Econometric Theory, **24**, 2008, 1404–1424. Printed in the United States of America.
doi:10.1017/S0266466608080559

ESTIMATION RISK IN GARCH VaR AND ES ESTIMATES

FENG GAO AND FENGMING SONG
Tsinghua University

Value-at-risk (VaR) and expected shortfall (ES) are now both widely used risk measures. However, users have not paid much attention to the estimation risk issues, especially in the case of heteroskedastic financial time series. The key challenge arises from the fact that the estimated generalized autoregressive conditional heteroskedasticity (GARCH) innovations are not the true independent innovations. The purpose of this work is to provide an analytical method to assess the precision of conditional VaR and ES in the GARCH model estimated by the filtered historical simulation (FHS) method based on the asymptotic behavior of the residual empirical distribution function in GARCH processes. The proposed method is evaluated by simulation and proved valid.

1. INTRODUCTION

Value-at-risk (VaR) is generally defined as the maximum loss of a financial position over a given time period such that there is a low, prespecified probability that the actual loss will be larger (Jorion, 2000). It has been a widely used measure in portfolio risk management and capital allocation because of its clear economic meaning. However, VaR lacks the subadditivity property and provides no information on the amount of losses when the "bad" events occur. Therefore in recent years another useful risk measure, i.e., expected shortfall (ES) (Artzner et al., 1999) has been proposed. The expected shortfall is defined as the expected loss when the loss is larger than the corresponding VaR and thus ES is informative on the extent of excess losses.

In earlier studies VaR and ES have usually been calculated based on models for the independent and identically distributed (i.i.d.) return. Recently, many sophisticated models have been applied in calculating VaR and ES, such as the generalized autoregressive conditional heteroskedasticity (GARCH) model (McNeil and Frey, 2000), the extreme value theory (Embrechts, Klüppelberg, and Mikosch, 1997), and the kernel smoothing method (Chen and Tang, 2005; Scaillet, 2004). Because of their ability to capture the fact of stochastic volatility in financial time series, the estimates based on GARCH models seem to

We thank Professor Bruce E. Hansen (the co-editor) and three anonymous referees for their helpful comments and suggestions, which led to substantial improvement of the presentation. Address correspondence to Feng Gao, Department of Finance, School of Economics and Management, Tsinghua University, Beijing, P.R. China; e-mail: gaof@sem.tsinghua.edu.cn.

 © 2008 Cambridge University Press 0266-4666/08 $15.00

https://doi.org/10.1017/S0266466608080559 Published online by Cambridge University Press

be the most widely used and have been adopted by many financial institutions and products, e.g., RiskMetrics. The typical procedure to estimate VaR and ES based on GARCH models is first to forecast the volatility by GARCH models and then to scale the volatility by some quantile or truncated mean of innovations. The quantile and truncated mean can be estimated based on some parametric distribution assumption or using their empirical counterparts. Our paper focuses on the latter, which is called the filtered historical simulation (FHS) method (Barone-Adesi, Giannopoulos, and Vosper, 1999) and which has been proved to perform well in previous studies (Kuester, Mittnik, and Paolella, 2006) by reducing the model misspecification risk.

The purpose of this paper is to derive the limiting distribution of VaR and ES estimated by the FHS method and provide confidence intervals around the VaR and ES measures. Because VaR and ES are often used by financial institutions to set capital reserves or trading limits, the knowledge of the confidence intervals along with the VaR and ES point estimates will be valuable. For example, consider an oil future trader who needs sufficient capital against unfavorable price moving. If the point estimate of daily VaR is 8% whereas the confidence band is 6–10%, the cautious trader should set the capital requirement rate at 10% to avoid the estimation risks. However, in practice users of VaR or ES often overlook the estimation risk issues, partly because of the difficulty in deriving the closed-form confidence intervals.

Recently, the issue of estimation risk has attracted some attention. The i.i.d. return case is examined in Jorion (1996) and Dowd (2000). Christoffersen and Gonçalves (2005) examine the issue in GARCH models using bootstrap techniques. However, they do not provide an asymptotic theory for calculating confidence intervals in this context. Furthermore, the bootstrap method they proposed is time consuming and not applicable when the GARCH innovations have infinite fourth moment, whereas our methods remain valid in this case. Another related work is Chan, Deng, Peng, and Xia (2007). They derive the limiting distribution of the conditional intermediate quantile of a GARCH process. Their work is based on extreme value theory and does not consider the ES estimation. Thus their results cannot be easily extended to obtain ours.

Because the delta method can be used to calculate confidence intervals for the GARCH volatility forecast (Duan, 1994), it is straightforward to obtain the standard errors of the VaR and ES estimates in the GARCH model with fully parametric distribution (e.g., normal). However it is not obvious how to calculate confidence bands for the estimates in the GARCH model with general distribution. Roughly speaking, there are three components in the variance of GARCH VaR or ES in the general case: the first is from the volatility forecast, the second is from the quantile or truncated mean of the estimated innovations, and the third is from the covariance between the volatility forecast and the estimated innovations. If the GARCH estimated innovations are i.i.d., the variance of sample quantile or ES is known (see, e.g., Serfling, 1980; Shorack, 1974), and hence the variance of GARCH VaR or ES is easily obtained. Unfortu-

https://doi.org/10.1017/S0266466608080559 Published online by Cambridge University Press

nately, the empirical residuals are not the true independent innovations but depend on the parameter estimations of the GARCH model. As shown in Berkes and Horvath (2003) and Vyazilov (1999), the limit behaviors of GARCH residuals, i.e., the estimated innovations, are affected by parameter estimations in the GARCH volatility process. Therefore, to derive the asymptotic variances of VaR and ES in GARCH models, we must take the interaction between the estimated parameters in the GARCH volatility equation and the estimated GARCH innovations into account.

The remainder of the paper is structured as follows. Section 2 introduces the return dynamic models and corresponding VaR and ES estimates. Section 3 investigates the statistical properties of GARCH VaR and ES estimates. Section 4 presents the simulation results. The final section concludes. All the proofs are presented in the Appendix.

2. ESTIMATORS OF VaR AND ES

To account for empirical regularities in financial data, we model the returns as a GARCH(p,q) process (Bollerslev, 1986) with a general continuous innovation, i.e.,

$$r_t = \sigma_t \varepsilon_t,$$

$$\sigma_t^2 = c + \sum_{i=1}^{p} a_i r_{t-i}^2 + \sum_{j=1}^{q} b_j \sigma_{t-j}^2, \tag{2.1}$$

where the innovations ε_t are a sequence of i.i.d. random variables with mean zero, variance one, and distribution function F and ε_t is independent of $\mathcal{F}_{t-1}$, the sigma field generated by $\{r_i; i \le t-1\}$. In particular, the distribution F is assumed to have a positive density everywhere, i.e., $f(x) = F'(x) > 0$. The parameters satisfy the conditions $c > 0$, $a_i \ge 0$, $b_j \ge 0$, and the Lyapunov exponent is negative to ensure that the return process $\{r_t : -\infty < t < \infty\}$ is strictly stationary (Bougerol and Picard, 1992).

Given model (2.1), the unit asset's VaR and ES for time $t+1$ conditional on information at time t are

$$V_{t+1}^{\alpha} = \sigma_{t+1} F^{-1}(\alpha) := \sigma_{t+1} \varepsilon_{\alpha}, \tag{2.2}$$

$$ES_{t+1}^{\alpha} = \sigma_{t+1} \frac{\int_{-\infty}^{\varepsilon_{\alpha}} x dF(x)}{\alpha} := \sigma_{t+1} \mu_{\alpha}, \tag{2.3}$$

where σ_{t+1} is the conditional volatility for time $t+1$ and $F^{-1}(\alpha)$ is the α-quantile of F, the (conditional) distribution of innovations. With additional assumption of $E \log \sigma_0^2 < \infty$, (2.1) implies (Berkes, Horvath, and Kokoszka, 2003; Hall and Yao, 2003)

https://doi.org/10.1017/S0266466608080559 Published online by Cambridge University Press

$$\sigma_t^2 = \frac{c}{1 - \sum_j b_j} + \sum_{i=1}^{p} a_i r_{t-i}^2 + \sum_{i=1}^{p} a_i \sum_{k=1}^{\infty} \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} b_{j_1} \cdots b_{j_k} r_{t-i-j_1-\cdots-j_k}^2. \quad \textbf{(2.4)}$$

That is to say, σ_t is completely determined by realized return before time t if the parameters (a, b, c) are known. We hence write $\sigma_t = \sigma_t(a, b, c)$, expressing σ_t as a function of $a = (a_1, \ldots, a_p)$, $b = (b_1, \ldots, b_q)$, c, and the observed data $\{r_i; i < t\}$. Because the true parameters in the volatility equation and the true innovation distribution are totally unknown, the VaR and ES that are of interest for time $n + 1$, V_{n+1}^{α} and ES_{n+1}^{α}, cannot be exactly computed and must be inferred from the observed data. One successful and commonly used estimation method is called the filtered historical simulation (FHS) method. The VaR estimator $\hat{V}_{n+1}^{\alpha}$ and the ES estimator ES_{n+1}^{α} are obtained through the following steps.

1. Estimate the GARCH parameters (a, b, c) by the quasi-maximum likelihood (QML) method, maximizing

$$L_n(a, b, c) = -\frac{1}{2} \sum_{t=1}^{n} \left\{ \ln \sigma_t(a, b, c)^2 + \frac{r_t^2}{\sigma_t(a, b, c)^2} \right\}. \quad \textbf{(2.5)}$$

   Given the QML estimates $(\hat{a}, \hat{b}, \hat{c})$, we can estimate the past variances $\{\hat{\sigma}_t^2, t \le n\}$ and predicted variance $\hat{\sigma}_{n+1}^2$ with estimated $(\hat{a}, \hat{b}, \hat{c})$.
2. After the volatility $\hat{\sigma}_t$ is estimated, the innovations can be estimated as

$$\hat{\varepsilon}_t = \frac{r_t}{\hat{\sigma}_t}, \qquad 1 \le t \le n.$$

   Then ε_α can be estimated by the sample α-quantile of the empirical distribution $\hat{F}_n$ of $\{\hat{\varepsilon}_t,\ 1 \le t \le n\}$, i.e.,

$$\hat{\varepsilon}_\alpha = \hat{F}_n^{-1}(\alpha) = \inf\{x : \hat{F}_n(x) \ge \alpha\}.$$

   The estimator of μ_α is

$$\hat{\mu}_\alpha = \frac{\sum_{t=1}^{n} \hat{\varepsilon}_t I(\hat{\varepsilon}_t < \hat{\varepsilon}_\alpha)}{\sum_{t=1}^{n} I(\hat{\varepsilon}_t < \hat{\varepsilon}_\alpha)},$$

   where

$$\hat{F}_n(x) = \frac{1}{n} \sum_{t=1}^{n} I(\hat{\varepsilon}_t \le x)$$

   and $I(\cdot)$ is the indicator function.

https://doi.org/10.1017/S0266466608080559 Published online by Cambridge University Press

As Christoffersen and Gonçalves (2005) demonstrated, the estimated innovations can be centered to mimic the true mean zero expectation of ε_t; then the centered innovations are used to estimate ε_α and μ_α. These estimators are denoted as $\hat{\varepsilon}_\alpha^0$ and $\hat{\mu}_\alpha^0$ in this paper. Though whether the innovations are centered has no impact on their consistency, it has an impact on their asymptotic variances.

3. The estimators of VaR and ES can be obtained as

$$\hat{V}_{n+1}^{\alpha} = \hat{\sigma}_{n+1}\hat{\varepsilon}_\alpha, \qquad \hat{ES}_{n+1}^{\alpha} = \hat{\sigma}_{n+1}\hat{\mu}_\alpha$$

or

$$\hat{V}_{n+1}^{0\alpha} = \hat{\sigma}_{n+1}\hat{\varepsilon}_\alpha^0, \qquad \hat{ES}_{n+1}^{0\alpha} = \hat{\sigma}_{n+1}\hat{\mu}_\alpha^0.$$

Because the VaR and ES are estimated from the data, they inevitably introduce estimation risk and may affect the decision efficiency. It is thus important to quantify the estimation risk by establishing their asymptotic standard errors.

3. ESTIMATION ERRORS OF VaR AND ES ESTIMATES

Before we state the main results on the standard errors of GARCH VaR and ES estimates, we first cite two useful lemmas related to the GARCH model parameter estimators and the empirical distribution of estimated innovations. The first lemma, which generalizes the results of Lee and Hansen (1994) and Lumsdaine (1996), states that quasi-maximum likelihood estimators (QMLE) of the GARCH model are $\sqrt{n}$-consistent and asymptotically normally distributed if the innovations have finite fourth moment. The proof of this lemma can be found in Berkes, Horvath, and Kokoszka (2003) and Hall and Yao (2003).

LEMMA 3.1 (Berkes et al., 2003; Hall and Yao, 2003). *If* $E(\varepsilon_t^4) < \infty$, *the true value of* $\theta = (a_1, \ldots, a_p, b_1, \ldots, b_q, c)^T$ *is an interior point in some compact parameter subspace,* $a(z)$ *and* $b(z)$ *have no common root,* $a(1) > 0$, $a_p + b_q > 0$, *where* $a(z) = \sum_{i=1}^{p} a_i z^i$, $b(z) = 1 - \sum_{i=1}^{q} b_i z^i$, *then*

$$\hat{\theta}_n - \theta = \frac{1}{n} M^{-1} \sum_{t=1}^{n} (\varepsilon_t^2 - 1)\sigma_t^{-2} U_t + o_p(n^{-1/2}) \tag{3.1}$$

and

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, (E(\varepsilon_1^4) - 1)M^{-1}), \tag{3.2}$$

where $\hat{\theta}_n$ *denote the corresponding QMLE,* $U_t = U_t(a, b, c)$ *denotes the* $(p + q + 1)$*-vector of first derivatives of* $\sigma_t^2 = \sigma_t(a, b, c)^2$ *with respect to the components of a, b, and c, i.e.,*

https://doi.org/10.1017/S0266466608080559 Published online by Cambridge University Press

$$\frac{\partial \sigma_t^2}{\partial c} = \frac{1}{1 - \sum_{j=1}^{q} b_j},$$
$$\frac{\partial \sigma_t^2}{\partial a_i} = r_{t-i}^2 + \sum_{k=1}^{\infty} \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} b_{j_1} \ldots b_{j_k} r_{t-i-j_1-\cdots-j_k}^2, \tag{3.3}$$
$$\frac{\partial \sigma_t^2}{\partial b_j} = \frac{c}{\left(1 - \sum_{j=1}^{q} b_j\right)^2} + \sum_{i=1}^{p} a_i \sum_{k=0}^{\infty} (k+1) \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} b_{j_1} \ldots b_{j_k} r_{t-i-j_1-\cdots-j_k}^2,$$

and $M = E_0(\sigma_1^{-4} U_1 U_1^T)$ *is a* $(p + q + 1) \times (p + q + 1)$ *expectation matrix when a, b, and c take their true values.*

The second lemma states that the estimated innovations $\hat{\varepsilon}_t$ are not independent and are affected by the parameters' estimation of volatility equation. Although the lemma's formulation is slightly different than in the original papers, the proof is essentially identical.

LEMMA 3.2 (Vyazilov, 1999; Berkes and Horvath, 2003). *Assume that the innovation distribution* $F(x)$ *has a differentiable density function* $f(x)$ *and* $\sup_{x \in \mathbf{R}} x^2 |f'(x)| < \infty$*; then as* $n \to \infty$*,*

$$\sup_{x \in \mathbf{R}} \left| \sqrt{n}(\hat{F}_n(x) - F_n(x)) - \frac{1}{2}\sqrt{n} x f(x)(\hat{\theta}_n - \theta)^T \mathbf{e} \right| = o_p(1), \tag{3.4}$$

where $\mathbf{e} = E_0(\sigma_1^{-2} U_1)$ *and* $\hat{F}_n(x)$ *and* $F_n(x)$ *are empirical distributions of estimated and true innovations, i.e.,* $\hat{F}_n(x) = n^{-1} \sum_{t=1}^{n} I(\hat{\varepsilon}_t \le x)$, $F_n(x) = n^{-1} \sum_{t=1}^{n} I(\varepsilon_t \le x)$.

Based on the previous two lemmas, the estimation errors of volatility forecast $\hat{\sigma}_{n+1}$ and quantile $\hat{\varepsilon}_\alpha$ can be obtained as the following theorem.

THEOREM 3.1. *Suppose that the assumptions of Lemmas 3.1 and 3.2 hold; then* $\hat{\sigma}_{n+1}$, $\hat{\varepsilon}_\alpha(\hat{\varepsilon}_\alpha^0)$ *and* $\hat{V}_{n+1}^{\alpha}(\hat{V}_{n+1}^{0\alpha})$ *are all* $\sqrt{n}$*-consistent and asymptotically normal, and the asymptotical variance of the VaR estimator is*

$$\mathrm{Var}(\hat{V}_{n+1}^{\alpha}) = \mathrm{Var}(\hat{\sigma}_{n+1})(\varepsilon_\alpha)^2 + (\sigma_{n+1})^2 \mathrm{Var}(\hat{\varepsilon}_\alpha) + 2\sigma_{n+1}\varepsilon_\alpha \mathrm{Cov}(\hat{\sigma}_{n+1}, \hat{\varepsilon}_\alpha), \tag{3.5}$$

$$\mathrm{Var}(\hat{V}_{n+1}^{0\alpha}) = \mathrm{Var}(\hat{\sigma}_{n+1})(\varepsilon_\alpha)^2 + (\sigma_{n+1})^2 \mathrm{Var}(\hat{\varepsilon}_\alpha^0) + 2\sigma_{n+1}\varepsilon_\alpha \mathrm{Cov}(\hat{\sigma}_{n+1}, \hat{\varepsilon}_\alpha^0), \tag{3.6}$$

https://doi.org/10.1017/S0266466608080559 Published online by Cambridge University Press

where

$$\mathrm{Var}(\hat{\sigma}_{n+1}) = \frac{E(\varepsilon_1^4) - 1}{4n\sigma_{n+1}^2} U_{n+1}^T M^{-1} U_{n+1}, \tag{3.7}$$

$$\begin{aligned}\mathrm{Var}(\hat{\varepsilon}_\alpha) = {} & \frac{E(\varepsilon_1^4) - 1}{4n} \varepsilon_\alpha^2 \mathbf{e}^T M^{-1} \mathbf{e} + \frac{\alpha(1-\alpha)}{nf^2(\varepsilon_\alpha)} \\ & + \frac{\varepsilon_\alpha \int_{-\infty}^{\varepsilon_\alpha} (x^2 - 1) f(x)\, dx}{nf(\varepsilon_\alpha)} \mathbf{e}^T M^{-1} \mathbf{e},\end{aligned} \tag{3.8}$$

$$\mathrm{Var}(\hat{\varepsilon}_\alpha^0) = \mathrm{Var}(\hat{\varepsilon}_\alpha) + \frac{1}{n} + \frac{E(\varepsilon_1^3)}{n} \varepsilon_\alpha \mathbf{e}^T M^{-1} \mathbf{e} + \frac{2\int_{-\infty}^{\varepsilon_\alpha} x f(x)\, dx}{nf(\varepsilon_\alpha)}, \tag{3.9}$$

and

$$\mathrm{Cov}(\hat{\sigma}_{n+1}, \hat{\varepsilon}_\alpha) = -\frac{U_{n+1}^T M^{-1} \mathbf{e}}{2n\sigma_{n+1}} \left(\frac{\varepsilon_\alpha}{2} (E(\varepsilon_1^4) - 1) + \frac{\int_{-\infty}^{\varepsilon_\alpha} (x^2 - 1) f(x)\, dx}{f(\varepsilon_\alpha)} \right), \tag{3.10}$$

$$\mathrm{Cov}(\hat{\sigma}_{n+1}, \hat{\varepsilon}_\alpha^0) = \mathrm{Cov}(\hat{\sigma}_{n+1}, \hat{\varepsilon}_\alpha) - \frac{U_{n+1}^T M^{-1} \mathbf{e}}{2n\sigma_{n+1}} E(\varepsilon_1^3). \tag{3.11}$$

Remark 3.1. In the GARCH model with specific parametric distribution (e.g., normal), the maximum likelihood estimator (MLE) (not QMLE) $\hat{\theta}_n$ can be obtained, and the GARCH VaR or ES can be expressed as $q(\hat{\theta}_n)\hat{\sigma}_t$ for some specified function $q(\cdot)$ (a constant in the normal GARCH case). Thus the corresponding estimation error can be directly derived by the delta method (e.g., $q\sqrt{\mathrm{Var}(\hat{\sigma}_t)}$ in the normal GARCH case).

Remark 3.2. The preceding theorem implies that the variance of VaR estimated by the FHS method is roughly proportional to $1/f^2(\varepsilon_\alpha)$. Therefore for extreme quantiles the FHS VaR estimator might have large estimation risks. One could consider a method based on extreme value theory such as Chan et al. (2007) to somewhat reduce estimation risks.

Based on the preceding theorem, we can construct the confidence interval and the variance of the conditional VaR. However, the expression of asymptotic variance (3.8) contains some unknown quantities such as $f(\varepsilon_\alpha)$, $E(\varepsilon_1^4)$, ε_α, $\int_{-\infty}^{\varepsilon_\alpha} x f(x)\, dx$, $\int_{-\infty}^{\varepsilon_\alpha} (x^2 - 1) f(x)\, dx$, $\mathbf{e}$, and M. These quantities should be replaced by their consistent estimators. Here we only discuss how to consistently estimate $f(\varepsilon_\alpha)$ because the consistent estimators of other quantities are

https://doi.org/10.1017/S0266466608080559 Published online by Cambridge University Press

obvious. According to the following lemma, the density function $f(x)$ can be consistently estimated by the kernel method:

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_{t=1}^{n} K\left(\frac{x - \hat{\varepsilon}_t}{h_n}\right). \tag{3.12}$$

LEMMA 3.3. *Assume that $h_n > 0$, $h_n \to 0$, $\sqrt{n}h_n^2 \to \infty$, and K is some chosen unimodal density, symmetric about zero and Lipschitz, i.e., there exists a positive constant k such that*

$$|K(x) - K(y)| \le k|x - y|, \quad \forall x, y \in \mathbf{R}$$

and

$$\int x^2 K(x)\, dx < \infty.$$

Then

$$\hat{f}_n(\hat{\varepsilon}_\alpha) - f(\varepsilon_\alpha) = o_p(1).$$

Summarizing the previous analysis, we get the following theorem.

THEOREM 3.2. *Suppose that the assumptions of Theorem 3.1 and Lemma 3.3 hold, and define*

$$\hat{U}_{n+1} = U_{n+1}(\hat{a}, \hat{b}, \hat{c}), \qquad \hat{\mathbf{e}} = \frac{1}{n} \sum_{t=1}^{n} \hat{\sigma}_t^{-2} \hat{U}_t, \qquad \hat{M} = \frac{1}{n} \sum_{t=1}^{n} \hat{\sigma}_t^{-4} \hat{U}_t \hat{U}_t^T,$$

$$\hat{\mu}_4 = \frac{1}{n} \sum_{t=1}^{n} \hat{\varepsilon}_t^4, \qquad \hat{\mu}_2^\alpha = \frac{1}{n} \sum_{t=1}^{n} (\hat{\varepsilon}_t^2 - 1) I(\hat{\varepsilon}_t \le \hat{\varepsilon}_\alpha),$$

$$D_1 = \frac{\hat{\mu}_4 - 1}{4\hat{\sigma}_{n+1}^2} \hat{U}_{n+1}^T \hat{M}^{-1} \hat{U}_{n+1},$$

$$D_2 = \frac{\hat{\mu}_4 - 1}{4} \hat{\varepsilon}_\alpha^2 \hat{\mathbf{e}}^T \hat{M}^{-1} \hat{\mathbf{e}} + \frac{\alpha(1-\alpha)}{\hat{f}_n^2(\hat{\varepsilon}_\alpha)} + \frac{\hat{\varepsilon}_\alpha \hat{\mu}_2^\alpha}{\hat{f}_n(\hat{\varepsilon}_\alpha)} \hat{\mathbf{e}}^T \hat{M}^{-1} \hat{\mathbf{e}},$$

$$D_{12} = -\frac{\hat{U}_{n+1}^T \hat{M}^{-1} \hat{\mathbf{e}}}{2\hat{\sigma}_{n+1}} \left(\frac{\hat{\varepsilon}_\alpha}{2} (\hat{\mu}_4 - 1) + \frac{\hat{\mu}_2^\alpha}{\hat{f}_n(\hat{\varepsilon}_\alpha)} \right).$$

Then $D_1 \xrightarrow{p} n \operatorname{Var}(\hat{\sigma}_{n+1})$, $D_2 \xrightarrow{p} n \operatorname{Var}(\hat{\varepsilon}_\alpha)$, $D_{12} \xrightarrow{p} n \operatorname{Cov}(\hat{\sigma}_{n+1}, \hat{\varepsilon}_\alpha)$, *and* $D_1 \hat{\varepsilon}_\alpha^2 + D_2 \hat{\sigma}_{n+1}^2 + 2\hat{\sigma}_{n+1} \hat{\varepsilon}_\alpha D_{12} \xrightarrow{p} n \operatorname{Var}(\hat{V}_{n+1}^\alpha)$.

Now we turn to the issue of standard errors of GARCH ES estimates. Because the estimated error of $\hat{\sigma}_{n+1}$ has already been obtained, we only need to consider the error of $\hat{\mu}_\alpha$ ($\hat{\mu}_\alpha^0$).

https://doi.org/10.1017/S0266466608080559 Published online by Cambridge University Press

THEOREM 3.3. *Suppose the assumptions of Lemmas 3.1 and 3.2 hold; then* $\hat{\mu}_\alpha(\hat{\mu}_\alpha^0)$ *and* $\hat{ES}_{n+1}^\alpha(\hat{ES}_{n+1}^{0\alpha})$ *are* $\sqrt{n}$*-consistent and asymptotically normal, and the asymptotical variance of the ES estimator is*

$$\begin{aligned}\mathrm{Var}(\hat{ES}_{n+1}^\alpha) &= \mathrm{Var}(\hat{\sigma}_{n+1})(\mu_\alpha)^2 + (\sigma_{n+1})^2\,\mathrm{Var}(\hat{\mu}_\alpha)\\ &\quad + 2\sigma_{n+1}\mu_\alpha\,\mathrm{Cov}(\hat{\sigma}_{n+1},\hat{\mu}_\alpha), \qquad (3.13)\\ \mathrm{Var}(\hat{ES}_{n+1}^{0\alpha}) &= \mathrm{Var}(\hat{\sigma}_{n+1})(\mu_\alpha)^2 + (\sigma_{n+1})^2\,\mathrm{Var}(\hat{\mu}_\alpha^0)\\ &\quad + 2\sigma_{n+1}\mu_\alpha\,\mathrm{Cov}(\hat{\sigma}_{n+1},\hat{\mu}_\alpha^0), \qquad (3.14)\end{aligned}$$

where

$$\begin{aligned}\mathrm{Var}(\hat{\mu}_\alpha) &= \frac{E(\varepsilon_1^4)-1}{4n}\,\mu_\alpha^2\mathbf{e}^TM^{-1}\mathbf{e} - \frac{\mu_\alpha\displaystyle\int_{-\infty}^{\varepsilon_\alpha}(x-\varepsilon_\alpha)(x^2-1)f(x)\,dx}{n\alpha}\,\mathbf{e}^TM^{-1}\mathbf{e}\\ &\quad + \frac{\displaystyle\int_{-\infty}^{\varepsilon_\alpha}x^2f(x)\,dx + \alpha(1-\alpha)\varepsilon_\alpha^2 - \alpha^2\mu_\alpha^2 - 2\alpha(1-\alpha)\varepsilon_\alpha\mu_\alpha}{n\alpha^2}, \qquad (3.15)\end{aligned}$$

$$\mathrm{Var}(\hat{\mu}_\alpha^0) = \mathrm{Var}(\hat{\mu}_\alpha) + \frac{1}{n} + \frac{\mu_\alpha E(\varepsilon_1^3)}{n}\,\mathbf{e}^TM^{-1}\mathbf{e} - \frac{2\displaystyle\int_{-\infty}^{\varepsilon_\alpha}(x-\varepsilon_\alpha)xf(x)\,dx}{n\alpha}, \qquad (3.16)$$

and

$$\begin{aligned}\mathrm{Cov}(\hat{\sigma}_{n+1},\hat{\mu}_\alpha) &= -\frac{U_{n+1}^TM^{-1}\mathbf{e}}{2n\sigma_{n+1}}\\ &\quad\times\left(\frac{\mu_\alpha}{2}\,(E(\varepsilon_1^4)-1) - \frac{\displaystyle\int_{-\infty}^{\varepsilon_\alpha}(x-\varepsilon_\alpha)(x^2-1)f(x)\,dx}{\alpha}\right), \qquad (3.17)\end{aligned}$$

$$\mathrm{Cov}(\hat{\sigma}_{n+1},\hat{\mu}_\alpha^0) = \mathrm{Cov}(\hat{\sigma}_{n+1},\hat{\mu}_\alpha) - \frac{U_{n+1}^TM^{-1}\mathbf{e}}{2n\sigma_{n+1}}\,E(\varepsilon_1^3). \qquad (3.18)$$

Remark 3.3. From the preceding theorem, we can find that the effective sample size for the ES estimation is only $n\alpha$. Therefore in practice the ES estimator is less accurate even for moderate sample sizes because the commonly used α is merely 1%.

https://doi.org/10.1017/S0266466608080559 Published online by Cambridge University Press

Similar to Theorem 3.2 the following theorem shows how to consistently estimate the estimation errors in GARCH ES. In contrast to the case of GARCH VaR, here the estimator does not involve the estimation of the innovation density function, and hence it is easier to implement.

THEOREM 3.4. *Suppose that the assumptions of Theorem 3.1 and Lemma 3.2 hold and define*

$$\hat{U}_{n+1} = U_{n+1}(\hat{a}, \hat{b}, \hat{c}), \qquad \hat{\mathbf{e}} = \frac{1}{n}\sum_{t=1}^{n} \hat{\sigma}_t^{-2}\hat{U}_t, \qquad \hat{M} = \frac{1}{n}\sum_{t=1}^{n} \hat{\sigma}_t^{-4}\hat{U}_t\hat{U}_t^T,$$

$$\hat{\mu}_4 = \frac{1}{n}\sum_{t=1}^{n} \hat{\varepsilon}_t^4, \qquad \hat{\mu}_3^{\alpha} = \frac{\sum_{t=1}^{n} (\hat{\varepsilon}_t - \hat{\varepsilon}_\alpha)(\hat{\varepsilon}_t^2 - 1) I(\hat{\varepsilon}_t < \hat{\varepsilon}_\alpha)}{\sum_{t=1}^{n} I(\hat{\varepsilon}_t < \hat{\varepsilon}_\alpha)},$$

$$D_1 = \frac{\hat{\mu}_4 - 1}{4\hat{\sigma}_{n+1}^2}\hat{U}_{n+1}^T\hat{M}^{-1}\hat{U}_{n+1},$$

$$D_2 = \left(\frac{\hat{\mu}_4 - 1}{4}\hat{\mu}_\alpha^2 - \hat{\mu}_\alpha\hat{\mu}_3^{\alpha}\right)\hat{\mathbf{e}}^T\hat{M}^{-1}\hat{\mathbf{e}} + \frac{n\left(\sum_{t=1}^{n}\hat{\varepsilon}_t^2 I(\hat{\varepsilon}_t < \hat{\varepsilon}_\alpha) + \sum_{t=1}^{n}\hat{\varepsilon}_\alpha^2 I(\hat{\varepsilon}_t \geq \hat{\varepsilon}_\alpha)\right)}{\left(\sum_{t=1}^{n} I(\hat{\varepsilon}_t < \hat{\varepsilon}_\alpha)\right)^2} - \frac{\left(\sum_{t=1}^{n}\hat{\varepsilon}_t I(\hat{\varepsilon}_t < \hat{\varepsilon}_\alpha) + \sum_{t=1}^{n}\hat{\varepsilon}_\alpha I(\hat{\varepsilon}_t \geq \hat{\varepsilon}_\alpha)\right)^2}{\left(\sum_{t=1}^{n} I(\hat{\varepsilon}_t < \hat{\varepsilon}_\alpha)\right)^2},$$

$$D_{12} = -\frac{\hat{U}_{n+1}^T\hat{M}^{-1}\hat{\mathbf{e}}}{2\hat{\sigma}_{n+1}}\left(\frac{\hat{\mu}_\alpha}{2}(\hat{\mu}_4 - 1) - \hat{\mu}_3^{\alpha}\right).$$

Then $D_1 \xrightarrow{p} n\,\mathrm{Var}(\hat{\sigma}_{n+1})$, $D_2 \xrightarrow{p} n\,\mathrm{Var}(\hat{\mu}_\alpha)$, $D_{12} \xrightarrow{p} n\,\mathrm{Cov}(\hat{\sigma}_{n+1}, \hat{\mu}_\alpha)$, *and* $D_1\hat{\mu}_\alpha^2 + D_2\hat{\sigma}_{n+1}^2 + 2\hat{\sigma}_{n+1}\hat{\mu}_\alpha D_{12} \xrightarrow{p} n\,\mathrm{Var}(\hat{ES}_{n+1}^{\alpha})$.

Remark 3.4. In the preceding analysis an infinite series of $\{r_t\}_{-\infty<t<n}$ is observed and used to calculate σ_t^2 as defined in (2.4). In practice, however, the data are observed only over a finite time interval, say, $1 \leq t \leq n$, and σ_t^2 has to be approximated by a truncated series or using the unconditional expectation as the required initial value (Berkes et al., 2003; Hall and Yao, 2003). The equation (2.4) thus is replaced by the following equation:

https://doi.org/10.1017/S0266466608080559 Published online by Cambridge University Press

$$\tilde{\sigma}_t^2 = \frac{c}{1-\sum_j b_j} + \sum_{i=1}^{p} a_i r_{t-i}^2 + \sum_{i=1}^{p} a_i \sum_{k=1}^{\infty} \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} b_{j_1} \cdots b_{j_k}$$

$$\times r_{t-i-j_1-\cdots-j_k}^2 I(t-i-j_1-\cdots-j_k \geq 1). \tag{3.19}$$

Remark 3.5. Our main results will also apply to the GARCH model with heavy-tailed errors, i.e., $E(\varepsilon_t^4) = \infty$ and $\sum_{i=1}^{p} a_i + \sum_{j=1}^{q} b_j < 1$, because the validity of our results relies on the $\sqrt{n}$-consistency and asymptotic normality of estimators. Certainly, in this case we need to use least absolute deviations (LAD) estimators (Peng and Yao, 2003) and make some appropriate modifications to previous theorems. The relevant theorem is listed in Appendix A.4.

Remark 3.6. Our analysis could be extended to some other versions of GARCH models such as the threshold GARCH (TGARCH) model (Glosten, Jagannathan, and Runkle, 1993) to allow for the leverage effects when the corresponding GARCH model has $\sqrt{n}$-consistent and asymptotic normal parameter estimators. Similar analysis is also applicable in the autoregressive moving average GARCH (ARMA-GARCH) model, though the required lemmas corresponding to Lemmas 3.1 and 3.2 are more complicated (Ling, 2007; Koul and Ling, 2006).

4. SIMULATION RESULTS

In this section, we report results from some simulations designed to evaluate the finite-sample performance of standard error estimation for 1-day-ahead 1% GARCH VaR and ES. Four versions of GARCH(1,1) models are considered. There are two possible parameterizations, i.e.,

1. High persistence: $c = 0.08$, $a = 0.15$, $b = 0.8$;
2. Low persistence: $c = 0.16$, $a = 0.4$, $b = 0.5$.

Thus the unconditional volatility is about 20% per year, and there are two possible innovation distributions, i.e.,

1. Student t distribution with degrees of freedom 50: $t(50)$;
2. Student t distribution with degrees of freedom 5: $t(5)$.

For each version of the GARCH model, we draw 5,000 samples and consider two estimation sample sizes, $N = \{500, 1{,}000\}$. In all the experiments we estimate the model parameters by the QMLE method because the LAD method involves nonsmooth optimization and is computationally demanding. For the innovation density estimation, we adopt the Gaussian kernel and choose the bandwidth using a rule-of-thumb method suggested by Scott (1992). To assess the precision of the standard errors estimator, we report the corresponding coverage probabilities of the two-sided confidence intervals based on normal approx-

https://doi.org/10.1017/S0266466608080559 Published online by Cambridge University Press

imation with a nominal level of 90% (NCP). As a comparison, the corresponding coverage probabilities of bootstrap prediction intervals[1] based on 999 replications (BTCP) and confidence intervals of VaR and ES based on the normal GARCH model (NGCP; see Remark 3.1) are also reported. For better understanding which component has more effect on the accuracy of VaR V_{n+1}^{α} or ES ES_{n+1}^{α}, we also consider the properties of volatility prediction σ_{n+1}, innovation quantile ε_{α}, and innovation ES μ_{α}. All the results associated with uncentered estimators, i.e., $\hat{V}_{n+1}^{\alpha}$ and $\hat{ES}_{n+1}^{\alpha}$, are summarized in Table 1. The results of the centered estimators $\hat{V}_{n+1}^{0\alpha}$ and $\hat{ES}_{n+1}^{0\alpha}$ are similar and hence omitted.

Consistent with the results found in Christoffersen and Gonçalves (2005), the bootstrap intervals have approximately correct coverage for GARCH VaR

TABLE 1. Coverage probabilities of analytical method versus bootstrap method

		N					
		500			1,000		
	Component	NCP	BTCP	NGCP	NCP	BTCP	NGCP
H-t(50)	V_{n+1}^{α}	0.8416	0.8910	0.8694	0.8546	0.9096	0.8624
	σ_{n+1}	0.8842	0.8752	0.8842	0.8882	0.8868	0.8882
	ε_{α}	0.8368	0.9198	—	0.8440	0.9186	—
	ES_{n+1}^{α}	0.7810	0.7788	0.8390	0.8442	0.8368	0.8042
	μ_{α}	0.7702	0.7642	—	0.8366	0.8282	—
H-t(5)	V_{n+1}^{α}	0.8394	0.8930	0.4532	0.8496	0.8938	0.3128
	σ_{n+1}	0.8742	0.8600	0.8742	0.8828	0.8734	0.8828
	ε_{α}	0.8290	0.9214	—	0.8402	0.9210	—
	ES_{n+1}^{α}	0.7550	0.7322	0.0958	0.8004	0.7810	0.0312
	μ_{α}	0.7396	0.7210	—	0.7692	0.7808	—
L-t(50)	V_{n+1}^{α}	0.8468	0.8972	0.8704	0.8532	0.9084	0.8576
	σ_{n+1}	0.8906	0.8888	0.8906	0.8922	0.8936	0.8922
	ε_{α}	0.8450	0.9222	—	0.8510	0.9200	—
	ES_{n+1}^{α}	0.7886	0.7824	0.8366	0.8484	0.8386	0.8070
	μ_{α}	0.7670	0.7662	—	0.8328	0.8250	—
L-t(5)	V_{n+1}^{α}	0.8384	0.8902	0.4534	0.8554	0.8940	0.3128
	σ_{n+1}	0.8748	0.8634	0.8748	0.8882	0.8834	0.8882
	ε_{α}	0.8354	0.9198	—	0.8398	0.9176	—
	ES_{n+1}^{α}	0.7564	0.7340	0.0766	0.8000	0.7850	0.0310
	μ_{α}	0.7406	0.7236	—	0.7932	0.7854	—

The first two rows represent the sample size, and the third row represents the methods used to calculate coverage probabilities; i.e., NCP, BTCP, and NGCP are abbreviations for "coverage probability of normal approximation confidence interval," "coverage probability of bootstrap interval," and "coverage probability of normal approximation confidence interval with normal GARCH assumption." The first column represents the data generating process used; e.g., H-t(50) represents the GARCH(1,1) model with high volatility persistence and innovation t(50).

https://doi.org/10.1017/S0266466608080559 Published online by Cambridge University Press

TABLE 2. Coverage probabilities of analytical method under various sample sizes

		N						
		500	1,000	2,000	5,000	7,500	10,000	15,000
	Component	NCP	NCP	NCP	NCP	NCP	NCP	NCP
H-t(5)	V_{n+1}^{α}	0.8482	0.8624	0.8746	0.8844	0.8880	0.8908	0.8946
	σ_{n+1}	0.8774	0.8828	0.8852	0.8916	0.8960	0.8902	0.8946
	ε_{α}	0.8342	0.8554	0.8660	0.8796	0.8850	0.8798	0.8830
	ES_{n+1}^{α}	0.7446	0.8116	0.8474	0.8714	0.8844	0.8840	0.8882
	μ_{α}	0.7398	0.7984	0.8460	0.8706	0.8894	0.8918	0.8964

The first two rows represent the sample size, and the third row represents the method used to calculate coverage probabilities. The first column represents the data generating process.

but have considerable undercoverage for GARCH ES. The bootstrap methods perform better for computing GARCH VaR intervals, whereas the analytical methods perform slightly better for ES. Though the bootstrap method can provide satisfactory coverage for GARCH VaR, it is not so good for the two components of GARCH VaR, i.e., volatility prediction σ_{n+1} and innovation quantile ε_{α}. It appears that the bootstrap intervals' overcoverage for ε_{α} and undercoverage for σ_{n+1} together lead to an almost exact coverage for VaR. Finally, we find that the normal GARCH model is sensitive to the distribution specification and is unsatisfactory.

The undercoverage of the asymptotic intervals may indicate that the sample size is too small. We thus examine how the asymptotic coverage probabilities vary with the sample size, and we present the results associated with the H-t(5) GARCH model in Table 2. The other three models' results are similar and are omitted. As expected, the coverage probabilities do increase with the sample size and get close to the nominal level 90% when there are more than 7,500 observations.

However the sample size of about 7,500 is too large and hardly satisfied in practice; we hence try some sort of adjustment to get reasonably accurate coverage with moderate sample size. Here we only partially solve the problem in the case of VaR. Noticing that the kernel density estimation mentioned in Lemma 3.3 usually has significant bias (Hall, 1992), we use the variable-bandwidth kernel method (Abramson, 1982) to reduce the bias. More precisely, the estimator is defined as

$$\hat{f}_n^a(x) = \sum_{t=1}^{n} \frac{1}{nh_n\lambda_t} K\left(\frac{x-\hat{\varepsilon}_t}{h_n\lambda_t}\right),$$

where the local bandwidth factors λ are estimated based on a preliminary fixed bandwidth density estimate:

https://doi.org/10.1017/S0266466608080559 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466608080559 Published online by Cambridge University Press

TABLE 3. Coverage probabilities of analytical methods using various bandwidths

		$N = 500$									
		Bandwidth									
		h		$0.5h$		$0.8h$		$1.2h$		$1.5h$	
	Component	NCP	NCPV	NCP	NCPV	NCP	NCPV	NCP	NCPV	NCP	NCPV
H-t(50)	V_{n+1}^{α}	0.8440	0.8832	0.8522	0.8950	0.8512	0.8926	0.8322	0.8720	0.8060	0.8542
	σ_{n+1}	0.8840	0.8840	0.8840	0.8840	0.8840	0.8840	0.8840	0.8840	0.8840	0.8840
	ε_{α}	0.8326	0.8858	0.8358	0.8980	0.8412	0.8970	0.8198	0.8708	0.7878	0.8518
H-t(5)	V_{n+1}^{α}	0.8482	0.8910	0.8328	0.8892	0.8456	0.8944	0.8434	0.8854	0.8358	0.8732
	σ_{n+1}	0.8774	0.8774	0.8774	0.8774	0.8774	0.8774	0.8774	0.8774	0.8774	0.8774
	ε_{α}	0.8342	0.8968	0.8070	0.8870	0.8288	0.8986	0.8320	0.8908	0.8268	0.8782
L-t(50)	V_{n+1}^{α}	0.8400	0.8816	0.8516	0.8932	0.8490	0.8888	0.8272	0.8712	0.8050	0.8502
	σ_{n+1}	0.8874	0.8874	0.8874	0.8874	0.8874	0.8874	0.8874	0.8874	0.8874	0.8874
	ε_{α}	0.8392	0.8900	0.8392	0.9010	0.8430	0.8990	0.8224	0.8784	0.7908	0.8512
L-t(5)	V_{n+1}^{α}	0.8396	0.8812	0.8272	0.8812	0.8398	0.8858	0.8368	0.8738	0.8300	0.8640
	σ_{n+1}	0.8706	0.8706	0.8706	0.8706	0.8706	0.8706	0.8706	0.8706	0.8706	0.8706
	ε_{α}	0.8340	0.9024	0.8054	0.8880	0.8268	0.9026	0.8352	0.8964	0.8274	0.8824

The first two rows represent the sample size. The third row represents the bandwidth used in the kernel estimation of innovation density. The fourth row represents the methods used to calculate coverage probabilities (NCPV is the abbreviation for "coverage probability of analytical method using variable-bandwidth kernel density estimation"). The first column represents the data generating process.

$$\lambda_t = \sqrt{\frac{\left(\prod_{t=1}^{n} \hat{f}_n(\hat{\varepsilon}_t)\right)^{1/n}}{\hat{f}_n(\hat{\varepsilon}_t)}}.$$

The asymptotic coverage probabilities of the GARCH VaR confidence intervals using the bias reduced kernel density estimation are reported in Table 3 (NCPV, the fourth column) when the sample size is 500. Obviously the results are fairly good because now the coverage probability based on the asymptotic theory is very close to the nominal level 90%.

Finally we investigate how the choices of the bandwidth h_n affect the coverage probabilities of asymptotic GARCH VaR intervals. The results are also reported in Table 3, where the third row stands for the bandwidth used in the kernel estimation of innovation density, i.e., $0.5h$ represents a bandwidth equal to $0.5h$; here h is the bandwidth calculated using a rule-of-thumb method suggested by Scott (1992). In comparison with the coverage performance of GARCH VaR interval estimation with bandwidth h, the results indicate that the coverage performance is worse when the bandwidth is larger whereas the coverage performance is still good when the bandwidth is smaller. Therefore the results imply that the rule-of-thumb method is reasonable, though the choices of bandwidth do affect the coverage probabilities of confidence intervals.

5. CONCLUSIONS

Both value-at-risk (VaR) and expected shortfall (ES) are widely used tools in risk management. Quantifying the precision of their estimates is important because it can provide risk managers with more information for decision making. The aim of this work is to provide an analytical method to assess the precision of VaR and ES estimates in a heteroskedastic series setting. By considering the asymptotic behavior of the residual empirical distribution function in GARCH models, we derive the limiting distributions of the GARCH VaR and ES estimated by the FHS method and provide their confidence intervals. Based on the confidence intervals, risk managers may define an estimation-risk-free VaR or ES or identify a risk-prone region (Chan et al., 2007).

In this paper we only discuss 1-day-ahead VaR and ES measures. Future work may consider extending the horizon from 1 to multiple days. However, when going from 1 to multiple days the n-day ($n > 1$) conditional variance is stochastic, and the n-day conditional distribution is generally unknown. Therefore the n-day ahead VaR does not admit a simple expression. One possible method is to estimate the n-day VaR through simulation. Certainly the issue of asymptotic distributions for these measures would be quite challenging.

NOTES

1. The bootstrap method is similar to that of Christoffersen and Gonçalves (2005), but here the bootstrap series are resampled from the raw residuals. It is worth pointing out that the bootstrap confidence intervals would not be consistent if the series were resampled from the centered residuals.

https://doi.org/10.1017/S0266466608080559 Published online by Cambridge University Press

REFERENCES

Abramson, I.S. (1982) On bandwidth variation in kernel estimates—a square root law. *Annals of Statistics* 10, 1217–1223.

Artzner, P., F. Delbaen, J.M. Eber, & D. Heath (1999) Coherent measures of risk. *Mathematical Finance* 9(3), 203–228.

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

Berkes, I. & L. Horvath (2003) Limit results for the empirical process of squared residuals in GARCH models. *Stochastic Processes and Their Applications* 105, 271–298.

Berkes, I., L. Horvath, & P. Kokoszka (2003) GARCH processes: Structure and estimation. *Bernoulli* 9, 201–228.

Billingsley, P. (1999) *Convergence of Probability Measures*, 2nd ed. Wiley.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* 31, 307–327.

Bougerol, P. & N. Picard (1992) Stationarity of GARCH processes and of some non-negative time series. *Journal of Econometrics* 52, 115–127.

Chan, N.H., S.J. Deng, L. Peng, & Z. Xia (2007) Interval estimation for the conditional value-at-risk based on GARCH models with heavy tailed innovations. *Journal of Econometrics* 137, 556–576.

Chen, S.X. & C.Y. Tang (2005) Nonparametric inference of value-at-risk for dependent financial returns. *Journal of Financial Econometrics* 3, 227–255.

Christoffersen, P. & S. Gonçalves (2005) Estimation risk in financial risk management. *Journal of Risk* 7, 1–28.

Dowd, K. (2000) Assessing VaR accuracy. *Derivatives Quarterly* 6, 61–63.

Duan, J.C. (1994) Maximum likelihood estimation using price data of the derivative contract. *Mathematical Finance* 4, 155–167.

Embrechts, P., C. Klüppelberg, & T. Mikosch (1997) *Modelling Extremal Events for Insurance and Finance*. Springer.

Ghosh, J.K. (1971) A new proof of the Bahadur representation of quantiles and an application. *Annals of Mathematical Statistics* 42, 1957–1961.

Glosten, L.R., R. Jagannathan, & D.E. Runkle (1993) On the relation between the expected value and the volatility of the nominal excess returns on stocks. *Journal of Finance* 48, 1779–1801.

Hall, P. (1992) Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *Annals of Statistics* 20, 675–694.

Hall, P. & Q. Yao (2003) Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica* 71, 285–317.

Jorion, P. (1996) Risk^2: Measuring the risk in value-at-risk. *Financial Analysts Journal* 52, 47–56.

Jorion, P. (2000) *Value at Risk: The New Benchmark for Managing Financial Risk*, 2nd ed. McGraw-Hill.

Koul, H. & S. Ling (2006) Fitting an error distribution in some heteroscedastic time series models. *Annals of Statistics* 34, 994–1012.

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

Lee, S.W. & B.E. Hansen (1994) Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory* 10, 29–52.

Ling, S. (2007) Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models. *Journal of Econometrics* 140(2), 849–873.

Lumsdaine, R.L. (1996) Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models. *Econometrica* 64, 575–596.

McNeil, A.J. & R. Frey (2000) Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance* 7, 271–300.

Peng, L. & Q. Yao (2003) Least absolute deviations estimator for ARCH and GARCH models. *Biometrika* 90, 967–975.

https://doi.org/10.1017/S0266466608080559 Published online by Cambridge University Press

Scaillet, O. (2004) Nonparametric estimation and sensitivity analysis of expected shortfall. *Mathematical Finance* 14, 115–129.
Scott, D.W. (1992) *Multivariate Density Estimation. Theory, Practice and Visualization*. Wiley.
Serfling, R. (1980) *Approximation Theorems of Mathematical Statistics*. Wiley.
Shorack, G.R. (1974) Random means. *Annals of Statistics* 2, 661–675.
Vyazilov, A.E. (1999) The residual empirical distribution function in GARCH(1,1) and its application in hypothesis testing. *Russian Mathematical Surveys* 54, 856–857.

APPENDIX A

A.1. Proof of Theorem 3.1. We complete the proof in two steps. First we will prove that $(\hat{\varepsilon}_\alpha, \hat{\theta}_n)$ or $(\hat{\varepsilon}^0_\alpha, \hat{\theta}_n)$ is asymptotically multivariate normal. Then we will show the other results associated with $\hat{\sigma}_{n+1}$ and $\hat{V}^{\alpha}_{n+1}$ $(\hat{V}^{0\alpha}_{n+1})$ with the delta method.

The key of the first step is to derive the Bahadur representation of $\hat{\varepsilon}_\alpha$ based on Ghosh (1971),

$$\hat{\varepsilon}_\alpha = \varepsilon_\alpha + \frac{F(\varepsilon_\alpha) - \hat{F}_n(\varepsilon_\alpha)}{f(\varepsilon_\alpha)} + o_p\left(\frac{1}{\sqrt{n}}\right). \tag{A.1}$$

Let $t \in \mathbf{R}$, $\varepsilon_{nt} = \varepsilon_\alpha + tn^{-1/2}$, $Z_n(t) = \sqrt{n}[F(\varepsilon_{nt}) - \hat{F}_n(\varepsilon_{nt})]/f(\varepsilon_\alpha)$, and $U_n(t) = \sqrt{n}[F(\varepsilon_{nt}) - \hat{F}_n(\hat{\varepsilon}_\alpha)]/f(\varepsilon_\alpha)$.

Consider $E(Z_n(t) - Z_n(0))^2$,

$$E(Z_n(t) - Z_n(0))^2 \le \frac{2n}{f^2(\varepsilon_\alpha)}\left[E(F(\varepsilon_{nt}) - F(\varepsilon_\alpha) + F_n(\varepsilon_{nt}) - F_n(\varepsilon_\alpha))^2 + E\left(\frac{1}{2}(\varepsilon_{nt}f(\varepsilon_{nt}) - \varepsilon_\alpha f(\varepsilon_\alpha))(\hat{\theta}_n - \theta)^T\mathbf{e}\right)^2\right] + o(1)$$

$$= \frac{2P^*_n(1 - P^*_n)}{f^2(\varepsilon_\alpha)} + \frac{E(\varepsilon_1^4) - 1}{2f^2(\varepsilon_\alpha)}\,\mathbf{e}^T M^{-1}\mathbf{e} \times (\varepsilon_{nt}f(\varepsilon_{nt}) - \varepsilon_\alpha f(\varepsilon_\alpha))^2 + o(1),$$

where $P^*_n = F(\varepsilon_{nt}) - F(\varepsilon_\alpha)$ and we have used the result of Lemma 3.2. When $n \to \infty$, $\varepsilon_{nt} \to \varepsilon_\alpha$ and $P^*_n \to 0$; thus $E(Z_n(t) - Z_n(0))^2 \to 0$. It follows that $Z_n(t) - Z_n(0) = o_p(1)$.

Because $|\alpha - \hat{F}_n(\hat{\varepsilon}_\alpha)| \le n^{-1}$, we have

$$U_n(t) = \sqrt{n}[F(\varepsilon_{nt}) - \alpha + \alpha - \hat{F}_n(\hat{\varepsilon}_\alpha)]/f(\varepsilon_\alpha)$$

$$= \sqrt{n}[F(\varepsilon_{nt}) - \alpha]/f(\varepsilon_\alpha) + O(n^{-1/2})$$

$$= \sqrt{n}[f(\varepsilon_\alpha)tn^{-1/2} + O(n^{-1})]/f(\varepsilon_\alpha) + O(n^{-1/2}) \to t.$$

https://doi.org/10.1017/S0266466608080559 Published online by Cambridge University Press

Let $\zeta_n = \sqrt{n}(\hat{\varepsilon}_\alpha - \varepsilon_\alpha)$. Then, for any $t \in \mathbf{R}$ and $\delta > 0$,

$$\begin{aligned}
P(\zeta_n \le t, Z_n(0) \ge t + \delta) &= P(\hat{\varepsilon}_\alpha \le \varepsilon_\alpha + tn^{-1/2} = \varepsilon_{nt}, Z_n(0) \ge t + \delta) \\
&\le P(Z_n(t) \le U_n(t), Z_n(0) \ge t + \delta) \\
&\le P(Z_n(t) - Z_n(0) \le U_n(t) - t - \delta) \\
&\le P(|(Z_n(t) - Z_n(0)) - (U_n(t) - t)| \ge \delta) \\
&\le P(|Z_n(t) - Z_n(0)| + |U_n(t) - t| \ge \delta) \\
&\le P\left(|Z_n(t) - Z_n(0)| \ge \frac{\delta}{2}\right) + P\left(|U_n(t) - t| \ge \frac{\delta}{2}\right) \to 0.
\end{aligned}$$

Similarly, $P(\zeta_n \ge t + \epsilon, Z_n(0) \le t) \to 0$. Then by Lemma 1 of Ghosh (1971), we have $\zeta_n - Z_n(0) = o_p(1)$, which is just equation (A.1).

Now substituting the results of Lemmas 3.1 and 3.2, i.e., equations (3.1) and (3.4), in equation (A.1), we have

$$\begin{aligned}
\hat{\varepsilon}_\alpha &= \varepsilon_\alpha + \frac{F(\varepsilon_\alpha) - F_n(\varepsilon_\alpha) - \frac{1}{2}\varepsilon_\alpha f(\varepsilon_\alpha)(\hat{\theta}_n - \theta)^T \mathbf{e}}{f(\varepsilon_\alpha)} + o_p\left(\frac{1}{\sqrt{n}}\right) \\
&= \varepsilon_\alpha + \sum_{t=1}^{n} \frac{\alpha - I(\varepsilon_t \le \varepsilon_\alpha)}{nf(\varepsilon_\alpha)} - \frac{\varepsilon_\alpha}{2n}\left(M^{-1}\sum_{t=1}^{n}(\varepsilon_t^2 - 1)\sigma_t^{-2}U_t\right)^T \mathbf{e} + o_p\left(\frac{1}{\sqrt{n}}\right) \\
&= \varepsilon_\alpha + \sum_{t=1}^{n}\left\{\frac{\alpha - I(\varepsilon_t \le \varepsilon_\alpha)}{nf(\varepsilon_\alpha)} - \frac{\varepsilon_\alpha}{2n}(M^{-1}(\varepsilon_t^2 - 1)\sigma_t^{-2}U_t)^T \mathbf{e}\right\} + o_p\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

Thus, we have

$$\begin{pmatrix} \hat{\varepsilon}_\alpha - \varepsilon_\alpha \\ \hat{\theta}_n - \theta \end{pmatrix} = \begin{pmatrix} \sum_{t=1}^{n}\left\{\frac{\alpha - I(\varepsilon_t \le \varepsilon_\alpha)}{nf(\varepsilon_\alpha)} - \frac{\varepsilon_\alpha}{2n}(M^{-1}(\varepsilon_t^2 - 1)\sigma_t^{-2}U_t)^T \mathbf{e}\right\} \\ \frac{1}{n}\sum_{t=1}^{n}\{M^{-1}(\varepsilon_t^2 - 1)\sigma_t^{-2}U_t\} \end{pmatrix} + o_p\left(\frac{1}{\sqrt{n}}\right).$$

To prove the asymptotic normality of $(\hat{\varepsilon}_\alpha, \hat{\theta}_n)$ it is sufficient to prove the asymptotic normality of $k_0\sum_{t=1}^{n}[(\alpha - I(\varepsilon_t \le \varepsilon_\alpha))/nf(\varepsilon_\alpha)] + \mathbf{k}^T(1/n)\sum_{t=1}^{n}\{(M^{-1}(\varepsilon_t^2 - 1)\sigma_t^{-2}U_t)\} := (1/n)\sum_{t=1}^{n} x_t$ for any $(k_0, \mathbf{k}^T) \in \mathbb{R}^{p+q+2}$ by the Cramér–Wold device. Because $\{\varepsilon_t\}_{-\infty \le t \le \infty}$ is stationary and ergodic, $\sigma_t^{-2}U_t = G(\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots)$ for some measurable function G, and $E[x_t|\mathcal{F}_{t-1}] = 0$, $\{x_t\}_{-\infty \le t \le \infty}$ is a stationary, ergodic martingale difference sequence. Thus $(1/n)\sum_{t=1}^{n} x_t$ converges in distribution to the normal law according to Theorem 18.3 of Billingsley (1999). Then it is easy to apply the delta method to get the asymptotic variance of $\hat{\varepsilon}_\alpha$, $\hat{\sigma}_{n+1}$, and $\hat{V}_{n+1}^{\alpha}$.

https://doi.org/10.1017/S0266466608080559 Published online by Cambridge University Press

Finally, observing that $(1/n)\sum_{t=1}^{n}\hat{\varepsilon}_t = (1/n)\sum_{t=1}^{n}\varepsilon_t + o_p(1/\sqrt{n})$ and $\hat{\varepsilon}_\alpha^0 = \hat{\varepsilon}_\alpha - (1/n)\sum_{t=1}^{n}\hat{\varepsilon}_t$, we have

$$\hat{\varepsilon}_\alpha^0 = \varepsilon_\alpha + \sum_{t=1}^{n}\left\{\frac{\alpha - I(\varepsilon_t \le \varepsilon_\alpha)}{nf(\varepsilon_\alpha)} - \frac{\varepsilon_t}{n} - \frac{\varepsilon_\alpha}{2n}(M^{-1}(\varepsilon_t^2 - 1)\sigma_t^{-2}U_t)^T\mathbf{e}\right\} + o_p\left(\frac{1}{\sqrt{n}}\right).$$

Thus the results associated with $\hat{\varepsilon}_\alpha^0$ are direct consequences. ■

A.2. Proof of Lemma 3.3. Observe that

$$\hat{f}_n(\hat{\varepsilon}_\alpha) - f(\varepsilon_\alpha) = [\hat{f}_n(\hat{\varepsilon}_\alpha) - \hat{f}_n(\varepsilon_\alpha)] + [\hat{f}_n(\varepsilon_\alpha) - f_n(\varepsilon_\alpha)] + [f_n(\varepsilon_\alpha) - f(\varepsilon_\alpha)],$$

where $f_n(x) = (1/nh_n)\sum_{t=1}^{n}K((x - \varepsilon_t)/h_n)$. Because $\hat{f}_n(\hat{\varepsilon}_\alpha) - \hat{f}_n(\varepsilon_\alpha) = o_p(1)$ as a result of the continuity of $\hat{f}_n$, and $f_n(\varepsilon_\alpha) - f(\varepsilon_\alpha) = o_p(1)$ as a result of the consistency of kernel density estimators with i.i.d. observations, it is sufficient to prove that $\hat{f}_n(\varepsilon_\alpha) - f_n(\varepsilon_\alpha) = o_p(1)$.

Noticing that

$$|\hat{f}_n(\varepsilon_\alpha) - f_n(\varepsilon_\alpha)| = \left|\frac{1}{nh_n}\sum_{t=1}^{n}\left[K\left(\frac{\varepsilon_\alpha - \hat{\varepsilon}_t}{h_n}\right) - K\left(\frac{\varepsilon_\alpha - \varepsilon_t}{h_n}\right)\right]\right|$$

$$\le \frac{k}{nh_n^2}\sum_{t=1}^{n}|\hat{\varepsilon}_t - \varepsilon_t| = \frac{k}{nh_n^2}\sum_{t=1}^{n}\left|\frac{\sigma_t}{\hat{\sigma}_t} - 1\right||\varepsilon_t|$$

and

$$\left|\frac{\sigma_t}{\hat{\sigma}_t} - 1\right| = \left|\frac{1}{1 - \dfrac{\hat{\sigma}_t - \sigma_t}{\sigma_t}} - 1\right| = \left|\frac{\hat{\sigma}_t - \sigma_t}{\sigma_t}\right| + o\left(\left|\frac{\hat{\sigma}_t - \sigma_t}{\sigma_t}\right|\right)$$

$$= \left|\frac{(\hat{\theta}_n - \theta)^T U_t}{2\sigma_t^2} + o(|\hat{\theta}_n - \theta|)\right| + o\left(\left|\frac{(\hat{\theta}_n - \theta)^T U_t}{2\sigma_t^2} + o(|\hat{\theta}_n - \theta|)\right|\right)$$

$$\le \left|\frac{(\hat{\theta}_n - \theta)^T U_t}{2\sigma_t^2}\right|(1 + \delta_n) \quad \text{for some positive } \delta_n \to 0,$$

we have

$$|\hat{f}_n(\varepsilon_\alpha) - f_n(\varepsilon_\alpha)| \le \frac{k(1 + \delta_n)}{\sqrt{n}h_n^2}\sqrt{n}|\hat{\theta}_n - \theta|^T\left(\frac{1}{n}\sum_{t=1}^{n}\left|\frac{U_t\varepsilon_t}{2\sigma_t^2}\right|\right) = o_p(1).$$

The last equality follows because $\sqrt{n}h_n^2 \to \infty$, $\sqrt{n}|\hat{\theta}_n - \theta|^T\mathbf{1} = O_p(1)$, and $(1/n)\sum_{t=1}^{n}|(\varepsilon_t U_t^T\mathbf{1}/2\sigma_t^2)| = O_p(1)$ by the ergodic theorem because $\{(\varepsilon_t U_t^T\mathbf{1}/2\sigma_t^2)\}_{t\ge1}$ is a stationary sequence. Here $\mathbf{1} \equiv (1, 1, \ldots, 1)^T$. ■

https://doi.org/10.1017/S0266466608080559 Published online by Cambridge University Press

A.3. Proof of Theorem 3.3. Write $\hat{\mu}_\alpha$ as $(1/\alpha)\int_{-\infty}^{\hat{\varepsilon}_\alpha} x d\hat{F}_n(x)$. Then

$$\begin{aligned}\hat{\mu}_\alpha &= \mu_\alpha + \frac{1}{\alpha}\left(\int_{-\infty}^{\hat{\varepsilon}_\alpha} x d\hat{F}_n(x) - \int_{-\infty}^{\varepsilon_\alpha} x dF(x)\right)\\ &= \mu_\alpha + \frac{1}{\alpha}\left(\int_{-\infty}^{\varepsilon_\alpha} (F(x) - \hat{F}_n(x))\, dx + \alpha(\hat{\varepsilon}_\alpha - \varepsilon_\alpha) - \int_{\varepsilon_\alpha}^{\hat{\varepsilon}_\alpha} \hat{F}_n(x)\, dx\right)\\ &= \mu_\alpha + \frac{1}{\alpha}\left(\int_{-\infty}^{\varepsilon_\alpha} (F(x) - F_n(x) + F_n(x) - \hat{F}_n(x))\, dx\right) + o_p\left(\frac{1}{\sqrt{n}}\right).\end{aligned}$$

With the aid of Lemma 3.2 and noticing that

$$\int_{-\infty}^{\varepsilon_\alpha} F_n(x)\, dx = \int_{-\infty}^{\varepsilon_\alpha} \frac{\sum_{t=1}^{n} I(\varepsilon_t \le x)}{n}\, dx = \frac{1}{n}\sum_{t=1}^{n} (\varepsilon_\alpha - \varepsilon_t) I(\varepsilon_t \le \varepsilon_\alpha),$$

we have

$$\hat{\mu}_\alpha = \mu_\alpha + \sum_{t=1}^{n} \frac{(\varepsilon_t - \varepsilon_\alpha) I(\varepsilon_t \le \varepsilon_\alpha) - \int_{-\infty}^{\varepsilon_\alpha} F(x)\, dx}{n\alpha} - \frac{1}{2}\mu_\alpha(\hat{\theta}_n - \theta)^T \mathbf{e} + o_p\left(\frac{1}{\sqrt{n}}\right)$$

and

$$\begin{aligned}\hat{\mu}_\alpha^0 = \mu_\alpha + \sum_{t=1}^{n} \frac{(\varepsilon_t - \varepsilon_\alpha) I(\varepsilon_t \le \varepsilon_\alpha) - \int_{-\infty}^{\varepsilon_\alpha} F(x)\, dx - \alpha\varepsilon_t}{n\alpha} - \frac{1}{2}\mu_\alpha(\hat{\theta}_n - \theta)^T \mathbf{e}\\ + o_p\left(\frac{1}{\sqrt{n}}\right).\end{aligned}$$

By an argument similar to the argument in Theorem 3.1 we can easily derive the required results. ■

A.4. Asymptotic of Heavy-Tailed GARCH VaR and ES.

THEOREM A.1. *As in Peng and Yao (2003), suppose that the GARCH model is reparameterized such that the median of ε_t^2 is equal to 1 whereas $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) < \infty$, and $E(\varepsilon_t^4) = \infty$. Then the GARCH VaR and ES can be estimated a method similar to that introduced in Section 2 except that the GARCH parameters are estimated by the LAD method. In such a case, $\hat{\sigma}_{n+1}$, $\hat{\varepsilon}_\alpha$, $\hat{V}_{n+1}^\alpha$, $\hat{\mu}_\alpha$, and $\hat{ES}_{n+1}^\alpha$ are all $\sqrt{n}$-consistent and asymptotically normal, and the asymptotic variance of the VaR estimator is*

https://doi.org/10.1017/S0266466608080559 Published online by Cambridge University Press

$$\mathrm{Var}(\hat{V}_{n+1}^{\alpha}) = \mathrm{Var}(\hat{\sigma}_{n+1})(\varepsilon_\alpha)^2 + (\sigma_{n+1})^2\,\mathrm{Var}(\hat{\varepsilon}_\alpha) + 2\sigma_{n+1}\varepsilon_\alpha\,\mathrm{Cov}(\hat{\sigma}_{n+1}, \hat{\varepsilon}_\alpha), \tag{A.2}$$

$$\mathrm{Var}(\hat{ES}_{n+1}^{\alpha}) = \mathrm{Var}(\hat{\sigma}_{n+1})(\mu_\alpha)^2 + (\sigma_{n+1})^2\,\mathrm{Var}(\hat{\mu}_\alpha) + 2\sigma_{n+1}\mu_\alpha\,\mathrm{Cov}(\hat{\sigma}_{n+1}, \hat{\mu}_\alpha), \tag{A.3}$$

$$\mathrm{Var}(\hat{\sigma}_{n+1}) = \frac{1}{4ng^2(0)\sigma_{n+1}^2}\,U_{n+1}^T M^{-1} U_{n+1}, \tag{A.4}$$

$$\begin{aligned}\mathrm{Var}(\hat{\varepsilon}_\alpha) = {} & \frac{1}{4ng^2(0)}\,\varepsilon_\alpha^2 \mathbf{e}^T M^{-1}\mathbf{e} + \frac{\alpha(1-\alpha)}{nf^2(\varepsilon_\alpha)} \\ & + \frac{\varepsilon_\alpha \displaystyle\int_{-\infty}^{\varepsilon_\alpha} \mathrm{sgn}(\log(x^2)) f(x)\,dx}{ng(0)f(\varepsilon_\alpha)}\,\mathbf{e}^T M^{-1}\mathbf{e},\end{aligned} \tag{A.5}$$

$$\begin{aligned}\mathrm{Var}(\hat{\mu}_\alpha) = {} & \frac{1}{4ng^2(0)}\,\mu_\alpha^2 \mathbf{e}^T M^{-1}\mathbf{e} \\ & - \frac{\mu_\alpha \displaystyle\int_{-\infty}^{\varepsilon_\alpha} (x-\varepsilon_\alpha)\mathrm{sgn}(\log(x^2)) f(x)\,dx}{n\alpha g(0)}\,\mathbf{e}^T M^{-1}\mathbf{e} \\ & + \frac{\displaystyle\int_{-\infty}^{\varepsilon_\alpha} x^2 f(x)\,dx + \alpha(1-\alpha)\varepsilon_\alpha^2 - \alpha^2\mu_\alpha^2 - 2\alpha(1-\alpha)\varepsilon_\alpha\mu_\alpha}{n\alpha^2},\end{aligned} \tag{A.6}$$

$$\mathrm{Cov}(\hat{\sigma}_{n+1}, \hat{\varepsilon}_\alpha) = -\frac{U_{n+1}^T M^{-1}\mathbf{e}}{2n\sigma_{n+1}}\left(\frac{\varepsilon_\alpha}{2g(0)} + \frac{\displaystyle\int_{-\infty}^{\varepsilon_\alpha} \mathrm{sgn}(\log(x^2)) f(x)\,dx}{f(\varepsilon_\alpha)g(0)}\right), \tag{A.7}$$

$$\mathrm{Cov}(\hat{\sigma}_{n+1}, \hat{\mu}_\alpha) = -\frac{U_{n+1}^T M^{-1}\mathbf{e}}{2n\sigma_{n+1}}\left(\frac{\mu_\alpha}{2g(0)} - \frac{\displaystyle\int_{-\infty}^{\varepsilon_\alpha} (x-\varepsilon_\alpha)\mathrm{sgn}(\log(x^2)) f(x)\,dx}{\alpha g(0)}\right), \tag{A.8}$$

where sgn( ) *is the sign function and* $g(\cdot)$ *is the density function of* $\log(\varepsilon_t^2)$.

Proof. Notice that the LAD estimator of the GARCH model satisfies

$$\hat{\theta}_n - \theta = \frac{1}{ng(0)}\,M^{-1}\sum_{t=1}^{n}\mathrm{sgn}(\log(\varepsilon_t^2))\sigma_t^{-2}U_t + o_p(n^{-1/2}). \tag{A.9}$$

■

https://doi.org/10.1017/S0266466608080559 Published online by Cambridge University Press