Interval Estimation of Value-at-Risk Based on GARCH Models with Heavy Tailed Innovations

Ngai Hang Chan[a 2], Shi-Jie Deng[b 1], Liang Peng[c 3] and Zhendong Xia[b 4]

[a] Department of Statistics, The Chinese University of Hong Kong, Hong Kong, China.

[b] School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA.

[c] School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332, USA.

Abstract

ARCH and GARCH models are widely used to model financial market volatilities in risk management applications. Considering a GARCH model with heavy-tailed innovations, we characterize the limiting distribution of an estimator of the conditional Value-at-Risk (VaR), which corresponds to the extremal quantile of the conditional distribution of the GARCH process. We propose two methods, the normal approximation method and the data tilting method, for constructing confidence intervals for the conditional VaR estimator and assess their accuracies by simulation studies. Finally, we apply the proposed approach to an energy market data set.

Key words and Phrases: Data tilting, GARCH models, heavy tail, tail empirical process, Value-at-Risk.

Classification: C13

[1] Corresponding author. School of Industrial and Systems Engineering, Georgia Institute of Technology. Email: deng@isye.gatech.edu. Tel: (404) 894-6519. Fax: (404) 894-2301.

[2] E-mail: nhchan@sta.cuhk.edu.hk.

[3] Email: peng@math.gatech.edu.

[4] E-mail: dengie@isye.gatech.edu.

1 Introduction

Two important empirical features about financial return series have drawn considerable attentions in the field of financial econometrics, namely, heteroscedasticity and heavy-tailed phenomenon. For example, the recent Séminaire Européen de Statistique reported in Finkenstädt and Rootzén (2004) consists of excellent reviewing articles on a variety of research topics related to these two features. As an attempt for capturing these stylized empirical findings in financial data, ARCH and generalized ARCH (GARCH) models were proposed to explicitly model the conditional second moments and their long-range dependence structure. The classical ARCH/GARCH models are based on conditional Gaussian innovations (see Engle (1982) and Bollerslev (1986)). They can be used to model risk attributes such as volatility clustering and the long-range dependence structure that exist in equity prices, financial indices, and foreign exchange rates (see Bollerslev et al. (1992) and Taylor (1986)).

There is a growing literature on applications of ARCH/GARCH models in asset pricing and risk management. With ubiquitous risks in financial markets, one of the most important tasks of financial institutions is to evaluate the exposure to market risks. This is commonly done by estimating the so-called Value-at-Risk (VaR). Market risks experienced during extreme market movements can cause dramatic changes in portfolio values. This can create huge profits or losses for financial institutions and may lead to financial pitfalls as demonstrated in the Long Term Capital Management case. VaR measures market risks by providing a single estimate of the worst possible financial loss to a portfolio over a fixed time horizon for a given confidence (or, probability) level (see Jorion (1997), Rachev and Mittnik (2000) and Duffie and Pan (1997) for a general introduction and exposition of VaR). Mathematically, VaR is defined as a quantile of a probability distribution, which is used to model an underlying portfolio value or its return. Financial institutions and regulators use VaR to quantify market risks and set capital reserves for market risks. For instance, traders at financial institutions often have their trading limits specified in terms of daily VaR of their trading books. Another appealing implication of VaR is that it can be utilized as a vehicle for corporate self-insurance since VaR can be interpreted as the amount of uninsured loss acceptable to a corporation (see Shimko (1997)). A corporation should buy external insurance when the self-insurance losses, as reflected by VaR measures, are greater than the cost of insurance by hedging.

In practice, a key risk measure for financial institutions based on the VaR concept is the conditional VaR, which is the worst possible loss due to adverse market movements over the next reporting period (e.g., a day or a week) *conditional* on current portfolio volatility and market information. This quantity corresponds to the tails of the conditional profit-and-loss (P&L) distribution of a portfolio. It is essentially the basis for setting portions of the day-to-day operating capital reserves for many financial institutions. As the GARCH models have been successfully applied in modeling the P&L distribution and the volatility structure of a portfolio of securities and other financial assets, the conditional VaR of a GARCH model becomes an important quantity to study. An additional important information of the conditional VaR is the robustness property of the conditional VaR estimator. When financial institutions utilize conditional VaR for setting capital reserves, they first need to estimate it based on some statistical models, either parametrically or non-parametrically. While a large amount of efforts have been focused on producing new and better conditional VaR estimates, two sources of errors may affect the estimation accuracy significantly: model mis-specification error and estimation error due to the inherent noise in the data. We address these two problems by considering non-parametric heavy-tailed distributions for the conditional innovations of a GARCH model. We then obtain the confidence intervals for the conditional VaR estimators of the heavy-tailed GARCH model. The knowledge of the confidence interval of the conditional VaR can be highly valuable in applications such as setting prudent capital reserve requirements for banks and conservative trading limits for traders or evaluating corporate self-insurance exposures by providing upper and lower bounds, rather than a single point estimate, of the VaR estimator at certain confidence level. For instance, in the 1996 amendment to the 1988 Basel Accord, a fudge factor of at least 3 is recommended for multiplying the historical VaR in setting the market risk capital requirements to ensure a safety margin for risk capitals. If the confidence interval bounds of VaR were employed instead, then it would provide a better understanding and justification for setting the safety margin.

Empirical evidence has demonstrated that the conditional normal time series models (e.g., the classical GARCH models) are inadequate in estimating the tail quantiles of conditional return distributions (see Danielssson and de Vries (1997) for instance). This prompts the gradual adoption of models with heavy-tailed innovations in risk modelling practice. Many extensions of the classical GARCH models with heavy-tailed innovations have been proposed. McNeil and Frey (2000) consider a GARCH model with generalized Pareto distributed innovations and propose a two-step approach to estimate the conditional VaR. While their idea seems intuitive, important statistical properties such as confidence interval estimation and asymptotic properties remain largely unexplored.

There are two main objectives in this paper. We first derive the limiting distribution of the extreme conditional VaR estimator in McNeil and Frey (2000). Instead of working within the framework of generalized Pareto distribution as in McNeil and Frey, we deal with the heavy-tailed innovations. In particular, besides the heavy-tailed feature, no specific parametric distributional assumptions on the GARCH innovations are imposed. A major advantage of this non-parametric approach is that it is applicable regardless of the true data-generating mechanism of the GARCH innovations, as long as it has heavy tails. As pointed out by Rachev and Mittnik (2000), one weakness of the VaR methodology comes from model (mis-)specification risk. With a non-parametric model, this model risk may be mitigated. Another advantage is that in addition to a VaR estimator, we can provide a VaR interval so that financial institutions can calculate the risks of loss exceeding the upper boundary of the VaR interval. This information amends the ability of quantifying and controlling risks.

Some existing work proposes to use the bootstrap for constructing confidence intervals of a conditional VaR estimator (for instance, Dowd (2002) and Christoffersen and Goncalves (2004)). Note that the bootstrap method is very computationally intensive because it requires repetitively solving non-linear optimizations in fitting GARCH models. Moreover, the bootstrap method fails when the innovations have infinite fourth moment, for example, when the innovation has a t-distribution with degrees of freedom 3 or 4. In this case, a subsample bootstrap method is needed (see Hall and Yao (2003a)). However, our methods are valid regardless of finite or infinite fourth moment of the innovation.

We propose two methods for constructing confidence intervals of a conditional VaR estimator developed from the extreme value theory. One is the traditional normal approximation method based on the asymptotic normality of the VaR estimator and the other one is the recent data tilting method studied in Hall and Yao (2003b) and Peng and Qi (2003).

The rest of the paper is organized as follows. In Section 2, we study the asymptotic behavior of the conditional VaR estimator by deriving its limiting distribution. We then present two methods for constructing confidence intervals for the conditional VaR. We perform a simulation study and test our approach on a real data set from energy commodity markets in Section 3, while proofs are given in Section 5. We conclude in Section 4.

2 Model Specification and Estimation Methodology

Suppose the data generating process for observations $\{X_t :\ t = \cdots, -1, 0, 1, 2, \cdots, n, \cdots\}$ follows a GARCH(p, q) model, namely,

$$X_t = \sigma_t \epsilon_t,\ \sigma_t^2 = c + \sum_{i=1}^{p} b_i X_{t-i}^2 + \sum_{j=1}^{q} a_j \sigma_{t-j}^2, \tag{1}$$

where $c > 0$, $b_1 \geq 0, \cdots, b_p \geq 0$, $a_1 \geq 0, \cdots, a_q \geq 0$ are constants, $\{\epsilon_t\}$ are a sequence of independent identically distributed random variables with mean 0 and variance 1 (i.e., $IID(0,1)$'s), and ϵ_t is independent of $\{X_{t-k}, k \geq 1\}$ for all t. Further assume that (1) uniquely defines a strictly stationary process with $EX_t^2 < \infty$, i.e.,

$$\sum_{i=1}^{p} b_i + \sum_{j=1}^{q} a_j < 1. \tag{2}$$

The 100α $(0 < \alpha < 1)$ percent one-step ahead conditional Value-at-Risk, based on observations $\{X_1, \cdots, X_n\}$, is defined as

$$x_{\alpha,n} = \inf\{x : P(X_{n+1} \leq x | X_{n+1-k}, k \geq 1) \geq \alpha\}.$$

It is a straightforward derivation from (1) that $x_{\alpha,n} = \sigma_{n+1} x_\alpha^0$, where x_α^0 is the $100\alpha\%$ quantile of ϵ_{n+1}. Our aim is to construct a confidence interval for the extreme conditional quantile $x_{\alpha,n}$ (i.e., $\alpha = \alpha(n)$ tends to zero or one as $n \to \infty$) through deriving the limiting distribution of an estimator of $x_{\alpha,n}$ and then applying two interval estimation methods.

2.1 Point Estimation

In this subsection, we study the asymptotic behavior of an estimator for $x_{\alpha,n}$ used in McNeil and Frey (2000) under the assumption that ϵ_t in (1) has heavy tails. Specifically, the distribution function G of ϵ_t satisfies

$$\lim_{x\to\infty} \frac{1 - G(xy)}{1 - G(x)} = y^{-\gamma} \; and \; \lim_{x\to\infty} \frac{G(-x)}{1 - G(x)} = d, \tag{3}$$

for all $y > 0$, where $\gamma > 2$ ensures that $E\epsilon_t^2 < \infty$ and d is some constant in $[0, \infty)$.

Note that (2) implies that

$$\begin{aligned}\sigma_t^2(a,b,c) &= \tfrac{c}{1-\sum_{j=1}^q a_j} + \textstyle\sum_{i=1}^p b_i X_{t-i}^2 \\ &+ \textstyle\sum_{i=1}^p b_i \sum_{k=1}^\infty \sum_{j_1=1}^q \cdots \sum_{j_k=1}^q a_{j_1}\cdots a_{j_k} X_{t-i-j_1-\cdots-j_k}^2,\end{aligned}$$

where $a=(a_1,\cdots,a_q)$ and $b=(b_1,\cdots,b_p)$. In practice we replace the above expression by a truncated version

$$\begin{aligned}\tilde\sigma_t^2(a,b,c) &= \tfrac{c}{1-\sum_{j=1}^q a_j} + \textstyle\sum_{i=1}^p b_i X_{t-i}^2 + \sum_{i=1}^p b_i \sum_{k=1}^\infty \sum_{j_1=1}^q \cdots \sum_{j_k=1}^q a_{j_1}\cdots a_{j_k} \\ &\times X_{t-i-j_1-\cdots-j_k}^2 I(t-i-j_1-\cdots-j_k\ge 1),\end{aligned}$$

where $I(\cdot)$ is an indicator function.

Define

$$L_\nu(a,b,c)=\sum_{t=\nu}^n\{X_t^2/\tilde\sigma_t^2(a,b,c)+\log\tilde\sigma_t^2(a,b,c)\},$$

where $\nu=\nu(n)\to\infty$ and $\nu/n\to 0$ as $n\to\infty$. Then the quasi maximum likelihood estimator of (a,b,c) is defined as

$$(\hat a,\hat b,\hat c)=\mathrm{argmin}_{(a,b,c)}L_\nu(a,b,c).$$

Set

$$\lambda_n=\begin{cases}\inf\{\lambda>0: nP(\epsilon_t^2\ge\lambda)\le 1\} & \text{if } \ 2<\gamma<4\\ \inf\{\lambda>0: nE(\epsilon_t^4 I(\epsilon_t^2\le\lambda))\le\lambda^2\} & \text{if } \ \gamma\ge 4.\end{cases}$$

Then, it follows from Hall and Yao (2003a) that

$$\hat a-a=O_p(n^{-1}\lambda_n),\quad \hat b-b=O_p(n^{-1}\lambda_n),\quad \hat c-c=O_p(n^{-1}\lambda_n).$$

Thus, ϵ_t can be estimated by $\hat\epsilon_t=X_t/\tilde\sigma_t(\hat a,\hat b,\hat c)$ for $t=\nu,\cdots,n$. Next we use $\hat\epsilon_t$'s to estimate x_α^0 as follows. We only deal with the case $\alpha=\alpha(n)\to 1$ as $n\to\infty$.

Let $\hat\epsilon_{m,1}\le\cdots\le\hat\epsilon_{m,m}$ denote the order statistics of $\hat\epsilon_\nu,\cdots,\hat\epsilon_n$ with $m\equiv n-\nu+1$.

Then γ can be estimated by the Hill estimator

$$\hat{\gamma} = \{\frac{1}{k}\sum_{i=1}^{k} \log \frac{\hat{\epsilon}_{m,m-i+1}}{\hat{\epsilon}_{m,m-k}}\}^{-1},$$

where $k = k(m) \to \infty$ and $k/m \to 0$ as $n \to \infty$ (see Hill (1975)). Replacing x, $1 - G(x)$, γ in (3) by $\hat{\epsilon}_{m,m-k}$, $\frac{1}{m}\sum_{i=1}^{n} I(\hat{\epsilon}_i > x)$ and $\hat{\gamma}$, respectively, we have $1-G(y\hat{\epsilon}_{m,m-k}) \sim \frac{k}{m}y^{-\hat{\gamma}}$. Since $1 - G(x_\alpha^0) = 1 - \alpha$, we solve $\frac{k}{m}y^{-\hat{\gamma}} = 1 - \alpha$ to obtain $y = (1-\alpha)^{-1/\hat{\gamma}}(\frac{k}{m})^{1/\hat{\gamma}}$, i.e., $x_\alpha^0 \sim y\hat{\epsilon}_{m,m-k}$. So we estimate x_α^0 by

$$\hat{x}_\alpha^0 = (1-\alpha)^{-1/\hat{\gamma}}(\frac{k}{m})^{1/\hat{\gamma}}\hat{\epsilon}_{m,m-k},$$

i.e.,

$$\hat{x}_{\alpha,n} = \tilde{\sigma}_{n+1}(\hat{a}, \hat{b}, \hat{c})\hat{x}_\alpha^0$$

is an estimator of $x_{\alpha,n}$.

Let $U(x)$ denote the inverse function of $\frac{1}{1-G(x)}$. Suppose there exists some function $A(x) \to 0$, as $x \to \infty$, such that

$$\lim_{x\to\infty} \frac{U(xy)/U(x) - y^{1/\gamma}}{A(x)} = y^{1/\gamma}\frac{y^\rho - 1}{\rho}, \tag{4}$$

for all $y > 0$, where $\rho < 0$.

The following result characterizes the limiting distribution of the estimator $\hat{x}_{\alpha,n}$.

Theorem 1. *Suppose (1), (2), (3), (4) and the conditions in Theorem 2.2 of Hall and Yao (2003a) hold. Assume*

$$k = k(m) \to \infty, k/m \to 0, \sqrt{k}A(m/k) \to 0, n^{-1}\lambda_n/A(m/k) \to 0, \log(\frac{k}{m(1-\alpha)})/\sqrt{k} \to 0$$

as $n \to \infty$. Then

$$\frac{\hat{\gamma}\sqrt{k}}{|\log(k/(m(1-\alpha)))|}\{\frac{\hat{x}_{\alpha,n}}{x_{\alpha,n}} - 1\} \xrightarrow{d} N(0,1),$$

i.e.,

$$\frac{\hat{\gamma}\sqrt{k}\log(\hat{x}_{\alpha,n}/x_{\alpha,n})}{|\log(k/(m(1-\alpha)))|} \xrightarrow{d} N(0,1).$$

Remark. As shown in Peng and Yao (2003), we can re-parameterize the model (1) in such a way that the median of ϵ_t^2 is equal to 1 while keeping $E(\epsilon_t) = 0$ unchanged. Under this new parametrization the parameters c and $b_1, \cdots, b_p$ differ from those in the old setting by a common positive constant factor while the parameters $a_1, \cdots, a_q$ remain unchanged. More importantly, the estimator $\hat{x}_{\alpha,n}$ remains the same, but now the parameters can be estimated with convergence rate $n^{-1/2}$ whenever $E\epsilon_t^4 = \infty$ or $< \infty$. Therefore, with this parameter estimation, the condition $n^{-1}\lambda_n/A(m/k) \to 0$ in Theorem 1 can be removed.

2.2 Interval Estimation

In this subsection we propose two methods to construct confidence intervals for the conditional VaR $x_{\alpha,n}$ as follows.

Method I: Normal approximation method. Based on Theorem 1 above, a confidence interval with level β for $x_{\alpha,n}$ is

$$I_\beta^n = (\hat{x}_{\alpha,n}\exp\{-z_\beta|\log\frac{k}{m(1-\alpha)}|/(\hat{\gamma}\sqrt{k})\}, \quad \hat{x}_{\alpha,n}\exp\{z_\beta|\log\frac{k}{m(1-\alpha)}|/(\hat{\gamma}\sqrt{k})\}),$$

where z_β satisfies $P(|N(0,1)| \le z_\beta) = \beta$.

Method II: Data tilting method. The general data tilting method was proposed by Hall and Yao (2003b) to tilt time series data which includes the empirical likelihood method as a special case. The empirical likelihood method, introduced in Owen (1988, 1990), is a nonparametric approach for constructing confidence regions. Like the bootstrap and the jackknife, the empirical likelihood method does not need to specify a family of distributions for the data. One of the advantages of empirical likelihood is that it enables the shape of a region, such as the degree of asymmetry in a confidence interval, to be determined automatically by the sample. In certain regular cases, confidence regions based on empirical likelihood are Bartlett correctable; see Hall & La Scala (1991) and DiCiccio et al. (1991). For a more complete disclosure of recent references and development we refer to the book by Owen (2001). As a generalization of the empirical likelihood method, the data tilting method not only has all of those nice properties of the empirical likelihood method, but also admits a wide range of distance functions. Recently Peng and Qi (2003) applied the data tilting method in Hall and Yao (2003b) to construct a confidence interval for the high quantile of a heavy tailed distribution based on iid observation. Here we apply the data tilting method in Peng and Qi (2003) to the estimated innovations as follows.

Define $\delta_i = I(\hat{\epsilon}_i \geq \hat{\epsilon}_{m,m-k})$. First, for any fixed $w = (w_\nu, \cdots, w_n)$ such that $w_i \geq 0$ and $\sum_{i=\nu}^n w_i = 1$, we solve

$$(\hat{\gamma}(w), \hat{c}(w)) = \mathrm{argmin}_{(\gamma,c)} \sum_{i=\nu}^{n} w_i \log((c\gamma\hat{\epsilon}_i^{-\gamma-1})^{\delta_i}(1 - c\hat{\epsilon}_{m,m-k}^{-\gamma})^{1-\delta_i}).$$

This results in

$$\hat{\gamma}(w) = \frac{\sum_{i=\nu}^n w_i\delta_i}{\sum_{i=\nu}^n w_i\delta_i(\log\hat{\epsilon}_i - \log\hat{\epsilon}_{m,m-k})}$$

and

$$\hat{c}(w) = \hat{\epsilon}_{m,m-k}^{\hat{\gamma}(w)} \sum_{i=\nu}^{n} w_i\delta_i.$$

Define

$$D_l(w) = \begin{cases} (l(1-l))^{-1}(1 - m^{-1}\sum_{i=\nu}^n (mw_i)^l) & \text{if } l \neq 0,1 \\ -m^{-1}\sum_{i=\nu}^n \log(mw_i) & \text{if } l = 0 \\ \sum_{i=\nu}^n w_i \log(mw_i) & \text{if } l = 1. \end{cases}$$

Next, solve

$$(2m)^{-1}L(x_{\alpha,n}) = \min_w D_l(w)$$

subject to

$$w_i \geq 0, \sum_{i=\nu}^{n} w_i = 1, \hat{\gamma}(w)\log(x_{\alpha,n}/(\tilde{\sigma}_{n+1}(\hat{a},\hat{b},\hat{c})\hat{\epsilon}_{m,m-k})) = \log((\sum_{i=\nu}^{n} w_i\delta_i)/(1-\alpha)).$$

Here we only consider the case $l = 1$ since other cases are similar and the case $l = 1$ gives good robustness properties. Put

$$A_1(\lambda_1) = 1 - \frac{m-k}{m} e^{-1-\lambda_1}, \quad A_2(\lambda_1) = A_1(\lambda_1) \frac{\log(x_{\alpha,n}/(\tilde{\sigma}_{n+1}(\hat{a}, \hat{b}, \hat{c})\hat{\epsilon}_{m,m-k}))}{\log(A_1(\lambda_1)/(1-\alpha))}.$$

Then, by the standard method of Lagrange multipliers, we have

$$w_i = \begin{cases} \frac{1}{m} e^{-1-\lambda_1}, & \text{if } \delta_i = 0 \\ \frac{1}{m} \exp\{-1 - \lambda_1 + \lambda_2 (\frac{\log(x_{\alpha,n}/(\tilde{\sigma}_{n+1}(\hat{a},\hat{b},\hat{c})\hat{\epsilon}_{m,m-k}))}{A_2(\lambda_1)} - \frac{1}{A_1(\lambda_1)} & \\ -\frac{A_1(\lambda_1)}{A_2^2(\lambda_1)} \log(\hat{\epsilon}_i/\hat{\epsilon}_{m,m-k}) \log(x_{\alpha,n}/(\tilde{\sigma}_{n+1}(\hat{a}, \hat{b}, \hat{c})\hat{\epsilon}_{m,m-k})))\}, & \text{if } \delta_i = 1, \end{cases}$$

where λ_1 and λ_2 satisfy

$$\sum_{i=\nu}^{n} w_i = 1 \quad \hat{\gamma}(w) \log(x_{\alpha,n}/(\tilde{\sigma}_{n+1}(\hat{a}, \hat{b}, \hat{c})\hat{\epsilon}_{m,m-k})) = \log(\sum_{i=\nu}^{n} w_i \delta_i/(1-\alpha)).$$

Theorem 2. *Under the conditions of Theorem 1,*

$$L(x_{\alpha,n}^0) \xrightarrow{d} \chi^2(1)$$

as $n \to \infty$, where $x_{\alpha,n}^0$ denotes the true value of $x_{\alpha,n}$.

Based on this theorem, a confidence interval with level β for $x_{\alpha,n}^0$ can be constructed as

$$I_\beta^t = \{x_{\alpha,n} : L(x_{\alpha,n}) \le u_\beta\},$$

where u_β is the β-level critical point of $\chi^2(1)$.

3 Simulation Study and Application

In this section, we investigate the finite sample behavior of our methods in constructing confidence intervals for the extreme conditional Value-at-Risk. We also apply the methodology to a real data set taken from the energy (e.g., electricity) markets.

3.1 Simulation Study

We draw 1,000 samples of size $1,000$ from GARCH(1,1) models with $c = 1.0, b_1 = 0.2, a_1 = 0.3$ and $c = 1.0, b_1 = 0.4, a_1 = 0.5$, respectively. We choose the errors ϵ_t to have Student's t distribution with degrees of freedom $d = 3, 5, 7, 9$. We truncate the likelihood functions defined in section 2.1 at $\nu = 20$. We compute the coverage probabilities of the confidence intervals of high quantile $\alpha = 0.99$ based on both method I (Normal approximation) and method II (data tilting) with a confidence level of 0.90. These coverage probabilities are plotted against different sample fractions $k = 20, 22, \cdots, 120$ in Figures 1-4 with the upper/lower plots corresponding to the low/high persistence cases, respectively. From Figures 1-4, the following are observed:

1. These two methods behave similarly while the normal approximation method appears to perform slightly better. This seems a bit surprising since the data tilting method is better than the normal approximation methods in general. This may be due to the fact that the data tilting method is much more sensitive to the accuracy of estimating innovations than the normal approximation method. This reasoning is confirmed by simulation studies under true innovations, i.e., using ϵ_t instead of $\hat{\epsilon}_t$ in these two methods, which are not reported here.

2. Both methods become accurate when d becomes large. This is because the tail probability is small for a large value of d, i.e., we only need to extrapolate data slightly to reach a high quantile when d is large.

3. In contrast to point estimation, the choice of k for interval estimation is more important. This is always a difficult task to handle both theoretically and practically. One way to deal with this issue is to develop more comparable different approaches and hope to be able to choose k such that intervals derived from those methods

are similar. Here we propose $k = 1.5(\log n)^2$, where n is the sample size. The coverage probabilities corresponding to this particular choice of k are plotted by the starred points in Figures 1–4, which indicate that such choice of k works for the cases studied here.

3.2 Application

We apply method I to a historical time series data set: the log returns of real time (RT) electricity locational marginal prices (LMPs) in the Pennsylvania-New Jersey-Maryland (PJM) power market from April 1998 to September 2003; see Figure 5. Electricity markets are relatively new markets. Electricity prices in the emerging power markets are much more volatile than prices in other financial markets due to the almost non-storable nature and the physical production characteristics of electricity.

In electricity markets, market participants such as utility companies are especially concerned about the risks of electricity prices rising too high, since they have natural short positions in electricity due to their obligations to provide electric power to customers. In this example, we examine the 99 percent VaR at the right tails (i.e., the positive-return side) of the conditional distribution of the 1-day electricity price return. Like the back test in McNeil and Frey (2000), we calculate $\hat{x}^t_{\alpha,n}$ on day t in the set $T = \{n, \cdots, N-1\}$ using a time window of $n = 500$ days each time where $N >> n$. We run our program for a high quantile $\alpha = 0.99$. For each day $t \in T$, we compute the confidence interval based on method I with confidence level 0.90 by taking $k = 30$ and $k = [1.5(\log n)^2] = 57$. Furthermore, we pay particular attention to the points which are less than the estimator $\hat{x}^t_{\alpha,n}$, but greater than its left endpoint of I^n_β because these points are perceived to have high risks even though they do not exceed our estimated conditional VaR. Thus, we may call the area between the estimator $\hat{x}^t_{\alpha,n}$ and its left endpoint of I^n_β a "risk-prone" region. Knowing the risk-prone region can be quite valuable in applications such as setting trading limits for traders or evaluating corporate self-insurance exposures since it provides bounds of the conditional VaR estimator at certain confidence level. We plot the log returns of PJM real time LMPs, the VaR estimator, and its confidence interval in Figure 6 and mark the points in the risk-prone region with squares. In a highly volatile market, conservative market participants may want to employ the interval estimation instead of the point estimator for their VaR estimation and then use the risk-prone region to monitor and control their risks.

4 Conclusions

In this paper, we derive the limiting distribution of a high conditional VaR estimator of a family of GARCH models with heavy-tailed innovations. With the limiting distribution, a traditional normal approximation method is proposed to construct a confidence interval of the conditional VaR estimator. An alternative method for constructing a confidence interval based on the data tilting method is proposed as well. Monte Carlo simulation studies with the GARCH models with Student-t innovations indicate that both methods yield valid confidence intervals for the VaR estimator while the normal approximation has a slightly higher coverage probability. Based on the confidence intervals, one can identify a risk-prone region, which is given by the area between the conditional VaR estimator and the left endpoint of its confidence interval. In practice, one should pay attention to this entire region since it signifies high risk scenarios even though individual points in the region may not exceed the estimated VaR threshold.

As a result of the non-parametric setting, the proposed methods are applicable to GARCH models with general innovations including those with asymmetric tails. For instance, they can be applied to asymmetric GARCH models as long as the tail balance assumption (3) holds. Future work is expected to extend these methods to other risk measures such as the expected shortfall probabilities with a broader class of time series models.

5 Proofs

Throughout this section we shall assume $p = q = 1$ since other cases can be shown in a similar way. Define

$$\hat{W}_n(u) = k^{-1/2} \sum_{t=\nu}^{n} \{I(1 - G(\hat{\epsilon}_t) \leq \frac{k}{m}u) - \frac{k}{m}u\}.$$

We first prove a lemma.

Lemma 1. *As $n \to \infty$,*

$$\hat{W}_n(u) \xrightarrow{d} B(u) \ \text{in}\ D[0,1],$$

where $D[0,1]$ denotes the space of functions on $[0,1]$ which is defined and equipped with the Skorokhod topology (see Billingsley (1968)) and $\{B(u), u \geq 0\}$ is a standard Brownian motion.

Proof. Define

$$\hat{\delta}_{n1} = n\lambda_n^{-1}(\hat{a}_1 - a_1), \quad \hat{\delta}_{n2} = n\lambda_n^{-1}(\hat{b}_1 - b_1), \quad \hat{\delta}_{n3} = n\lambda_n^{-1}(\hat{c} - c),$$

$$s_t(\delta_1, \delta_2, \delta_3) = [\tilde{\sigma}_t(a_1 + n^{-1}\lambda_n\delta_1, b_1 + n^{-1}\lambda_n\delta_2, c + n^{-1}\lambda_n\delta_3) - \sigma_t(a_1, b_1, c)]/\sigma_t(a_1, b_1, c),$$

$$E_{n1}(u, \delta_1, \delta_2, \delta_3) = k^{-1/2} \sum_{t=\nu}^{n} \{1 - G(U(\frac{m}{ku})(1 + s_t(\delta_1, \delta_2, \delta_3))) - \frac{k}{m}u\},$$

$$\begin{aligned} E_{n2}(u, \delta_1, \delta_2, \delta_3) \ &= k^{-1/2} \textstyle\sum_{t=\nu}^{n} \{I(\epsilon_t \geq U(\frac{m}{ku})(1 + s_2(\delta_1, \delta_2, \delta_3))) \\ &-(1 - G(U(\tfrac{m}{ku})(1 + s_t(\delta_1, \delta_2, \delta_3)))) + \tfrac{k}{m}u - I(\epsilon_t \geq U(\tfrac{m}{ku}))\} \end{aligned}$$

and

$$W_n(u) = k^{-1/2} \sum_{t=\nu}^{n} \{I(1 - G(\epsilon_t) \leq \frac{k}{m}u) - \frac{k}{m}u\}.$$

Since

$$\hat{W}_n(u) - W_n(u) = E_{n1}(u, \hat{\delta}_{n1}, \hat{\delta}_{n2}, \hat{\delta}_{n3}) + E_{n2}(u, \hat{\delta}_{n1}, \hat{\delta}_{n2}, \hat{\delta}_{n3}),$$

$\hat{\delta}_{n1} = O_p(1), \hat{\delta}_{n2} = O_p(1), \hat{\delta}_{n3} = O_p(1)$ and $W_n(u) \xrightarrow{D} B(u)$ in $D[0,1]$. To prove this lemma, it is sufficient to show that for any fixed $\Delta > 0$,

$$\sup_{-\Delta \le \delta_1, \delta_2, \delta_3 \le \Delta} \sup_{0 \le u \le 1} |E_{n1}(u, \delta_1, \delta_2, \delta_3)| = o_p(1) \tag{5}$$

and

$$\sup_{-\Delta \le \delta_1, \delta_2, \delta_3 \le \Delta} \sup_{0 \le u \le 1} |E_{n2}(u, \delta_1, \delta_2, \delta_3)| = o_p(1). \tag{6}$$

Define

$$\begin{aligned}
s_t^*(\Delta) \quad &= s_t(\Delta, \Delta, \Delta) \\
a_{nt}(u, \Delta) \quad &= I(\epsilon_t \ge U(\tfrac{m}{ku})(1 + s_t^*(\Delta))) \\
&\quad -(1 - G(U(\tfrac{m}{ku})(1 + s_t^*(\Delta)))) + \tfrac{k}{m}u - I(\epsilon_t \ge U(\tfrac{m}{ku})).
\end{aligned}$$

Let $N(n) = [M/A(m/k)]$ for any fixed $M > 0$, and $u_i = i/N(n)$, $i = 0, 1, \cdots, N(n)$. When $u \in [u_r, u_{r+1}]$, we have

$$\begin{aligned}
& k^{-1/2} \textstyle\sum_{t=\nu}^{n} a_{nt}(u, \Delta) \\
\le\ & k^{-1/2} \textstyle\sum_{t=\nu}^{n} a_{nt}(u_{r+1}, \Delta) \\
& + k^{-1/2} \textstyle\sum_{t=\nu}^{n} \{1 - G(U(\tfrac{m}{ku_{r+1}})(1 + s_t^*(\Delta))) - \tfrac{k}{m}u_{r+1}\} \\
& - k^{-1/2} \textstyle\sum_{t=\nu}^{n} \{1 - G(U(\tfrac{m}{ku_r})(1 + s_t^*(\Delta))) - \tfrac{k}{m}u_r\} \\
& + 2k^{-1/2} \textstyle\sum_{t=\nu}^{n} \{\tfrac{k}{m}u_{r+1} - \tfrac{k}{m}u_r\} \\
& + k^{-1/2} \textstyle\sum_{t=\nu}^{n} \{I(\epsilon_t \ge U(\tfrac{m}{ku_{r+1}})) - \tfrac{k}{m}u_{r+1} + \tfrac{k}{m}u_r - I(\epsilon_t \ge U(\tfrac{m}{ku_r}))\}
\end{aligned}$$

and

$$\begin{aligned}
& k^{-1/2} \textstyle\sum_{t=\nu}^{n} a_{nt}(u, \Delta) \\
\ge\ & k^{-1/2} \textstyle\sum_{t=\nu}^{n} a_{nt}(u_r, \Delta) \\
& + k^{-1/2} \textstyle\sum_{t=\nu}^{n} \{1 - G(U(\tfrac{m}{ku_r})(1 + s_t^*(\Delta))) - \tfrac{k}{m}u_r\} \\
& - k^{-1/2} \textstyle\sum_{t=\nu}^{n} \{1 - G(U(\tfrac{m}{ku_{r+1}})(1 + s_t^*(\Delta))) - \tfrac{k}{m}u_{r+1}\} \\
& + 3k^{-1/2} \textstyle\sum_{t=\nu}^{n} \{\tfrac{k}{m}u_r - \tfrac{k}{m}u_{r+1}\} \\
& + k^{-1/2} \textstyle\sum_{t=\nu}^{n} \{I(\epsilon_t \ge U(\tfrac{m}{ku_r})) - \tfrac{k}{m}u_r + \tfrac{k}{m}u_{r+1} - I(\epsilon_t \ge U(\tfrac{m}{ku_{r+1}}))\}.
\end{aligned}$$

Hence

$$
\begin{aligned}
& \sup_{0\le u\le 1} |k^{-1/2} \textstyle\sum_{t=\nu}^n a_{nt}(u,\Delta)| \\
\le\ & \sup_r |k^{-1/2} \textstyle\sum_{t=\nu}^n a_{nt}(u_r,\Delta)| \\
& +2\sup_r |k^{-1/2} \textstyle\sum_{t=\nu}^n \{1-G(U(\frac{m}{ku_r})(1+s_t^*(\Delta))) - \frac{k}{m}u_r\}| \\
& +3\sup_r k^{-1/2} \textstyle\sum_{t=\nu}^n \{\frac{k}{m}u_{r+1} - \frac{k}{m}u_r\} \\
& +\sup_r |k^{-1/2} \textstyle\sum_{t=\nu}^n \{I(\epsilon_t \ge U(\frac{m}{ku_r})) - \frac{k}{m}u_r + \frac{k}{m}u_{r+1} - I(\epsilon_t \ge U(\frac{m}{ku_{r+1}}))\}| \\
=\ & I_1 + I_2 + I_3 + I_4.
\end{aligned}
$$

Let $\mathcal{F}_s = \sigma(\epsilon_t, t \le s)$. Then

$$
\begin{aligned}
& P(I_1 > \epsilon) \\
\le\ & N(n)\sup_r P(|k^{-1/2} \textstyle\sum_{t=\nu}^n a_{nt}(u_r,\Delta)| > \epsilon) \\
\le\ & N(n)k^{-1}\epsilon^{-2}\sup_r E(\textstyle\sum_{t=\nu}^n a_{nt}(u_r,\Delta))^2 \\
=\ & N(n)k^{-1}\epsilon^{-2}\sup_r \textstyle\sum_{t=\nu}^n E\{E(a_{nt}^2(u_r,\Delta)|\mathcal{F}_{t-1})\} \\
\le\ & N(n)k^{-1}\epsilon^{-2}\sup_r \textstyle\sum_{t=\nu}^n E|\frac{k}{m}u_r - (1-G(U(\frac{m}{ku_r})(1+s_t^*(\Delta))))| \\
\le\ & N(n)k^{-1}\epsilon^{-2}\sup_r \textstyle\sum_{t=\nu}^n \frac{k}{m}u_r E|1-(1+s_t^*(\Delta))^{-\gamma}| \\
& +N(n)k^{-1}\epsilon^{-2}\sup_r \textstyle\sum_{t=\nu}^n \frac{k}{m}u_r E|\frac{1-G(U(m/(ku_r))(1+s_t^*(\Delta)))}{1-G(U(m/(ku_r)))} - (1+s_t^*(\Delta))^{-\gamma}| \\
=\ & II_1 + II_2.
\end{aligned}
$$

Set

$$
\begin{aligned}
s_{t1}(\delta_1,\delta_2,\delta_3) &= (\tilde\sigma_t(a_1+n^{-1}\lambda_n\delta_1, b_1+n^{-1}\lambda_n\delta_2, c+n^{-1}\lambda_n\delta_3) \\
&\quad -\tilde\sigma_t(a_1,b_1,c))/(\sigma_t(a_1,b_1,c)),
\end{aligned}
$$

$$
s_{t2} = (\tilde\sigma_t(a_1,b_1,c) - \sigma_t(a_1,b_1,c))/(\sigma_t(a_1,b_1,c)),
$$

and let D denote a generic positive constant. It is easy to check that

$$
\begin{cases}
s_t(\delta_1,\delta_2,\delta_3) = s_{t1}(\delta_1,\delta_2,\delta_3) + s_{t2} \\
|s_t^*(\Delta)| \le D \\
0 \le s_{t1}^*(\Delta) \le Dn^{-1}\lambda_n
\end{cases}
\tag{7}
$$

and

$$
\begin{aligned}
& \sup_{t\geq v} E|s_{t2}| \\
\leq\ & \sup_{t\geq \nu} DE\{b_1 \textstyle\sum_{j=1}^{\infty}(a_1)^j X_{t-i-j}^2 I(t-1-j<1)\} \\
\leq\ & D(a_1)^{\nu}.
\end{aligned} \tag{8}
$$

By (7), (8) and $\nu/\log n \to \infty$, we have

$$
II_1 \leq DN(n)(n^{-1}\lambda_n + (a_1)^{\nu}) \to 0. \tag{9}
$$

Using (4) and Lemma 2 of Draisma et al. (2001) we can show that

$$
\begin{aligned}
& \sup_r E\big|\big((1-G(U(\tfrac{m}{ku_r})(1+s_t^*(\Delta)))/(1-G(U(\tfrac{m}{ku_r})))\\
& -(1+s_t^*(\Delta))^{-\gamma}\big)A^{-1}(\tfrac{m}{ku_r})\big| \to 0.
\end{aligned} \tag{10}
$$

By Potter's inequality (see Geluk and de Haan (1987)), we have

$$
\sup_r |u_r A(\frac{m}{ku_r})/A(\frac{m}{k})| \leq D. \tag{11}
$$

So, by (10) and (11),

$$
II_2 \to 0. \tag{12}
$$

It follows from (9) and (12) that

$$
I_1 \xrightarrow{p} 0.
$$

Similarly, we can show that

$$
\begin{aligned}
I_2 \quad & \leq 2\sup_r k^{-1/2}\textstyle\sum_{t=\nu}^n \frac{k}{m}u_r\big|\frac{1-G(U(m/(ku_r))(1+s_t^*(\Delta)))}{1-G(U(m/(ku_r)))} - (1+s_t^*(\Delta))^{-\gamma}\big| \\
& +2\sup_r k^{-1/2}\textstyle\sum_{t=\nu}^n \frac{k}{m}u_r|(1+s_t^*(\Delta))^{-\gamma}-1| \\
& \xrightarrow{p} 0.
\end{aligned} \tag{13}
$$

It is easy to show that

$$
I_4 \xrightarrow{p} 0, \quad I_3 \to 0. \tag{14}
$$

So

$$\sup_{0\le u\le 1}|k^{-1/2}\sum_{t=\nu}^{n}a_{nt}(u,\Delta)| \xrightarrow{p} o.$$

Similarly,

$$\sup_{0\le u\le 1}|k^{-1/2}\sum_{t=\nu}^{n}a_{nt}(u,-\Delta)| \xrightarrow{p} 0.$$

Note that

$$\begin{aligned} & E_{n2}(u,\delta_1,\delta_2,\delta_3) \\ \le\ & k^{-1/2}\textstyle\sum_{t=\nu}^{n} a_{nt}(u,-\Delta) \\ & +k^{-1/2}\textstyle\sum_{t=\nu}^{n}\{(1-G(U(\frac{m}{ku})(1+s_t^*(-\Delta)))-(1-G(U(\frac{m}{ku})(1+s_t^*(\Delta)))\} \end{aligned}$$

and

$$\begin{aligned} & E_{n2}(u,\delta_1,\delta_2,\delta_3) \\ \ge\ & k^{-1/2}\textstyle\sum_{t=\nu}^{n} a_{nt}(u,\Delta) \\ & +k^{-1/2}\textstyle\sum_{t=\nu}^{n}\{(1-G(U(\frac{m}{ku})(1+s_t^*(\Delta))))-(1-G(U(\frac{m}{ku})(1+s_t^*(-\Delta)))\}. \end{aligned}$$

In a similar way to the proofs of (13) and (14), we can show that

$$\sup_{0\le u\le 1}|k^{-1/2}\sum_{t=\nu}^{n}G(U(\frac{m}{ku})(1+s_t^*(\Delta)))-G(U(\frac{m}{ku})(1+s_t^*(-\Delta))\}| \xrightarrow{p} 0.$$

Thus

$$\begin{aligned} & P(\sup\nolimits_{-\Delta\le\delta_1,\delta_2,\delta_3\le\Delta}\sup\nolimits_{0\le u\le 1}|E_{n2}(u,\delta_1,\delta_2,\delta_3)|\ge\epsilon) \\ \le\ & P(\sup\nolimits_{0\le u\le 1}|k^{-1/2}\textstyle\sum_{t=\nu}^{n}a_{nt}(u,\Delta)|\ge\epsilon/4) \\ & +P(\sup\nolimits_{0\le u\le 1}|k^{-1/2}\textstyle\sum_{t=\nu}^{n}a_{nt}(u,-\Delta)|\ge\epsilon/4) \\ & +2P(\sup\nolimits_{0\le u\le 1}|k^{-1/2}\textstyle\sum_{t=\nu}^{n}\{G(U(\frac{m}{ku})(1+s_t^*(\Delta))) \\ & -G(U(\frac{n}{ku})(1+s_t^*(-\Delta)))\}|\ge\epsilon/4) \\ \to\ & 0, \end{aligned}$$

i.e., (6) holds. Using the same arguments in the proofs of (13) and (14), it can be seen that (5) holds. Hence the lemma. □

Proof of Theorem 1. It can be shown by using Lemma 1 and the standard arguments in Ferreira, de Haan and Peng (2003). $\square$

Proof of Theorem 2. It can be shown by using Lemma 1, the arguments in Peng and Qi (2003) and the fact that

$$\frac{\tilde{\sigma}_{n+1}(\hat{a}, \hat{b}, \hat{c})}{\sigma_{n+1}(a, b, c)} - 1 = O_p(n^{-1}\lambda_n) = o_p(1/\sqrt{k}).$$

$\square$

Acknowledgments. Comments from three referees and an associate editor are greatly appreciated. This research is supported by HKRGC Grant 2160195 (Chan), NSF Grant ECS-0134210 (Deng and Xia), the Power System Engineering Research Center (PSerc) (Deng), and NSF grant DMS-0403443 (Peng).

References

Billingsley, P., 1968. Convergence of Probability Measures (Wiley, New York).

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

Bollerslev, T., Chou, R. and Kroner, K., 1992. ARCH Modeling in Finance: A review of the theory and empirical evidence, Journal of Econometrics 52, 5–59.

Christoffersen, P. and Gonçalves, S., 2004. Estimation Risk in Financial Risk Management, Working Paper, McGill University, Quebec.

Danielsson, J. and de Vries, C., 1997. Value-at-Risk and Extreme returns. FMG-Discussion Paper No.273, Financial Markets Group, London School of Economics.

DiCiccio, T.J., Hall, P. and Romano, J.P., 1991. Bartlett adjustment for empirical likelihood, Annals of Statistics 19, 1053–1061.

Dowd, K., 2002. Measuring Market Risk (Wiley, Chichester).

Draisma, G., de Haan, L., Peng, L. and Pereira, T.T., 1999. A bootstrap-based method to achieve optimality in estimating the extreme-value index, Extremes 2, 367–404.

Duffie, D. and Pan, J., 1997. An overview of value at risk, Journal of Derivative, Spring, 7–48.

Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation, Econometrica 50, 987–1008.

Ferreira, A., de Haan, L. and Peng, L., 2003. On optimising the estimation of high quantiles of a probability distribution, Statistics 37, 403–434.

Geluk, J. and De Haan, L., 1987. Regular Variation, Extensions and Tauberian Theorems, CWI Tract 40, Amsterdam.

Hall, P. and La Scala, B., 1990. Methodology and algorithms of empirical likelihood, International Statistical Review 58, 109–127.

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

Hall, P. and Yao, Q., 2003b. Data tilting for time series, Journal of the Royal Statistical Society, Series B 65, 425–442.

Hill, B., 1975. A simple approach to inference about the tail of a distribution, Annals of Statistics 3, 1163–1174.

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

McNeil, A. and Frey, R., 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach, Journal of Empirical Finance 7, 271–300.

Mittnik, S. and Rachev, S.T., 2000. Stable Paretian Models in Finance (Wiley, New York).

Owen, A., 1988. Empirical likelihood ratio confidence intervals for single functional, Biometrika 75, 237–249.

Owen, A., 1990. Empirical likelihood regions, Annals of Statistics 18, 90–120.

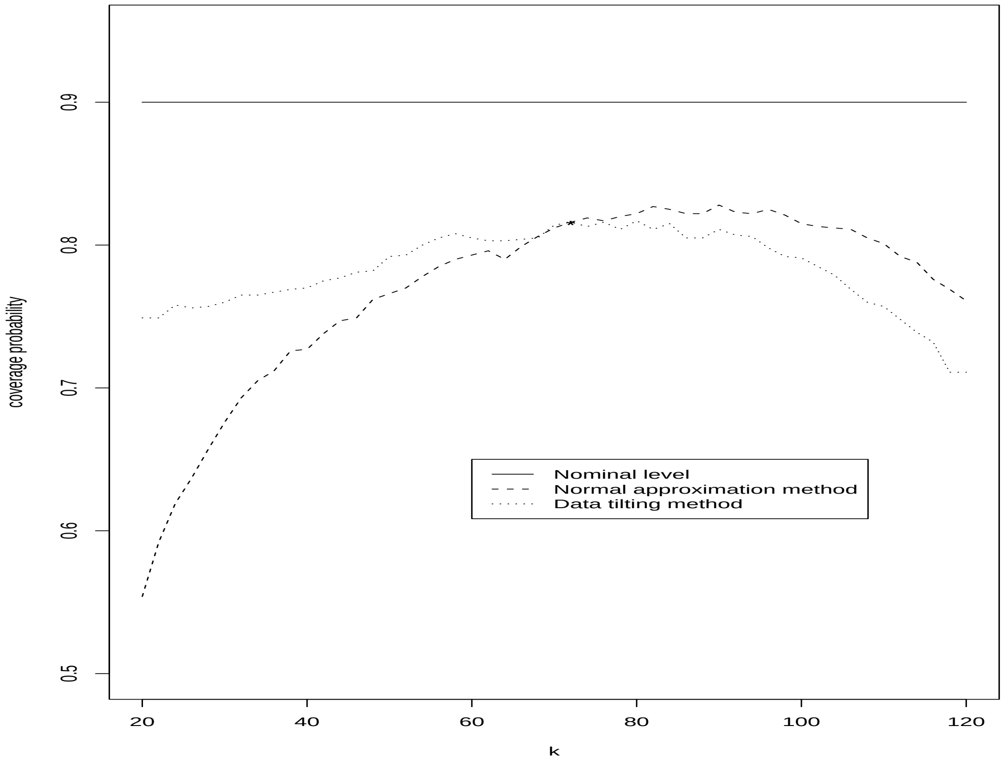
Owen, A., 2001. Empirical Likelihood (Chapman and Hall, New York).

Peng, L. and Qi, Y., 2003. Confidence intervals for high quantiles of a heavy tailed distribution, Technical Report, Georgia Institute of Technology, Atlanta.

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

Shimko, D., 1997. VaR for Corporates. VaR: Understanding and Applying Value-at-Risk, 345–347 (Risk Publications, London).

Taylor, S.J., 1986. Modelling Financial Time Series (Wiley, Chichester).



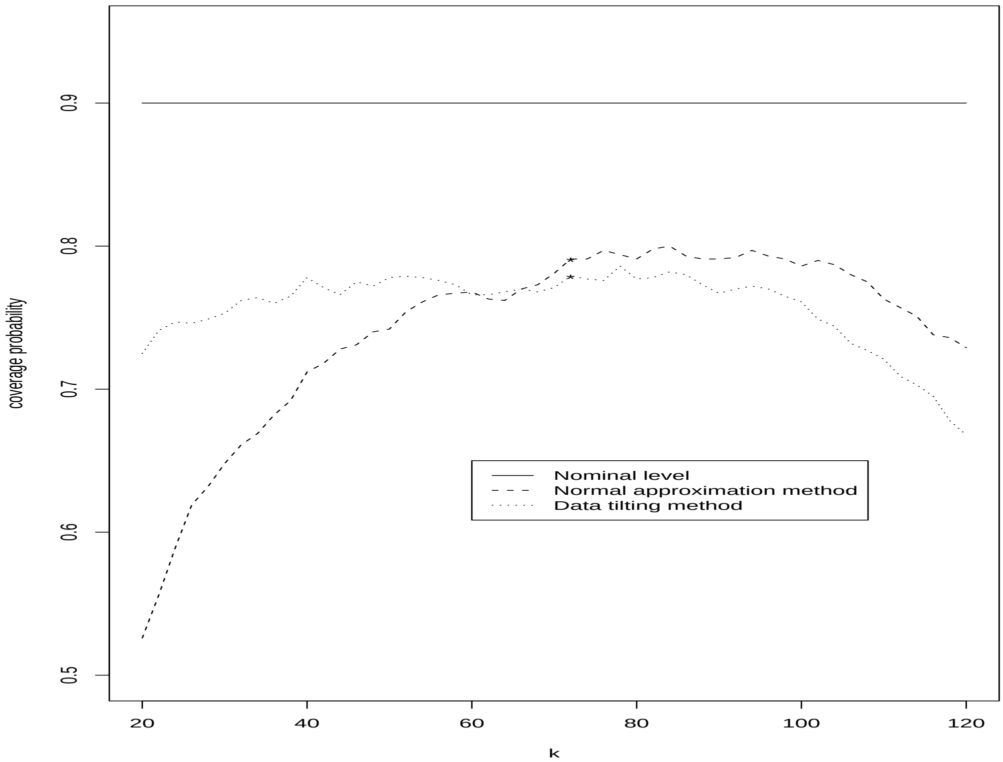


Figure 1: Coverage probability for Student t-distribution with degrees of freedom $d = 3$ as a function of k. Upper and lower panels plot the case $c = 1, b_1 = 0.2, a_1 = 0.3$ and the case $c = 1, b_1 = 0.4, a_1 = 0.5$, respectively. The $*$ point in the plots correspond to $k = 1.5(\log n)^2$, where n is the sample size.

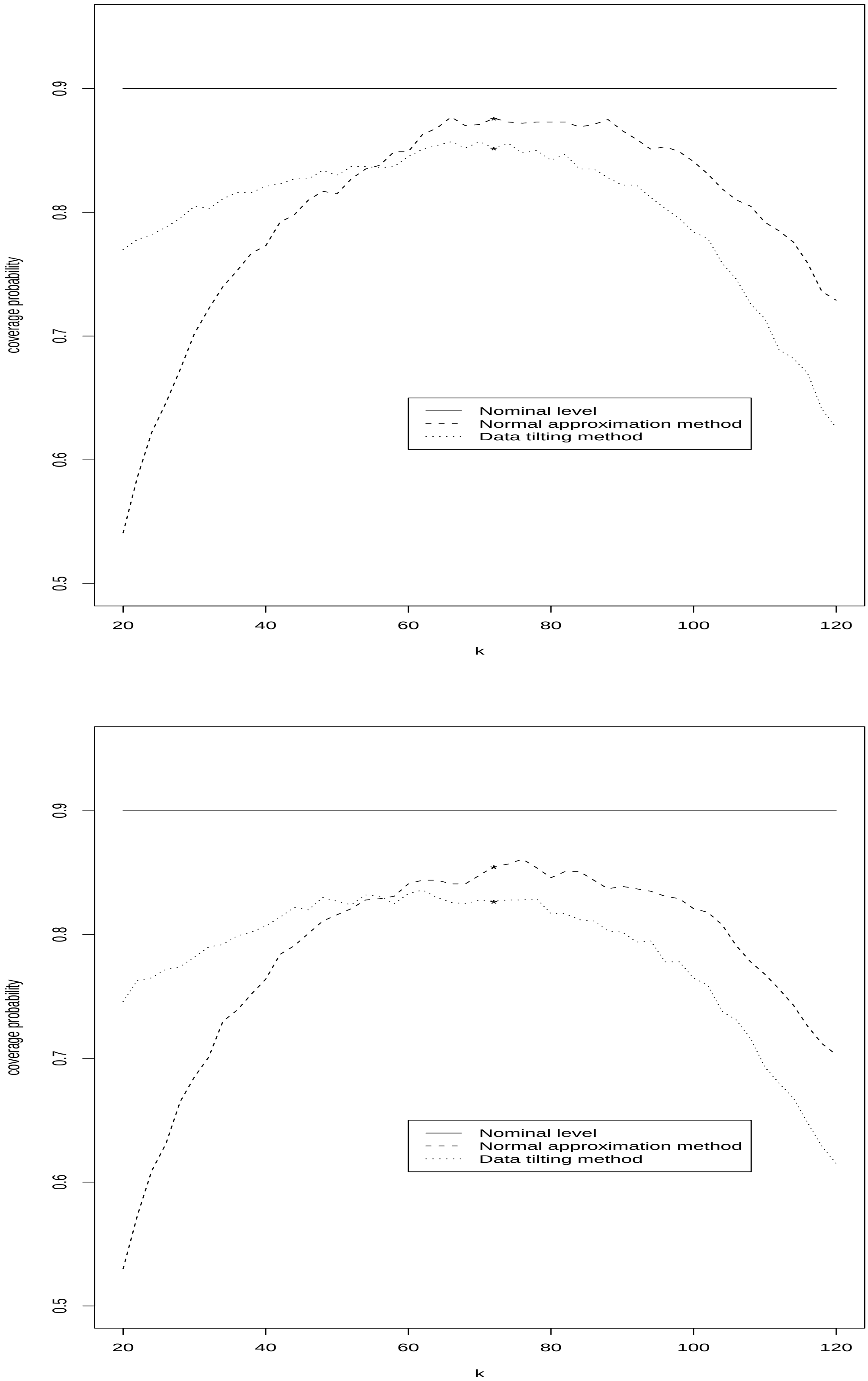
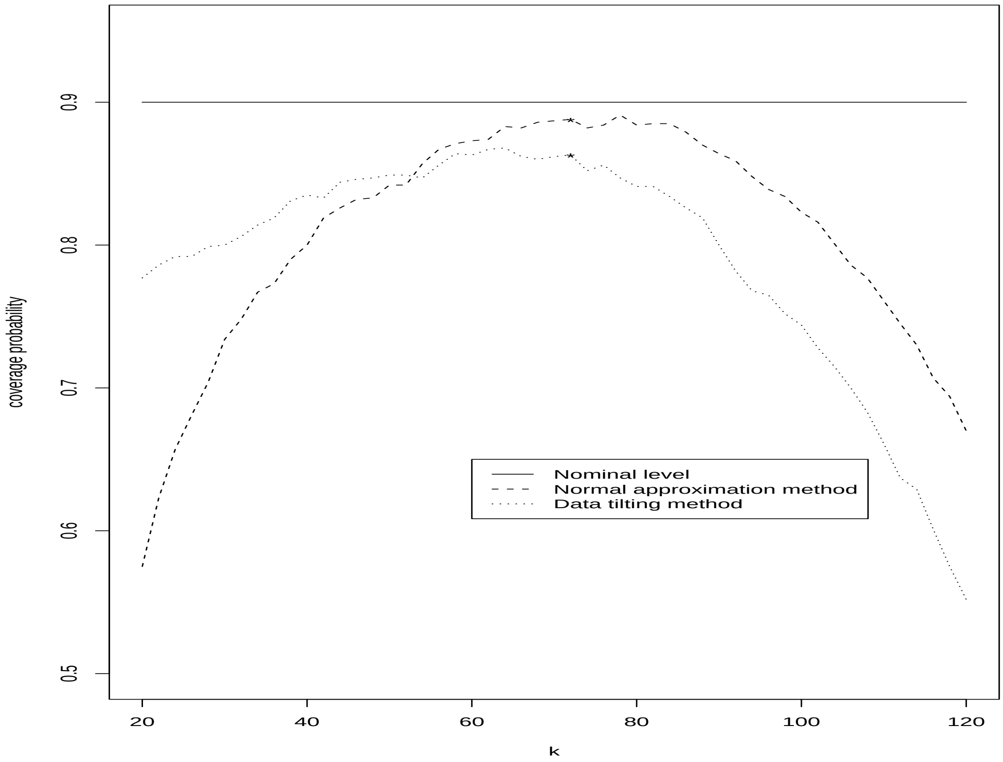


Figure 2: Coverage probability for Student t-distribution with degrees of freedom $d = 5$ as a function of k. Upper and lower panels plot the case $c = 1, b_1 = 0.2, a_1 = 0.3$ and the case $c = 1, b_1 = 0.4, a_1 = 0.5$, respectively. The $*$ point in the plots correspond to $k = 1.5(\log n)^2$, where n is the sample size.



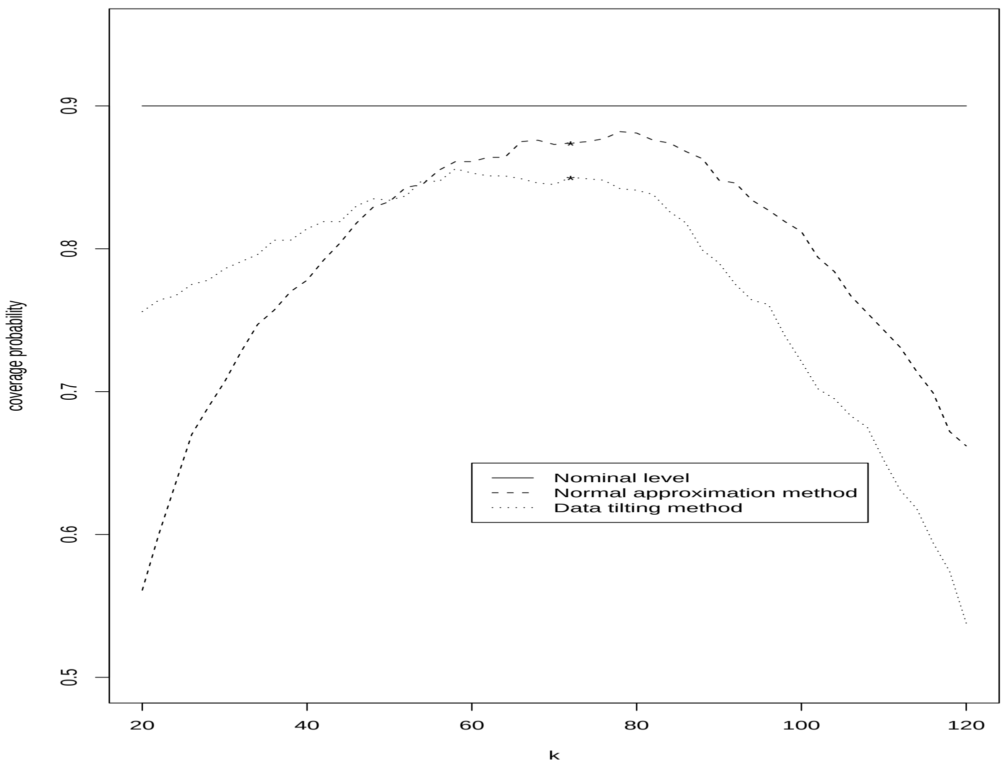
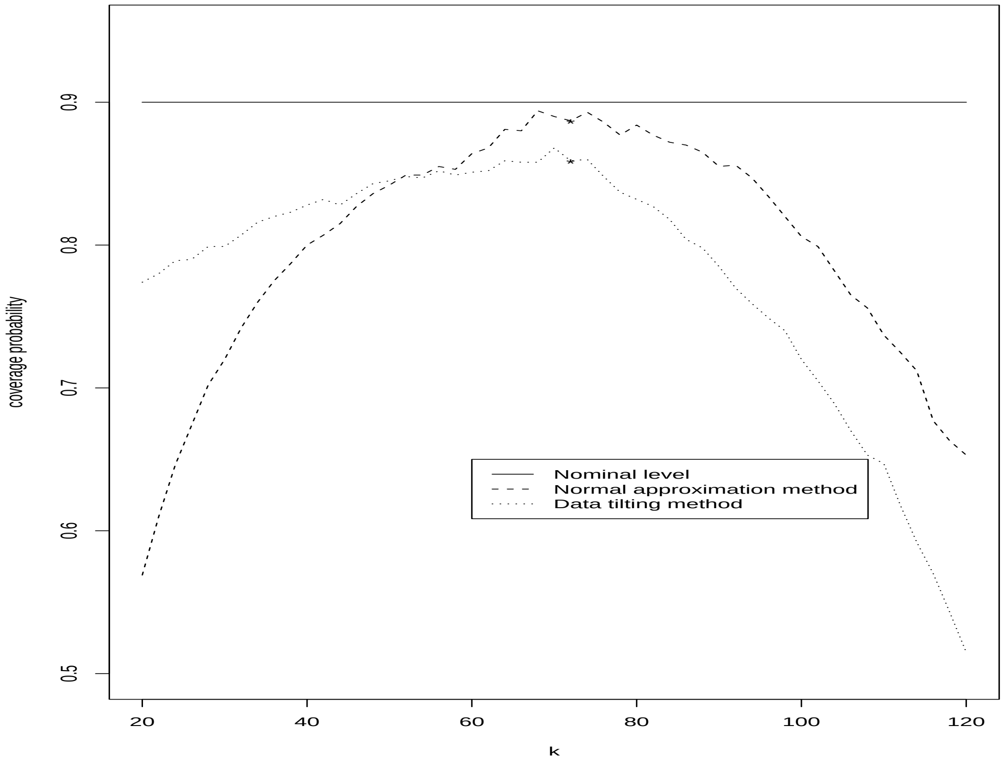


Figure 3: Coverage probability for Student t-distribution with degrees of freedom $d = 7$ as a function of k. Upper and lower panels plot the case $c = 1, b_1 = 0.2, a_1 = 0.3$ and the case $c = 1, b_1 = 0.4, a_1 = 0.5$, respectively. The $*$ point in the plots correspond to $k = 1.5(\log n)^2$, where n is the sample size.



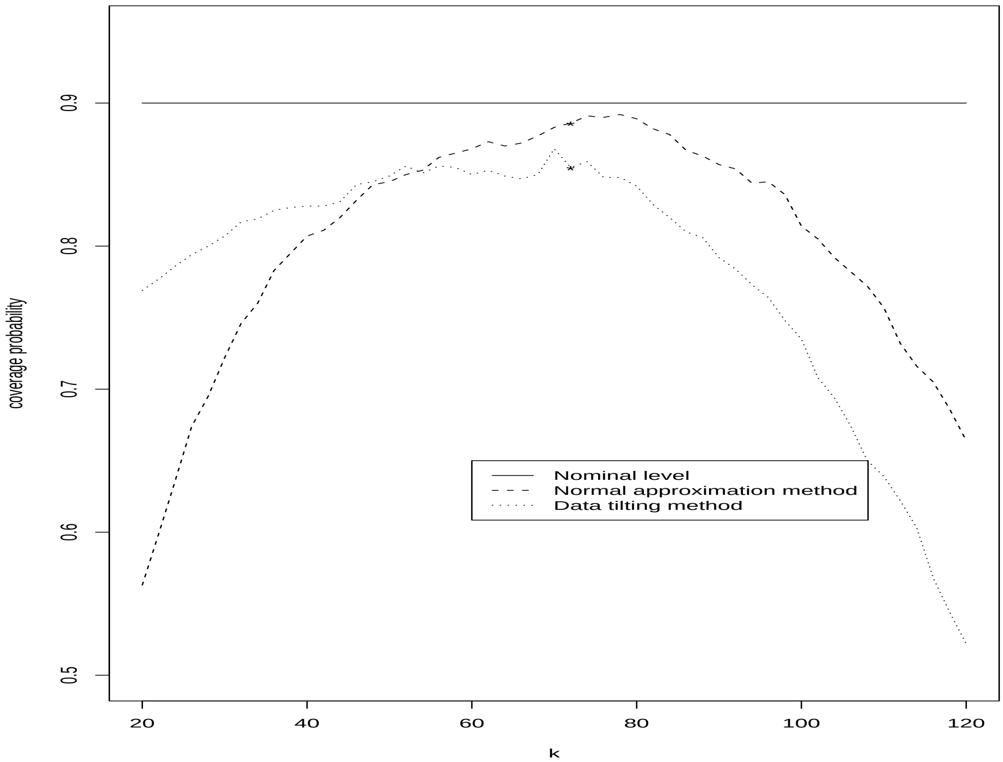


Figure 4: Coverage probability for Student t-distribution with degrees of freedom $d = 9$ as a function of k. Upper and lower panels plot the case $c = 1, b_1 = 0.2, a_1 = 0.3$ and the case $c = 1, b_1 = 0.4, a_1 = 0.5$, respectively. The $*$ point in the plots correspond to $k = 1.5(\log n)^2$, where n is the sample size.

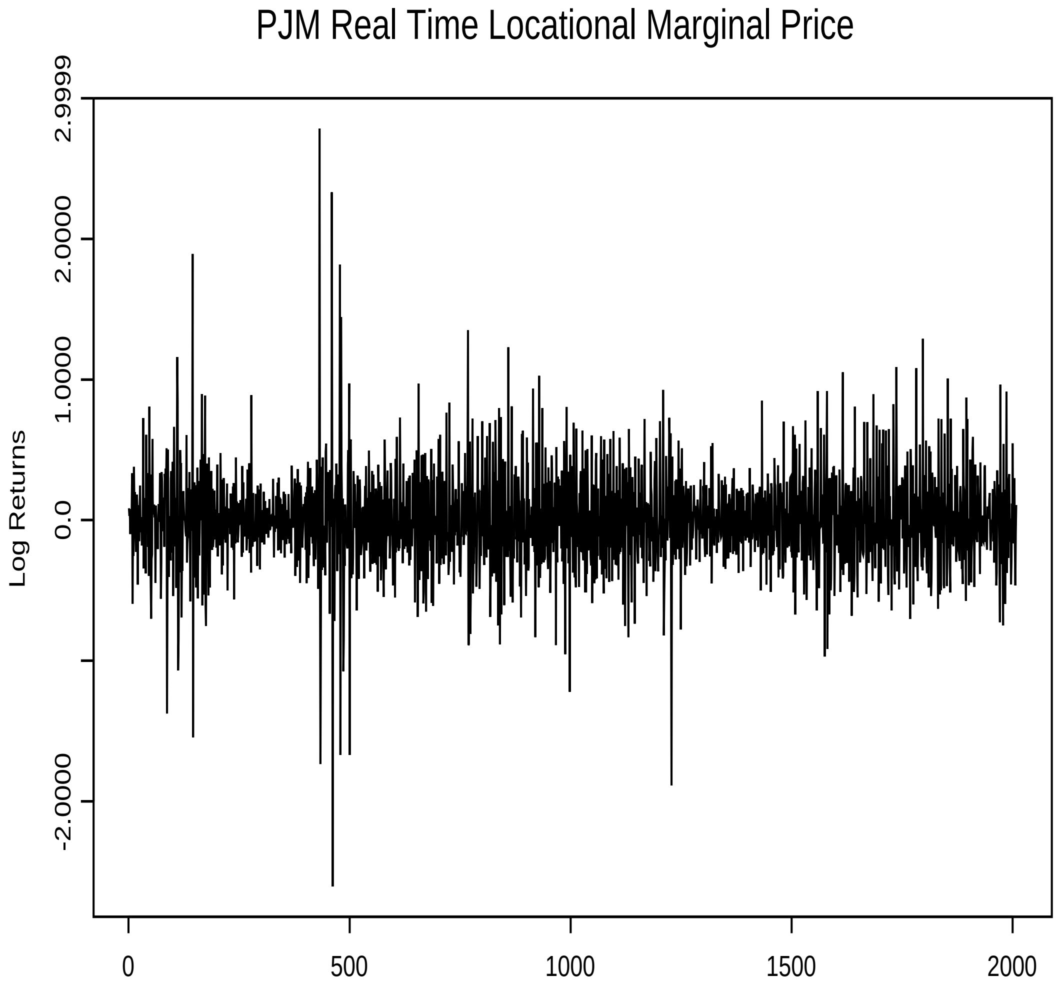


Figure 5: The log returns of real time electricity locational marginal price in the Pennsylvania-New Jersey-Maryland (PJM) power market from April 1998 to September 2003.

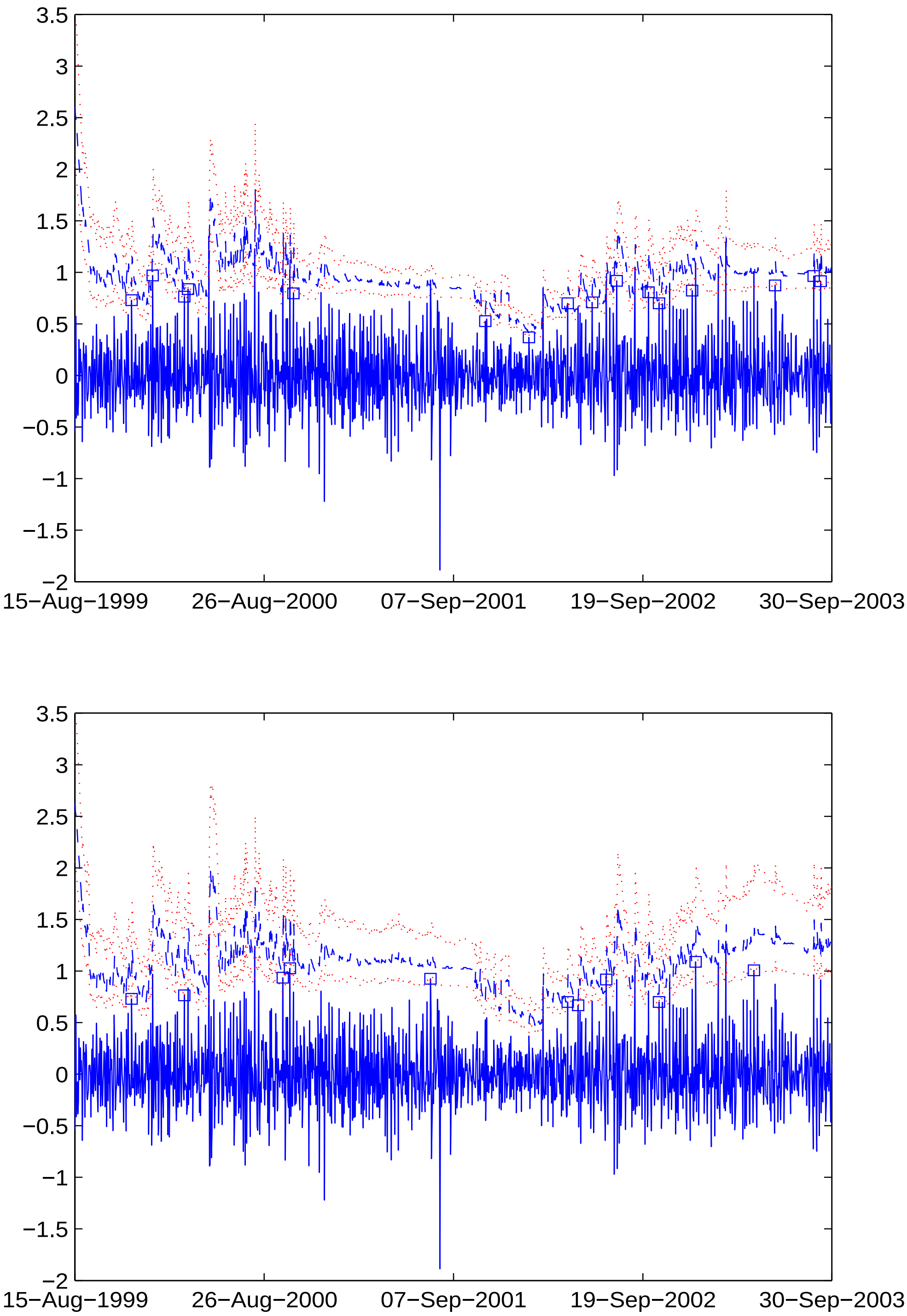


Figure 6: The estimator $\hat{x}^t_{\alpha,n}$ (broken line) and the endpoints (dotted lines) of its 90% confidence intervals (CI) are plotted against the log return X_{t+1} (solid line) of PJM electricity price. Points X_{t+1} such that $X_{t+1} < \hat{x}^t_{\alpha,n}$ but greater than the left endpoint of the CI I^n_β are marked by squares. Upper and lower panels correspond to $k = 30$ and $k = 1.5(\log n)^2 = 57$, respectively.