Centre de Recherche en Économie et Statistique

CREST

Série des Documents de Travail

n° 2012-16

Estimation Adjusted VaR

C. GOURIEROUX[1]

J. M. ZAKOIAN[2]

September 2010, Revised July 2012

[1] CREST and University of Toronto, Department of Economics. *Address CREST; 15 Boulevard Gabriel Péri, 92245 Malakoff Cedex, France*. Email : gourieroux@ensae.fr - Tel. : +33 1 41 17 35 93.

[2] CREST and University Lille 3 (EQUIPPE). *Address CREST; 15 Boulevard Gabriel Péri, 92245 Malakoff Cedex, France*. Email : zakoian@ensae.fr - Tel. : +33 1 41 17 78 25.

ESTIMATION ADJUSTED VaR*

Christian GOURIEROUX† and Jean-Michel ZAKOIAN‡

September 2010, revised July 2012

Abstract: Standard risk measures, such as the Value-at-Risk (VaR), or the Expected Shortfall, have to be estimated and their estimated counterparts are subject to estimation uncertainty. Replacing, in the theoretical formulas, the true parameter value by an estimator based on n observations of the Profit and Loss variable, induces an asymptotic bias of order $1/n$ in the coverage probabilities. This paper shows how to correct for this bias by introducing a new estimator of the VaR, called Estimation adjusted VaR (EVaR). This adjustment allows for a joint treatment of theoretical and estimation risks, taking into account for their possible dependence. The estimator is derived for a general parametric dynamic model and is particularized to stochastic drift and volatility models. The finite sample properties of the EVaR estimator are studied by simulation and an empirical study of the S&P Index is proposed.

KEYWORDS: Value-at-Risk, Estimation Risk, Bias Correction, ARCH Model.

*Forthcoming in *Econometric Theory.* The authors thank the Editor Peter C.B. Phillips and three anonymous referees for helpful comments and suggestions.

†CREST and University of Toronto, Department of Economics. Address: CREST, 15 Boulevard G. Péri, 92245 Malakoff Cedex, France. E-mail: gourieroux@ensae.fr, tel: 33.1.41.17.35.93

‡CREST and University Lille 3 (EQUIPPE). Address: CREST, 15 Boulevard G. Péri, 92245 Malakoff Cedex, France. E-mail: zakoian@ensae.fr, tel: 33.1.41.17.78.25

1 Introduction

The Value-at-Risk (VaR) and more generally the Distortion Risk Measures such as the Expected Shortfall are standard risk measures used in the current regulations introduced in Finance (Basel 2), or Insurance (Solvency 2) to fix the required capital (Pillar 1), or to monitor the risk by means of internal risk models (Pillar 2). These measures can be estimated nonparametrically such as in the so-called historical simulation used in the standard approaches of regulation. In the so-called advanced approaches, these measures can be conditional, that is, take into account the current available information. In such advanced approaches, the risk dynamic is usually represented by a parametric or semi-parametric model, which has to be estimated in a preliminary step. However, the estimated counterparts of risk measures are subject to estimation uncertainty. Replacing, in the theoretical formulas, the true parameter value by an estimator based on n observations of the Profit and Loss variable, induces an asymptotic bias of order $1/n$ in the coverage probabilities. This paper shows how to correct for this bias by introducing a new estimator of the VaR, called Estimation adjusted VaR (EVaR). This adjustment allows for a joint treatment of theoretical and estimation risks, taking into account for their possible dependence.

1.1 Parametric dynamic risk model

More precisely, let us consider a parametric dynamic model for a Markov Profit and Loss (P&L) process (y_t), with a parametric conditional cumulative distribution function (cdf) of y_t given y_{t-1}, denoted $F_{\theta_0}(\cdot \mid y_{t-1})$. The model can be written as :

$$y_t = g(y_{t-1}, \theta_0, \varepsilon_t), \qquad t > t_0, \quad t_0 \in \mathbb{Z}, \tag{1}$$

with an initial value y_{t_0} assumed to be independent of (ε_t), which is a sequence of independent and identically distributed (i.i.d.) variables. The distribution of ε_t can be assumed standard normal without loss of generality[1], $\theta_0 \in \mathbb{R}^d$ is the true parameter value and $g : \mathbb{R} \times \mathbb{R}^d \times \mathbb{R} \mapsto \mathbb{R}$ is a continuous function, strictly increasing with respect to the last component. We denote by a the inverse of the function g w.r.t. the last component, which provides the expression of the Gaussian nonlinear innovation [see e.g. Gouriéroux, Jasiak (2005)] in terms of unknown parameter and observed P & L :

[1] Let us denote by Φ the cdf of the standard normal. Then the autoregressive formula (1) is satisfied with $\varepsilon_t = \Phi^{-1}[F_{\theta_0}(y_t \mid y_{t-1})]$.

$$\varepsilon_t = a(y_{t-1}, \theta_0, y_t). \tag{2}$$

When θ_0 is known, the conditional VaR at risk level $\alpha \in (0, 1)$ is defined by :

$$P_{t-1}[y_t < -\mathrm{VaR}_t(\alpha)] = \alpha, \tag{3}$$

where P_{t-1} denotes the historical distribution conditional on y_{t-1}. This is the opposite of the quantile at level α of the conditional distribution P_{t-1}. Condition (3) is equivalent to :

$$\begin{aligned} & P_{t-1}[g(y_{t-1}, \theta_0, \varepsilon_t) < -\mathrm{VaR}_t(\alpha)] = \alpha \\ \Longleftrightarrow \quad & P_{t-1}[\varepsilon_t < a\{y_{t-1}, \theta_0, -\mathrm{VaR}_t(\alpha)\}] = \alpha, \end{aligned}$$

that is, to a VaR computed on the Gaussian nonlinear innovation ε_t. Let Φ denote the cumulative distribution function (cdf) of the standard normal distribution. We deduce that $a[y_{t-1}, \theta_0, -\mathrm{VaR}_t(\alpha)] = \Phi^{-1}(\alpha)$, or equivalently :

$$\mathrm{VaR}_t(\alpha) = -g[y_{t-1}, \theta_0, \Phi^{-1}(\alpha)].$$

For instance, in a conditionally Gaussian risk model with autoregressive drift and volatility:

$$y_t = \mu(y_{t-1}, \theta_0) + \sigma(y_{t-1}, \theta_0)\varepsilon_t, \quad \varepsilon_t \sim \mathrm{IIN}(0, 1), \tag{4}$$

the theoretical conditional VaR is:

$$\mathrm{VaR}_t(\alpha) = -\mu(y_{t-1}, \theta_0) - \sigma(y_{t-1}, \theta_0)\Phi^{-1}(\alpha).$$

1.2 Estimated VaR

In practice the true parameter value is unknown and replaced by an estimate $\hat{\theta}_n$, say, based on n observations of the P&L. Thus, the conventional plug-in VaR predictor is :

$$\mathrm{VaR}_{n,t}(\alpha) = -g[y_{t-1}; \hat{\theta}_n, \Phi^{-1}(\alpha)]. \tag{5}$$

For instance, in the risk model with autoregressive drift and volatility (4), we get: $\mathrm{VaR}_{n,t}(\alpha) = -\mu(y_{t-1}, \hat{\theta}_n) - \sigma(y_{t-1}, \hat{\theta}_n)\Phi^{-1}(\alpha)$. As observed by Hansen (2006), this practice does not provide an accurate approximation of the conditional coverage probability. Indeed, the inequality $y_t < -\mathrm{VaR}_{n,t}(\alpha)$ is equivalent to the inequality $\hat{\varepsilon}_t < \Phi^{-1}(\alpha)$, where $\hat{\varepsilon}_t = a(y_{t-1}, \hat{\theta}_n, y_t)$ is the nonlinear residual. But the residual distribution is no longer standard normal. Thus, in general

$$P_{t-1}[y_t < -\mathrm{VaR}_{n,t}(\alpha)] \neq \alpha + o_P(1/n),$$

for instance.

1.3 Estimation risk in the literature

Estimation risk in dynamic models has been considered by several authors. Berkowitz and O'Brien (2002) observed that the usual VaR estimates are too conservative. Figlewski (2004) examined the effect of estimation errors on the VaR by simulation. The bias of the VaR estimator, resulting from parameter estimation and misspecified errors distribution, was studied for ARCH(1) models by Bao and Ullah (2004). In the i.i.d. setting, Inui, Kijma and Kitano (2005) showed that the nonparametric VaR estimator (that is an empirical quantile) may have a strong positive bias when the distribution features fat tails. Christoffersen and Gonçalves (2005) studied the loss of accuracy in VaR and ES due to estimation error, and constructed bootstrap predictive confidence intervals for risk measures. Hartz, Mittnik and Paolella (2006) proposed a resampling method based on bootstrap to correct bias in VaR forecasts for the normal-GARCH model. For GARCH models with heavy-tailed errors distributions, Chan, Deng, Peng and Xia (2007) derived the asymptotic distributions of extremal estimated quantiles (that is, the estimated VaR with α tending to zero with n). Escanciano and Olmo (2010, 2011) studied the effects of estimation risk on backtesting procedures. They showed how to correct the critical values in standard tests used to assess VaR models.

These analyses are compatible with the approach of Basel 2 regulation, which distinguishes the reserve for the theoretical risk (corresponding to the estimated VaR) and the reserve for the estimation risk.

1.4 Outline of our paper

Our approach is different. We propose a method to directly adjust the VaR to estimation risk, by computing an Estimation adjusted VaR, denoted $\mathrm{EVaR}_{n,t}(\alpha)$, ensuring the right conditional coverage probability at order $1/n$, that is,

$$P_{t-1}[y_t < -\mathrm{EVaR}_{n,t}(\alpha)] = \alpha + o_P(1/n).$$

Our goal is similar to that of Hansen (2006), who derived adjustments of interval forecasts to account for parameter estimation (see also Lönnbark, (2010)). His assumptions and results will be compared to ours in Section 2.

In Section 2, we explain how the VaR can be adjusted when the parameter has been estimated on a base estimation period. We get an explicit form of the adjustment at order $1/n$, where n is the length of the estimation period. We also provide the adjustment at order $1/n$ of the conditional coverage probability. Applications to stochastic volatility models and drift-volatility models are presented in Section 3. Numerical illustrations are provided in Section 4. We first discuss the finite sample properties of the estimation adjusted VaR. Then, the methodology is applied to the analysis of extremes of the returns on S&P index. Section 5 concludes. The proofs are gathered in appendices.

2 Estimation adjusted VaR

As noted above, an Estimation adjusted VaR is directly derived from the conditional quantile of the residual distribution. We first derive an asymptotic expansion at order $1/n$ of the residual distribution. Then, this expansion is used to obtain the associated expansion of the residual quantile function.

2.1 Expansion of the residual

Let us assume that the parameter is estimated on a base estimation period $t = -n, \ldots, -1$, say, of large length n, with $-n = t_0 + 1$, and that the associated estimator of parameter θ_0 is consistent and asymptotically normal :

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \overset{d}{\to} N(0, \Omega), \tag{6}$$

where Ω is a positive definite matrix and $\overset{d}{\to}$ denotes the convergence in distribution.

Assuming that function a is twice continuously differentiable w.r.t parameter θ, we have :

$$\begin{aligned}
\hat{\varepsilon}_t &= a(y_{t-1}, \hat{\theta}_n, y_t) \\
&= a(y_{t-1}, \theta_0, y_t) + \frac{\partial a(y_{t-1}, \theta_0, y_t)}{\partial \theta'}(\hat{\theta}_n - \theta_0) \\
&\quad + \frac{1}{2n}\sqrt{n}(\hat{\theta}_n - \theta_0)' \frac{\partial^2 a}{\partial \theta \partial \theta'}(y_{t-1}, \theta_0, y_t)\sqrt{n}(\hat{\theta}_n - \theta_0) + o_P(1/n) \\
&= \varepsilon_t + W_{t,n} + o_P(1/n),
\end{aligned} \tag{7}$$

where:
$$\begin{aligned}
W_{t,n} &= \frac{1}{\sqrt{n}} \frac{\partial a}{\partial \theta'}(y_{t-1}, \theta_0, y_t)\sqrt{n}(\hat{\theta}_n - \theta_0) \\
&\quad + \frac{1}{2n}\sqrt{n}(\hat{\theta}_n - \theta_0)' \frac{\partial^2 a}{\partial \theta \partial \theta'}(y_{t-1}, \theta_0, y_t)\sqrt{n}(\hat{\theta}_n - \theta_0).
\end{aligned} \tag{8}$$

Thus the nonlinear residual $\hat{\varepsilon}_t$ is the sum of the Gaussian nonlinear innovation ε_t and a stochastic term of higher order equal to $1/\sqrt{n}$.

2.2 Expansion of the residual quantile

We deduce from (7) that the residual quantile is approximately equal to the quantile of the sum $\varepsilon_t + W_{t,n}$, where $W_{t,n}$ is negligible with respect to ε_t. The expansion of the quantile of such a sum is given in the following Lemma, which can be seen as a second-order Bahadur's expansion, in which the different elements receive an interpretation in terms of conditional moments of random variables (see Appendix A.1).

Lemma 1 *Suppose that $X_n = X + W_n$ where X_n, X, W_n are real random variables. For any n assume that the function $x \mapsto E(W_n^\ell \mid X = x)$ is twice differentiable, for $\ell = 1, 2$, and the pdf f of X is twice differentiable. In addition, assume $E(W_n^2 \mid X = x)$, $\frac{\partial}{\partial x} E[W_n^2 \mid X = x]$ (resp. $E(W_n \mid X = x)$ and $\frac{\partial}{\partial x} E[W_n \mid X = x]$) tend to zero and are of the same stochastic order as $n \to \infty$. Let f_{X,W_n} denote the joint pdf of (X, W_n). Let, for $z, w \in \mathbb{R}$ such that $f_{X,W_n}(z, w) \neq 0$,*

$$C(z, w) = \frac{1}{f_{X,W_n}(z, w)} \int_{-|w|}^{|w|} \left| \frac{\partial^2 f_{X,W_n}}{\partial x^2}(z + x, w) \right| dx.$$

For $z \in \mathbb{R}$ assume that $E|W_n|^{2(1+\nu)} < \infty$ for some $\nu > 0$, and

$$E[|W_n|^{1+\nu} \mid X = z] = o\{E(|W_n| \mid X = z)\}, \tag{9}$$

$$E[|W_n|^{2(1+\nu)} \mid X = z] = o\{E(W_n^2 \mid X = z)\} \quad \text{and} \tag{10}$$

$$\sup_n E\{C(z, W_n)^{1+1/\nu} \mid X = z\} < \infty. \tag{11}$$

Then the following expansions hold, for $z \in \mathbb{R}$ and $u \in (0, 1)$:

$$\begin{aligned}
F_n(z) - F(z) &= -E[W_n \mid X = z] f(z) \\
&\quad + \frac{1}{2} \left\{ \frac{\partial}{\partial x} E[W_n^2 \mid X = x] + E(W_n^2 \mid X = x) \frac{\partial \log f(x)}{\partial x} \right\}_{x=z} f(z) \\
&\quad + o\{E[W_n^2 \mid X = z]\}, \\
G_n(u) - G(u) &= E[W_n \mid X = G(u)] \\
&\quad - \frac{1}{2} \left\{ \frac{\partial}{\partial x} Var[W_n \mid X = x] + Var(W_n \mid X = x) \frac{\partial \log f(x)}{\partial x} \right\}_{x=G(u)} \\
&\quad + o\{E[W_n^2 \mid X = G(u)]\},
\end{aligned}$$

where F and G denote, respectively, the cdf and quantile function of the variable X, while F_n and G_n denote the same functions for X_n.

Proof: see Appendix A.2.

The assumptions of this lemma can be considerably reduced in particular cases. In particular, it is illustrated for Gaussian variables in Appendix A.3. In the special case where $W_n = \varepsilon_n W$, with ε_n a scalar tending to zero, the expansion for the quantile of X_n was derived by Gouriéroux, Laurent, Scaillet (2000), Martin, Wilde (2002). Our version of this lemma controls the residual term in this expansion.

Lemma 1 can be used to derive an asymptotic expansion of the conditional quantile of $\hat{\varepsilon}_t$, by using (7). We denote by $(\partial^k a/\partial y^k)(y_{t-1},\theta,\cdot)$ and $(\partial^k g/\partial x^k)(y_{t-1},\theta,\cdot)$ the k-th order derivatives of the functions $y \mapsto a(y_{t-1},\theta,y)$ and $x \mapsto g(y_{t-1},\theta,x)$, respectively. Let $\|\cdot\|$ denote any norm on $\mathbb{R}^d$.

Proposition 1 *Suppose that the estimator $\hat{\theta}_n$ based on observations $y_{-1},\dots,y_{-n}$ satisfies the asymptotic behavior (6), technical assumptions displayed in Appendix A.4 on the function a and its derivatives, and*

$$E[\hat{\theta}_n - \theta_0 \mid y_{t-1}] = o_P(1/n), \tag{12}$$

$$Var[\sqrt{n}(\hat{\theta}_n - \theta_0) \mid y_{t-1}] = \Omega + o_P(1), \tag{13}$$

$$E[\|\sqrt{n}(\hat{\theta}_n - \theta_0)\|^{4(1+\nu)} \mid y_{t-1}] = O_P(1), \tag{14}$$

where $t>0$ and $\nu>0$ is introduced in the technical assumptions. Then, the conditional quantile of $\hat{\varepsilon}_t$ has the following expansion:

$$q_{\hat{\varepsilon}_t}(\alpha,\theta_0) \quad = \quad \Phi^{-1}(\alpha) - \frac{b_t\{\theta_0,\Phi^{-1}(\alpha)\}}{2n} + o_P\left(1/n\right),$$

where

$$\begin{aligned} b_t(\theta_0,\varepsilon) \quad = \quad & \Bigg[2\frac{\partial^2 a}{\partial y \partial \theta'}(y_{t-1},\theta,y)\frac{\partial g}{\partial \varepsilon}(y_{t-1},\theta_0,\varepsilon)\Omega\frac{\partial a}{\partial \theta}(y_{t-1},\theta,y) \\ & -\varepsilon\frac{\partial a}{\partial \theta'}(y_{t-1},\theta,y)\Omega\frac{\partial a}{\partial \theta}(y_{t-1},\theta,y) \\ & - Tr\left\{\frac{\partial^2 a}{\partial\theta\partial\theta'}(y_{t-1},\theta,y)\Omega\right\}\Bigg]_{\theta=\theta_0,y=g(y_{t-1},\theta_0,\varepsilon)}. \end{aligned}$$

Moreover, we also have

$$\begin{aligned} & b_t(\theta_0,\varepsilon) \\ = \quad & \Bigg[\left\{\frac{\partial^2 a}{\partial y^2}(y_{t-1},\theta,y) - \varepsilon\left(\frac{\partial a}{\partial y}(y_{t-1},\theta,y)\right)^2\right\}\frac{\partial g}{\partial\theta'}(y_{t-1},\theta,\varepsilon)\Omega\frac{\partial g}{\partial\theta}(y_{t-1},\theta,\varepsilon) \\ & +\frac{\partial a}{\partial y}(y_{t-1},\theta,y)Tr\left(\Omega\frac{\partial^2 g}{\partial\theta\partial\theta'}(y_{t-1},\theta,\varepsilon)\right)\Bigg]_{\theta=\theta_0,y=g(y_{t-1},\theta_0,\varepsilon)}. \end{aligned} \tag{15}$$

Proof: see Appendix A.5.

Remark 1: When the first-order autoregressive model takes the additive form $y_t = \mu(y_{t-1}, \theta_0) + \varepsilon_t$, the adjustment term (15) reduces to:

$$b_t(\theta_0, \varepsilon) = \left[-\varepsilon \frac{\partial \mu}{\partial \theta'}(y_{t-1}, \theta) \Omega \frac{\partial \mu}{\partial \theta}(y_{t-1}, \theta) + Tr \left(\Omega \frac{\partial^2 \mu}{\partial \theta \partial \theta'}(y_{t-1}, \theta) \right) \right]_{\theta=\theta_0}.$$

It is interesting to note that $b_t\{\theta_0, \Phi^{-1}(\alpha)\} > 0$, at least for α sufficiently small. Indeed, the first term in the right-hand side is strictly positive for $\varepsilon < 0$ (because Ω is positive definite) and tends to infinity as ε decreases, whereas the second term does not depend on ε. This means, that asymptotically, the estimation effect is to lower the quantiles for small values of α.[2]

2.3 Conditional bias reduction

The regularity conditions in Proposition 1 concern the conditional distribution of the estimator $\hat{\theta}_n$ given y_{t-1}. Indeed, the estimator depends on observations on a base period and these observations are dependent of future variables. It is assumed that this dependence can be neglected at the first-order. It is important to note that even if the estimator is unbiased, or of order smaller than $1/n$, that is, if $E(\hat{\theta}_n - \theta_0) = o_P(1/n)$, this is not sufficient to ensure the negligibility of the *conditional* bias (the first equality in (12)). To illustrate this, let us consider the prediction problem in the AR(1) model $y_t = \rho y_{t-1} + u_t$, where (u_t) is an independent white noise. The optimal one-step ahead prediction of y_{t+1} in the L^2 sense is ρy_t, which can be estimated by $\hat{\rho} y_t$, where $\hat{\rho}$ is an estimator of ρ obtained from the observations $y_1, \ldots, y_n$ with $n \leq t$. The estimated prediction is thus unbiased if and only if $E[(\hat{\rho} - \rho) y_t] = 0$. The conditional moment assumption $E(\hat{\rho} - \rho \mid y_t) = 0$ implies prediction unbiasedness, but the unconditional one, $E(\hat{\rho} - \rho) = 0$, does not.

A preliminary automatic approach, such as a conditional jacknife, can be applied to remove the conditional bias before applying the formula in Proposition 1. The jacknife technique was introduced by Quenouille (1956). See Chambers (2012) for a recent investigation of the use of the jackknife as a method of estimation in stationary autoregressive models. Phillips and Yu (2005) proposed a method of bias reduction based on the jacknife technique for the pricing of bond options and other derivative securities. Our problem is not standard because the first part of (12) concerns the bias of

[2] By equation (15), the same conclusion holds for the general case under the assumption that g is convex in θ and a is convex in its last component around $(y_{t-1}, \theta, -\mathrm{VaR}_t(\alpha))$.

the estimator conditional on a future variable. The next result shows how a jacknife correction can be introduced in our framework to get satisfied the assumption of unbiasedness.

Proposition 2 *Let $\hat{\theta}_n = \arg\min_\Theta Q_n(\theta; y_{-1}, \ldots, y_{-n})$. Suppose that the following expansion holds for the conditional bias of the estimator*

$$E(\hat{\theta}_n - \theta_0 \mid y_t) = \frac{A(t)}{n} C_n(y_t, \theta_0) + o_P(1/n), \tag{16}$$

where $C_n(y_t, \theta_0) = O_P(1)$ in $\mathbb{R}^d$ and A is a known real valued function. Let, for $n = 2\ell$,

$$\hat{\theta}_\ell^{(1)} = \arg\min_\Theta Q_\ell(\theta; y_{-1}, \ldots, y_{-\ell}), \quad \hat{\theta}_\ell^{(2)} = \arg\min_\Theta Q_\ell(\theta; y_{-\ell-1}, \ldots, y_{-2\ell}).$$

A Jacknife estimator based on this subsampling scheme is given by

$$\hat{\theta}_n^{(J)} = \frac{1}{A(t) - A(t+\ell)} (A(t)\hat{\theta}_\ell^{(2)} - A(t+\ell)\hat{\theta}_\ell^{(1)}),$$

and we have:

$$E(\hat{\theta}_n^{(J)} - \theta_0 \mid y_t) = o_P(1/n).$$

The proof is straightforward. Therefore, Assumption (12) is satisfied for this conditional jacknife adjusted estimator, whenever condition (16) is satisfied. An illustrative example, in which the other assumptions of Proposition 1 are also satisfied, is developed in Appendix A.6.

2.4 Definition of the Estimation Adjusted VaR

We deduce from Proposition 1 the estimation adjusted VaR for P& L, denoted $\text{EVaR}_{n,t}(\alpha)$. A more precise terminology would distinguish the EVaR defined by:

$$\text{EVaR}_t(\alpha) = -g[y_{t-1}; \theta_0, q_{\hat{\varepsilon}_t}(\alpha, \theta_0)],$$

from the estimated EVaR, in which the estimate $\hat{\theta}_n$ is substituted to the true parameter value θ_0. For expository purpose, we denote the estimated EVaR by $\text{EVaR}_{n,t}$ and do not mention the term "estimated".

Definition 1 *The estimation adjusted VaR is given by :*

$$EVaR_{n,t}(\alpha) = -g[y_{t-1}; \hat{\theta}_n, q_{\hat{\varepsilon}_t}(\alpha, \hat{\theta}_n)], \tag{17}$$

where

$$q_{\hat{\varepsilon}_t}(\alpha, \hat{\theta}_n) = \Phi^{-1}(\alpha) - \frac{b_t(\hat{\theta}_n, \Phi^{-1}(\alpha))}{2n}. \tag{18}$$

Thus, $\text{EVaR}_{n,t}(\alpha)$ is obtained by substituting the estimate $\hat{\theta}_n$ to the true parameter value θ_0 (as for the standard estimated VaR, see Section 1), but also by substituting the estimated α-quantile of the nonlinear residual to the α-quantile of the Gaussian nonlinear innovation.

Remark 2: In view of Remark 1, for the additive model we have:

$$\text{EVaR}_{n,t}(\alpha) > \text{VaR}_{n,t}(\alpha), \tag{19}$$

at least for small α. The required capital is deduced from the estimated VaR by a formula of the type

$$\text{RC}_t = \max\left(\text{VaR}_{n,t}(\alpha), \frac{k}{60}\sum_{h=0}^{59}\text{VaR}_{n,t-h}(\alpha)\right),$$

where k is a trigger parameter, which can be controlled by the regulator. If the trigger parameter is fixed, inequality (19) shows that the required capital above can imply an insufficient risk coverage by $\text{VaR}_{n,t}(\alpha)$, since the estimation risk has not been taken into account. If this insufficient risk coverage is observed at several consecutive periods, the regulator might either increase the trigger parameter in the above formula to compensate the mishedging, or ask for substituting $\text{EVaR}_{n,t}$ to $\text{VaR}_{n,t}$.

Remark 3: There exists an alternative bootstrap approach to the estimation adjusted VaR. More precisely, let us denote by $\hat{\varepsilon}_{-n}, \ldots, \hat{\varepsilon}_{-1}$ the residuals computed on the base period. Let us consider n independent drawings in the set $\{\hat{\varepsilon}_{-n}, \ldots, \hat{\varepsilon}_{-1}\}$, denoted by $\hat{\varepsilon}^h_{-n}, \ldots, \hat{\varepsilon}^h_{-1}$, and compute recursively the set of simulated values $y^h_t = g(y^h_{t-1}, \hat{\theta}_n, \hat{\varepsilon}^h_t), \quad t = -n, \ldots, -1$, using an initial value. This simulated observation set can be used to deduce an ML or QML bootstrapped estimator $\hat{\theta}^h_n$, say, and thus a bootstrapped residual for date $t \geq 0$ by $\hat{\varepsilon}^h_t = a(y_{t-1}, \hat{\theta}^h_n, y_t)$. This procedure can be replicated $H = 100$ times, say, and the set of residuals $\hat{\varepsilon}^h_t, \quad h = 1, \ldots, 100$ can be ranked by increasing order as $\hat{\varepsilon}^{(1)}_t < \hat{\varepsilon}^{(2)}_t < \ldots < \hat{\varepsilon}^{(100)}_t$. A bootstrapped adjusted VaR, for $\alpha = 5\%$, say, can then be defined by:

$$\text{BVaR}_{n,t}(\alpha) = -g[y_{t-1}; \hat{\theta}_n, \hat{\varepsilon}^{(5)}_t].$$

This approach is in the spirit of the Monte-Carlo correction of testing procedures to satisfy the type I-error restriction in finite sample (see e.g. Dufour and Kiviet, 1997). Our approach with a closed form correction avoids the computational cumbersome $H = 100$ estimations of the dynamic model on the base period. Moreover, as seen below, the consequence of the adjustment on the coverage probabilities are known in our case, and is more difficult to derive in closed form for the bootstrap approach.

2.5 Expansion of the coverage probability

Let us now discuss the conditional coverage probabilities.[3] We know that the theoretical VaR satisfies the exact restriction on the coverage probability given by (3), but this theoretical VaR depends on the unknown parameter value and thus cannot be used in practice. The next proposition shows that the error on the conditional coverage probability is of order $1/n$ (resp. strictly smaller than $1/n$), when the standard estimated VaR (resp. the EVaR) is used. This justifies ex-post the estimation adjustment of the VaR.

Proposition 3 *Under the assumptions of Proposition 1, the conditional coverage probability of the standard estimated VaR is such that:*

$$P_{t-1}[y_t < -VaR_{n,t}(\alpha)] \quad = \quad \alpha + \frac{b_t\{\theta_0, \Phi^{-1}(\alpha)\}}{2n}\phi\left\{\Phi^{-1}(\alpha)\right\} + o_P\left(1/n\right), \tag{20}$$

where ϕ denotes the density function of the standard Gaussian distribution, and the conditional coverage probability of the estimation adjusted VaR is given by

$$P_{t-1}[y_t < -EVaR_{n,t}(\alpha)] = \alpha + o_P\left(1/n\right). \tag{21}$$

Proof: The conditional probabilities in the right-hand side of (20) can be written under the form $P_{t-1}[\varepsilon_t < z]$, allowing to apply Lemma 1. A Taylor expansion allows to handle the right-hand side of (21), which is written under the form $P_{t-1}[\varepsilon_t < z(\hat{\theta}_n)]$. See Appendix A.7 (in Appendix A.8 we give an alternative proof, following the lines of Hansen (2006)). □

Remark 4: Hansen (Theorem 2, 2006) derived a different expression for the conditional coverage probability, under the following (implicit) alternative assumption (see the last equality on p. 396, and Appendix A.8 below):

$$\begin{aligned} E[\mathrm{VaR}_{n,t}(\alpha) - \mathrm{VaR}_t(\alpha) \mid y_{t-1}] &= o_P(1/n), \\ \mathrm{Var}[\sqrt{n}\{\mathrm{VaR}_{n,t}(\alpha) - \mathrm{VaR}_t(\alpha)\} \mid y_{t-1}] &= \Sigma(\theta_0) + o_P(1), \end{aligned}$$

for some positive definite matrix $\Sigma(\theta_0)$. This conditional unbiasedness assumption differs from our assumption (12), since it is written on the estimated VaR and not on the estimated parameter. It seems more appropriate

[3] It is more appropriate for risk management to consider the accuracy of the conditional coverage probability than that of the VaR itself. Indeed, the coverage probability is the basic diagnostic tool used by the regulator to check ex-post the adequacy of the selected reserves.

to write a more primitive condition, that is, a condition on the estimator of the parameter itself. Indeed, as seen in Section 2.3, the conditional bias on the estimator can often be reduced by applying a conditional jacknife, whereas a similar approach does not exist to correct the conditional bias in the VaR estimation.

In Hansen (2006) the term $b_t\{\theta_0, \Phi^{-1}(\alpha)\}$ is replaced by:

$$\begin{aligned} & c_t\{\theta_0, \Phi^{-1}(\alpha)\} \\ = & \left\{ \frac{\partial^2 a}{\partial y^2}\{y_{t-1}, \theta_0, -\mathrm{VaR}_t(\alpha)\} - \Phi^{-1}(\alpha) \left(\frac{\partial a}{\partial y}\{y_{t-1}, \theta_0, -\mathrm{VaR}_t(\alpha)\} \right)^2 \right\} \\ & \times \frac{\partial g}{\partial \theta'}(y_{t-1}, \theta_0, \Phi^{-1}(\alpha))\Omega \frac{\partial g}{\partial \theta}(y_{t-1}, \theta_0, \Phi^{-1}(\alpha)). \end{aligned}$$

By comparison with (15), we see that the term

$$\frac{\partial a}{\partial y}\{y_{t-1}, \theta_0, -\mathrm{VaR}_t(\alpha)\} Tr \left(\Omega \frac{\partial^2 g}{\partial \theta \partial \theta'}\{y_{t-1}, \theta_0, \Phi^{-1}(\alpha)\} \right), \tag{22}$$

is missing in Hansen's result. If the asymptotic conditional unbiasedness assumption on $\hat{\theta}_n$ is satisfied, we generally have a bias on $\mathrm{VaR}_{n,t}(\alpha)$, except in the very special case where function g is linear in parameter θ. The additional term (22) appearing in the expression of $b_t\{\theta_0, \Phi^{-1}(\alpha)\}$ corrects for this bias on $\mathrm{VaR}_{n,t}(\alpha)$. Finally, the linearity of function g with respect to θ is generally not fulfilled, except in simple models such as $y_t = \mu + \rho y_{t-1} + \sigma \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0,1)$ with $\theta = (\mu, \rho, \sigma)'$.

2.6 Extension to higher-order autoregressive models

In practical situations, it may be worth considering dynamic models with a longer memory, such as:

$$y_t = g(y_{t-1}, \ldots, y_{t-p}, \theta_0, \varepsilon_t),$$

where $p \geq 1$ denotes the autoregressive order and (ε_t) is an i.i.d. sequence of standard normal variables. When function g is invertible w.r.t. ε_t, this model can be equivalently written as $\varepsilon_t = a(y_{t-1}, \ldots, y_{t-p}, \theta_0, y_t)$. The estimation adjusted VaR is now given by:

$$\mathrm{EVaR}_{n,t}(\alpha) = -g(y_{t-1}, \ldots, y_{t-p}; \hat{\theta}_n, q_{\hat{\varepsilon}_t}(\alpha, \hat{\theta}_n)),$$

where $q_{\hat{\varepsilon}_t}(\alpha, \hat{\theta}_n)$ is defined by (18), and b_t is defined as in Proposition 1 with y_{t-1} replaced by $y_{t-1}, \ldots, y_{t-p}$. It is easily checked that Propositions

1 and 3 continue to hold under the regularity conditions: $E[\sqrt{n}(\hat{\theta}_n - \theta_0) \mid y_{t-1}, , \dots, y_{t-p}, \varepsilon_t] = o_P(n^{-1/2})$, $\mathrm{Var}[\sqrt{n}(\hat{\theta}_n - \theta_0) \mid y_{t-1}, , \dots, y_{t-p}, \varepsilon_t] = \Omega + o_P(1)$, and $(\partial/\partial u \mathrm{Var}[\sqrt{n}(\hat{\theta}_n - \theta_0) \mid y_{t-1}, , \dots, y_{t-p}, u])_{u=\varepsilon_t} = o_P(1)$.

3 Applications to stochastic drift and volatility models

As an illustration, we consider below stochastic drift and volatility models, for which explicit formulas for the estimation adjusted VaR can be derived and interpreted .

3.1 Stochastic volatility model

The first class of models is of the form:

$$y_t = \sigma_t(\theta_0)H(\varepsilon_t), \tag{23}$$

where $\sigma_t(\theta_0)$ is a positive function of the past of y_t, depending on an unknown parameter θ_0, H is a continuous increasing function and (ε_t) is a sequence of i.i.d. standard normal variables. When H is the identity function, we get conditionally normal observations. By selecting another increasing function H, we can change the form of the conditional distribution of y_t and allow for heavy-tailed conditional distributions. The variable $\sigma_t(\theta_0)$ is a conditional scale factor. It is the conditional standard-deviation of y_t if the variance of $H(\varepsilon_t)$ exists and is equal to 1. This specification encompasses, in particular, the standard ARCH(q) model, with possibly non Gaussian conditional distribution.

In such models, the unknown parameter value θ_0 is usually estimated either by the Maximum Likelihood (ML) method, or by the Quasi-Maximum Likelihood (QML) applied as if the distribution of $H(\varepsilon_t)$ were standard normal. Other estimation methods can also be considered, as the quantile regression for instance (see Koenker and Zhao (1996) for an extension of quantile regression to linear ARCH models). For this method, the non differentiability of the optimization criterion would have to be taken into account. The strong consistency and asymptotic normality in GARCH models has been established for the QML estimator under mild conditions by Berkes, Horváth and Kokoszka (2003) and Francq and Zakoïan (2004), and for the ML estimator by Berkes and Horváth (2004); see also Francq and Zakoian (2010). Under regularity conditions (in particular $E\eta_t^2 = 1$, where $\eta_t = H(\varepsilon_t)$, for the QML method), these papers show that the ML (or QML) estimator $\hat{\theta}_n$ of θ_0 satisfies $\hat{\theta}_n \to \theta_0$, *a.s.*, and

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, \xi J^{-1}), \qquad J = E\left(\frac{1}{\sigma_t^4(\theta_0)} \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta} \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta'}\right), \quad (24)$$

where the coefficient ξ depends on the estimation method. More precisely, we have

$$\xi = \begin{cases} E(\eta_t^4) - 1, & \text{for the QML method,} \\ 4/E\left\{\left(1 + \frac{\partial \log f}{\partial y}(\eta_t)\eta_t\right)^2\right\}, & \text{for the ML method,} \end{cases}$$

where f denotes the density of $\eta_t = H(\varepsilon_t)$. Note that the QML approach requires rather thin tails for η_t, since $E(\eta_t^4)$ has to be finite.

For Model (23), we have

$$g(y_{t-1}, \theta, \varepsilon) = \sigma_t(\theta) H(\varepsilon), \quad a(y_{t-1}, \theta, y) = H^{-1}(y/\sigma_t(\theta)).$$

By applying Proposition 1, we get an explicit form of the EVaR.

Corollary 1 *For the volatility model (23),*

i) the estimated VaR is given by:

$$VaR_{n,t}(\alpha) = -\sigma_t(\hat{\theta}_n) H\{\Phi^{-1}(\alpha)\} = -\sigma_t(\hat{\theta}_n) G(\alpha),$$

where G is the quantile function of $\eta_t = H(\varepsilon_t)$;

ii) under the assumptions of Proposition 1, the estimation adjusted VaR is given by:

$$\begin{aligned} & EVaR_{n,t}(\alpha) \\ = \; & VaR_{n,t}(\alpha) \\ & -\frac{1}{2n} \frac{1}{\sigma_t(\hat{\theta}_n)} \frac{\partial \sigma_t(\hat{\theta}_n)}{\partial \theta'} \hat{\Omega}_n \frac{\partial \sigma_t(\hat{\theta}_n)}{\partial \theta} \left[\frac{H^2(\varepsilon)}{h(\varepsilon)} \left\{\frac{\partial \log h(\varepsilon)}{\partial \varepsilon} + \varepsilon\right\}\right]_{\varepsilon = \Phi^{-1}(\alpha)} \\ & \qquad + \frac{1}{2n} H\{\Phi^{-1}(\alpha)\}\, Tr\left[\frac{\partial^2 \sigma_t(\hat{\theta}_n)}{\partial \theta \partial \theta'} \hat{\Omega}_n\right] + o_P(1/n), \end{aligned}$$

where $h(\varepsilon) = \partial H(\varepsilon)/\partial \varepsilon$, and $\hat{\Omega}_n$ is a consistent estimator of the asymptotic variance of $\hat{\theta}_n$;

iii) in the standard ARCH case, the estimation adjusted VaR is given by:

$$\begin{aligned}
& EVaR_{n,t}(\alpha) \\
= \quad & VaR_{n,t}(\alpha) - \frac{1}{2n}\frac{1}{\sigma_t(\hat{\theta}_n)}\frac{\partial \sigma_t(\hat{\theta}_n)}{\partial \theta'}\hat{\Omega}_n\frac{\partial \sigma_t(\hat{\theta}_n)}{\partial \theta} \\
& \times \left[\frac{H^2(\varepsilon)}{h(\varepsilon)}\left\{\frac{\partial \log h(\varepsilon)}{\partial \varepsilon} + \varepsilon + \frac{\partial \log H(\varepsilon)}{\partial \varepsilon}\right\}\right]_{\varepsilon=\Phi^{-1}(\alpha)} + o_P(1/n),
\end{aligned}$$

Proof: see Appendix A.9.

In the Gaussian case we have $H(\varepsilon) = \varepsilon$ and the inequality:

$$\mathrm{EVaR}_{n,t}(\alpha) > \mathrm{VaR}_{n,t}(\alpha),$$

for $\alpha \leq 0.5$. For such values of α, taking into account the estimation step in the evaluation of the quantile increases the reserve. For very small α, the difference between the estimated EVaR and VaR can be large and is path-dependent. In standard GARCH models, the quantity $\frac{1}{\sigma_t^2(\hat{\theta}_n)}\frac{\partial \sigma_t}{\partial \theta'}(\hat{\theta}_n)\hat{\Omega}_n\frac{\partial \sigma_t}{\partial \theta}(\hat{\theta}_n)$ is bounded, and this difference is approximately proportional to the current volatility. In high-volatility periods, the increase of reserve due to the estimation risk is large.

The following result gives more insight on the mean asymptotic discrepancy between the two estimated VaR's in the standard ARCH case.

Corollary 2 *For a standard ARCH model, under the conditions ensuring the validity of (24), and if $\hat{\Omega}_n$ is a strongly consistent estimator of ξJ^{-1}, we have:*

$$\begin{aligned}
& E \lim_{n\to\infty} a.s. \frac{n(EVaR_{n,t}(\alpha) - VaR_{n,t}(\alpha))}{\sigma_t(\hat{\theta}_n)} \\
= \quad & -d\frac{\xi}{8}\left[\frac{H^2(\varepsilon)}{h(\varepsilon)}\left\{\frac{\partial \log h(\varepsilon)}{\partial \varepsilon} + \varepsilon + \frac{\partial \log H(\varepsilon)}{\partial \varepsilon}\right\}\right]_{\varepsilon=\Phi^{-1}(\alpha)} := \Delta(\alpha),
\end{aligned}$$

where d is the dimension of θ_0.

Proof: see Appendix A.10.

The adjustment scale factor $\Delta(\alpha)$ involves the number d of parameters to be estimated and the difference between the estimated VaR's is proportional to the ratio d/n. This is a function of the risk level α, which generally tends

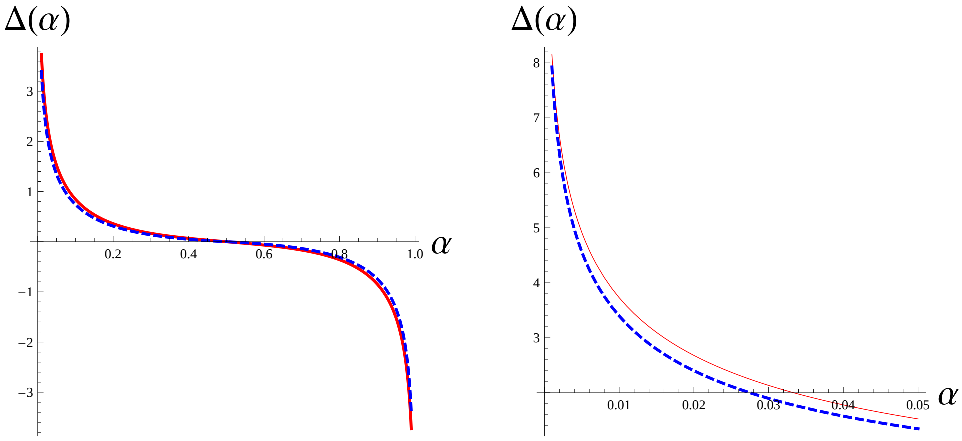

Figure 1: $\Delta(\alpha)$ with $d = 1$ and $\alpha \in (0, 1)$ (left panel), for the standard Gaussian distribution (red thick line) and the Laplace distribution (blue dashed line). The right panel is a zoom for $\alpha \in (0, 0.05)$.

to infinity when α is close to either 0 (VaR of a long investment in asset y), or 1 (VaR of a short investment in asset y). Since function H is increasing, the sign of $\Delta(\alpha)$ is the sign of the term into wide brackets. The pattern of function Δ is illustrated in Figure 1 for conditionally Gaussian and Laplace returns, respectively. As expected the adjustment is larger when the tails are fatter, that is, for Laplace returns. Finally, function Δ is symmetric w.r.t. $\alpha = 0.5$ due to the symmetry in the standard Gaussian and Laplace distributions.

3.2 Stochastic drift-volatility model

Suppose now that a conditional mean $\mu_t(\theta_0)$ is added to (23), as:

$$y_t = \mu_t(\theta_0) + \sigma_t(\theta_0) H(\varepsilon_t), \quad (\varepsilon_t) \sim \mathrm{IIN}(0, 1). \tag{25}$$

This specification encompasses the ARCH-M model (see Engle, Lilien and Robbins (1987)), with a risk premium in the drift. It also encompasses AR(p)-ARCH(q) models, with possibly non Gaussian conditional distribution. The strong consistency and asymptotic normality in ARMA-GARCH models has been established for the QML estimator by Francq and Zakoïan (2004). In this paper it is shown that, under some regularity conditions (in particular $E\eta_t = 0$ and $E\eta_t^2 = 1$, where $\eta_t = H(\varepsilon_t)$), the QML estimator $\hat{\theta}_n$ of θ_0 satisfies

$$\hat{\theta}_n \to \theta_0, \quad a.s., \quad \sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, \mathcal{J}^{-1}\mathcal{I}\mathcal{J}^{-1}), \tag{26}$$

where the matrices $\mathcal{I}$ and $\mathcal{J}$ are expectations of functions of first and second-order derivatives of $\mu_t(\theta)$ and $\sigma_t(\theta)$.

For Model (25) we have

$$g(y_{t-1}, \theta, \varepsilon) = \mu_t(\theta) + \sigma_t(\theta)H(\varepsilon), \quad a(y_{t-1}, \theta, y) = H^{-1}[\{y - \mu_t(\theta)\}/\sigma_t(\theta)].$$

By applying Proposition 1, we get an explicit form of the EVaR.

Corollary 3 *For the stochastic drift-volatility model (25),*

i) the estimated VaR is given by

$$VaR_{n,t}(\alpha) = -\mu_t(\hat{\theta}_n) - \sigma_t(\hat{\theta}_n)H\{\Phi^{-1}(\alpha)\} = -\mu_t(\hat{\theta}_n) - \sigma_t(\hat{\theta}_n)G(\alpha),$$

where G is the quantile function of $\eta_t = H(\varepsilon_t)$;

ii) under the assumptions of Proposition 1, the estimation adjusted VaR is given by

$$\begin{aligned}
& EVaR_{n,t}(\alpha) \\
= \; & VaR_{n,t}(\alpha) - \frac{1}{2n}\frac{1}{\sigma_t(\hat{\theta}_n)}\left(\frac{\partial \mu_t(\hat{\theta}_n)}{\partial \theta'} + H\left[\Phi^{-1}(\alpha)\right]\frac{\partial \sigma_t(\hat{\theta}_n)}{\partial \theta'}\right)\hat{\Omega}_n \\
& \times \left(\frac{\partial \mu_t(\hat{\theta}_n)}{\partial \theta} + H\left[\Phi^{-1}(\alpha)\right]\frac{\partial \sigma_t(\hat{\theta}_n)}{\partial \theta}\right)\left[\frac{1}{h(\varepsilon)}\left\{\frac{\partial \log h(\varepsilon)}{\partial \varepsilon} + \varepsilon\right\}\right]_{\varepsilon = \Phi^{-1}(\alpha)} \\
& + \frac{1}{2n} Tr\left[\left(H\{\Phi^{-1}(\alpha)\}\frac{\partial^2 \sigma_t(\hat{\theta}_n)}{\partial \theta \partial \theta'} + \frac{\partial^2 \mu_t(\hat{\theta}_n)}{\partial \theta \partial \theta'}\right)\hat{\Omega}_n\right] + o_P(1/n),
\end{aligned}$$

where $\hat{\Omega}_n$ is a consistent estimator of the asymptotic variance of $\hat{\theta}_n$.

Proof: see Appendix A.11.

4 Numerical Illustrations

4.1 Simulation experiments

To assess the performance of the VaR adjustment in finite sample, we computed the estimator on $N = 5,000$ independent simulated trajectories of an ARCH(1) model:

$$y_t = \sqrt{1 + a y_{t-1}^2}\,\eta_t, \qquad \eta_t \sim t_\nu, \tag{27}$$

where t_ν denotes the standardized Student distribution with ν degrees of freedom. The standardized Student is often employed for GARCH errors in applied works. The degree of freedom ν was chosen in the set $\{6, 7, 10, \infty\}$, corresponding to a kurtosis of 6, 5, 4 and 3, respectively, where the last value corresponds to the Gaussian distribution. The ARCH coefficient a was allowed to vary in $\{0.1, 0.5, 1, 1.4, 2, 2.5\}$. These values and error distributions satisfy the strict stationarity condition, $a < e^{-E \log \eta_t^2}$.

For each simulated trajectory of length $n + H$, the ARCH(1) model was estimated over the first n observations and the remaining H observations were reserved for VaR evaluations.

Table 1 shows, for $\alpha = 0.1, \alpha = 0.05$ and $\alpha = 0.01$, the percentages of violations for the theoretical VaR (computed with the true parameter value) and the estimated VaR's: $\mathrm{VaR}_{n,t}(\alpha)$ and $\mathrm{EVaR}_{n,t}(\alpha)$. For instance, for the standard VaR estimator, this percentage is defined as the proportion of the events:

$$y_t < -\mathrm{VaR}_{n,t}(\alpha), \quad t = n + 1, \ldots n + H,$$

among the simulated samples. The most striking result is that the adjusted VaR does a better job than the standard one in any situation. Of course, the theoretical VaR is often closer to the nominal probability than our estimator, but the difference is small. By comparison, the standard VaR estimator is generally twice or three times more distant to the nominal value than the adjusted VaR. It is also worth noting that our estimator provides satisfactory results even for fat tailed marginal distributions. Such fat tails occur when the degree of freedom ν is small and/or the ARCH coefficient a is large. Recall that a fourth-order moment is required for η_t, but that no moment condition is imposed on y_t.

Another comparison of the VaR estimators can be based on the expected shortfall when the VaR is violated, and on the expected excess of reserves when it is not violated, in percentages of the theoretical VaR. To this aim we introduce, for each simulated path, the quantities

$$\begin{aligned}
p_-^{VaR_n} &= \frac{1}{H_+^{VaR_n}} \sum_{t=n+1}^{n+H} 100 \frac{\{VaR_{n,t}(\alpha) + y_t\}^-}{VaR_t(\alpha)}, \\
p_+^{VaR_n} &= \frac{1}{H - H_+^{VaR_n}} \sum_{t=n+1}^{n+H} 100 \frac{\{VaR_{n,t}(\alpha) + y_t\}^+}{VaR_t(\alpha)}, \\
H_+^{VaR_n} &= \frac{1}{H} \sum_{t=n+1}^{n+H} \mathbb{1}_{VaR_{n,t}(\alpha) > y_t},
\end{aligned}$$

where $x^+ = \max(x, 0)$ and $x^- = \max(-x, 0)$. Similar quantities are defined for the EVaR estimator and for the VaR at the true parameter value. Table 2 displays the average of the $p_+^{VaR(\theta_0)}$, $p_+^{VaR_n}$, $p_+^{EVaR_n}$, $p_-^{VaR(\theta_0)}$, $p_-^{VaR_n}$ and $p_-^{EVaR_n}$ over the $5,000$ simulations. For this table and the next one, to save space, we only report results for $\nu = 7$ and for the Gaussian distribution ($\nu = \infty$), and for $\alpha = 0.1$ and $\alpha = 0.05$. The estimation adjusted VaR being more prudential, the expected shortfall is diminished and compensated by an increase of the expected excesses of reserves. These features are due to the complicated mix between the larger value of the EVaR in average and its behavior in high volatility periods.

To gauge the impact of the sample size used in the estimation step, we reproduced the experiment of Table 1 for $n = 200$ instead of $n = 100$. The results displayed in Table 3 show that, as expected, increasing n improves the accuracy of the VaR estimators: the percentages of violations are closer to the nominal probability, for both the standard estimator and the adjusted VaR. However, the superiority of the latter estimator remains. Table 4 shows that the same conclusion holds for $n = 250$ and $H = 250$, which corresponds to approximately one year of daily observation followed by one year of daily VAR evaluation.

Finally, our estimator is compared with the estimator proposed by Hansen (2006) (see Remark 4 above). For the same simulated paths as in Table 1, we show in Table 5 the results obtained with Hansen's estimator. Percentages are underlined when they are closer to the nominal probability than the percentage obtained with the EVaR estimator. For 55 over 72 cases, the estimator of this paper provides better results than Hansen's estimator. Similar findings, not reported here, were observed for $n = 200$ and $n = H = 250$.

4.2 Application to financial data

We apply the VaR adjustment to daily returns of the SP500 index. The data range from January 2, 1990 to March 25, 2008. As noted in Section 2.6, our method can be applied to finite-order Markov models. In particular, this approach does not apply to GARCH(p, q) models with $p > 0$. We therefore estimated an ARCH(6) model, by the QML method, from the first $n = 250$ observations (one year) of the series of centered log-returns. The estimated model is given by: $y_t = \hat{\sigma}_t \hat{\eta}_t$,

$$\hat{\sigma}_t^2 = 0.562 + 0.098 y_{t-1}^2 + 0.045 y_{t-2}^2 + 0.027 y_{t-3}^2 + 0.103 y_{t-4}^2 + 0.095 y_{t-5}^2 + 0.081 y_{t-6}^2.$$

Table 1: Percentages of violations computed over $5,000$ independent simulations of the ARCH(1) model (27). The model is estimated over a sample of length $n = 100$ and the violations are computed over the next $H = 30$ observations (out-of-sample).

α	a	VaR_t	$\mathrm{VaR}_{n,t}$	$\mathrm{EVaR}_{n,t}$	VaR_t	$\mathrm{VaR}_{n,t}$	$\mathrm{EVaR}_{n,t}$
			$\nu = 6$			$\nu = 7$	
0.01	0.1	0.0102	0.0118	0.0093	0.0100	0.0120	0.0096
	0.5	0.0102	0.0130	0.0099	0.0102	0.0128	0.0099
	1.0	0.0099	0.0130	0.0098	0.0099	0.0123	0.0096
	1.4	0.0103	0.0131	0.0101	0.0096	0.0121	0.0094
	2.0	0.0098	0.0122	0.0095	0.0098	0.0124	0.0099
	2.5	0.0100	0.0123	0.0098	0.0103	0.0127	0.0102
0.05	0.1	0.0500	0.0545	0.0486	0.0494	0.0533	0.0483
	0.5	0.0496	0.0555	0.0486	0.0497	0.0546	0.0489
	1.0	0.0501	0.0558	0.0493	0.0505	0.0548	0.0495
	1.4	0.0502	0.0557	0.0494	0.0502	0.0548	0.0493
	2.0	0.0494	0.0548	0.0486	0.0500	0.0552	0.0501
	2.5	0.0503	0.0552	0.0490	0.0497	0.0546	0.0494
0.1	0.1	0.1007	0.1065	0.0997	0.1004	0.1047	0.0993
	0.5	0.1001	0.1064	0.0990	0.1005	0.1057	0.0996
	1.0	0.1000	0.1069	0.0998	0.0994	0.1051	0.0997
	1.4	0.1003	0.1064	0.0994	0.0995	0.1054	0.0995
	2.0	0.0999	0.1065	0.0996	0.1010	0.1065	0.1004
	2.5	0.1007	0.1068	0.1000	0.0993	0.1044	0.0989
			$\nu = 10$			$\nu = \infty$	
0.01	0.1	0.0103	0.0122	0.0101	0.0102	0.0119	0.0098
	0.5	0.0094	0.0120	0.0095	0.0103	0.0127	0.0104
	1.0	0.0101	0.0125	0.0100	0.0101	0.0129	0.0105
	1.4	0.0101	0.0123	0.0101	0.0103	0.0126	0.0105
	2.0	0.0101	0.0126	0.0104	0.0101	0.0124	0.0105
	2.5	0.0099	0.0120	0.0100	0.0103	0.0122	0.0104
0.05	0.1	0.0504	0.0540	0.0500	0.0500	0.0522	0.0489
	0.5	0.0499	0.0537	0.0492	0.0498	0.0530	0.0497
	1.0	0.0499	0.0552	0.0506	0.0501	0.0537	0.0504
	1.4	0.0503	0.0545	0.0503	0.0509	0.0542	0.0511
	2.0	0.0505	0.0557	0.0511	0.0502	0.0539	0.0509
	2.5	0.0495	0.0538	0.0498	0.0517	0.0548	0.0515
0.1	0.1	0.1002	0.1036	0.0992	0.1008	0.1035	0.1005
	0.5	0.1003	0.1048	0.1005	0.1010	0.1044	0.1012
	1.0	0.1000	0.1041	0.0997	0.1000	0.1032	0.1002
	1.4	0.0997	0.1040	0.0998	0.1015	0.1043	0.1013
	2.0	0.0992	0.1041	0.0999	0.1006	0.1038	0.1008
	2.5	0.1005	0.1050	0.1009	0.1007	0.1032	0.1002

Table 2: Expected shortfall and expected excess of reserves, in percentages of the theoretical VaR, computed over the 5,000 simulated paths of Table 1 for $\nu = 7$.

α	a	$p_-^{VaR(\theta_0)}$	$p_-^{VaR_n}$	$p_-^{EVaR_n}$	$p_+^{VaR(\theta_0)}$	$p_+^{VaR_n}$	$p_+^{EVaR_n}$
0.05	0.1	37.09	37.30	37.50	106.99	106.77	108.37
	0.5	37.68	37.19	37.43	107.17	106.76	108.33
	1.0	36.43	36.13	36.42	107.19	106.95	108.50
	1.4	36.18	35.90	35.90	107.25	107.06	108.62
	2.0	37.23	36.97	36.88	107.13	106.94	108.53
	2.5	37.18	37.11	36.95	107.10	106.93	108.53
0.1	0.1	48.58	48.79	48.90	116.51	116.35	117.45
	0.5	49.00	49.08	48.99	116.44	116.44	117.47
	1.0	49.27	49.11	49.14	116.59	116.60	117.64
	1.4	49.01	49.27	49.12	116.57	116.41	117.48
	2.0	49.18	49.54	49.50	116.87	116.69	117.73
	2.5	49.65	49.62	49.48	116.41	116.29	117.37

Table 3: As Table 1 for $n = 200$.

α	a	VaR_t	$\mathrm{VaR}_{n,t}$	$\mathrm{EVaR}_{n,t}$	VaR_t	$\mathrm{VaR}_{n,t}$	$\mathrm{EVaR}_{n,t}$
			$\nu = 7$			$\nu = \infty$	
0.05	0.1	0.0502	0.0525	0.0497	0.0494	0.0508	0.0494
	0.5	0.0501	0.0530	0.0503	0.0499	0.0517	0.0500
	1.0	0.0500	0.0529	0.0501	0.0498	0.0511	0.0494
	1.4	0.0494	0.0524	0.0498	0.0502	0.0517	0.0502
	2.0	0.0506	0.0526	0.0501	0.0493	0.0512	0.0496
	2.5	0.0497	0.0522	0.0495	0.0504	0.0517	0.0502
0.1	0.1	0.1004	0.1032	0.1002	0.0988	0.1012	0.0998
	0.5	0.0996	0.1030	0.1002	0.1007	0.1024	0.1008
	1.0	0.0999	0.1025	0.0997	0.0994	0.1019	0.1001
	1.4	0.1001	0.1032	0.1003	0.0988	0.1006	0.0991
	2.0	0.0984	0.1013	0.0986	0.0995	0.1006	0.0992
	2.5	0.1011	0.1041	0.1012	0.0991	0.1008	0.0993

Table 4: As Table 1 for $n = H = 250$.

α	a	VaR_t	$\mathrm{VaR}_{n,t}$	$\mathrm{EVaR}_{n,t}$	VaR_t	$\mathrm{VaR}_{n,t}$	$\mathrm{EVaR}_{n,t}$
			$\nu = 7$			$\nu = \infty$	
0.05	0.1	0.0501	0.0518	0.0497	0.0496	0.0508	0.0495
	0.5	0.0498	0.0519	0.0498	0.0502	0.0517	0.0504
	1.0	0.0498	0.0519	0.0498	0.0498	0.0515	0.0502
	1.4	0.0502	0.0521	0.0500	0.0503	0.0514	0.0501
	2.0	0.0500	0.0520	0.0499	0.0499	0.0513	0.0500
	2.5	0.0502	0.0524	0.0503	0.0497	0.0512	0.0499
0.1	0.1	0.1003	0.1022	0.1000	0.0998	0.1012	0.0999
	0.5	0.1001	0.1024	0.1001	0.0994	0.1010	0.0998
	1.0	0.1000	0.1024	0.1002	0.1000	0.1015	0.1003
	1.4	0.1000	0.1021	0.0998	0.1000	0.1014	0.1002
	2.0	0.1001	0.1026	0.1003	0.1003	0.1014	0.1002
	2.5	0.1003	0.1026	0.1004	0.1001	0.1014	0.1002

Table 5: Percentages of violations using Hansen's estimator, computed over the $5,000$ independent simulations of Table 1. Percentages are underlined when they are closer to the nominal probability than those obtained with the EVaR estimator in Table 1.

α	a	$\nu = 6$	$\nu = 7$	$\nu = 10$	$\nu = \infty$
0.01	0.1	$\underline{0.0097}$	0.0099	0.0104	$\underline{0.0101}$
	0.5	0.0104	0.0103	$\underline{0.0098}$	0.0108
	1.0	0.0105	$\underline{0.0100}$	0.0103	0.0108
	1.4	0.0106	$\underline{0.0098}$	0.0104	0.0109
	2.0	$\underline{0.0099}$	0.0102	0.0108	0.0108
	2.5	$\underline{0.0101}$	0.0105	0.0102	0.0106
0.05	0.1	$\underline{0.0499}$	$\underline{0.0496}$	0.0509	$\underline{0.0498}$
	0.5	$\underline{0.0504}$	$\underline{0.0505}$	$\underline{0.0503}$	0.0505
	1.0	0.0509	0.0509	0.0518	0.0513
	1.4	0.0510	0.0507	0.0514	0.0519
	2.0	$\underline{0.0501}$	0.0515	0.0522	0.0517
	2.5	$\underline{0.0506}$	0.0508	0.0508	0.0524
0.1	0.1	0.1018	0.1012	$\underline{0.1007}$	0.1016
	0.5	0.1017	0.1017	0.1020	0.1024
	1.0	0.1023	0.1015	0.1013	0.1012
	1.4	0.1018	0.1015	0.1011	0.1024
	2.0	0.1020	0.1024	0.1014	0.1019
	2.5	0.1022	$\underline{0.1005}$	0.1025	0.1014

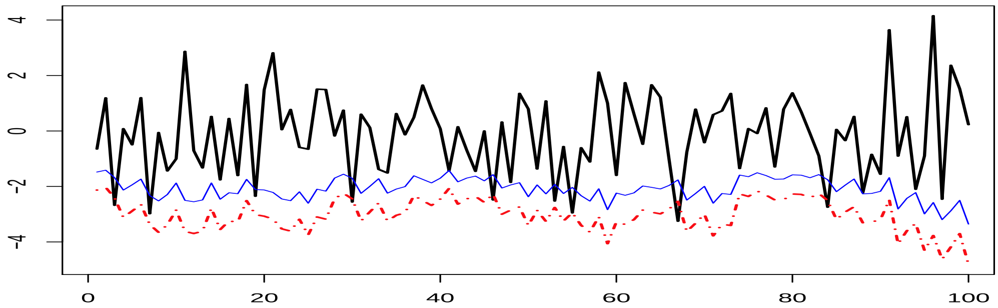

Figure 2: Log-returns of the SP500 and estimated $-$EVaR's at the 1% and 5% levels, from October 30, 2007 to March 25, 2008.

Figure 2 displays the series of the log-returns and the estimated EVaR's at the 1% and 5% levels, while Figure 3 displays the estimated VaR and EVaR at the 1% level, for all dates starting from September 16, 1998, assuming a Gaussian distribution for η_t (that is, choosing for H the identity function). The estimated adjusted VaR's are always larger than the standard estimated VaR's, which is not surprising in view of Corollary 1. The difference between the two estimated VaR's can be very large in more volatile periods, that is, in the more risky periods, with significant consequences in terms of required capital. This is seen more clearly in Figure 4, which shows the difference series $\mathrm{EVaR}_{n,t}(0.01) - \mathrm{VaR}_{n,t}(0.01)$. The proportions of log-returns that are below $\mathrm{VaR}_{n,t}(0.01)$ and $\mathrm{VaR}_{n,t}(0.05)$ are respectively equal to 98.45% and 94.74%. With the Estimation adjusted VaR, the proportions are equal to 98.79% and 95.15%, respectively, indicating the better coverage. Applying standard unconditional and independence backtesting procedures (see e.g. Berkowitz, Christoffersen and Pelletier (2009)) lead to accept the assumption that the violations form a martingale difference sequence with both estimators. Now, if we consider the mean of the differences between the returns and the VaR, when this difference is positive, that is the Expected Shortfall, we get 0.64 for the standard VaR(0.01) and 0.60 for the EVaR(0.01).

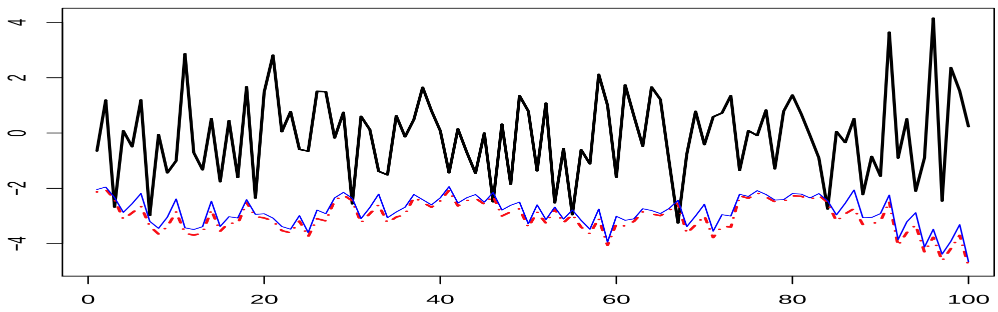

Figure 3: Log-returns of the SP500 (thick line), estimated −VaR (blue thin line) and estimated −EVaR (red dashed line) at the 1% level, from October 30, 2007 to March 25, 2008.

5 Concluding remarks

The substitution of an estimate to the unknown parameter value in the expression of the theoretical VaR can imply a bias in the coverage probability of the reserve, and a significant underestimation of the required capital. In the current regulation, this problem is circumvented in a rather ad hoc way by introducing an additional reserve to hedge the so-called estimation risk. However, the treatments of market and estimation risks are performed separately, without taking into account the possible dependence between these risks.

In this paper, we developed an alternative approach, which consists in jointly considering the two types of risks, by simply introducing an estimation adjustment to the VaR. This adjustment involves closed-form formulas, which were illustrated in the case of stochastic drift-volatility models. It represents a convenient approach compared to the numerical estimation of the residual quantile by bootstrap.

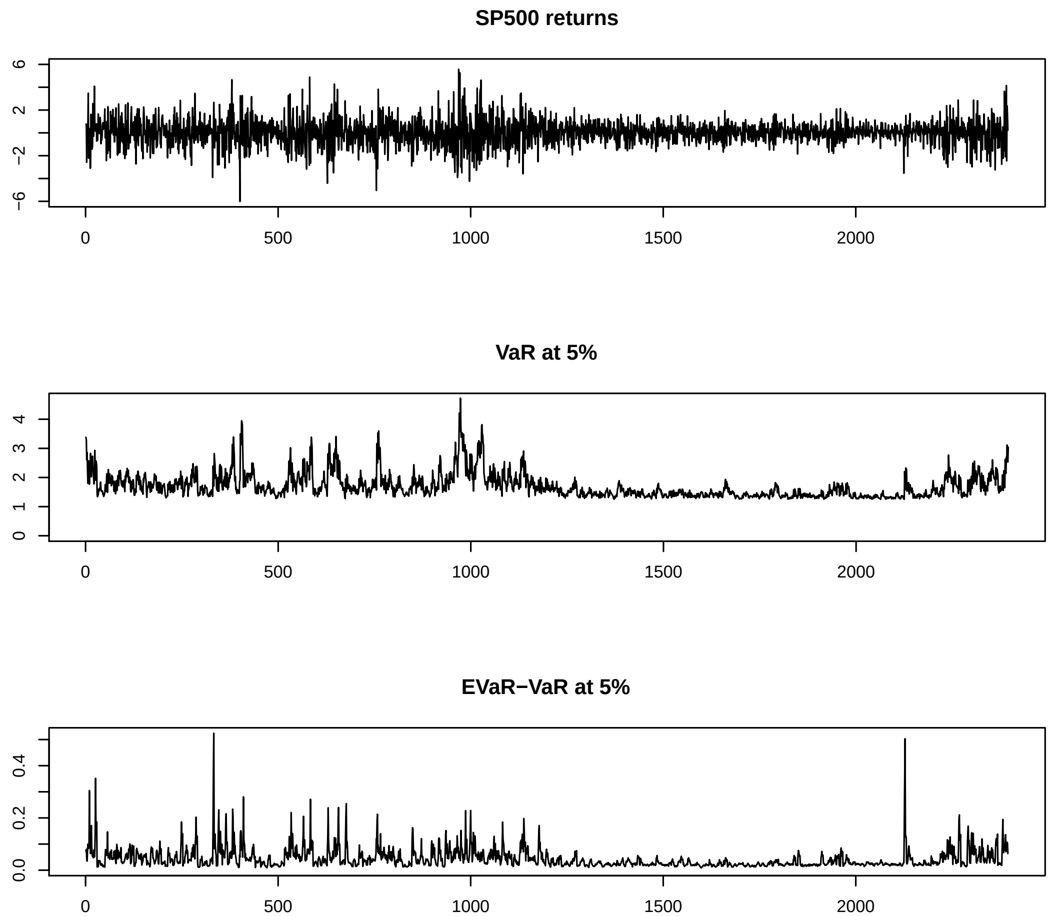

Figure 4: Log-returns of the SP500, estimated VaR's and difference $\mathrm{EVaR}_{n,t}$-$\mathrm{VaR}_{n,t}$ at the 5% level, from September 16, 1998 to March 25, 2008.

Appendices

A.1 Second-order Bahadur's expansion

Proposition 4 *Consider a sequence of one-dimensional continuous distributions with cdf F_n converging to a cdf F, and positive densities f_n converging to a positive density f, as n goes to infinity. Also assume that the first-order derivative of f_n converges to that of f. Let G_n (resp. G) denote the quantile function of F_n (resp. F). Then the following expansion holds, for $u \in (0,1)$,*

$$\begin{aligned} & G_n(u) - G(u) \\ = \; & -\left(\frac{F_n - F}{f}\right)[G(u)] + \left(\frac{F_n - F}{f}\right)[G(u)]\left(\frac{f_n - f}{f}\right)[G(u)] \\ & -\frac{1}{2}\frac{\partial \log f}{\partial x}[G(u)]\left[\left(\frac{F_n - F}{f}\right)[G(u)]\right]^2 \\ & + o\{(F_n - F)^2[G(u)]\} + o\{(F_n - F)[G(u)](f_n - f)[G(u)]\}. \end{aligned}$$

Proof. *i) First-order expansion.* The assumptions of continuous distributions with strictly positive densities entails that $F_n\{G_n(u)\} = F\{G(u)\} = u$, for all $u \in (0,1)$. Hence

$$\begin{aligned} F[G_n(u)] - F_n[G_n(u)] &= F[G_n(u)] - F[G(u)] \\ &= f[G(u)][G_n(u) - G(u)] + o[(G_n - G)(u)]. \end{aligned}$$

Moreover,

$$F(G_n(u)) = F[G(u)] + [G_n(u) - G(u)]f[G(u)] + o[(G_n - G)(u)], \quad (28)$$

$$F_n(G_n(u)) = F_n[G(u)] + [G_n(u) - G(u)]f_n[G(u)] + o[(G_n - G)(u)]. \quad (29)$$

Thus, we have

$$\begin{aligned} & F[G(u)] - F_n[G(u)] + [G_n(u) - G(u)]\{f[G(u)] - f_n[G(u)]\} \\ = \; & f[G(u)][G_n(u) - G(u)] + o[(G_n - G)(u)]. \end{aligned}$$

Since f_n converges to f, we deduce:

$$G_n(u) - G(u) = \frac{(F - F_n)[G(u)]}{f[G(u)]} + o[(G_n - G)(u)]. \quad (30)$$

ii) Second-order expansion. By similar arguments, the existence of quantile density functions g_n and g, defined as the derivatives of G_n and G, entails

$$(F_n - F)[G(u)] = \frac{(G - G_n)(u)}{g(u)} + o\{(F_n - F)[G(u)]\}.$$

This shows that $o\{(F_n - F)[G(u)]\}) = o[(G_n - G)(u)]$, allowing to write (30) as

$$G_n(u) - G(u) = \frac{(F - F_n)[G(u)]}{f[G(u)]} + o\{(F_n - F)[G(u)]\}. \quad (31)$$

A second-order expansion is similarly obtained and is given by

$$\begin{aligned} F[G_n(u)] - F_n[G_n(u)] &= F[G_n(u)] - F[G(u)] \\ &= f[G(u)][(G_n - G)(u)] \\ &\quad + \frac{1}{2}\frac{\partial f}{\partial x}[G(u)][(G_n - G)(u)]^2 + o[(G_n - G)(u)]^2. \end{aligned}$$

It follows that

$$\begin{aligned} & G_n(u) - G(u) \\ = \; & [f[G(u)]]^{-1}\left(1 - \frac{1}{2}\frac{\partial \log f}{\partial x}[G(u)][(G_n - G)(u)] + o[(G_n - G)(u)]\right) \\ & \times\{F[G_n(u)] - F_n[G_n(u)]\}. \end{aligned} \tag{32}$$

Next, similar to (28)-(29),

$$\begin{aligned} F(G_n(u)) &= F[G(u)] + [G_n(u) - G(u)]f[G(u)] + \frac{1}{2}\frac{\partial f}{\partial x}[G(u)][G_n(u) - G(u)]^2 \\ &\quad + o[(G_n - G)^2(u)], \\ F_n(G_n(u)) &= F_n[G(u)] + [G_n(u) - G(u)]f_n[G(u)] + \frac{1}{2}\frac{\partial f_n}{\partial x}[G(u)][G_n(u) - G(u)]^2 \\ &\quad + o[(G_n - G)^2(u)]. \end{aligned}$$

Thus

$$\begin{aligned} (F - F_n)[G_n(u)] &= F[G(u)] - F_n[G(u)] + \{f[G(u)] - f_n[G(u)]\}(G_n(u) - G(u)) \\ &\quad + o[(G_n - G)^2(u)]. \end{aligned}$$

Thus, using (32) and (31), the conclusion follows from:

$$\begin{aligned} & G_n(u) - G(u) \\ = \; & [f[G(u)]]^{-1}\left(1 - \frac{1}{2}\frac{\partial \log f}{\partial x}[G(u)][(G_n - G)(u)] + o[(G_n - G)(u)]\right) \\ & \times\left[(F - F_n)[G(u)] + (G_n - G)(u)\{(f - f_n)[G(u)]\} + o[(G_n - G)^2(u)]\right] \\ = \; & [f[G(u)]]^{-1}\left(1 - \frac{1}{2}\frac{\partial \log f}{\partial x}[G(u)][(G_n - G)(u)] + o[(G_n - G)(u)]\right) \\ & \times\Bigg[(F - F_n)[G(u)] + \left(\frac{(F - F_n)[G(u)]}{f[G(u)]} + o\{(F_n - F)[G(u)]\}\right) \\ & \times\{(f - f_n)[G(u)]\} + o[(G_n - G)^2(u)]\Bigg]. \end{aligned}$$

□

A.2 Proof of Lemma 1.

We shall prove the following intermediate results.

$$i) \qquad -\frac{F_n(z)-F(z)}{f(z)} = E(W_n \mid X=z) - \frac{1}{2}E\left(W_n^2 \frac{\partial \log f_{X,W_n}(z,W_n)}{\partial x} \mid X=z\right) + o[E(W_n^2 \mid X=z)],$$

$$ii) \qquad E\left(W_n^2 \frac{\partial \log f_{X,W_n}(z,W_n)}{\partial x} \mid X=z\right) = \frac{\partial \log f(z)}{\partial z} E\left(W_n^2 \mid X=z\right) + \left[\frac{\partial}{\partial x} E\left(W_n^2 \mid X=x\right)\right]_{x=z},$$

$$iii) \qquad \frac{f(z)-f_n(z)}{f(z)} = \frac{\partial}{\partial z}E(W_n \mid X=z) + \left(\frac{\partial \log f(z)}{\partial z}\right) E(W_n \mid X=z) + o[E(|W_n| \mid X=z)].$$

To prove i) we note that

$$\begin{aligned} F(z)-F_n(z) &= P(X<z) - P(X+W_n<z) \\ &= \int \left(\int_{z-w}^{z} f_{X,W_n}(x,w)dx\right) dw. \end{aligned} \tag{33}$$

Hence, by a Taylor expansion of the joint pdf with remainder in the integral form we get

$$\begin{aligned} F(z)-F_n(z) &= \int \left(\int_{z-w}^{z} f_{X,W_n}(z,w)dx\right) dw \\ &\quad + \int \left(\int_{z-w}^{z} \{f_{X,W_n}(x,w) - f_{X,W_n}(z,w)\}dx\right) dw \\ &= \int w f_{X,W_n}(z,w)dw + \int \frac{\partial f_{X,W_n}}{\partial x}(z,w) \left(\int_{z-w}^{z} (x-z)dx\right) dw \\ &\quad + \int \left(\int_{z-w}^{z} \int_{z}^{x} \frac{\partial^2 f_{X,W_n}}{\partial x^2}(u,w)(x-u)dudx\right) dw, \end{aligned}$$

By inverting the integrals in u and x, the latter term can be written as

$$\frac{1}{2}\int \left(\int_0^w \frac{\partial^2 f_{X,W_n}}{\partial x^2}(z-w+x,w)x^2dx\right) dw$$

and is bounded in absolute value by

$$
\begin{aligned}
& \frac{1}{2}\int w^2 \left(\int_0^{|w|} \left| \frac{\partial^2 f_{X,W_n}}{\partial x^2}(z+x,w) \right| dx \right) dw \\
= \quad & \frac{f(z)}{2} \int w^2 f_{W_n|X=z}(w) \left(\frac{1}{f_{X,W_n}(z,w)} \int_0^{|w|} \left| \frac{\partial^2 f_{X,W_n}}{\partial x^2}(z+x,w) \right| dx \right) dw \\
\leq \quad & \frac{f(z)}{2} E\{W_n^2 C(W_n,z) \mid X=z\} \\
\leq \quad & \frac{f(z)}{2} E(W_n^{2(1+\nu)} \mid X=z) E\{C(z,W_n)^{1+1/\nu} \mid X=z\},
\end{aligned}
$$

by Hölder's inequality, and using the decomposition of the joint density $f_{X,W_n}(z,w) = f_{W_n|X=z}(w)f(z)$ into the product of the conditional density of W_n given $X=z$ and the marginal density of X. In view of (10) and (11) the conclusion of i) follows.

Next, we show ii). We have

$$
\begin{aligned}
& E\left(W_n^2 \frac{\partial \log f_{X,W_n}(z,W_n)}{\partial x} \mid X=z \right) \\
= \quad & \frac{\partial \log f(z)}{\partial x} E\left(W_n^2 \mid X=z\right) + \int w^2 \left(\frac{\partial}{\partial x} f_{W_n|X=x}(w) \right)_{x=z} dw \\
= \quad & \frac{\partial \log f(z)}{\partial x} E\left(W_n^2 \mid X=z\right) + \frac{\partial}{\partial x} \left(\int w^2 f_{W_n|X=x}(w) dw \right)_{x=z},
\end{aligned}
$$

and ii) is proven.

Now we turn to iii). By differentiation of (33) and another Taylor expansion we obtain, for some z^* between $z-w$ and z,

$$
\begin{aligned}
f(z) - f_n(z) \quad &= \quad \int \{f_{X,W_n}(z,w) - f_{X,W_n}(z-w,w)\} dw, \\
&= \quad \int \frac{\partial f_{X,W_n}}{\partial z}(z,w) w dw - \int \int_z^{z-w} \frac{\partial^2 f_{X,W_n}}{\partial x^2}(u,w)(z-w-u) du dw \\
&:= \quad I_1 + I_2.
\end{aligned}
$$

We have

$$
\begin{aligned}
I_1 \quad &= \quad f(z) \frac{\partial}{\partial z} \left(\int f_{W_n|X=z}(w) w dw \right) + \frac{\partial f}{\partial z}(z) \int f_{W_n|X=z}(w) w dw \\
&= \quad f(z) \left\{ \frac{\partial}{\partial z} E(W_n \mid X=z) + \left(\frac{\partial \log f(z)}{\partial z} \right) E(W_n \mid X=z) \right\}.
\end{aligned}
$$

Moreover, for $w > 0$, by arguments used to prove i),

$$\begin{aligned}
|I_2| &\le \int w \int_{z-w}^{z} \left| \frac{\partial^2 f_{X,W_n}}{\partial x^2}(u,w) \right| dudw \\
&= \int w \int_{-w}^{0} \left| \frac{\partial^2 f_{X,W_n}}{\partial x^2}(x+z,w) \right| dxdw \\
&\le f(z) E\{|W_n| C(W_n, z) \mid X = z\} \\
&\le f(z) E(|W_n|^{1+\nu} \mid X = z) E\{C(z, W_n)^{1+1/\nu} \mid X = z\}
\end{aligned}$$

In view of (9) and (11) the conclusion of iii) follows.

The first expansion in Lemma 1 is a consequence of i) and ii). To establish the second expansion, we use Proposition 4. By substituting the approximations in i)-iii) into Bahadur's expansion we have:

$$\begin{aligned}
& G_n(u) - G(u) \\
= \; & E[W_n \mid X = G(u)] \\
& + \Bigg(-\frac{1}{2} \frac{\partial \log f(z)}{\partial z} E(W_n^2 \mid X = z) - \frac{1}{2} \frac{\partial}{\partial z} E(W_n^2 \mid X = z) \\
& + E(W_n \mid X = z) \frac{\partial}{\partial z} E(W_n \mid X = z) + \frac{\partial \log f(z)}{\partial z} \{E(W_n \mid X = z)\}^2 \\
& - \frac{1}{2} \frac{\partial \log f(z)}{\partial z} \{E(W_n \mid X = z)\}^2 \Bigg)_{z=G(u)} + o[\{E[W_n \mid X = G(u)]\}^2],
\end{aligned}$$

from which the conclusion follows. $\square$

A.3 Illustration of Lemma 1 for Gaussian variables

Suppose that (X, W_n) is a Gaussian vector with

$$X \sim \mathcal{N}(m, \sigma^2), \quad W_n \sim \mathcal{N}(\tau_n, \xi_n^2), \quad \mathrm{Cov}(X, W_n) = \rho_n \sigma \xi_n,$$

where $\rho_n \in [0, 1)$, $\tau_n \to 0$ and $\xi_n^2 \to 0$ as $n \to \infty$. Then we have

$$\begin{aligned}
F_n(z) &= \Phi\left(\frac{z - m - \tau_n}{\sqrt{\sigma^2 + \xi_n^2 + 2\rho_n \sigma \xi_n}} \right) := \Phi(z_n), \\
F(z) &= \Phi\left(\frac{z - m}{\sigma} \right), \\
G_n(u) &= m + \tau_n + \sqrt{\sigma^2 + \xi_n^2 + 2\rho_n \sigma \xi_n}\, \Phi^{-1}(u), \qquad G(u) = m + \sigma \Phi^{-1}(u).
\end{aligned}$$

Given $X = z$ the distribution of W_n is $\mathcal{N}\left(\tau_n + \rho_n \xi_n \frac{z-m}{\sigma}, (1 - \rho_n^2)\xi_n^2\right)$. Thus,

$$\begin{aligned}
E(W_n \mid X = z) &= \tau_n + \rho_n \xi_n \frac{(z - m)}{\sigma}, \qquad \mathrm{Var}(W_n \mid X = z) = (1 - \rho_n^2)\xi_n^2, \\
E(W_n^2 \mid X = z) &= \left(\tau_n + \rho_n \xi_n \frac{z - m}{\sigma} \right)^2 + (1 - \rho_n^2)\xi_n^2.
\end{aligned}$$

Moreover, $\frac{\partial \log f(x)}{\partial x} = -\frac{x-m}{\sigma^2}$. Hence,

$$
\begin{aligned}
& G_n(u) - G(u) - E[W_n \mid X = G(u)] \\
& + \frac{1}{2}\left\{\frac{\partial}{\partial x}\mathrm{Var}[W_n \mid X = x] + \mathrm{Var}(W_n \mid X = x)\frac{\partial \log f(x)}{\partial x}\right\}_{x=G(u)} \\
= & \left\{(\sqrt{\sigma^2 + \xi_n^2 + 2\rho_n\sigma\xi_n} - \sigma) - \rho_n\xi_n - \frac{1}{2\sigma}(1-\rho_n^2)\xi_n^2\right\}\Phi^{-1}(u) \\
= & \; o(\xi_n^2).
\end{aligned}
$$

Now, letting $a_n = \xi_n/\sigma$, a Taylor expansion yields:

$$
\begin{aligned}
& -\frac{F_n(z) - F(z)}{f(z)} \\
= & -\sigma(z_n - z_0) + \frac{1}{2}(z_n - z_0)^2(z-m) + o(z_n - z_0)^2 \\
= & \; \tau_n(1 - \rho_n a_n) + (z-m)\left\{\rho_n a_n + \frac{1}{2}a_n^2(1 - 3\rho_n^2) + \frac{\tau_n^2}{2\sigma^2}\right\} \\
& + \frac{(z-m)^2}{\sigma^2}\rho_n a_n \tau_n + \frac{(z-m)^3}{2\sigma^2}\rho_n^2 a_n^2 + o(a_n^2 + \tau_n^2 + \tau_n a_n) \\
= & -E[W_n \mid X = z] + \frac{1}{2}\left\{\frac{\partial}{\partial x}E[W_n^2 \mid X = x] + E(W_n^2 \mid X = x)\frac{\partial \log f(x)}{\partial x}\right\}_{x=z} \\
& + o\{E[W_n^2 \mid X = z]\}.
\end{aligned}
$$

The result of Lemma 1 is then verified by direct computation in the Gaussian case.

Finally, it can be checked that the assumptions of this lemma, in particular (11), are satisfied in this case.

A.4 Technical assumptions used in Proposition 1

Let us denote

$$
\begin{aligned}
Z_1(u, y_{t-1}) &= \left(\frac{\partial a}{\partial \theta'}(y_{t-1}, \theta, g(y_{t-1}, \theta_0, u))\right)_{\theta=\theta_0}, \\
Z_2(u, y_{t-1}) &= \left(\frac{\partial^2 a}{\partial \theta \partial \theta'}(y_{t-1}, \theta, g(y_{t-1}, \theta_0, u))\right)_{\theta=\theta_0}.
\end{aligned}
$$

We have

$$
W_{t,n} = Z_1(\varepsilon_t, y_{t-1})(\hat{\theta}_n - \theta_0) + \frac{1}{2}(\hat{\theta}_n - \theta_0)' Z_2(\varepsilon_t, y_{t-1})(\hat{\theta}_n - \theta_0).
$$

The following assumptions are made in Proposition 1. For some $\nu > 0$,

(a) $E[\|Z_i(\varepsilon_t, y_{t-1})\|^{2(1+\nu)} \mid y_{t-1}] < \infty$, for $i = 1, 2$.

(b) The functions $a(y_{t-1}, \theta, \cdot)$, $\dfrac{\partial a}{\partial \theta'}[y_{t-1}, \theta, \cdot]$ and $\dfrac{\partial^2 a}{\partial \theta \partial \theta'}[y_{t-1}, \theta, \cdot]$ are twice differentiable.

(c) The joint density of $(\varepsilon_t, W_{t,n})$ conditional on y_{t-1} satisfies condition (11).

A.5 Proof of Proposition 1

i) Let us derive the first expression for the quantile adjustment. The proof relies on expansion (7) and Lemma 1, with $X = \varepsilon_t$, $W_n = W_{t,n}$ and expectations replaced by expectations conditional on y_{t-1}.

Hence, for any multiplicative norm,

$$\begin{aligned} |W_{t,n}|^{2(1+\nu)} \quad \leq \quad & \frac{2^{2(1+\nu)}}{n^{1+\nu}} \|Z_1(\varepsilon_t, y_{t-1})\|^{2(1+\nu)} \|\sqrt{n}(\hat{\theta}_n - \theta_0)\|^{2(1+\nu)} \\ & + \|Z_2(\varepsilon_t, y_{t-1})\|^{2(1+\nu)} \|\sqrt{n}(\hat{\theta}_n - \theta_0)\|^{4(1+\nu)}. \end{aligned}$$

Thus:

$$\begin{aligned} & E[|W_{t,n}|^{2(1+\nu)} \mid y_{t-1}] \\ \leq \quad & \frac{2^{2(1+\nu)}}{n^{1+\nu}} E[\|Z_1(\varepsilon_t, y_{t-1})\|^{2(1+\nu)} \mid y_{t-1}] E[\|\sqrt{n}(\hat{\theta}_n - \theta_0)\|^{2(1+\nu)} \mid y_{t-1}] \\ & + \frac{1}{n^{2(1+\nu)}} E[\|Z_2(\varepsilon_t, y_{t-1})\|^{2(1+\nu)} \mid y_{t-1}] E[\|\sqrt{n}(\hat{\theta}_n - \theta_0)\|^{4(1+\nu)} \mid y_{t-1}], \end{aligned}$$

which is finite by Assumptions (a), in Appendix A.4, and (14).

Now, since the conditional moments of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ given y_{t-1} and ε_t do not depend on the latter, it is clear that the assumptions on terms of the same stochastic orders in Lemma 1 are satisfied. Moreover, we have:

$$\begin{aligned} & E[|W_{t,n}|^{1+\nu} \mid \varepsilon_t, y_{t-1}] \\ \leq \quad & \frac{2^{1+\nu}}{n^{(1+\nu)/2}} \|Z_1(\varepsilon_t, y_{t-1})\|^{2(1+\nu)} E[\|\sqrt{n}(\hat{\theta}_n - \theta_0)\|^{2(1+\nu)} \mid \varepsilon_t, y_{t-1}] \\ & + \frac{1}{n^{1+\nu}} \|Z_2(\varepsilon_t, y_{t-1})\|^{2(1+\nu)} E[\|\sqrt{n}(\hat{\theta}_n - \theta_0)\|^{4(1+\nu)} \mid \varepsilon_t, y_{t-1}] = o_p(n^{-1/2}), \end{aligned}$$

and similarly, $E[|W_{t,n}|^{2(1+\nu)} \mid \varepsilon_t, y_{t-1}] = o_p(n^{-1})$, which establishes (9) and (10). Thus the assumptions of Lemma 1.

Next we turn to the computation of the conditional quantile of $\hat{\varepsilon}_t$ given y_{t-1}. By

$$\begin{aligned} E[W_{t,n} \mid \varepsilon_t, y_{t-1}] \quad = \quad & Z_1(\varepsilon_t, y_{t-1}) E[\hat{\theta}_n - \theta_0 \mid \varepsilon_t, y_{t-1}] \\ & + \frac{1}{2} \mathrm{Tr} \left[Z_2(\varepsilon_t, y_{t-1}) E[(\hat{\theta}_n - \theta_0)(\hat{\theta}_n - \theta_0)' \mid \varepsilon_t, y_{t-1}] \right], \\ \mathrm{Var}[W_{t,n} \mid \varepsilon_t, y_{t-1}] \quad = \quad & \frac{1}{n} Z_1(\varepsilon_t, y_{t-1}) \mathrm{Var}[\sqrt{n}(\hat{\theta}_n - \theta_0) \mid \varepsilon_t, y_{t-1}] Z_1(\varepsilon_t, y_{t-1})' \\ & + o_P(1/n), \end{aligned}$$

we get:

$$\begin{aligned}
& q_{\hat{\varepsilon}_t}(\alpha,\theta_0) - \Phi^{-1}(\alpha) \\
= \quad & E[W_{t,n} \mid \varepsilon_t, y_{t-1}] \\
& -\frac{1}{2}\left\{\frac{\partial}{\partial x}\mathrm{Var}[W_{t,n} \mid \varepsilon_t = x, y_{t-1}] - x\mathrm{Var}(W_{t,n} \mid \varepsilon_t = x, y_{t-1})\right\}_{x=\Phi^{-1}(\alpha)} \\
& + o_P\left(1/n\right).
\end{aligned}$$

The announced result follows.

ii) The second expression for $b_t(\theta_0,\varepsilon)$ is a consequence of the following links between partial derivatives. To simplify notations, write $a(y_{t-1},\theta,y) = a(\theta,y)$ and $g(y_{t-1},\theta,x) = g(\theta,x)$. By assumption, $y = g(\theta,x) \Leftrightarrow x = a(\theta,y)$. It follows from $y = g(\theta,a(\theta,y))$ and $x = a(\theta,g(\theta,x))$ that

$$\begin{aligned}
\frac{\partial g}{\partial \theta}(\theta,a(\theta,y)) + \frac{\partial g}{\partial x}(\theta,a(\theta,y))\frac{\partial a}{\partial \theta}(\theta,y) &= 0, \\
\frac{\partial a}{\partial \theta}(\theta,g(\theta,x)) + \frac{\partial a}{\partial y}(\theta,g(\theta,x))\frac{\partial g}{\partial \theta}(\theta,x) &= 0, \qquad (34)
\end{aligned}$$

and

$$\begin{aligned}
& \frac{\partial^2 g}{\partial\theta\partial\theta'}(\theta,a(\theta,y)) + \frac{\partial^2 g}{\partial x\partial\theta}(\theta,a(\theta,y))\frac{\partial a}{\partial\theta'}(\theta,y) + \frac{\partial a}{\partial\theta}(\theta,y)\frac{\partial^2 g}{\partial x\partial\theta'}(\theta,a(\theta,y)) \\
& +\frac{\partial^2 g}{\partial x^2}(\theta,a(\theta,y))\frac{\partial a}{\partial\theta}(\theta,y)\frac{\partial a}{\partial\theta'}(\theta,y) + \frac{\partial g}{\partial x}(\theta,a(\theta,y))\frac{\partial^2 a}{\partial\theta\partial\theta'}(\theta,y) \\
= \quad & 0, \qquad (35) \\
& \frac{\partial^2 a}{\partial\theta\partial\theta'}(\theta,g(\theta,x)) + \frac{\partial^2 a}{\partial y\partial\theta}(\theta,g(\theta,x))\frac{\partial g}{\partial\theta'}(\theta,x) + \frac{\partial g}{\partial\theta}(\theta,x)\frac{\partial^2 a}{\partial y\partial\theta'}(\theta,g(\theta,x)) \\
& +\frac{\partial^2 a}{\partial y^2}(\theta,g(\theta,x))\frac{\partial g}{\partial\theta}(\theta,x)\frac{\partial g}{\partial\theta'}(\theta,x) + \frac{\partial a}{\partial y}(\theta,g(\theta,x))\frac{\partial^2 g}{\partial\theta\partial\theta'}(\theta,x) \\
= \quad & 0. \qquad (36)
\end{aligned}$$

Now, in view of (36) we have

$$\begin{aligned}
& \frac{\partial a}{\partial y}(\theta_0,g(\theta,\varepsilon))Tr\left(\Omega\frac{\partial^2 g}{\partial\theta\partial\theta'}(\theta,\varepsilon)\right) \\
= \quad & -Tr\left(\Omega\frac{\partial^2 a}{\partial\theta\partial\theta'}(\theta_0,g(\theta,\varepsilon))\right) - 2\frac{\partial g}{\partial\theta'}(\theta,\varepsilon)\Omega\frac{\partial^2 a}{\partial\theta\partial y}(\theta_0,g(\theta,\varepsilon)) \\
& -\frac{\partial^2 a}{\partial y^2}(\theta_0,g(\theta,\varepsilon))\frac{\partial g}{\partial\theta'}(\theta,\varepsilon)\Omega\frac{\partial g}{\partial\theta}(\theta,\varepsilon).
\end{aligned}$$

It follows, by using (34), that the right-hand side of (15) is given by:

$$
\begin{aligned}
& -\varepsilon \left(\frac{\partial a}{\partial y}(\theta_0, g(\theta, \varepsilon)) \right)^2 \frac{\partial g}{\partial \theta'}(\theta, \varepsilon)\Omega \frac{\partial g}{\partial \theta}(\theta, \varepsilon) \\
& -2\frac{\partial g}{\partial \theta'}(\theta, \varepsilon)\Omega \frac{\partial^2 a}{\partial \theta \partial y}(\theta_0, g(\theta, \varepsilon)) - Tr\left(\Omega \frac{\partial^2 a}{\partial \theta \partial \theta'}(\theta_0, g(\theta, \varepsilon)) \right) \\
= \quad & -\varepsilon \frac{\partial a}{\partial \theta'}(\theta_0, g(\theta, \varepsilon))\Omega \frac{\partial a}{\partial \theta}(\theta_0, g(\theta, \varepsilon)) \\
& +2\frac{\partial g}{\partial x}(\theta, \varepsilon)\frac{\partial a}{\partial \theta'}(\theta_0, g(\theta, \varepsilon))\Omega \frac{\partial^2 a}{\partial \theta \partial y}(\theta_0, g(\theta, \varepsilon)) \\
& -Tr\left(\Omega \frac{\partial^2 a}{\partial \theta \partial \theta'}(\theta_0, g(\theta, \varepsilon)) \right) = b_t\{\theta, \varepsilon\},
\end{aligned}
$$

and the proof of Proposition 1 is completed. □

A.6 The jacknife for bias reduction

To illustrate the use of the jacknife technique to remove the conditional bias of the estimator, consider the AR(1) model with intercept θ_0

$$
y_t = \theta_0 + \rho y_{t-1} + \varepsilon_t, \qquad |\theta_0| < 1, \qquad (\varepsilon_t) \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1) \tag{37}
$$

where, for simplicity, the autoregressive coefficient ρ is known. Given observations, $y_{-1}, \ldots, y_{-n-1}$, θ_0 can be consistently estimated by

$$
\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n}(y_{-i} - \rho y_{-i-1}) = \theta_0 + \frac{1}{n}\sum_{i=1}^{n}\varepsilon_{-i}.
$$

The expansion $y_t = \frac{\theta_0}{1-\rho} + \sum_{i=-t}^{\infty} \rho^{t+i}\varepsilon_{-i}$ shows that the joint distribution of $(y_t, \sqrt{n}(\hat{\theta}_n - \theta_0))$, for $t \geq 0$ is

$$
\mathcal{N}\left(\begin{pmatrix} m_y \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{1}{1-\rho^2} & \frac{1}{\sqrt{n}}\frac{1-\rho^n}{1-\rho}\rho^{t+1} \\ \frac{1}{\sqrt{n}}\frac{1-\rho^n}{1-\rho}\rho^{t+1} & 1 \end{pmatrix} \right), \quad m_y = \frac{\theta_0}{1-\rho},
$$

from which we deduce

$$
E(\sqrt{n}(\hat{\theta}_n - \theta_0) \mid y_t) \quad = \quad \frac{\rho^{t+1}}{\sqrt{n}}(1-\rho^n)(1+\rho)(y_t - m_y),
$$

which is of the form (16) with $A(t) = \rho^t$. Thus the conditional bias of the estimator is of order $1/n$ and Assumption (12) is not satisfied. We also have

$$
\mathrm{Var}[\sqrt{n}(\hat{\theta}_n - \theta_0) \mid y_t] = 1 - \left(\frac{1}{\sqrt{n}}(1-\rho^n)\rho^{t+1} \right)^2 = 1 + o_P(1),
$$

showing that Assumption (13) is satisfied. Because the distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ conditional on y_t is Gaussian, Assumption (14) is also satisfied. Finally, the technical assumptions displayed in Appendix A.4 are verified: we have $a(y_{t-1}, \theta, y_t) = y_t - \theta - \rho y_{t-1},\ Z_1(u, y_{t-1}) = -1, Z_2(u, y_{t-1}) = 0$ showing that (a) and (b) hold true; moreover, it can be verified that condition (11) is satisfied, by noting that $C(z, w) \leq (|w| + |z|)^2/\phi(z)$ and by using again the fact that the distribution of $\hat{\theta}_n - \theta_0$ conditional on y_{t-1} is Gaussian and thus admits moments at any order.

To adjust for the bias, a Jacknife estimator can be constructed. For simplicity assume that $n = 2\ell$ and let

$$\hat{\theta}_n^{(1)} = \frac{1}{\ell}\sum_{i=1}^{\ell}(y_{-i} - \rho y_{-i-1}), \quad \hat{\theta}_n^{(2)} = \frac{1}{\ell}\sum_{i=\ell+1}^{n}(y_{-i} - \rho y_{-i-1}).$$

A Jacknife estimator based on this subsampling scheme is given by

$$\hat{\theta}_n^{(J)} = \frac{1}{1-\rho^\ell}(\hat{\theta}_n^{(2)} - \rho^\ell \hat{\theta}_n^{(1)}).$$

Indeed, we have

$$E(\hat{\theta}_n^{(J)} - \theta_0 \mid y_t) = \frac{\rho^{t+1+\ell} - \rho^\ell \rho^{t+1}}{\ell(1-\rho^\ell)}(1-\rho^\ell)(1+\rho)(y_t - m_y) = 0.$$

The joint distribution of $(y_t, \sqrt{n}(\hat{\theta}_n^{(J)} - \theta_0))$, for $t \geq 0$ is

$$\mathcal{N}\left(\left(\begin{array}{c} m_y \\ 0 \end{array}\right), \left(\begin{array}{cc} \frac{1}{1-\rho^2} & 0 \\ 0 & 2\frac{1+\rho^{2\ell}}{(1-\rho^\ell)^2} \end{array}\right)\right),$$

showing that, in this model, the jacknife estimator is independent from the y_t's for $t \geq 0$. We also have

$$\mathrm{Var}[\sqrt{n}(\hat{\theta}_n^{(J)} - \theta_0) \mid y_t] = 2\frac{1+\rho^{2\ell}}{(1-\rho^\ell)^2} = 2 + o_P(1),$$

showing that the bias reduction comes at a price, namely an increase of the variance. Such a bias and variance trade-off in jacknife estimators has often be noted (see Phillips and Yu (2005) for a discussion of this issue in the context of option prices).

A.7 Proof of Proposition 3.

The conditional probability of violation with the standard VaR estimator given by (5) is

$$\begin{aligned} P_{t-1}[y_t < -\mathrm{VaR}_{n,t}(\alpha)] &= P_{t-1}[g(y_{t-1}; \hat{\theta}_n, \hat{\varepsilon}_t) < g(y_{t-1}; \hat{\theta}_n, \Phi^{-1}(\alpha))] \\ &= P_{t-1}[\hat{\varepsilon}_t < \Phi^{-1}(\alpha)], \end{aligned} \tag{38}$$

because g is strictly increasing in its last argument. Similarly, with the estimation adjusted VaR estimator defined in (17) we have

$$P_{t-1}[y_t < -\mathrm{EVaR}_{n,t}(\alpha)] \quad = \quad P_{t-1}[\hat{\varepsilon}_t < q_{\hat{\varepsilon}_t}(\alpha, \hat{\theta}_n)].$$

Lemma 1 entails the following expansion for the c.d.f. of $\hat{\varepsilon}_t$, keeping the notations of the proof of Proposition 1:

$$\begin{aligned}
& P_{t-1}[\hat{\varepsilon}_t < z] \\
= \; & \Phi(z) - E[W_{t,n} \mid \varepsilon_t, y_{t-1}]\phi(z) \\
& + \frac{1}{2}\left\{\frac{\partial}{\partial x}E[W_{t,n}^2 \mid \varepsilon_t = x, y_{t-1}] - xE(W_{t,n}^2 \mid \varepsilon_t = x, y_{t-1})\right\}_{x=z} \phi(z) + \; o_P\left(1/n\right) \\
= \; & \Phi(z) + \frac{b_t(\theta_0, z)}{2n}\phi(z) + \; o_P\left(1/n\right),
\end{aligned}$$

since we have seen that $E[W_{t,n} \mid \varepsilon_t, y_{t-1}]$ is of order $o_P\left(1/n\right)$. It follows that

$$P_{t-1}[\hat{\varepsilon}_t < \Phi^{-1}(\alpha)] \quad = \quad \alpha + \frac{b_t\{\theta_0, \Phi^{-1}(\alpha)\}}{2n}\phi\left(\Phi^{-1}(\alpha)\right) + \; o_P\left(1/n\right),$$

which, from (38), proves (20). Now, using a Taylor expansion of Φ around $\Phi^{-1}(\alpha)$,

$$\begin{aligned}
& P_{t-1}[\hat{\varepsilon}_t < q_{\hat{\varepsilon}_t}(\alpha, \hat{\theta}_n) \mid \hat{\theta}_n] \\
= \; & \Phi(q_{\hat{\varepsilon}_t}(\alpha, \hat{\theta}_n)) + \frac{b_t(\theta_0, q_{\hat{\varepsilon}_t}(\alpha, \hat{\theta}_n))}{2n}\phi(q_{\hat{\varepsilon}_t}(\alpha, \hat{\theta}_n)) + \; o_P\left(1/n\right) \\
= \; & \Phi(\Phi^{-1}(\alpha)) - \frac{b_t\{\hat{\theta}_n, \Phi^{-1}(\alpha)\}}{2n}\phi(\Phi^{-1}(\alpha)) + \frac{b_t\{\theta_0, \Phi^{-1}(\alpha)\}}{2n}\phi(\Phi^{-1}(\alpha)) \\
& + o_P\left(1/n\right) \\
= \; & \alpha + \; o_P\left(1/n\right)
\end{aligned}$$

in view of the convergence of $\hat{\theta}_n$ to θ_0. Thus (21) is proven. □

A.8 Another way to derive the coverage probability

i) Let us consider the type of proof followed by Hansen (2006) (proof of Theorem 2), which differs from that of Appendix A.7. We have, for any real variable $Z(\hat{\theta}_n)$

$$\begin{aligned}
P[y_t < Z(\hat{\theta}_n) \mid y_{t-1}] \quad & = \quad P[g(\theta_0, \varepsilon_t) < Z(\hat{\theta}_n) \mid y_{t-1}] \\
& = \quad P[\varepsilon_t < a(\theta_0, Z(\hat{\theta}_n)) \mid y_{t-1}] \\
& = \quad E\{P[\varepsilon_t < a(\theta_0, Z(\hat{\theta}_n)) \mid y_{t-1}, \hat{\theta}_n] \mid y_{t-1}\} \\
& = \quad E\{\Phi[a(\theta_0, Z(\hat{\theta}_n))] \mid y_{t-1}\}, \qquad (39)
\end{aligned}$$

using the simplified notation of Appendix A.5 *ii)*. Let $a(\theta_0, Z(\theta_0)) = u$. A Taylor expansion around θ_0 gives

$$
\begin{aligned}
& \Phi[a(\theta_0, Z(\hat{\theta}_n))] \\
= \;& \Phi(u) + \phi(u)\frac{\partial a}{\partial y}(\theta_0, Z(\theta_0))\frac{\partial Z}{\partial \theta'}(\theta_0)(\hat{\theta}_n - \theta_0) \\
& +\frac{1}{2}(\hat{\theta}_n - \theta_0)' \left[\left\{ \phi'(u)\left(\frac{\partial a}{\partial y}(\theta_0, Z(\theta_0))\right)^2 + \phi(u)\frac{\partial^2 a}{\partial y^2}(\theta_0, Z(\theta_0)) \right\} \frac{\partial Z}{\partial \theta}\frac{\partial Z}{\partial \theta'}(\theta_0) \right. \\
& \qquad \left. +\phi(u)\frac{\partial a}{\partial y}(\theta_0, Z(\theta_0))\frac{\partial^2 Z}{\partial \theta \partial \theta'}(\theta_0) \right] (\hat{\theta}_n - \theta_0) \\
& +o_p(\|\hat{\theta}_n - \theta_0\|^2).
\end{aligned}
$$

If $Z(\theta_0)$ is a function of y_{t-1}, it follows from (39) and Assumption (12) that

$$
\begin{aligned}
& P[y_t < Z(\hat{\theta}_n) \mid y_{t-1}] \\
= \;& \Phi(u) + \frac{\phi(u)}{2n}\left[\left\{ \frac{\partial^2 a}{\partial y^2}(\theta_0, Z(\theta_0)) - u\left(\frac{\partial a}{\partial y}(\theta_0, Z(\theta_0))\right)^2 \right\} \frac{\partial Z}{\partial \theta'}\Omega\frac{\partial Z}{\partial \theta}(\theta_0) \right. \\
& \qquad \left. +\frac{\partial a}{\partial y}(\theta_0, Z(\theta_0))Tr\left(\Omega\frac{\partial^2 Z}{\partial \theta \partial \theta'}(\theta_0)\right) \right] + o_p(1/n). \qquad (40)
\end{aligned}
$$

ii) In the special case of the VaR, we have

$$
Z(\hat{\theta}_n) = -\mathrm{VaR}_{n,t}(\alpha) = g[\hat{\theta}_n, \Phi^{-1}(\alpha)], \quad \frac{\partial Z}{\partial \theta}(\theta_0) = \frac{\partial g}{\partial \theta}[\theta_0, \Phi^{-1}(\alpha)], \quad u = \Phi^{-1}(\alpha)
$$

and we get (20), using formula (15) for $b_t\{\theta_0, \Phi^{-1}(\alpha)\}$. Similarly, we obtain (21).

iii) Comparison with Hansen's (2006) expansion for the coverage probability. Starting from (39), Hansen makes a Taylor expansion around $Z(\theta_0)$ (instead of θ_0). Doing this we get

$$
\begin{aligned}
& \Phi[a(\theta_0, Z(\hat{\theta}_n))] \\
= \;& \Phi(u) + \phi(u)\frac{\partial a}{\partial y}(\theta_0, Z(\theta_0))(Z(\hat{\theta}_n) - Z(\theta_0)) \\
& +\frac{\phi(u)}{2}\left\{ \frac{\partial^2 a}{\partial y^2}(\theta_0, Z(\theta_0)) - u\left(\frac{\partial a}{\partial y}(\theta_0, Z(\theta_0))\right)^2 \right\} (Z(\hat{\theta}_n) - Z(\theta_0))^2 \\
& +o_p([Z(\hat{\theta}_n) - Z(\theta_0)]^2).
\end{aligned}
$$

Assuming that

$$
\begin{aligned}
E[Z(\hat{\theta}_n) - Z(\theta_0) \mid y_{t-1}] &= o_P(1/n), \\
\mathrm{Var}[\sqrt{n}(Z(\hat{\theta}_n) - Z(\theta_0)) \mid y_{t-1}] &= \sigma^2(\theta_0) + o_P(1),
\end{aligned}
$$

we find

$$\begin{aligned}
& P[y_t < Z(\hat{\theta}_n) \mid y_{t-1}] \\
= \quad & \Phi(u) + \frac{\phi(u)}{2n}\left\{\frac{\partial^2 a}{\partial y^2}(\theta_0, Z(\theta_0)) - u\left(\frac{\partial a}{\partial y}(\theta_0, Z(\theta_0))\right)^2\right\}\sigma^2(\theta_0) + o_p(1/n).
\end{aligned}$$

Comparing with (40) we see that the term

$$\frac{\phi(u)}{2n}\frac{\partial a}{\partial y}(\theta_0, Z(\theta_0))Tr\left(\Omega\frac{\partial^2 Z}{\partial\theta\partial\theta'}(\theta_0)\right),$$

is missing. This term corresponds to the first-order bias of $Z(\hat{\theta}_n)$, since we have

$$Z(\hat{\theta}_n) = Z(\theta_0) + \frac{\partial Z}{\partial\theta'}(\theta_0)(\hat{\theta}_n - \theta_0) + (\hat{\theta}_n - \theta_0)'\frac{\partial^2 Z}{\partial\theta'\partial\theta}(\theta_0)(\hat{\theta}_n - \theta_0) + o_p(\|\hat{\theta}_n - \theta_0\|^2).$$

Thus, under our assumption (12),

$$E[Z(\hat{\theta}_n) - Z(\theta_0) \mid y_{t-1}] = \frac{1}{n}Tr\left(\Omega\frac{\partial^2 Z}{\partial\theta\partial\theta'}(\theta_0)\right) + o_p(1/n).$$

In general, this expectation cannot be neglected at order $1/n$ if (12) holds.

A.9 Proof of Corollary 1

We have

$$a(y_{t-1}, \theta, g(y_{t-1}, \theta_0, \varepsilon)) = H^{-1}\left(\frac{\sigma_t(\theta_0)}{\sigma_t(\theta)}H(\varepsilon)\right).$$

We deduce

$$\begin{aligned}
\frac{\partial a}{\partial\theta}(y_{t-1}, \theta, g(y_{t-1}, \theta_0, \varepsilon)) \quad &= \quad \frac{-1}{h\left\{H^{-1}\left(\frac{\sigma_t(\theta_0)}{\sigma_t(\theta)}H(\varepsilon)\right)\right\}}\frac{\sigma_t(\theta_0)}{\sigma_t^2(\theta)}\frac{\partial\sigma_t}{\partial\theta}H(\varepsilon), \\
\left(\frac{\partial a}{\partial\theta}(y_{t-1}, \theta, g(y_{t-1}, \theta_0, \varepsilon))\right)_{\theta=\theta_0} \quad &= \quad -K(\varepsilon)\frac{1}{\sigma_t(\theta_0)}\frac{\partial\sigma_t}{\partial\theta}(\theta_0),
\end{aligned}$$

where $K(\varepsilon) = 1/(\partial\log H(\varepsilon)/\partial\varepsilon)$. By differentiating again w.r.t. ε, we get

$$\left(\frac{\partial^2 a}{\partial\theta\partial\varepsilon}(y_{t-1}, \theta, g(y_{t-1}, \theta_0, \varepsilon))\right)_{\theta=\theta_0} \quad = \quad -\frac{\partial K(\varepsilon)}{\partial\varepsilon}\frac{1}{\sigma_t(\theta_0)}\frac{\partial\sigma_t}{\partial\theta}(\theta_0),$$

and w.r.t. θ', we get

$$\begin{aligned}
& \left(\frac{\partial^2 a}{\partial\theta\partial\theta'}(y_{t-1}, \theta, g(y_{t-1}, \theta_0, \varepsilon))\right)_{\theta=\theta_0} \\
= \quad & K(\varepsilon)\left[-\frac{\partial\log h(\varepsilon)}{\partial\varepsilon}K(\varepsilon) + 2\right]\frac{1}{\sigma_t^2(\theta_0)}\frac{\partial\sigma_t}{\partial\theta}\frac{\partial\sigma_t}{\partial\theta'}(\theta_0) \\
& -K(\varepsilon)\frac{1}{\sigma_t(\theta_0)}\frac{\partial^2\sigma_t}{\partial\theta\partial\theta'}(\theta_0).
\end{aligned}$$

By applying the formula in Proposition 1 we get

$$\begin{aligned} b_t(\theta_0,\varepsilon) &= \frac{-1}{\sigma_t^2(\theta_0)}\frac{\partial\sigma_t(\theta_0)}{\partial\theta'}\Omega\frac{\partial\sigma_t(\theta_0)}{\partial\theta}K^2(\varepsilon)\left\{\frac{\partial\log h(\varepsilon)}{\partial\varepsilon}+\varepsilon\right\} \\ &\quad +\frac{K(\varepsilon)}{\sigma_t(\theta_0)}\mathrm{Tr}\left\{\frac{\partial^2\sigma_t}{\partial\theta\partial\theta'}(\theta_0)\Omega\right\}. \end{aligned}$$

It follows from (17)-(18) that

$$\begin{aligned} \mathrm{EVaR}_{n,t}(\alpha) &= -g\left[y_{t-1};\hat{\theta}_n,\Phi^{-1}(\alpha)-b_t(\hat{\theta}_n,\Phi^{-1}(\alpha))/2n\right] \\ &= -\sigma_t(\hat{\theta}_n)H\left[\Phi^{-1}(\alpha)-b_t(\hat{\theta}_n,\Phi^{-1}(\alpha))/2n\right] \\ &= -\sigma_t(\hat{\theta}_n)H[\Phi^{-1}(\alpha)]+\sigma_t(\hat{\theta}_n)h[\Phi^{-1}(\alpha)]b_t(\hat{\theta}_n,\Phi^{-1}(\alpha))/2n+o_P(1/n). \end{aligned}$$

It suffices to replace (θ_0,ε) by $(\hat{\theta}_n,\Phi^{-1}(\alpha))$ in the above formula for $b_t(\theta_0,\varepsilon)$ to get the announced result. □

A.10 Proof of Corollary 2

In the standard ARCH model we have

$$\frac{\partial^2\sigma_t}{\partial\theta\partial\theta'}=\frac{-1}{\sigma_t}\frac{\partial\sigma_t}{\partial\theta}\frac{\partial\sigma_t}{\partial\theta'}.$$

Thus, in view of Corollary 1, we have

$$\mathrm{EVaR}_{n,t}(\alpha)-\mathrm{VaR}_{n,t}(\alpha)=-a[\Phi^{-1}(\alpha)]\frac{\sigma_t(\hat{\theta}_n)}{8n}\frac{1}{\sigma_t^4(\hat{\theta}_n)}\frac{\partial\sigma_t^2}{\partial\theta'}(\hat{\theta}_n)\hat{\Omega}_n\frac{\partial\sigma_t^2}{\partial\theta}(\hat{\theta}_n).$$

where

$$a(\varepsilon)=\frac{H^2(\varepsilon)}{h(\varepsilon)}\left\{\frac{\partial\log h(\varepsilon)}{\partial\varepsilon}+\varepsilon+\frac{\partial\log H(\varepsilon)}{\partial\varepsilon}\right\}.$$

Hence, from the strong consistency of $\hat{\theta}_n$ to θ_0, almost surely

$$\lim_{n\to\infty}n\frac{\mathrm{EVaR}_{n,t}(\alpha)-\mathrm{VaR}_{n,t}(\alpha)}{\sigma_t(\hat{\theta}_n)}=\frac{\xi}{8}a(\alpha)\frac{1}{\sigma_t^4(\theta_0)}\frac{\partial\sigma_t^2}{\partial\theta'}(\theta_0)J^{-1}\frac{\partial\sigma_t^2}{\partial\theta}(\theta_0).$$

The conclusion follows by

$$\begin{aligned} E\left(\frac{1}{\sigma_t^4(\theta_0)}\frac{\partial\sigma_t^2}{\partial\theta'}(\theta_0)J^{-1}\frac{\partial\sigma_t^2}{\partial\theta}(\theta_0)\right) &= \mathrm{Tr}\left\{E\left(\frac{1}{\sigma_t^4(\theta_0)}\frac{\partial\sigma_t^2}{\partial\theta'}(\theta_0)J^{-1}\frac{\partial\sigma_t^2}{\partial\theta}(\theta_0)\right)\right\} \\ &= \mathrm{Tr}\left\{E\left(\frac{1}{\sigma_t^4(\theta_0)}\frac{\partial\sigma_t^2}{\partial\theta}(\theta_0)\frac{\partial\sigma_t^2}{\partial\theta'}(\theta_0)J^{-1}\right)\right\} \\ &= \mathrm{Tr}(I_d)=d. \end{aligned}$$

A.11 Proof of Corollary 3

We have

$$a(y_{t-1},\theta,g(y_{t-1},\theta_0,\varepsilon)) = H^{-1}\left(\frac{\mu_t(\theta_0)-\mu_t(\theta)}{\sigma_t(\theta)}+\frac{\sigma_t(\theta_0)}{\sigma_t(\theta)}H(\varepsilon)\right).$$

We deduce

$$\begin{aligned}
\frac{\partial a}{\partial\theta}(y_{t-1},\theta,g(y_{t-1},\theta_0,\varepsilon)) &= \frac{-\left\{\frac{\mu_t(\theta_0)-\mu_t(\theta)+\sigma_t(\theta_0)H(\varepsilon)}{\sigma_t^2(\theta)}\frac{\partial\sigma_t}{\partial\theta}+\frac{1}{\sigma_t(\theta)}\frac{\partial\mu_t}{\partial\theta}\right\}}{h\left\{H^{-1}\left(\frac{\mu_t(\theta_0)-\mu_t(\theta)}{\sigma_t(\theta)}+\frac{\sigma_t(\theta_0)}{\sigma_t(\theta)}H(\varepsilon)\right)\right\}},\\
\left(\frac{\partial a}{\partial\theta}(y_{t-1},\theta,g(y_{t-1},\theta_0,\varepsilon))\right)_{\theta=\theta_0} &= \frac{-1}{\sigma_t(\theta_0)}\left[K(\varepsilon)\frac{\partial\sigma_t}{\partial\theta}(\theta_0)+\frac{1}{h(\varepsilon)}\frac{\partial\mu_t}{\partial\theta}(\theta_0)\right].
\end{aligned}$$

By differentiating again w.r.t. ε, we get

$$\begin{aligned}
&\left(\frac{\partial^2 a}{\partial\theta\partial\varepsilon}(y_{t-1},\theta,g(y_{t-1},\theta_0,\varepsilon))\right)_{\theta=\theta_0}\\
&= \frac{-1}{\sigma_t(\theta_0)}\left[\frac{\partial K(\varepsilon)}{\partial\varepsilon}\frac{\partial\sigma_t}{\partial\theta}(\theta_0)-\frac{\partial h(\varepsilon)}{\partial\varepsilon}\frac{1}{h^2(\varepsilon)}\frac{\partial\mu_t}{\partial\theta}(\theta_0)\right].
\end{aligned}$$

By applying the formula in Proposition 1 we get

$$\begin{aligned}
b_t(\theta_0,\varepsilon) &= \frac{1}{\sigma_t^2(\theta_0)}\left[\frac{\partial\sigma_t(\theta_0)}{\partial\theta'}\Omega\frac{\partial\sigma_t(\theta_0)}{\partial\theta}K^2(\varepsilon)\left\{2\frac{\partial\log K(\varepsilon)}{\partial\varepsilon}-\varepsilon\right\}\right.\\
&\quad+2\frac{\partial\sigma_t(\theta_0)}{\partial\theta'}\Omega\frac{\partial\mu_t(\theta_0)}{\partial\theta}\frac{K(\varepsilon)}{h(\varepsilon)}\left\{\frac{\partial\log(K/h)(\varepsilon)}{\partial\varepsilon}-\varepsilon\right\}\\
&\quad\left.+\frac{\partial\mu_t(\theta_0)}{\partial\theta'}\Omega\frac{\partial\mu_t(\theta_0)}{\partial\theta}\frac{1}{h^2(\varepsilon)}\left\{2\frac{\partial\log h(\varepsilon)}{\partial\varepsilon}-\varepsilon\right\}\right].
\end{aligned}$$

It follows from (17)-(18) that

$$\begin{aligned}
&\mathrm{EVaR}_{n,t}(\alpha)\\
&= -g\left[y_{t-1};\hat\theta_n,\Phi^{-1}(\alpha)-b_t(\hat\theta_n,\Phi^{-1}(\alpha))/2n\right]\\
&= -\mu_t(\hat\theta_n)-\sigma_t(\hat\theta_n)H\left[\Phi^{-1}(\alpha)-b_t(\hat\theta_n,\Phi^{-1}(\alpha))/2n\right]\\
&= -\mu_t(\hat\theta_n)-\sigma_t(\hat\theta_n)H[\Phi^{-1}(\alpha)]+\sigma_t(\hat\theta_n)h[\Phi^{-1}(\alpha)]b_t(\hat\theta_n,\Phi^{-1}(\alpha))/2n+o_P(1/n).
\end{aligned}$$

It suffices to replace (θ_0,ε) by $(\hat\theta_n,\Phi^{-1}(\alpha))$ in the above formula for $b_t(\theta_0,\varepsilon)$ to get the announced result. $\square$

References

Bao, Y. and A. Ullah (2004) Bias of a Value-at-Risk Estimator. *Finance Research Letters* 1, 241–249.

Berkes, I., Horváth, L. and P. Kokoszka (2003) GARCH Processes: Structure and Estimation. *Bernoulli* 9, 201–227.

Berkes, I. and L. Horváth (2004) The Efficiency of the Estimators of the Parameters in GARCH Processes. *The Annals of Statistics* 32, 633–655.

Berkowitz, J., Christoffersen, P. and D. Pelletier (2011) Evaluating Value-at-Risk Models with Desk-Level Data. *Management Science* 57, 2213–2227.

Berkowitz, J. and J. O'Brien (2002) How Accurate Are Value-at-Risk Models at Commercial Banks? *Journal of Finance* 57, 1093–1111.

Chambers, M.J. (2012) Jacknife Estimation of Stationary Autoregressive Models. *Discussion paper, University of Essex.*

Chan, N.H., Deng, S-J., Peng, L. and Z. Xia (2007) Interval Estimation of Value-at-Risk Based on GARCH Models with Heavy-Tailed Innovations. *Journal of Econometrics* 137, 556–576.

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

Dufour, J-M. and J. Kiviet (1997) Exact Tests in Single Equation Autoregressive Distributed Lag Models. *Journal of Econometrics* 80, 325–353.

Engle, R.F., Lillien, D.M. et R.P. Robins (1987) Estimating Time Varying Risk Premia in the Term Structure: the ARCH-M Model. *Econometrica* 55, 391–407.

Escanciano, J.C. and J. Olmo (2010) Backtesting Parametric Value-at-Risk with Estimation Risk, *Journal of Business and Economics Statistics* 28, 36–51.

Escanciano, J.C. and J. Olmo (2011) Robust Backtesting Tests for Value-at-Risk Models. *Journal of Financial Econometrics* 9, 132–161.

Figlewski, S. (2004) Estimation Error in the Assessment of Financial Risk Exposure. Working paper, New-York University.

Francq, C. and J-M. Zakoïan (2004) Maximum Likelihood Estimation of Pure GARCH and ARMA-GARCH processes. *Bernoulli* 10, 605-637.

Francq, C. and J-M. Zakoïan (2010) *GARCH Models: Structure, Statistical Inference and Financial Applications.* John Wiley.

Gouriéroux, C., and J., Jasiak (2005) Nonlinear Impulse Response Function. *Annales d'Economie et de Statistique* 78, 1–33.

Gouriéroux, C., Laurent, J-P., and O., Scaillet (2000) Sensitivity Analysis of Value-at-Risk. *Journal of Empirical Finance* 7, 225–246.

Hansen, B.E. (2006) Interval Forecasts and Parameter Uncertainty. *Journal of Econometrics* 135, 377–398.

Hartz, C., Mittnik, S., and M. Paolella (2006) Accurate Value-at-Risk Forecasting Based on the Normal-GARCH Model. *Computational Statistics & Data Analysis* 51, 2295–2312.

Inui, K., Kijima, M. and A. Kitano (2005) VaR is Subject to a Significant Positive Bias. *Statistics & Probability Letters* 72, 299–311.

Koenker, R. and Q. Zhao (1986) Conditional Quantile Estimation and Inference for ARCH Models. *Econometric Theory* 12, 793–813.

Lönnbark, C. (2010) A Corrected Value-at-Risk Predictor. *Applied Economics Letters* 17, 1193–1196.

Martin, R., and T. Wilde (2002) Unsystematic Credit Risk. *Risk Magazine* 15, 123–128.

Phillips, P.C.B., and J. Yu (2005) Jacknifing Bond Option Prices. *Review of Financial Studies* 18, 707–742.

Quenouille, M.H. (1956) Notes on Bias in Estimation. *Biometrika* 43, 353–360.