1}\left(\omega_0 + \alpha_0\epsilon_{t-i}^2\right) + {}_\beta\sigma_t^2\left(1 - \frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_{t,T_n}^2}\right) \\
&= -\frac{\widetilde{\sigma}_{t,T_n}^2}{\sigma_{t,T_n}^2}{}_\beta\sigma_t^2 + \frac{\alpha_0\sum_{i=1}^{T_n-2}i\beta_0^{i-1}\widetilde{\epsilon}_{t-i,T_n-i}^2}{2\sigma_{t,T_n}^2} + \frac{\sum_{i=T_n-1}^{\infty}i\beta_0^{i-1}\left(\omega_0 + \alpha_0\epsilon_{t-i}^2\right)}{2\sigma_{t,T_n}^2}. \quad (B.2)
\end{aligned}
$$

In view of (B.1), the first term of the right-hand side of the equality tends to zero in probability. Using Lemma 2.2 in Francq and Zakoïan (2019), the strict stationarity condition $E\log a(u_1) < 0$ entails the existence of $s \in (0, 1)$ such that $\rho := Ea^s(u_1) < 1$. We then have $E\widetilde{\epsilon}_{t,T_n}^{2s} \le K\rho^{T_n}$, which entails

$$E\left|\sum_{i=1}^{T_n-2}i\beta_0^{i-1}\widetilde{\epsilon}_{t-i,T_n-i}^2\right|^s \le K\sum_{i=1}^{T_n-2}i^s\beta_0^{s(i-1)}\rho^{T_n-i} \to 0 \quad \text{as} \quad n \to \infty.$$

Noting that $E|X_n|^s \to 0$ for some $s > 0$ entails that $X_n \to 0$ in probability, we conclude that the second term of the right-hand side of the equality (B.2) tends to zero in probability. Let $s \in (0, 1)$ such that $E|\epsilon_t|^{2s} < \infty$. We have

$$E\left|\sum_{i=T_n-1}^{\infty}i\beta_0^{i-1}\left(\omega_0 + \alpha_0\epsilon_{t-i}^2\right)\right|^s \le \left(\omega_0^s + \alpha_0^s E|\epsilon_1|^{2s}\right)\sum_{i=T_n-1}^{\infty}i^s(\beta_0^s)^{i-1} \to 0$$

as $n \to \infty$. If follows that the third term of the right-hand side of the equality (B.2) tends to zero in probability. We thus have shown that $_\beta\widetilde{\sigma}_{t,T_n}^2 = o_p(1)$. It follows that $_\beta\sigma_{t,T_n}^2$ can be chosen as being the last component of $\boldsymbol{D}_{t,T_n}$. As already argued, the two other components are handled more easily. This shows that **A3** is satisfied for any sequence T_n tending to infinity. Turning to **A4**, assume that $\sigma_{t+1}(\boldsymbol{\theta}_1) = \sigma_{t+1}(\boldsymbol{\theta}_2)$ a.s. for $\boldsymbol{\theta}_i = (\omega_i, \alpha_i, \beta_i)$. We thus have $\omega_1 + \alpha_1 u_t^2\sigma_t^2(\boldsymbol{\theta}_0) + \beta_1\sigma_t^2(\boldsymbol{\theta}_1) = \omega_2 + \alpha_2 u_t^2\sigma_t^2(\boldsymbol{\theta}_0) + \beta_2\sigma_t^2(\boldsymbol{\theta}_2)$ a.s. Thus, if $\alpha_1 \ne \alpha_2$, u_t^2 can be written as a variable belonging to $\mathcal{F}_{t-1}$. This variable is in fact degenerate and equal to 1 because $E(u_t^2 \mid \mathcal{F}_{t-1}) = 1$, in contradiction with (ii). It follows that $\alpha_1 = \alpha_2$, and thus $\omega_1 + \beta_1\sigma_t^2(\boldsymbol{\theta}_1) = \omega_2 + \beta_2\sigma_t^2(\boldsymbol{\theta}_2)$. Proceeding in the same way, by expressing u_{t-1}^2 as a $\mathcal{F}_{t-2}$-measurable variable, allows to conclude that $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$, that is that **A4** is satisfied. The other assumptions are easily verified. In particular, **A7** can be handled using (7.51) and (7.54) in Francq and Zakoïan (2019).[7] □

B.3. Proof of Theorem 3.1

We start by showing the following lemma, which only requires a small part of the assumptions of Theorem 3.1.

Lemma B.1. *Assume that $E(u_t^2 \mid \mathcal{F}_{t-1}) = 1$, and that Assumptions **A2**, **A4** and **A5** hold. Then $\hat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0$ a.s. as $n \to \infty$.*

---

[7] For the latter equation, examination of the proof shows that the iidness of the innovation is not used.

Proof. Let

$$Q_n(\theta) = \frac{1}{n} \sum_{t=1}^n \ell_t, \qquad \ell_t = \ell_t(\boldsymbol{\theta}) = \frac{\epsilon_t^2}{\sigma_t^2(\boldsymbol{\theta})} + \log \sigma_t^2(\boldsymbol{\theta}).$$

The strong consistency of $\hat{\boldsymbol{\theta}}_n$ is a consequence of the following results:

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

(ii) $\mathbb{E}|\ell_t(\theta_0)| < \infty$, and if $\theta \neq \theta_0$, $\mathbb{E}\ell_t(\theta) > \mathbb{E}\ell_t(\theta_0)$;

(iii) any $\theta \neq \theta_0$ has a neighborhood $V(\theta)$ such that $\lim_{n \to \infty} \sup \inf_{\theta^* \in V(\theta)} Q_n(\theta^*) > \mathbb{E}\ell_t(\theta_0)$, a.s.

To prove (i) we note that

$$\sup_{\theta \in \Theta} |Q_n(\theta) - \tilde{Q}_n(\theta)| \leq n^{-1} \sum_{t=1}^n \sup_{\theta \in \Theta} \left\{ \left| \frac{\sigma_t^2(\boldsymbol{\theta}) - \tilde{\sigma}_t^2(\boldsymbol{\theta})}{\sigma_t^2(\boldsymbol{\theta})\tilde{\sigma}_t^2(\boldsymbol{\theta})} \right| \epsilon_t^2 + \left| \log \left(\frac{\sigma_t^2(\boldsymbol{\theta})}{\tilde{\sigma}_t^2(\boldsymbol{\theta})} \right) \right| \right\}$$

$$\leq \frac{2C}{\underline{\omega}^3} n^{-1} \sum_{t=1}^n \rho^t \epsilon_t^2 + \frac{2C}{\underline{\omega}} n^{-1} \sum_{t=1}^n \rho^t,$$

where the latter inequality is deduced from **A4–A5**. The right-hand side goes to 0 a.s., by the Cesàro lemma and the existence of a small-order moment for ϵ_t (Assumption **A2**). The proof of (ii) uses the identifiability assumption in **A4**, and the fact that $E(u_t^2 \mid \mathcal{F}_{t-1}) = 1$, enabling us to write

$$\mathbb{E}\ell_t(\boldsymbol{\theta}) = \mathbb{E} \frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2(\boldsymbol{\theta})} + \mathbb{E} \log \sigma_t^2(\boldsymbol{\theta}),$$

where the existence of the latter expectation, in $\mathbb{R} \cup \{+\infty\}$, follows from the first part of **A4**. At the true value we have $\mathbb{E}|\ell_t(\boldsymbol{\theta}_0)| < \infty$ using again $E|\epsilon_1|^{s_0} < \infty$. The proof of (iii) uses the ergodic theorem and a standard compactness argument (see for instance Francq and Zakoïan (2019, Proof of Theorem 7.1) for details). □

Turning to the asymptotic distribution, we establish the following lemma.

Lemma B.2. *Under* **A1** *and* **A3–A4**, *we have*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \begin{pmatrix} (u_t^2 - 1)\boldsymbol{D}_t \\ \mathbf{1}_{\{u_t < \xi_\alpha\}} - \alpha \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \boldsymbol{S}_\alpha), \quad \text{where} \quad \boldsymbol{S}_\alpha = \begin{pmatrix} \boldsymbol{S}^{11} & \boldsymbol{S}_\alpha^{12} \\ \boldsymbol{S}_\alpha^{21} & \boldsymbol{S}_\alpha^{22} \end{pmatrix}$$

is a positive definite matrix.

Proof. Let $c_0 \in \mathbb{R}$, $\boldsymbol{c}_1 \in \mathbb{R}^m$, $\boldsymbol{c} = (\boldsymbol{c}_1', c_0)'$ and

$$x_t = (u_t^2 - 1)\boldsymbol{c}_1'\boldsymbol{D}_t + c_0(\mathbf{1}_{\{u_t < \xi_\alpha\}} - \alpha), \quad x_{t,n} = (u_t^2 - 1)\boldsymbol{c}_1'\boldsymbol{D}_{t,T_n} + c_0(\mathbf{1}_{\{u_t < \xi_\alpha\}} - \alpha).$$

We will apply a central limit theorem for the mixing triangular array $(x_{t,n})$. For convenience, we reproduce it below.

Theorem B.1 (*Francq and Zakoïan, 2005*)**.** *Let* $(x_{t,n})$ *be a triangular array of centered real-valued random variables. For each* $n \geq 2$ *and* $h = 1, \ldots, n - 1$, *let the strong mixing coefficients of* $x_{1,n}, \ldots, x_{n,n}$ *be defined by*

$$\alpha_n(h) = \sup_{1 \leq t \leq n-h} \sup_{A \in \mathcal{A}_{t,n}, B \in \mathcal{B}_{t+h,n}} |P(A \cap B) - P(A)P(B)|,$$

where $\mathcal{A}_{t,n} = \sigma(x_{u,n}, 1 \leq u \leq t)$ *and* $\mathcal{B}_{t,n} = \sigma(x_{u,n}, t \leq u \leq n)$ *and, by convention,* $\alpha_n(h) = 1/4$ *for* $h \leq 0$, $\alpha_n(h) = 0$ *for* $h \geq n$. *Let* $S_n = \sum_{t=1}^n x_{t,n}$. *Under the following assumptions*

(1) $\sup_{n \geq 1} \sup_{1 \leq t \leq n} \|x_{t,n}\|_{2+\nu^*} < \infty$ *for some* $\nu^* \in (0, \infty]$,

(2) $\lim_{n \to \infty} n^{-1} Var S_n = \sigma^2 > 0$,

(3) *there exists a sequence of integers* (T_n) *such that* $T_n = O(n^\kappa)$ *for some* $\kappa \in [0, \nu^*/\{4(1+\nu^*)\})$ *and a sequence* $\{\alpha(h)\}_{h \geq 1}$ *such that*

$$\alpha_n(h) \leq \alpha(h - T_n), \quad \text{for all} \ \ h > T_n, \tag{B.3}$$

$$\sum_{h=1}^\infty h^{r^*} \alpha^{\nu^*/(2+\nu^*)}(h) < \infty \ \text{for some} \ r^* > \frac{2\kappa(1+\nu^*)}{\nu^* - 2\kappa(1+\nu^*)}, \tag{B.4}$$

we have $n^{-1/2} S_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$.

Let $\|\cdot\|$ denote any norm on $\mathbb{R}^d$ and, for any random vector $\boldsymbol{X} \in \mathbb{R}^d$, let $\|\boldsymbol{X}\|_p = E(\|\boldsymbol{X}\|^p)^{1/p}$ for $p \geq 1$. Let $\nu^* = (\nu - \epsilon)/2$ where ν and ϵ are defined in **A1**. By **A1** and **A3**, we have

$$\sup_n \|u_t^2 \boldsymbol{D}_{t,T_n}\|_{2+\nu^*}^{2+\nu^*} \leq \|u_t^{4+2\nu^*}\|_p \left\| \sup_n \|\boldsymbol{D}_{t,T_n}\|^{2+\nu^*} \right\|_q < \infty$$

provided that $p = q/(q-1)$ satisfies $1 < p \leq (4+\nu)/(4+\nu-\epsilon)$. Therefore (1) holds. Now, letting $\widetilde{\boldsymbol{D}}_{t,T_n} = \boldsymbol{D}_t - \boldsymbol{D}_{t,T_n}$, we note that $\|\widetilde{\boldsymbol{D}}_{t,T_n}\|^r \to 0$ in probability by **A3**, and the sequence $\|\widetilde{\boldsymbol{D}}_{t,T_n}\|^r$ is uniformly integrable because

$$\sup_n \|\widetilde{\boldsymbol{D}}_{t,T_n}\|_r \leq \sup_n \|\boldsymbol{D}_{t,T_n}\|_r + \|\boldsymbol{D}_t\|_r < \infty.$$

From Theorem 3.5 in Billingsley (1999) it follows that $E\|\widetilde{\boldsymbol{D}}_{t,T_n}\|^r \to 0$ as $n \to \infty$, for any $r \geq 1$. We thus have

$$\mathrm{Var} \frac{1}{\sqrt{n}} \sum_{t=1}^n (x_t - x_{t,n}) = E\left[(u_t^2 - 1)^2 \{\boldsymbol{c}_1' \widetilde{\boldsymbol{D}}_{t,T_n}\}^2\right]$$

$$\leq E\left[(u_t^2 - 1)^{2(1+\frac{\nu}{4})}\right]^{\frac{4}{4+\nu}} E\left[(\boldsymbol{c}_1' \widetilde{\boldsymbol{D}}_{t,T_n})^{2(1+\frac{4}{\nu})}\right]^{\frac{\nu}{4+\nu}} \to 0 \tag{B.5}$$

as $n \to \infty$. Therefore, for $\boldsymbol{c} \neq 0$,

$$\lim_{n\to\infty} n^{-1} \mathrm{Var} \sum_{t=1}^n x_{t,n} = \lim_{n\to\infty} n^{-1} \mathrm{Var} \sum_{t=1}^n x_t \to \sigma^2 = \boldsymbol{c}' \boldsymbol{S}_\alpha \boldsymbol{c} > 0$$

provided that $\boldsymbol{S}_\alpha$ is positive definite. To show the latter, let $x_0 \in \mathbb{R}$, $\boldsymbol{x}_1 \in \mathbb{R}^m$, $\boldsymbol{x} = (\boldsymbol{x}_1', x_0)'$ such that $\boldsymbol{x}' \boldsymbol{S}_\alpha \boldsymbol{x} = 0$. We then have $(u_t^2 - 1)\boldsymbol{x}_1' \boldsymbol{D}_t + x_0(\mathbf{1}_{\{u_t < \xi_\alpha\}} - \alpha) = 0$ a.s. Thus, conditional on $\mathcal{F}_{t-1}$, u_t takes at most three different values when $\boldsymbol{x} \neq 0$, in contradiction with the existence of the density f_{t-1}. Thus (2) is satisfied. By **A3**, $x_{t,n} \in \mathcal{F}_{t:t-T_n}$. Thus, $\boldsymbol{S}_\alpha$ is positive definite and (B.3) is satisfied, where the sequence $\alpha(h)$ is defined in **A1**. The conditions of (3) are also satisfied by **A1**. It follows that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n x_{t,n} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2).$$

The conclusion then follows from (B.5) and the Cramér–Wold device. $\square$

Now we turn to the proof of Theorem 3.1. Let $u_t(\boldsymbol{\theta}) = \epsilon_t/\sigma_t(\boldsymbol{\theta})$ and $\hat{u}_t = \epsilon_t/\tilde{\sigma}_t(\hat{\boldsymbol{\theta}}_n)$. Note that, by **A4** and **A5**, for n large enough

$$\left|\hat{u}_t - u_t(\hat{\boldsymbol{\theta}}_n)\right| = \left|\epsilon_t \frac{\sigma_t(\hat{\boldsymbol{\theta}}_n) - \tilde{\sigma}_t(\hat{\boldsymbol{\theta}}_n)}{\tilde{\sigma}_t(\hat{\boldsymbol{\theta}}_n)\sigma_t(\hat{\boldsymbol{\theta}}_n)}\right| \leq \frac{C}{\omega} \rho^t u_t \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left|\frac{\sigma_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta})}\right|. \tag{B.6}$$

A Taylor expansion around $\boldsymbol{\theta}_0$ and **A4**, **A5** yields

$$\hat{u}_t = u_t - u_t \boldsymbol{D}_t'(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + r_{n,t}$$

with

$$r_{n,t} = \frac{1}{2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' \frac{\partial^2 u_t(\boldsymbol{\theta}_{n,t})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + \hat{u}_t - u_t(\hat{\boldsymbol{\theta}}_n),$$

where $\boldsymbol{\theta}_{n,t}$ is between $\hat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}_0$. Following the approach of Knight (1998) and Koenker and Xiao (2006) (see also Francq and Zakoïan (2015)), we then obtain

$$\sqrt{n}(\xi_{n,\alpha} - \xi_\alpha) = \arg\min_{z \in \mathbb{R}} \mathcal{Q}_n(z), \quad \text{where} \quad \mathcal{Q}_n(z) = zX_n + I_n(z) + J_n(z) + K_n(z),$$

with

$$X_n = \frac{1}{\sqrt{n}} \sum_{t=1}^n (\mathbf{1}_{\{u_t < \xi_\alpha\}} - \alpha),$$

$$I_n(z) = \sum_{t=1}^n \int_0^{z/\sqrt{n}} (\mathbf{1}_{\{u_t \leq \xi_\alpha + s\}} - \mathbf{1}_{\{u_t < \xi_\alpha\}}) ds,$$

$$J_n(z) = \sum_{t=1}^n \int_0^{R_{t,n}/\sqrt{n}} \left(\mathbf{1}_{\{u_t - \xi_\alpha - z/\sqrt{n} \leq u\}} - \mathbf{1}_{\{u_t - \xi_\alpha - z/\sqrt{n} < 0\}}\right) du,$$

$$K_n(z) = \sum_{t=1}^{n} \frac{R_{t,n}}{\sqrt{n}} \mathbf{1}^*_{\{u_t - \xi_\alpha \in (0, z/\sqrt{n})\}},$$

$\mathbf{1}^*_{\{X \in (a,b)\}} = \mathbf{1}_{\{X<b\}} - \mathbf{1}_{\{X<a\}}$ for any real numbers a, b and any real random variable X, and $R_{t,n} = u_t D'_t \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) - \sqrt{n} r_{n,t}$. We will show that

$$\mathcal{Q}_n(z) = \frac{z^2}{2} E f_0(\xi_\alpha) + z\{X_n + \xi_\alpha \boldsymbol{\Psi}'_\alpha \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\} + O_P(1). \tag{B.7}$$

Noting that

$$\begin{aligned}
K_n(z) &= \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} u_t \mathbf{1}^*_{\{u_t - \xi_\alpha \in (0, z/\sqrt{n})\}} D'_t \right) \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \\
&\quad - \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' \frac{1}{2n} \sum_{t=1}^{n} \frac{\partial^2 u_t(\boldsymbol{\theta}_{n,t})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \mathbf{1}^*_{\{u_t - \xi_\alpha \in (0, z/\sqrt{n})\}} \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \\
&\quad - \sum_{t=1}^{n} \left\{ \hat{u}_t - u_t(\hat{\boldsymbol{\theta}}_n) \right\} \mathbf{1}^*_{\{u_t - \xi_\alpha \in (0, z/\sqrt{n})\}} \\
&=: K_{n1}(z) + K_{n2}(z) + K_{n3}(z),
\end{aligned}$$

the proof of (B.7) will be divided in the following steps.

  (i) $K_{ni}(z) \to 0$ in probability as $n \to \infty$, for $i = 2, 3$.
  (ii) $K_{n1}(z) = z\xi_\alpha \boldsymbol{\Psi}'_\alpha \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o_P(1)$ in probability as $n \to \infty$.
  (iii) $J_n(z)$ does not depend on z asymptotically.
  (iv) $I_n(z) \to \frac{z^2}{2} E f_0(\xi_\alpha)$ in probability as $n \to \infty$.

To prove (i) for $i = 2$, note that

$$\frac{\partial^2 u_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = -u_t \frac{\sigma_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta})} \frac{1}{\sigma_t(\boldsymbol{\theta})} \frac{\partial^2 \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} + 2u_t \frac{\sigma_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta})} \frac{1}{\sigma_t^2(\boldsymbol{\theta})} \frac{\partial \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}. \tag{B.8}$$

Let $0 < \delta < \frac{2+\nu}{6+\nu}$. By the Cauchy–Schwartz inequality we have, for $p, q, r > 0$ such that $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1$,

$$\begin{aligned}
&E \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\| u_t \frac{\sigma_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta})} \frac{1}{\sigma_t(\boldsymbol{\theta})} \frac{\partial^2 \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right\|^{1+\delta} \\
&\le \{E|u_t|^{p(1+\delta)}\}^{1/p} \left(E \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left| \frac{\sigma_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta})} \right|^{q(1+\delta)} \right)^{1/q} \left(E \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\| \frac{1}{\sigma_t(\boldsymbol{\theta})} \frac{\partial^2 \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right\|^{r(1+\delta)} \right)^{1/r}.
\end{aligned} \tag{B.9}$$

In view of **A1** and **A7**, the first and third expectation in the right-hand side are finite if we choose $p = \frac{4+\nu}{1+\delta}$ and $r = \frac{2}{1+\delta}$. We then have $q(1+\delta) = \frac{2(4+\nu)(1+\delta)}{2+\nu - \delta(6+\nu)}$. The latter term increases, when δ varies in $(0, \frac{2+\nu}{6+\nu})$, from $\frac{2(4+\nu)}{2+\nu}$ to infinity. It is thus possible to choose δ small enough such that $q(1+\delta) < \frac{2(4+\nu)(1+\tau)}{2+\nu}$. For such δ, by **A7**, the second expectation and finally the right-hand side of (B.9) are finite. The second summand in the right-hand side of (B.8) can be handled similarly. Thus we have

$$E \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\| \frac{\partial^2 u_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right\|^{1+\delta} < \infty. \tag{B.10}$$

By the Hölder inequality, for $\boldsymbol{\theta}_{n,t} \in V(\boldsymbol{\theta}_0)$,

$$\begin{aligned}
&\left\| \frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2 u_t(\boldsymbol{\theta}_{n,t})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \mathbf{1}^*_{\{u_t - \xi_\alpha \in (0, z/\sqrt{n})\}} \right\| \\
&\le \left\{ \frac{1}{n} \sum_{t=1}^{n} \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\| \frac{\partial^2 u_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right\|^{1+\delta} \right\}^{1/(1+\delta)} \left\{ \frac{1}{n} \sum_{t=1}^{n} \mathbf{1}^*_{\{u_t - \xi_\alpha \in (0, z/\sqrt{n})\}} \right\}^{\delta/(1+\delta)}.
\end{aligned}$$

By (B.10) and the ergodic theorem, the limit of the first term of the latter product is almost surely finite. Letting $v_{t,n} = \mathbf{1}^*_{\{u_t - \xi_\alpha \in (0, z/\sqrt{n})\}}$ and $\bar{v}_{t,n} = v_{t,n} - E(v_{t,n} \mid \mathcal{F}_{t-1})$, we have

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \mathbf{1}^*_{\{u_t - \xi_\alpha \in (0, z/\sqrt{n})\}} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \bar{v}_{t,n} + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} E(v_{t,n} \mid \mathcal{F}_{t-1}).$$

First note that

$$E(v_{t,n} \mid \mathcal{F}_{t-1}) = \int_{\xi_\alpha}^{\xi_\alpha + z/\sqrt{n}} f_{t-1}(x)dx = \frac{z}{\sqrt{n}} f_{t-1}(\xi_\alpha) + \frac{k_{t,n}}{\sqrt{n}}$$

where

$$\left| k_{t,n} \right| = \sqrt{n} \left| \int_{\xi_\alpha}^{\xi_\alpha + z/\sqrt{n}} \{ f_{t-1}(x) - f_{t-1}(\xi_\alpha) \} \, dx \right| \leq K_{t-1} \frac{z^2}{\sqrt{n}},$$

by **A1**. Now, note that we have $E\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \bar{v}_{t,n} = 0$ and

$$\mathrm{Var}\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \bar{v}_{t,n} \right) = E\bar{v}_{1,n}^2 = \int_{\xi_\alpha}^{\xi_\alpha + z/\sqrt{n}} E\left\{ 1 - \frac{z}{\sqrt{n}} f_0(\xi_\alpha) - \frac{k_{t,n}}{\sqrt{n}} \right\}^2 f_0(x)dx \to 0,$$

using again **A1**. Moreover, almost surely

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} E(v_{t,n} \mid \mathcal{F}_{t-1}) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \int_{\xi_\alpha}^{\xi_\alpha + z/\sqrt{n}} f_{t-1}(x)dx \to zEf_{t-1}(\xi_\alpha).$$

We thus have shown that $\sum_{t=1}^{n} \mathbf{1}^*_{\{u_t - \xi_\alpha \in (0, z/\sqrt{n})\}} = O_P(\sqrt{n})$. Thus, (i) for $i = 2$ is established. By the same arguments and (B.6), it can be shown that (i) for $i = 3$ holds.

Turning to (ii), we have

$$E\left(u_t \mathbf{1}^*_{\{u_t - \xi_\alpha \in (0, z/\sqrt{n})\}} D'_t \mid \mathcal{F}_{t-1} \right) = \int_{\xi_\alpha}^{\xi_\alpha + z/\sqrt{n}} x f_{t-1}(x)dx D'_t$$

$$= \int_{\xi_\alpha}^{\xi_\alpha + z/\sqrt{n}} x f_{t-1}(\xi_\alpha)dx D'_t + \int_{\xi_\alpha}^{\xi_\alpha + z/\sqrt{n}} x \{ f_{t-1}(x) - f_{t-1}(\xi_\alpha) \} dx D'_t$$

$$= \xi_\alpha f_{t-1}(\xi_\alpha) \frac{z}{\sqrt{n}} D'_t + \frac{k^*_{t,n}}{\sqrt{n}} D'_t$$

with

$$\left| k^*_{t,n} \right| = \sqrt{n} \left| f_{t-1}(\xi_\alpha) \frac{z^2}{2n} + \int_{\xi_\alpha}^{\xi_\alpha + z/\sqrt{n}} x \{ f_{t-1}(x) - f_{t-1}(\xi_\alpha) \} dx \right|$$

$$\leq \frac{z^2}{\sqrt{n}} \left\{ 2K_{t-1}\xi_\alpha + \frac{f_{t-1}(\xi_\alpha)}{2} + \frac{|z|}{\sqrt{n}} K_{t-1} \right\}.$$

Denoting by d_t a generic element of D_t, we also have

$$EE\left[\left\{ u_t \mathbf{1}^*_{\{u_t - \xi_\alpha \in (0, z/\sqrt{n})\}} d_t - E\left(u_t \mathbf{1}^*_{\{u_t - \xi_\alpha \in (0, z/\sqrt{n})\}} d_t \mid \mathcal{F}_{t-1} \right) \right\}^2 \mid \mathcal{F}_{t-1} \right]$$

$$= \int_{\xi_\alpha}^{\xi_\alpha + z/\sqrt{n}} E\left(x - \xi_\alpha f_{t-1}(\xi_\alpha) \frac{z}{\sqrt{n}} + \frac{k^*_{t,n}}{\sqrt{n}} \right)^2 d_t^2 f_{t-1}(x)dx = o(1), \tag{B.11}$$

as $n \to \infty$. To show that the expectation inside the latter integral is finite, we used in particular the fact that

$$E \sup_{x \in [\xi_\alpha, \xi_\alpha + z/\sqrt{n}]} d_t^2 f_{t-1}^2(\xi_\alpha) f_{t-1}(x) \leq \sqrt{E d_t^8 E \sup_{\xi \in V(\xi_\alpha)} f_{t-1}^4(\xi)} < \infty$$

for sufficiently large n under **A1** and **A7**. Hence, (ii) is established.

To prove (iii), write $J_n(z) = \sum_{t=1}^{n} J_{n,t}$. Write $r_{n,t} = r_{n,t}(\hat{\theta}_n)$, $R_{n,t} = R_{n,t}(\hat{\theta}_n)$, $J_{n,t} = J_{n,t}(\hat{\theta}_n)$ and $J_n(z) = J_n(z, \hat{\theta}_n)$. Let (θ_n) be a deterministic sequence such that $\sqrt{n}(\theta_n - \theta_0) = O(1)$. By the change of variable $u = u_t v$, we have

$$E\left(J_{n,t}(\theta_n) \mid \mathcal{F}_{t-1} \right) = \int_0^{D'_t(\theta_n - \theta_0) + o_P(n^{-1/2})} E\left(u_t \mathbf{1}^*_{\{u_t \in (\xi_\alpha + z/\sqrt{n}, (\xi_\alpha + z/\sqrt{n})(1-v)^{-1})\}} \mid \mathcal{F}_{t-1} \right) dv$$

$$= \frac{\xi_\alpha^2}{2} f_{t-1}(\xi_\alpha)(\theta_n - \theta_0)' D_t D'_t (\theta_n - \theta_0) + o_P(n^{-1}).$$

By the arguments used to show (B.11), we can show that

$$E\left[J_{n,t}(\theta_n) - E\left\{ J_{n,t}(\theta_n) \mid \mathcal{F}_{t-1} \right\} \right]^2 = o(n^{-1}).$$

Table 4
Design of Monte Carlo experiments.

	$\boldsymbol{\omega}_0'$	$(\text{vec}\boldsymbol{A}_0)'$	$\text{diag}\boldsymbol{B}_0$	$\boldsymbol{S}_0(1,2)$	α	β	P_η
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_η follows an AEPD.

We thus have

$$J_n(z, \boldsymbol{\theta}_n) = \frac{\xi_\alpha^2}{2} \sqrt{n}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0)' E\{f_0(\xi_\alpha)D_1 D_1'\} \sqrt{n}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0) + o(1), \quad a.s.$$

It follows that $J_n(z, \boldsymbol{\theta}_n)$ does not depend of z asymptotically. Since this is true for any sequence such that $\sqrt{n}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0) = O(1)$, this is also true almost surely for $J_n(z)$ and (iii) is established.

By the previously used arguments, it can be shown that (iv) holds which completes the proof of (B.7). By Lemma 2.2 in Davis et al. (1992) and convexity arguments, we can conclude that

$$\sqrt{n}(\xi_\alpha - \xi_{n,\alpha}) = \frac{\xi_\alpha}{Ef_0(\xi_\alpha)} \boldsymbol{\Psi}_\alpha' \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + \frac{1}{Ef_0(\xi_\alpha)} \frac{1}{\sqrt{n}} \sum_{t=1}^n (\mathbf{1}_{\{u_t < \xi_\alpha\}} - \alpha) + o_P(1).$$

We have the following Taylor expansion

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = \frac{-\boldsymbol{J}^{-1}}{2\sqrt{n}} \sum_{t=1}^n (1 - u_t^2)\boldsymbol{D}_t + o_P(1).$$

By the CLT for martingale differences we get the announced result, noting that

$$\text{Cov}_{as} \left(\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0), \frac{1}{\sqrt{n}} \sum_{t=1}^n (\mathbf{1}_{\{u_t < \xi_\alpha\}} - \alpha) \right) = \frac{1}{2} \boldsymbol{J}^{-1} \boldsymbol{S}_\alpha^{12},$$

which entails

$$\text{Var}_{as}\{\sqrt{n}(\xi_{n,\alpha} - \xi_\alpha)\} = \zeta_\alpha, \qquad \text{Cov}_{as} \left(\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0), \sqrt{n}(\xi_\alpha - \xi_{n,\alpha}) \right) = \boldsymbol{\Lambda}_\alpha. \quad \square$$

Appendix C. Designs for the cDCC-GARCH model

The cDCC-GARCH(1,1) model is defined by $\boldsymbol{\Sigma}_t = \boldsymbol{D}_t \boldsymbol{R}_t^{1/2}$ where the diagonal matrix $\boldsymbol{D}_t = \text{diag}(\sigma_{1t}, \dots, \sigma_{mt})$ is assumed to satisfy the GARCH(1,1) equation

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

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

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

where $\alpha_0, \beta_0 \geq 0, \alpha_0 + \beta_0 < 1, \boldsymbol{S}_0$ is a correlation matrix, $\boldsymbol{Q}_t^*$ is the diagonal matrix with the same diagonal elements as $\boldsymbol{Q}_t$, and $\boldsymbol{\eta}_t^* = \boldsymbol{D}_t^{-1} \boldsymbol{y}_t$. The unknown parameter $\boldsymbol{\theta}_0$ includes the volatility parameters $\boldsymbol{\omega}_0, \boldsymbol{A}_0$ and $\text{diag}(\boldsymbol{B}_0)$, and the conditional correlation parameters α_0, β_0 and the sub-diagonal elements of $\boldsymbol{S}_0$.

The parameters used in the Monte-Carlo experiments of Section 4.1.2 are displayed in Table 4. 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.

References

Aielli, G.P., 2013. Dynamic conditional correlation: on properties and estimation. J. Bus. Econom. Statist. 31, 282–299.

Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–858.

Barone-Adesi, G., Giannopoulos, K., Vosper, L., 1999. VaR without correlations for nonlinear portfolios. J. Futures Mark. 19, 583–602.

Bauwens, L., Hafner, C.M., Laurent, S., 2012. Handbook of Volatility Models and their Applications. Wiley.

Bauwens, L., Laurent, S., Rombouts, J.V.K., 2006. Multivariate GARCH models: a survey. J. Appl. Econometrics 21, 79–109.

Beutner, B., Heinemann, A., Smeekes, S., 2018. A Residual Bootstrap for Conditional Value-At-Risk. Discussion paper, arXiv:1808.09125.

Beutner, B., Heinemann, A., Smeekes, S., 2019. A Justification of Conditional Confidence Intervals. Discussion paper, arXiv:1710.00643v2.

Billingsley, P., 1999. Convergence of Probability Measures, second ed. Wiley.

Boudt, K., Laurent, S., Quaedvlieg, R., Sauri, O., 2017. Positive semidefinite integrated covariance estimation, factorizations and asynchronicity. J. Econometrics 196, 347–367.

Boyd, S., Vandenberghe, L., 2018. Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares. Cambridge University Press.

Christoffersen, P.F., 2003. Elements of Financial Risk Management. Academic Press, London.

Christoffersen, P.F., Gonçalves, S., 2005. Estimation risk in financial risk management. J. Risk 7, 1–28.

Davis, R.A., Knight, K., Liu, J., 1992. M-estimation for autoregressions with infinite variance. Stochastic Process. Appl. 40, 145–180.

Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. J. Bus. Econom. Statist. 13, 253–263.

Engle, R.F., Ledoit, O., Wolf, M., 2017. Large dynamic covariance matrices. J. Bus. Econom. Statist. http://dx.doi.org/10.1080/07350015.2017.1345683.

Escanciano, J.C., Olmo, J., 2010. Backtesting parametric VaR with estimation risk. J. Bus. Econom. Statist. 28, 36–51.

Escanciano, J.C., Olmo, J., 2011. Robust backtesting tests for value-at-risk models. J. Financ. Econ. 9, 132–161.

Farkas, W., Fringuellotti, F., Tunaru, R., 2016. Regulatory Capital Requirements: Saving too Much for Rainy Days? Unpublished Document. University of Zürich.

Francq, C., Zakoïan, J.M., 2005. A central limit theorem for mixing triangular arrays of variables whose dependence is allowed to grow with the sample size. Econom. Theory 21, 1165–1171.

Francq, C., Zakoïan, J.M., 2013. Optimal predictions of powers of conditionally heteroskedastic processes. J. R. Stat. Soc. B 75, 345–367.

Francq, C., Zakoïan, J.M., 2015. Risk-parameter estimation in volatility models. J. Econometrics 184, 158–173.

Francq, C., Zakoïan, J.M., 2016. Estimating multivariate GARCH models equation by equation. J. R. Stat. Soc. Ser. B Stat. Methodol. 78, 613–635.

Francq, C., Zakoïan, J.M., 2018. Estimation risk for the VaR of portfolios driven by semi-parametric multivariate models. J. Econometrics 205, 381–401.

Francq, C., Zakoïan, J.M., 2019. GARCH Models: Structure, Statistical Inference and Financial Applications, second ed. John Wiley, Chichester.

Giacomini, R., Komunjer, I., 2005. Evaluation and combination of conditional quantile forecasts. J. Bus. Econom. Statist. 23, 416–431.

Gneiting, T., 2011. Making and evaluating point forecasts. J. Amer. Statist. Assoc. 106, 746–762.

Gong, Y., Li, Z., Peng, L., 2010. Empirical likelihood intervals for conditional value-at-risk in ARCH/GARCH models. J. Time Series Anal. 31, 65–75.

Gouriéroux, C., Zakoïan, J.M., 2013. Estimation adjusted VaR. Econom. Theory 29, 735–770.

Hansen, B.E., 2004. Nonparametric Conditional Density Estimation. Discussion paper, University of Wisconsin.

Holton, G.A., 2014. Value-At-Risk: Theory and Practice, second ed. Academic press.

Hurlin, C., Laurent, S., Quaedvlieg, R., Smeekes, S., 2017. Risk measure inference. J. Bus. Econom. Statist. 35, 499–512.

Knight, K., 1998. Limiting distributions for L_1 regression estimators under general conditions. Ann. Statist. 26, 755–770.

Koenker, R., Xiao, Z., 2006. Quantile autoregression. J. Amer. Statist. Assoc. 101, 980–990.

Laurent, S., Lecourt, C., Palm, F.C., 2016. Testing for jumps in conditionally Gaussian ARMA-GARCH models, a robust approach. Comput. Statist. Data Anal. 100, 383–400.

Laurent, J.P., Omidi Firouzi, H., 2017. Market risk and volatility weighted historical simulation after Basel III. Preprint available at https://www.msci.com/documents/10199/5915b101-4206-4ba0-aee2-3449d5c7e95a.

Mancini, L., Trojani, F., 2011. Robust value-at-risk prediction. J. Financ. Econ. 9, 281–313.

Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Nieto, M.R., Ruiz, E., 2016. Frontiers in VaR forecasting and backtesting. Int. J. Forecast. 32, 475–501.

Pérignon, C., Deng, Z.Y., Wang, Z.J., 2008. Do banks overstate their Value-at-Risk? J. Bank. Financ. 32, 783–794.

Rombouts, J.V.K., Verbeek, M., 2009. Evaluating portfolio Value-at-Risk using semi-parametric GARCH models. Quant. Finance 9, 737–745.

Santos, A.A.P., Nogales, F.J., Ruiz, E., 2013. Comparing univariate and multivariate models to forecast portfolio Value-at-Risk. J. Financ. Econ. 11, 400–441.

Spierdijk, L., 2016. Confidence intervals for ARMA-GARCH Value-at-Risk: the case of heavy tails and skewness. Comput. Statist. Data Anal. 100, 545–559.

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