Statistica Sinica **20** (2010), 1239-1256

SADDLE POINT APPROXIMATION AND VOLATILITY ESTIMATION OF VALUE-AT-RISK

Maozai Tian and Ngai Hang Chan

Renmin University of China and Chinese University of Hong Kong

Abstract: Value-at-Risk (VaR) is a commonly used risk measure adopted by financial engineers and regulators alike. Many of the techniques used in calculating VaR, however, rely on simulations that can be difficult and time consuming. One of the objectives of this paper is to conduct statistical inference for VaR based on saddle point approximation and volatility estimation. Specifically, by assuming that the loss distribution is a generalized hyperbolic, we propose a quasi-residual based volatility estimate. Because saddle point approximation furnishes a fast and accurate means to approximate the loss distribution and its percentiles, including the VaR in particular, it is then used to approximate the loss distribution of the quasi-residuals from which VaR can be estimated. Simulation studies and data analysis confirm that the proposed methodology works well both in theory and practice.

Key words and phrases: GARCH models, generalized hyperbolic distribution, heteroscedasticity, quasi-residuals, saddle point approximations, volatility estimate and value-at-Risk.

1. Introduction

One of the challenging tasks of modern risk management is fast and accurate calculation of the loss distribution so that value-at-risk (VaR) can be computed. The main objective of this study is to propose a fast and practical means for calculating the VaR when the return process $\{R_t\}$ of an underlying financial asset follows the conditional heteroscedastic model. Specifically, let S_t be the observed price of an asset at discrete time, $t = 1, \ldots, T$, and let $R_t = \log(S_t/S_{t-1})$ be the log returns of the asset. Consider the conditional heteroscedastic model

$$R_t = \sigma_t \varepsilon_t \,, \tag{1.1}$$

where σ_t is the volatility process which is assumed to be $\mathcal{F}_{t-1} = \sigma\{R_1, \ldots, R_{t-1}\}$ measurable, and $\{\varepsilon_t\}$ are assumed to be independently and identically distributed with $\mathrm{E}(\varepsilon_t)$=0 and $\mathrm{Var}(\varepsilon_t)$=1. No assumption has been imposed on the parametric form of the volatility process at this stage. Once an appropriate estimate $\hat{\sigma}_t$ of the volatility process $\{\sigma_t\}$ becomes available, the VaR of the return process can be easily calculated by virtue of the identity $V\hat{a}R_{p,t} = \hat{\sigma}_t q_p$, where q_p is the pth quantile of the innovation process $\{\varepsilon_t\}$.

As far as the modelling of the volatility process is concerned, there are two main approaches: parametric and nonparametric. Typical examples of the parametric approach includes the ARCH-GARCH family (see Chan et al. (2007) and Chan et al. (2009), Engle (1995), Eberlein, Kallsen and Kristen (2003)), the EGARCH (Nelson (1991)), and the QGARCH (Sentana (1995)). For a more comprehensive survey of this topic, see Rossi (1996) and the references therein. Roughly speaking, all these methods can be applied to estimate or to forecast the volatility at certain specified periods. They are not suitable for modelling unstable time series in the long run. For the nonparametric approach, there are the time-inhomogeneous volatility models (Mercurio and Spokoiny (2004)) and the nonparametric generalized hyperbolic distributions (Chen, Härdle and Spokoiny (2007)), among others. There are pros and cons for both approaches, see, for example, Engle and Manganell (2004), Chen, Härdle and Spokoiny (2007), and the references therein. In this paper, we propose a new quasi-residual based method for estimating the volatility process. This method imposes no assumption on the parametric form of the volatility process and avoids the difficulty of specifying the homogeneous intervals, as required in Mercurio and Spokoiny (2004)).

As for the innovations $\{\varepsilon_t\}$, they are usually assumed to be Gaussian or to follow some simple parametric form; here we take $\{\varepsilon_t\}$ to follow a generalized hyperbolic distribution (GH). One of the reasons for this assumption is that the GH provides a better fit to the observed log-returns than the Gaussian distributions. Figure 1 illustrates that the conditional Gaussian model fails to capture the semi-heavy-tailed nature of the extreme values (bigger than 95% or less than 5%) of the standardized log returns of foreign exchange rates between the Deutsch Mark and the US Dollar (DEM/USD) from 1979/12/01 to 1994/04/01 that was studied in Chen, Härdle and Spokoiny (2007). Table 1 further indicates that the generalized hyperbolic (GH) distribution seems to match the first four empirical moments better than the Gaussian distribution. Each given GH consists of five parameters, however. Their estimation can be time consuming and tricky. To circumvent this difficulty, in conjunction with the quasi-residuals, we propose to use the saddle point approximation of the GH density from which VaR can be efficiently calculated. It will be shown that this idea works well both in simulation studies and in the analysis of the foreign exchange data.

This paper is organized as follows. In Section 2, quasi-residuals based volatility modelling is introduced and properties of the proposed volatility process are examined. In Section 3, the saddle point approximations to VaR are derived, and confidence intervals are constructed for the estimator $\hat{VaR}_{p,t}$. In Section 4, simulation studies are conducted and data analysis of two foreign exchange rates are used to illustrate the newly proposed methodology.

Table 1. Comparisons for the nonparametric kernel, Gaussian and generalized hyperbolic distributions of the daily DEM/USD log returns between 1979/12/01 and 1994/04/01.

Distribution	mean	standard deviation	skewness	kurtosis
Empirical	0.007734	1.000171	-0.0560728	1.513616
Gaussian	0.000000	1.000000	0.0000000	0.000000
GH	0.004534	1.002781	-0.041999	1.627289

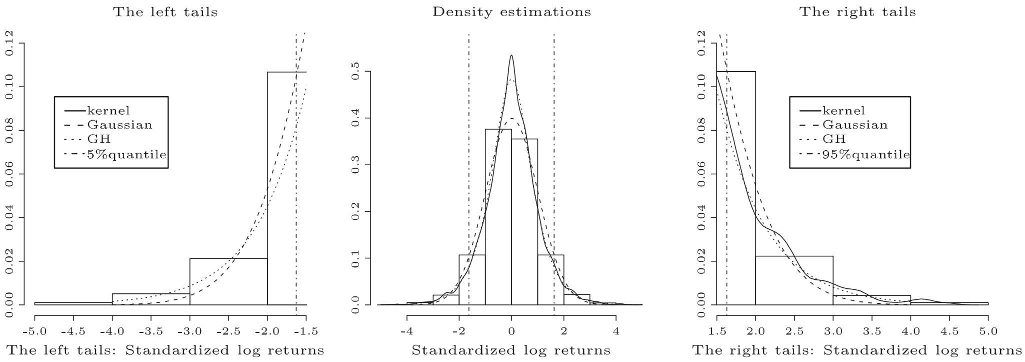

Figure 1. Comparison of density estimations of the log returns of DEM/USD exchange rates. The kernel density estimations are denoted by solid lines, the normal density estimations are indicated by the dashed lines, and the generalized hyperbolic distributions are denoted by dotted lines. The vertical lines that are denoted by dashed-dotted lines correspond to the quantiles at 5% on the left panel and 95% on the right, respectively.

2. Quasi-residuals and Volatility

It is well known that volatility modelling plays an important role in financial economics; statistical modelling of volatility has received considerable attention in both theoretical and empirical research. In this section, we develop a new method for estimating volatility using quasi-residuals that has been found to be a useful device in estimating heteroscedasticity, further details can be found in Müller and Stadtmüller (1987), Tian and Wu (2001), and Tian and Li (2004), and the references therein.

To introduce the quasi-residual idea, let q_p denote the pth-quantile of the distribution of ε_t, i.e., $\mathrm{P}(\varepsilon_t < q_p) = p$. Then $\mathrm{P}(R_t < \sigma_t q_p | \mathcal{F}_{t-1}) = p$. VaR is now defined as

$$VaR_{p,t} = \sigma_t q_p \, . \tag{2.1}$$

To obtain an estimate of $VaR_{p,t}$, we have to estimate the volatility σ_t and the quantile q_p. To this end, the notion of quasi-residual estimate for volatility turns out to be most relevant.

Definition 1. *The class of local quasi-residual based volatility estimates at time t is defined by*

$$\hat{\sigma}_t^2 \equiv \Big(\sum_{j=1}^{m} \omega_j R_{t-j}\Big)^2. \tag{2.2}$$

where $m > 1$ is a fixed integer, $\{\omega_1, \ldots, \omega_m\}$ are weights.

One of the distinctive features of the volatility process σ_t is that it varies little within a short time interval; although it is heteroscedastic in the long-run. The distinctive feature is known as "time homogeneity". It is therefore reasonable to assume that σ_t^2 can be locally approximated by a constant in a short-time interval $[t-m, t]$ for m small.

Next, we establish properties of the estimate $\widehat{\sigma}_t^2$, under general assumptions. Let $\mu_0 = \mathrm{E}(\varepsilon_t^2)$, $Z_t = \left\{\varepsilon_t^2 - \mathrm{E}(\varepsilon_t^2)\right\}/\mathrm{Var}^{1/2}(\varepsilon_t^2)$, $V_0 = \mathrm{Var}(\varepsilon_t^2)$, and $\Omega_m = \sum_{j=1}^m \omega_j$.

Assumption 1. *The volatility σ_t in* (1.1) *is a predictable process satisfying the condition that σ_t is $\mathscr{F}_{t-1}$ measurable, where $\mathscr{F}_{t-1} = \sigma(R_1, \ldots, R_{t-1})$, the $\sigma-$field generated by the first $t-1$ observations. Further, σ_t^2 is homogeneous in a short time interval $I = [t-m, t]$, for some m.*

Assumption 2. *$\{\varepsilon_t\}$ in* (1.1) *are independent and identically distributed variables with the generalized hyperbolic distribution prescribed in* (3.1).

Assumption 3. *In* (2.2)*, the positive weights $\{\omega_1, \ldots, \omega_m\}$ satisfy $\mu_0(\sum_{j=1}^m \omega_j)^2 = 1$.*

We have the following results regarding the volatility estimates. Theorem 1 shows that $\widehat{\sigma}_t^2$ is a conditional unbiased estimate and provides a closed form for the variance of the estimate. Theorem 2 offers a probability bound for the estimation error from which statistical testing for homogeneity can be conducted. Their proofs are given in the appendix.

Theorem 1. *Suppose Assumptions* 1 *and* 3 *hold. Then*

(1) $$\mathrm{E}(\widehat{\sigma}_t^2 | \mathscr{F}_{t-1}) = \sigma_t^2, \tag{2.3}$$

(2) $$\mathrm{Var}(\widehat{\sigma}_t^2 | \mathscr{F}_{t-1}) = V_0^2 \sigma_t^4 \Omega_m^4. \tag{2.4}$$

Theorem 2. *Let $\Sigma_I \equiv \sup\limits_{1 \le j < k \le m} |\sigma_{t-j}\sigma_{t-k} - \sigma_t^2|$. Under Assumptions* 1*–*3*, if the volatility process σ_t satisfies the condition $\delta\mu_0 V_0^{-1} \le \sigma_t^2 \le \delta\Delta\mu_0 V_0^{-1}$, for some positive constants δ and Δ, then there exists $\eta > 0$ such that, for every $\lambda \ge 1$,*

$$\mathrm{P}\Big\{|\widehat{\sigma}_t^2 - \sigma_t^2| > \Sigma_I + \lambda\mu_0^{-1} V_0 \sigma_t^2,\ \delta \le \mu_0^{-1} V_0 \sigma_t^2$$

$$\leq \delta\Delta\Big\} \leq 4\sqrt{e}\eta^{-1}\lambda(1+\log\Delta)\exp(-\frac{\lambda^2}{2\eta}). \tag{2.5}$$

Remark. The value of Σ_I can be viewed as a measure of departure from local homogeneity within the interval $I = [t-m, t]$. Theorem 2 indicates that if σ_t is homogeneous in the interval $I = [t-m, t]$, then the bias Σ_I is negligible. Consequently, a test for the homogeneity hypothesis in the interval $I = [t-m, t]$ for some $m > 0$ can be conducted. To perform the test, $I = [t-j, t]$ is split into two subintervals: Ξ and $I - \Xi$. If σ_t is homogeneous in I, then the estimates based on the two subintervals will be close. Further details of this idea can be found in Mercurio and Spokoiny (2004).

3. Saddle Point Approximations of VaR

From a statistical perspective, VaR is simply a quantile of the loss distribution. In what follows, we employ the saddle point approximation method for constructing a fast and accurate approximation to the tail of the loss distribution of assets. We also demonstrate how to obtain an accurate VaR without resorting to Monte Carlo simulations.

Suppose that $\{\varepsilon_t\}$ is an independent sequence of random variables with a generalized hyperbolic (GH) distribution, specified by five parameters $\boldsymbol{\theta} = (\lambda, \alpha, \beta, \delta, \mu)^{\mathrm{T}}$, with the probability density function

$$f_{GH}(x; \lambda, \alpha, \beta, \delta, \mu) = \frac{(\iota/\delta)^\lambda}{\sqrt{2\pi}K_\lambda(\delta\iota)} \cdot \frac{K_{\lambda-1/2}\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right)}{\left\{\sqrt{\delta^2+(x-\mu)^2}/\alpha\right\}^{1/2-\lambda}} \cdot e^{\beta(x-\mu)}. \tag{3.1}$$

Here $\mu \in R$ is the location parameter, $\alpha \in R$ is the shape parameter (kurtosis), $\beta \in R$, is the asymmetry parameter (skewness), $\delta \in R$ is the scale parameter, $\iota = \sqrt{\alpha^2 - \beta^2}$, $\lambda \in R$, and the modified Bessel function K_λ is given by

$$K_\lambda(\omega) = \frac{1}{2}\int_0^\infty x^{\lambda-1} e^{1/2\omega(x+x^{-1})}\, dx\,, \qquad \omega > 0.$$

Denote the MLE of the parameter vector $\boldsymbol{\theta} = (\lambda, \alpha, \beta, \delta, \mu)^{\mathrm{T}}$ by $\hat{\boldsymbol{\theta}} = (\widehat{\lambda}, \widehat{\alpha}, \widehat{\beta}, \widehat{\delta}, \widehat{\mu})^{\mathrm{T}}$.

3.1. Saddle point approximation

One of the main challenges in VaR is to find a fast and accurate means to compute the value VaR_t such that $\mathrm{P}(R_t > VaR_t|\mathcal{F}_{t-1}) = p$, where $0 < p < 1$. Note that

$$VaR\ (R_t|\mathcal{F}_{t-1}) = \sigma_t VaR\ (\varepsilon_t)\,. \tag{3.2}$$

Therefore, an estimator for VaR of R_t can be computed via

$$V\hat{a}R_{p,t} = \hat{\sigma}_t \hat{q}_p\,, \tag{3.3}$$

where $\hat{\sigma}_t$ is the quasi-residuals based volatility estimator given in (2.2), and $\hat{q}_p$ is the estimator for the p-quantile of the distribution of ε_t following (3.1) with the unknown parameter $\boldsymbol{\theta}$ being replaced by its maximum likelihood estimator $\hat{\boldsymbol{\theta}}$. Three methods for computing the estimate of $\hat{q}_p$ are widely used in the literature: enumerate the exact probabilities; use a normal density approximation; use brute force simulation. The first of these is usually intractable, the second may not result in the desired accuracy, and the third can be time consuming even with the speed of modern computers. Instead, we consider a saddle point approximation method that is shown to be highly accurate and efficient. To achieve this goal, we propose the following saddle point approximation algorithm.

1. Find the saddle point $s = \hat{t}$ such that $\kappa'_\varepsilon(\hat{t}) = t$, where $\kappa'(\cdot)$ is defined below.
2. Evaluate the pth quantile q_p of the distribution of $\{\varepsilon\}$ as

$$\begin{aligned} p &= \mathrm{P}(\varepsilon > t) \qquad (3.4) \\ &= \begin{cases} \exp\left\{\kappa_\varepsilon(\hat{t}) - \hat{t}t + \frac{1}{2}\hat{t}^2\kappa''_\varepsilon(\hat{t})\right\}\Phi\left(-\sqrt{\hat{t}^2\kappa''_\varepsilon(\hat{t})}\right), & t > \mathrm{E}(\varepsilon), \\ \frac{1}{2}, & t = \mathrm{E}(\varepsilon), \\ 1 - \exp\left\{\kappa_\varepsilon(\hat{t}) - \hat{t}t + \frac{1}{2}\hat{t}^2\kappa''_\varepsilon(\hat{t})\right\}\Phi\left(-\sqrt{\hat{t}^2\kappa''_\varepsilon(\hat{t})}\right), & t < \mathrm{E}(\varepsilon), \end{cases} \end{aligned}$$

where $\Phi(\cdot)$ denotes the cumulative normal distribution function. In (3.4),

$$\begin{aligned} \mathrm{E}(\varepsilon) &= \mu + \frac{\delta\beta}{\iota}\cdot\frac{K_{\lambda+1}(\delta\iota)}{K_\lambda(\delta\iota)}, \\ \kappa(z) &= \mu z + \log\iota^\lambda - \lambda\log\iota_z + \log K_\lambda(\delta\iota_z) - \log K_\lambda(\delta\iota), \\ \kappa'(z) &= \mu + \frac{\delta(\beta+z)}{\iota_z}\cdot\frac{K_{\lambda+1}(\delta\iota_z)}{K_\lambda(\delta\iota_z)}, \\ \kappa''(z) &= \frac{\delta}{\iota_z}\cdot\frac{K_{\lambda+1}(\delta\iota_z)}{K_\lambda(\delta\iota_z)} + \frac{\delta^2(\beta+z)^2}{\iota_z^2}\cdot\frac{K_{\lambda+2}(\delta\iota_z)}{K_\lambda(\delta\iota_z)} - \frac{\delta^2(\beta+z)^2}{\iota_z^2}\cdot\frac{K^2_{\lambda+1}(\delta\iota_z)}{K^2_\lambda(\delta\iota_z)}. \end{aligned}$$

(Once $\kappa_\varepsilon(s), \kappa'_\varepsilon(s)$ and $\kappa''_\varepsilon(s)$ are calculated, it is straightforward to calculate the VaR. First use expression (3.4) with t being replaced by $\kappa'_\varepsilon(\hat{t})$ and then adjust $\hat{t}$ until the right-hand side of (3.4) equals a given p. Note that this step is a simple root-finding problem. After obtaining $\hat{t}$, it is easy to calculate the value t, which is labelled as $\hat{q}_p$.)

3. Calculate the VaR of the return process R_t by means of $V\hat{a}R_{p,t} = \hat{\sigma}_t\widehat{q_p}$.

3.2. Confidence intervals for VaR

In this subsection, two ways to construct confidence intervals for the VaR are considered. The Wald-type confidence interval based on MLE (WM), and the saddle point approximation confidence interval (SA).

For the Wald-type confidence intervals, consider the log-likelihood function. Using the result of Pawitan (2001), we have the following.

Theorem 3. *Under Assumptions* 1–3 *and* (3.3)*, as* $n \to \infty$*, we have*

$$\left\{\mathrm{Var}(\log V\hat{a}R_{p,t})\right\}^{-1/2} \cdot \log \frac{V\hat{a}R_{p,t}}{VaR_{p,t}} \xrightarrow{\mathbb{D}} \mathrm{N}(0,1)\,. \tag{3.5}$$

Hence, a $100(1-\alpha)\%$ confidence interval for VaR is

$$\left[V\hat{a}R_{p,t}\exp\left\{-z_{\alpha/2}\sqrt{\mathrm{Var}(\log V\hat{a}R_{p,t})}\right\}, V\hat{a}R_{p,t}\exp\left\{z_{\alpha/2}\sqrt{\mathrm{Var}(\log V\hat{a}R_{p,t})}\right\}\right], \tag{3.6}$$

where z_q is the $100q$th upper percentile of the standard normal distribution.

The Wald-type confidence interval considered is based on large sample theory and its performance under small sample sizes remains to be determined. As an alternative, consider the saddle point approach to approximating the tail probability of the distribution. It is well known that a saddle point approximation provides a good approximations to the tail probabilities or to the density in the tail of the distribution (see Daniels (1954, 1987) and Jensen (1995)).

Theorem 4. *Let* $\hat{F}_{sd}(VaR|p)$ *be the saddle point approximation function of the cumulative distribution function* $F(VaR|p)$ *of* R_t*, i.e.,* $F(VaR|p) = P(R_t < VaR|\mathcal{F}_{t-1}) = p$*. Let* $0<\alpha<1$ *be a fixed value. For a given* VaR*, let* $\hat{F}_{sd}(VaR_U(p)|p)) = 1-\alpha/2$ *and* $\hat{F}_{sd}(VaR_L(p)|p) = \alpha/2$*. Then the interval* $[VaR_L(p), VaR_U(p)]$ *is a* $(1-\alpha)\%$ *confidence interval for* VaR*.*

Proof. The proof directly follows from Tian, Tang and Chan (2008).

4. Monte Carlo Simulations and Applications

4.1. Simulations

We first evaluate the performance of the saddle point approximation in conjunction with the quasi-residuals volatility estimates by means of Monte Carlo simulations. Specifically, the following algorithm is computed.

1. Find an estimator $\hat{\sigma}_t$ for the volatility process using (2.2).
2. Estimate the GH parameters based on $\{R_t/\hat{\sigma}_t\}$ using the MLE.
3. Calculate the p-quantile of the value $\hat{q}_p$ based on the saddle point approximation method (3.4).
4. Calculate $V\hat{a}R_t = \hat{\sigma}_t\hat{q}_p$.

Observe that by (3.3), the accuracy of the VaR estimate depends on two factors: the accuracy of the estimate of the pth quantile q_p and the accuracy of predicting the volatility process $\hat{\sigma}_t$. We focus on estimating the volatility. To examine the performance of the quasi-residuals method, two estimators of the first step are considered in this simulation study: the quasi-residuals based estimator and the GARCH (1,1) based estimator, see for example McNeil, Frey and Embrechts (2005). Three sets of weights ($m = 3, 4, 5$) are used in (2.2) and two typical models (see Mercurio and Spokoiny (2004)) for the volatility processes are considered.

I. *Small jumps model:*

$$\sigma_1(t) = \begin{cases} 0.1, & 1 \leq t \leq 120, \\ 0.2, & 120 < t \leq 240, \\ 0.1, & 240 < t \leq 360. \end{cases}$$

II. *High frequency model:*

$$\sigma_2(t) = \begin{cases} |0.001t^2 - 7|, & 1 \leq t \leq 120, \\ |0.007t - 0.2|, & 120 < t \leq 240, \\ |0.002t - 0.5|, & 240 < t \leq 360. \end{cases}$$

Consider the generalized hyperbolic distribution with $\lambda = 1$ in (3.1). Two hyperbolic return series were generated by multiplying the 360 simulated generalized hyperbolic random variables by the volatility processes $\sigma_1(t)$ and $\sigma_2(t)$. Specifically, for model I, we generated $\varepsilon^i(1), \ldots, \varepsilon^i(360)$ from the hyperbolic distribution with parameters $\alpha = 1, \beta = 0, \delta = 1, \mu = 0$, and $\sigma_1(1), \ldots, \sigma_1(360)$ from model I. Then we computed $r_1^i(t) = \sigma_1(t)\varepsilon^i(t),\ t = 1, \ldots, 360$. Next we repeated this step 5,000 times independently to obtain the series $\{r_1^i(t) : t = 1, \ldots, 360 \text{ and } i = 1, \ldots, 5,000\}$. Similarly, we computed $\sigma_2(1), \ldots, \sigma_2(360)$ from model II to obtain $r_2^i(1), \ldots, r_2^i(360)$, $i = 1, \ldots, 5,000$.

For a criterion independent of the scale of $\sigma(t)$, consider the relative error criterion defined by

$$\sum_{t=1}^{360} \sum_{i=1}^{5,000} \left(\frac{\hat{\sigma}_{ti} - \sigma_t}{\sigma_t} \right)^2. \tag{4.1}$$

This was used by Mercurio and Spokoiny (2004). To simplify the presentation, we only report results for the quasi-residual method in Table 2 and Figure 2 for the case of $m = 4$. Other results are available from the authors upon request.

As mentioned earlier, Theorems 1−2 constitute the theoretical basis for choosing the parameter m and the weights ω_i. For this particular example, results

Table 2. Estimation relative errors for different weights.

Weight	$m = 3$	$m = 4$	$m = 5$
Model I	19,231	17,171	17,254
Model II	44,087	42,032	42,008

Table 3. Summary statistics of DEM/USD daily exchange rates and German bank portfolio (GBP) from 1979/12/01 to 1994/04/01.

DESCRIPTION	n	Mean	Std	Skewness	Kurtosis
DEM/USD	3,720	2.061	0.466	-0.408	0.810
GBP	3,720	0.610	0.097	0.309	0.573

in Table 2 indicate that $m = 4$ is the best choice. Figure 2 depicts the graphical comparison of volatility estimations based on the quasi-residual approach on the left side and the GARCH model on the right. The solid line represents the true volatility process for Models I and II (straight line), the empirical median process among all estimates (thick dotted lines), and the 90% confidence confidence bands (two dashed lines). From Figure 2, it is clear that the behavior of the quasi-residual based estimate was stable here, whereas the GARCH-based approach overestimated the volatility process.

4.2. An application

We now demonstrate the saddle point approximation approach in the calculation of the VaR of two foreign exchange rate data sets: DEM/USD exchange rates and a German bank portfolio. These are daily exchange rates between DEM/USD from 1979/12/01 to 1994/04/01. Each series consists of 3,720 observations and is available at `http://www.quantlet.org/mdbase/`. Table 3 gives descriptive statistics for the two data sets. As shown in Figure 1 and Table 1, the GH distribution fits the foreign exchange rates DEM/USD better than the conditional Gaussian models because it can capture heavy tails.

4.2.1. Comparisons of different volatility estimators

Comparisons of the proposed quasi-residual approach with five commonly used volatility estimation approaches is pursued in this subsection. The five approaches are the following.

1. The equally weighted moving average approach (Hendricks (1996)).
2. The exponentially weighted moving average approach with the decay factor $\lambda = 0.97$ (Hendricks (1996)).
3. RiskMetrics with the decay factor $\lambda = 0.94$ (Morgan (1996)).

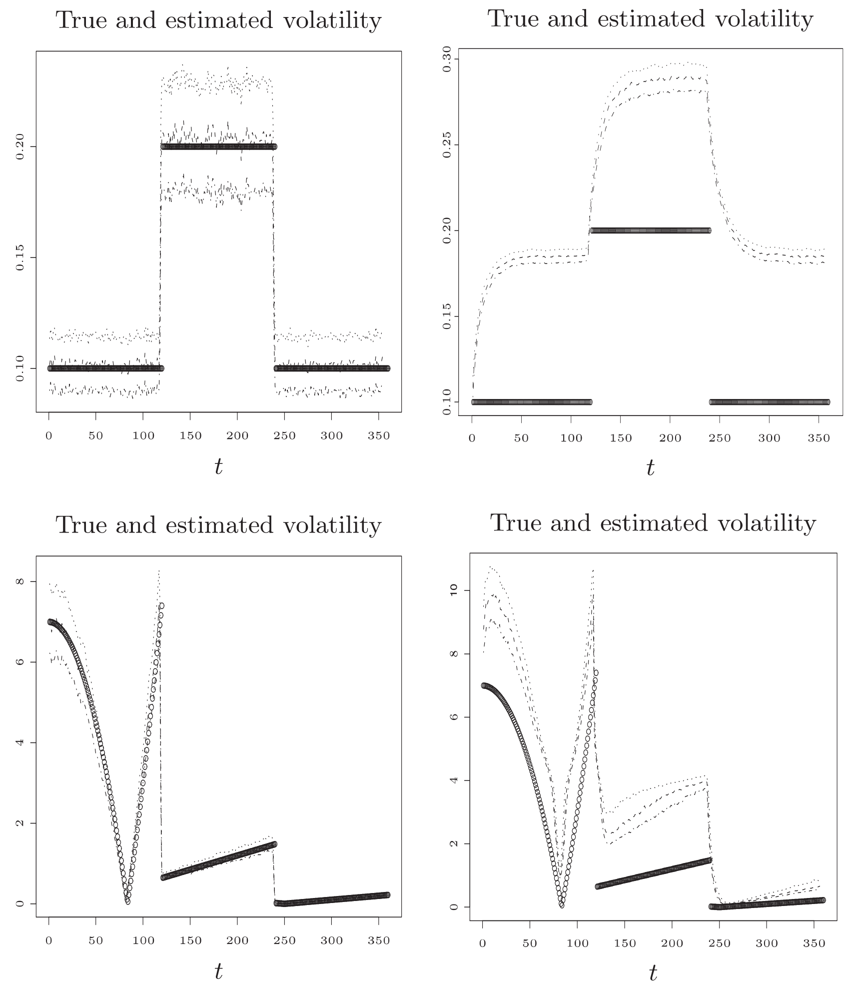

Figure 2. Graphical comparison of volatility estimations based on the quasi-residual approach on the left side, and the GARCH model on the right. The solid line represents the true volatility process. The thick dotted line is the median of all estimates. The two dashed lines are the lower and upper confidence bounds. The area between the lower and upper confidence bounds is a 90% point wise confidence band. The left column is the result of using quasi-residual method with $m = 4$. The right column corresponds to using the GARCH(1,1) method in estimating the volatility process.

4. Historical simulation (Hendricks (1996)) in which the estimation of volatility is defined as the sample standard deviation of the return process for the past 500 days.
5. The GARCH(1,1) model (Engle (1995)) using the quasi-maximum likelihood method in the case of estimating the volatility with a holding period of 1 day.

The initial period was set to $t_0 = 500$ for both series. To compare the performance of the different volatility estimators, the following three measures (see for example, Mercurio and Spokoiny (2004) and Fan and Gu (2003)) were adopted.

I. Empirical Mean Forecast Deviations (EMFD). Since $\mathrm{E}(R_{t+1}^2|\mathscr{F}_t) = \sigma_{t+1}^2$, for a given forecast $\widehat{\sigma}_{t+1|t}^2$, the empirical mean value of $|R_{t+1}^2 - \widehat{\sigma}_{t+1|t}^2|^p$ can be used to measure the quality of this forecast. As a result, the criterion

$$\frac{1}{T - t_0 - 1} \sum_{T-t_0-1}^{T} |R_{t+1}^2 - \widehat{\sigma}_{t+1|t}^2|^p, \tag{4.2}$$

is used to evaluate forecasting performance, see also Mercurio and Spokoiny (2004). In this study, p was 0.5.

II. Mean Absolute Deviations Error (MADE).

$$MADE = \frac{1}{n} \sum_{t=T+1}^{T+n} |R_t^2 - \widehat{\sigma}_t^2|. \tag{4.3}$$

III. Square-root Absolute Deviations Error (SADE).

$$SADE = \frac{1}{n} \sum_{t=T+1}^{T+n} \left| |R_t| - \sqrt{\frac{2}{\pi}} \widehat{\sigma}_t \right|. \tag{4.4}$$

More explanations on the motivations of choosing MADE and SADE as measures can be found in Fan and Gu (2003).

It is seen from Table 4 that the smallest values of EMFD, MADE and SADE were always attained by the quasi-residual saddle point approach. Based on these experiments, it is reasonable to argue that the quasi-residual approach performed the best among the commonly used competitors. On the other hand, the equally weighted moving average approach performed the worst in our study.

4.2.2. Comparisons of different VaR estimators

We compared six VaR estimators generated from the six approaches discussed in Section 4.2.1. Again, the daily returns of DEM/USD and GBP were used. We examined the absolute deviations errors (ADE), defined as the absolute difference between the p-value and a given confidence level. To compute the p-value, the following criterion (FM) was used.

$$FM = \frac{1}{n} \sum_{t=T+1}^{T+n} I(R_t < \widehat{q}_p \widehat{\sigma}_t). \tag{4.5}$$

Table 4. Comparisons of six volatility estimation methods for the three measures.

Index	Methods	EMFD ($\times 10^{-3}$)	MADE ($\times 10^{-5}$)	SADE ($\times 10^{-3}$)
	Equally	6.694	5.583	3.735
	Exponentially	6.457	5.373	3.597
DEM	RiskMetrics	6.376	5.335	3.560
/USD	Historical	6.679	5.562	3.725
	GARCH(1,1)	6.481	5.383	3.587
	Quasi-residuals	6.254	5.307	3.540
	Equally	6.627	5.546	3.722
	Exponentially	6.381	5.321	3.566
GBP	RiskMetrics	6.321	5.321	3.566
	Historical	6.614	5.528	3.714
	GARCH(1,1)	6.381	5.293	3.554
	Quasi-residuals	6.107	5.201	3.522

This quantity measures the number of events for which the loss exceeds the loss predicted by (3.3) at a given confidence level. The quantity FM is similar to the exceedance ratio at a given confidence level discussed in Fan and Gu (2003).

We report only the confidence levels 0.95 and 0.999 in Table 5. It is clear from this table that the ADE varied among the six approaches. But, at both confidence levels, the saddle point approach outperformed all its competitors. The improvement was even more conspicuous at an extreme quantile (0.001), where the saddle point method clearly distinguished itself from the other methods for both series.

5. Conclusion

A new method that combines quasi-residual estimation and the saddle point approximation is proposed for one-step ahead VaR forecasting. This method not only furnishes a fast and efficient means to calculate the VaR, as demonstrated by the simulated studies, it also allows one to conduct statistical inference.

Although some consider multi-step ahead VaR forecasting more useful, there exists empirical evidence suggesting that mult-step ahead forecast of VaR may not be that relevant at all; see for example Christoffersen and Diebold (2000), where it was argued that there is scant evidence of volatility prediction at horizons longer than ten days. Furthermore, one-step ahead forecasting of VaR is being conducted day in and day out on Wall Street and in major financial markets elsewhere. As multi-step ahead forecasting of VaR constitutes a well-known unsolved and challenging problem; see McNeil, Frey and Embrechts (2005), it

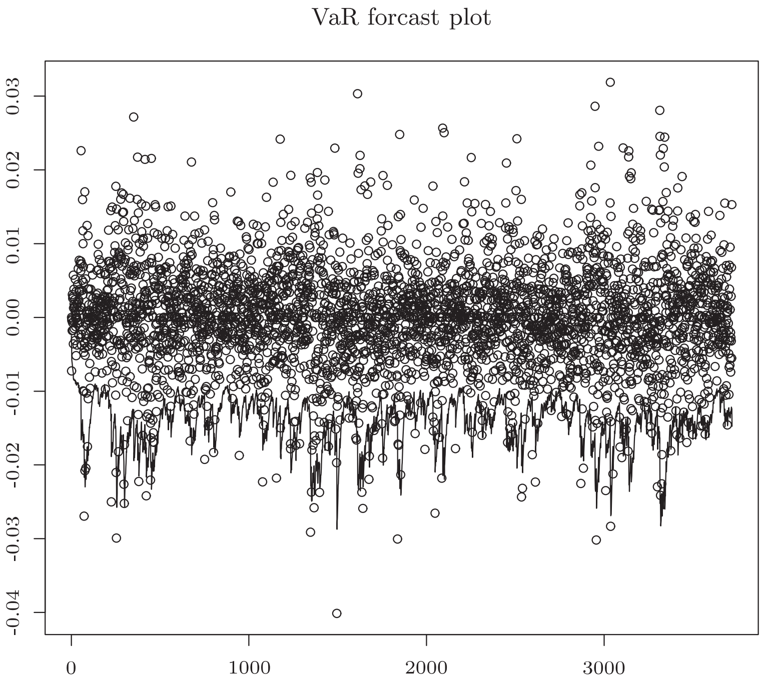

Figure 3. VaR forecast for the exchange rate DEM/USD at quantile $p = 0.025$. The circles are the log returns, the solid line is the VaR forecast based on the generalized hyperbolic distribution with parameters $\hat{\lambda} = 0.9845, \hat{\alpha} = 1.74385, \hat{\beta} = -0.01856, \hat{\delta} = 0.78177, \hat{\mu} = 0.01194$, which are the MLE based on the saddle point approximations method.

would be interesting to see how well the saddle point approximation would work in this context. Results in this paper furnish an intermediate step to solving this general, albeit much more difficult and challenging problem.

Acknowledgements

The authors gratefully acknowledge Professor Wolfgang Härdle for his helpful comments and for providing the data sets. Helpful comments from the Co-Editor, an associate editor and two referees are gratefully acknolwedged. This research was supported in part by HKSAR-RGC-GRF 400306, 400408 and the National Philosophy and Social Science Foundation Grants 07BTJ002 and NSFC-No.10871201.

Appendix

Proof of Theorem 1. Note that $R_t^2 = \mu_0\sigma_t^2 + V_0\sigma_t^2 Z_t$, which is a transform of (1.1). Obviously, this transformed model is also a heteroscedastic regression

Table 5. Comparisons of exceedance ratios of six VaR estimators using FM at two given confidence levels: $p_1 = 0.05$ and $p_2 = 0.001$.

Index	Methods	$p = 0.05$ p-value $(\times 10^{-2})$	$p = 0.05$ ADE-I $(\times 10^{-2})$	$p = 0.001$ p-value $(\times 10^{-3})$	$p = 0.001$ ADE-II $(\times 10^{-3})$
DEM /USD	Equally	4.536	0.464	5.592	4.592
	Exponentially	5.592	0.592	5.592	4.592
	RiskMetrics	5.592	0.592	6.213	5.213
	Historical	4.536	0.464	5.592	4.592
	GARCH(1,1)	4.908	0.092	4.970	3.970
	Saddle point	4.978	0.022	4.011	3.011
GBP	Equally	4.567	0.433	5.902	4.902
	Exponentially	4.753	0.247	2.796	1.796
	RiskMetrics	4.753	0.247	2.796	1.796
	Historical	4.567	0.433	5.902	4.902
	GARCH(1,1)	4.567	0.433	3.107	2.107
	Saddle point	4.981	0.019	2.545	1.545

* ADE-I and ADE-II denote the absolute deviations errors at nominal levels: 5% and 99.9%, respectively.

model. Consider the transformation of (2.2) as

$$\begin{aligned}
\widehat{\sigma}_t^2 &= \Big(\sum_{j=1}^{m} \omega_j R_{t-j}\Big)^2 \\
&= \sum_{j=1}^{m} \omega_j^2 R_{t-j}^2 + 2 \sum_{1\le j<} \sum_{k\le m} \omega_j \omega_k R_{t-j} R_{t-k} \\
&= \sum_{j=1}^{m} \omega_j^2 (\mu_0 \sigma_{t-j}^2 + V_0 \sigma_{t-j}^2 Z_{t-j}) + 2 \sum_{1\le j<} \sum_{k\le m} \omega_j \omega_k \sigma_{t-j} \sigma_{t-k} \epsilon_{t-j} \epsilon_{t-k} \\
&= \sum_{j=1}^{m} \omega_j^2 (\mu_0 \sigma_{t-j}^2 + V_0 \sigma_{t-j}^2 Z_{t-j}) + 2 \sum_{1\le j<} \sum_{k\le m} \omega_j \omega_k \sigma_{t-j} \sigma_{t-k} (V_0 Z_{t-j} + \mu_0) \\
&= \sum_{j=1}^{m} \omega_j^2 (\mu_0 \sigma_t^2 + V_0 \sigma_t^2 Z_t) + 2 \sum_{1\le j<} \sum_{k\le m} \omega_j \omega_k \sigma_t^2 (V_0 Z_t + \mu_0) \\
&= \mu_0 \sigma_t^2 \Omega_m^2 + V_0 \sigma_t^2 \Omega_m^2 Z_t.
\end{aligned}$$

The last approximation follows from Assumption 1. It follows that

(1) $\mathrm{E}(\widehat{\sigma}_t^2 | \mathscr{F}_{t-1}) = \mathrm{E}\Big(\mu_0 \sigma_t^2 \Omega_m^2 + V_0 \sigma_t^2 \Omega_m^2 Z_t \Big| \mathscr{F}_{t-1}\Big) = \mu_0 \sigma_t^2 \Omega_m^2.$

(2) $\mathrm{Var}(\widehat{\sigma}_t^2 | \mathscr{F}_{t-1}) = \mathrm{Var}\Big(\mu_0 \sigma_t^2 \Omega_m^2 + V_0 \sigma_t^2 \Omega_m^2 Z_t \Big| \mathscr{F}_{t-1}\Big) = V_0^2 \sigma_t^4 \Omega_m^4.$

The proof of Theorem 1 is complete.

Proof of Theorem 2. The proof of the theorem can be divided into three parts.

1. For $Z_t = \left\{\varepsilon_t^2 - \mathrm{E}(\varepsilon_t^2)\right\} / \mathrm{Var}^{1/2}(\varepsilon_t^2)$, there exist a constant $\eta > 0$, such that

$$\log \mathrm{E} \exp(sZ_t) \le \frac{\eta s^2}{2}. \tag{A.1}$$

   This can be established by direct calculation.

2. To show that the process

$$\mathbb{U}_t = \exp\left(\sum_{j=0}^{m} \sigma_{t-j} Z_{t-j} - \frac{\eta}{2}\sum_{j=0}^{m} \sigma_{t-j}^2\right) \tag{A.2}$$

   is a supermartingale, that is, $\mathrm{E}(\mathbb{U}_t|\mathscr{F}_{t-1}) \le \mathbb{U}_{t-1}$. To this end, we have

$$\begin{aligned}
&\mathrm{E}(\mathbb{U}_t|\mathscr{F}_{t-1}) - \mathbb{U}_{t-1}\\
&\quad = \mathrm{E}(\mathbb{U}_t|\mathscr{F}_{t-1}) - \mathrm{E}(\mathbb{U}_{t-1}|\mathscr{F}_{t-1})\\
&\quad = \mathrm{E}\bigg\{ \exp\Big(\sum_{j=0}^{m} \sigma_{t-j} Z_{t-j} - \frac{\eta}{2}\sum_{j=0}^{m}\sigma_{t-j}^2\Big)\\
&\qquad - \exp\Big(\sum_{j=1}^{m-1} \sigma_{t-j} Z_{t-j} - \frac{\eta}{2}\sum_{j=1}^{m-1}\sigma_{t-j}^2\Big)\Big|\mathscr{F}_{t-1}\bigg\}\\
&\quad = \mathrm{E}\bigg\{ \exp\Big(\sum_{j=1}^{m} \sigma_{t-j} Z_{t-j} - \frac{\eta}{2}\sum_{j=1}^{m}\sigma_{t-j}^2\Big)\Big\{\exp\Big(\sigma_t Z_t - \frac{\eta}{2}\sigma_t^2\Big) - 1\Big\}\Big|\mathscr{F}_{t-1}\bigg\}\\
&\quad = \Big(\prod_{j=1}^{m} \frac{\exp(\sigma_{t-j}Z_{t-j})}{\exp(\frac{\eta}{2}\sigma_{t-j}^2)}\Big)\mathrm{E}\bigg\{\exp\Big(\sigma_t Z_t - \frac{\eta}{2}\sigma_t^2\Big) - 1\Big|\mathscr{F}_{t-1}\bigg\}\\
&\quad \le 0.
\end{aligned}$$

   Assertion 2 immediately follows from (A.1).

3. It remains to show that

$$\begin{aligned}
&\mathrm{P}\Big\{|\hat{\sigma}_t^2 - \sigma_t^2| > \Sigma_I + \lambda\mu_0^{-1}V_0\sigma_t^2,\ \ \delta \le \mu_0^{-1}V_0\sigma_t^2 \le \delta\Delta\Big\}\\
&\quad = \mathrm{P}\Big\{\Big(\sum_{j=1}^{m}\omega_j R_{t-j}\Big)^2 - \sigma_t^2| > \Sigma_I + \lambda\mu_0^{-1}V_0\sigma_t^2,\ \ \delta \le \mu_0^{-1}V_0\sigma_t^2 \le \delta\Delta\Big\}\\
&\quad = \mathrm{P}\Big\{\sum_{j=1}^{m}\omega_j^2 R_{t-j}^2 + 2\sum_{1\le j<}\sum_{k\le m}\omega_j\omega_k R_{t-j}R_{t-k} - \sigma_t^2 > \Sigma_I + \lambda\mu_0^{-1}V_0\sigma_t^2,
\end{aligned}$$

$$\delta \le \mu_0^{-1} V_0 \sigma_t^2 \le \delta\Delta \Big\}$$

$$= \mathrm{P}\Big\{ \sum_{j=1}^{m} \omega_j^2 (\mu_0 \sigma_{t-j}^2 + V_0 \sigma_{t-j}^2 Z_{t-j}) + 2 \sum_{1\le j<} \sum_{k \le m} \omega_j \omega_k \sigma_{t-j} \sigma_{t-k} \epsilon_{t-j} \epsilon_{t-k} - \sigma_t^2$$

$$> \Sigma_I + \lambda \mu_0^{-1} V_0 \sigma_t^2, \quad \delta \le \mu_0^{-1} V_0 \sigma_t^2 \le \delta\Delta \Big\}$$

$$= \mathrm{P}\Big\{ \sum_{j=1}^{m} \omega_j^2 (\mu_0 \sigma_{t-j}^2 + V_0 \sigma_{t-j}^2 Z_{t-j}) + 2 \sum_{1\le j<} \sum_{k \le m} \omega_j \omega_k \sigma_{t-j} \sigma_{t-k} (V_0 Z_{t-j} + \mu_0) - \sigma_t^2$$

$$> \Sigma_I + \lambda \mu_0^{-1} V_0 \sigma_t^2, \quad \delta \le \mu_0^{-1} V_0 \sigma_t^2 \le \delta\Delta \Big\}$$

$$= \mathrm{P}\Big\{ \mu_0 \sum_{j=1}^{m} \omega_j^2 \sigma_{t-j}^2 + 2\mu_0 \sum_{1\le j<} \sum_{k \le m} \omega_j \omega_k \sigma_{t-j} \sigma_{t-k} + V_0 \sum_{j=1}^{m} \omega_j^2 \sigma_{t-j}^2 Z_{t-j}$$

$$+ 2V_0 \sum_{1\le j<} \sum_{k \le m} \omega_j \omega_k \sigma_{t-j} \sigma_{t-k} Z_{t-j} - \sigma_t^2 > \Sigma_I + \lambda \mu_0^{-1} V_0 \sigma_t^2,$$

$$\delta \le \mu_0^{-1} V_0 \sigma_t^2 \le \delta\Delta \Big\}$$

$$= \mathrm{P}\Big\{ \mu_0 \sum_{j=1}^{m} \omega_j^2 (\sigma_{t-j}^2 - \sigma_t^2) + 2\mu_0 \sum_{1\le j<} \sum_{k \le m} \omega_j \omega_k (\sigma_{t-j} \sigma_{t-k} - \sigma_t^2)$$

$$+ V_0 \sum_{j=1}^{m} \omega_j^2 \sigma_{t-j}^2 Z_{t-j} + 2V_0 \sum_{1\le j<} \sum_{k \le m} \omega_j \omega_k \sigma_{t-j} \sigma_{t-k} Z_{t-j} > \Sigma_I + \lambda \mu_0^{-1} V_0 \sigma_t^2,$$

$$\delta \le \mu_0^{-1} V_0 \sigma_t^2 \le \delta\Delta \Big\}$$

$$\le \mathrm{P}\Big\{ \mu_0 \sum_{j=1}^{m} \omega_j^2 \Sigma_I + 2\mu_0 \sum_{1\le j<} \sum_{k \le m} \omega_j \omega_k \Sigma_I + V_0 \sum_{j=1}^{m} \omega_j^2 \sigma_{t-j}^2 Z_{t-j}$$

$$+ 2V_0 \sum_{1\le j<} \sum_{k \le m} \omega_j \omega_k \sigma_{t-j} \sigma_{t-k} Z_{t-j} > \Sigma_I + \lambda \mu_0^{-1} V_0 \sigma_t^2,$$

$$\delta \le \mu_0^{-1} V_0 \sigma_t^2 \le \delta\Delta \Big\}$$

$$\le \mathrm{P}\Big\{ \Sigma_I + \Omega_m^2 V_0 \sigma_t^2 Z_t > \Sigma_I + \lambda \mu_0^{-1} V_0 \sigma_t^2, \;\; \delta \le \mu_0^{-1} V_0 \sigma_t^2 \le \delta\Delta \Big\}$$

$$\le \mathrm{P}\Big\{ \mu_0^{-1} V_0 \sigma_t^2 Z_t > \lambda \mu_0^{-1} V_0 \sigma_t^2, \; \delta \le \mu_0^{-1} V_0 \sigma_t^2 \le \delta\Delta \Big\}$$

$$\leq 4\sqrt{e}\eta^{-1}\lambda(1+\log\Delta)\exp(-\frac{\lambda^2}{2\eta}).$$

Theorem 2 now follows from the results of Lipster and Spokoiny (2000) and Mercurio and Spokoiny (2004).

References

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

Chan, N. H., Chen, J., Chen, X., Fan, Y. and Peng, L. (2009). Statistical inference for multivariate residual copula for GARCH models. *Statist. Sinica* **19**, 53-70.

Chen, Y., Härdle, W. and Spokoiny, V. (2007). Portfolio value at risk based on independent components analysis. *J. Comput. Appl. Math.* **205**, 594-607.

Christoffersen, P. and Diebold, F. (2000). How relevant is volatility forecasting for financial risk management? *Rev. Econom. Statist.* **82**, 12-22.

Daniels, H. E. (1954). Saddle point approximations in statistics. *Ann. Math. Statist.* **25**, 614-649.

Daniels, H. E. (1987). Tail probability approximations. *Internat. Statist. Rev.* **55**, 37-48.

Eberlein, E., Kallsen, J. and Kristen, J. (2003). Risk management based on stochastic volatility. *J. Risk* **5**, 19-44.

Engle, R. F., ed. (1995). *ARCH, Selected Readings.* Oxford University Press, Oxford.

Engle, R. F. and Manganell, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *J. Bus. Econom. Statist.* **22**, 233-234.

Fan, J. and Gu, J. (2003). Semiparametric estimation of Value at Risk. *Econometrics J.* **6**, 261-290.

Hendricks, D. (1996). Evaluation of value-at-risk models using historical data. *Federal Reserve Bank of New York Economic Policy Review*, 39-69.

Jensen, J. L. (1995). *Saddlepoint Approximations.* Oxford University Press, New York.

Lipster, R. and Spokoiny, V. (2000). Deviation probability bound for martingales with applications to statistical estimation. *Statist. Probab. Lett.* **46**, 347-357.

Mercurio, D. and Spokoiny, V. (2004). Statistical inference for time-inhomogeneous volatility models. *Ann. Statist.* **32**, 577-602.

Müller, H. and Stadtmüller, U. (1987). Estimation of heteroscedasticity in regression analysis. *Ann. Statist.* **15**, 610-625.

McNeil, A. J., Frey, R. and Embrechts, P. (2005). *Quantitative Risk Management: Concepts Techniques and Tools.* Princeton University Press, Princeton.

Morgan, J. P. (1996). *RiskMetrics Technical Document*, 4th edition. New York.

Nelson, D. B. (1991). Conditional heteroscedasticity in asset returns: A new approach. *Econometrica* **59**, 347-370.

Pawitan, Y. (2001). *In All Likelihood: Statistical Modelling and Inference Using Likelihood.* Oxford University Press, Oxford.

Rossi, P. E. (1996). *Modelling Stock Market Volatility: Bridging the Gap to Continuous Time.* Academic Press, San Diego.

Sentana, E. (1995). Quadratic ARCH models. *Rev. Econom. Stud.* **62**, 639-661.

Tian, M. Z. and Li, G. Y. (2004), Quasi-residuals method in sliced inverse regression. *Statist. Probab. Lett.* **66**, 205-211.

Tian, M. Z., Tang, M. L. and Chan, P. S. (2008). Confidence interval for epidemiologic rate based on saddle point approximations approach under inverse sampling. *Statist. Medicine* **27**, in press.

Tian, M. Z. and Wu, X. Z. (2001). A quasi-residuals method. *Adv. Math.* **30**, 182-184.

Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing, 100872, China.

E-mail: mztian@ruc.edu.cn

Department of Statistics, Chinese University of Hong Kong, Shatin, NT, Hong Kong.

E-mail: nhchan@sta.cuhk.edu.hk

(Received May 2008; accepted February 2009)