IMES DISCUSSION PAPER SERIES

Stochastic Volatility Models with Heavy-Tailed Distributions: A Bayesian Analysis

Toshiaki WATANABE and Manabu ASAI

Discussion Paper No. 2001-E-17

IMES

INSTITUTE FOR MONETARY AND ECONOMIC STUDIES

BANK OF JAPAN

C.P.O BOX 203 TOKYO
100-8630 JAPAN

IMES Discussion Paper Series 2001-E-17

November 2001

Stochastic Volatility Models with Heavy-Tailed Distributions: A Bayesian Analysis

Toshiaki WATANABE* and Manabu ASAI**

Abstract

Stochastic volatility (SV) models usually assume that the distribution of asset returns conditional on the latent volatility is normal. This article analyzes SV models with the Student-*t* distribution or generalized error distribution (GED). A Bayesian method via Markov-chain Monte Carlo (MCMC) techniques is used to estimate parameters and Bayes factors are calculated to compare the fit of distributions. Our method is illustrated by analyzing daily data from the Yen/Dollar exchange rate and the TOPIX. According to Bayes factors, it is found that the SV-*t* model fits the both data better than the SV-normal and the SV-GED. The effects of the specification of error distributions on the autocorrelation functions of squared returns and the Bayesian confidence intervals of future returns are also examined.

Key words: Bayes factor, Generalized error distribution, Marginal Likelihood, Markov-chain Monte Carlo, Multi-move sampler, Student-*t* distribution.

JEL classification: C51, G12, F31

* Faculty of Economics, Tokyo Metropolitan University (E-mail: watanabe@ bcomp.metro-u.ac.jp)
** Faculty of Economics, Ritsumeikan University (asai@ec.ritsumei.ac.jp)

The authors are deeply indebted to the following list of people in Bank of Japan who sacrificed their precious time for reading our paper and gave us helpful comments: Yoshifumi Hisata, Akira Ieda, Nobuyuki Oda, Hironobu Tozaka, and Toshinao Yoshiba. Their comments have made our paper more readable, less flawed, and more comprehensive. Thanks are also due to Siddhartha Chib for making some useful suggestions on the marginal likelihood evaluation. Remaining errors are ours alone.

1. INTRODUCTION

It has long been recognized that daily asset returns are leptokurtic, and hence some authors model stock returns as i.i.d. draws from fat-tailed distributions (see Mandelbrot (1963) and Fama (1965)). It is also a well-known phenomenon that the asset return volatility changes randomly over time. If so, the unconditional distribution is leptokurtic even though the conditional distribution is normal (see Bollerslev et al. (1994, p2963)). Note, however, that it does not mean that the leptokurtosis of asset returns can fully be explained by changes in volatility. Actually, several authors have found that the conditional distribution is also leptokurtic by assuming leptokurtic distributions for the conditional distribution in ARCH-type models. For instance, Bollerslev (1987) uses the Student-t distribution, while Nelson (1991) uses the generalized error distribution (GED). Bollerslev et al. (1994) and Watanabe (2000) have found that the Student t-distribution is adequate for capturing the excess kurtosis of the conditional distribution for daily US and Japanese stock returns respectively.

Such studies using the stochastic volatility (SV) model, which is a different model of changing volatility, are scarce because the volatility in the SV model is a latent variable and hence it is difficult to estimate the parameters in the SV model using the conventional maximum likelihood method. The only exception is Liesenfeld and Jung (2000), who fit a Student-t distribution and a GED as well as a normal distribution to the error distribution in the SV model using the simulated maximum likelihood method developed by Danielsson and Richard (1993) and Danielsson (1994). This article also considers a normal distribution, a Student-t distribution, and a GED as the error distribution in the SV model and compares which distribution is the most adequate. Unlike Liesenfeld and Jung (2000), we extend the Bayesian method introduced by Jacquier et al. (1994) and developed by Shephard and Pitt (1997). Specifically, both the model parameters and the latent volatility are sampled from their posterior distribution using Markov-chain Monte Carlo (MCMC) techniques, and simulated draws are used for Bayesian posterior analysis. We sample the latent volatility using the multi-move sampler proposed by Shephard and Pitt (1997) to improve the convergence rate of the MCMC. We calculate Bayes factors using the method proposed by Chib (1995) and Chib and Jeliazkov (2001) to compare the fit of distributions.

Using the MCMC Bayesian method, the SV models with a normal distribution, a Student-t distribution, and a GED are fitted to daily data from the Yen/Dollar exchange rate and the TOPIX. According to Bayes factors, we find that the Student-t distribution fits the both data better than the normal and the GED. We also examine how the specification of error distributions influences the autocorrelation functions of squared returns and the confidence intervals of future returns.

The rest of this article is organized as follows. Section 2 briefly explains the SV model with a normal distribution, a Student-t distribution, and a GED. Section 3 develops the Bayesian method for the analysis of the SV models with these distributions. These SV models are fitted to daily data from the Yen/Dollar exchange rate and the TOPIX in Section 4. Conclusions and extensions are given in Section 5.

2. STOCHASTIC VOLATILITY MODELS WITH HEAVY-TAILED DISTRIBUTIONS

The stochastic volatility (SV) model analyzed in this article is the standard one given by

$$\begin{aligned} r_t &= \exp(h_t/2)\epsilon_t, \quad \epsilon_t \sim i.i.d.,\ E(\epsilon_t)=0,\ V(\epsilon_t)=1, && (1) \\ h_t &= \mu + \phi(h_{t-1}-\mu) + \eta_t, \quad \eta_t \sim i.i.d. N(0,\sigma_\eta^2), && (2) \end{aligned}$$

where r_t is the asset return on day t from which the mean and autocorrelations are removed. We call $\exp(h_t/2)$ as volatility, so that h_t represents the log of squared volatility. We assume that ϵ_t and η_s are mutually independent for all t and s. We also assume that $|\phi| < 1$.

It is a well-known phenomenon that daily asset returns have leptokurtic distributions. The kurtosis of r_t following the above SV model is given by

$$k = \frac{E(r_t^4)}{E(r_t^2)^2} = E(\epsilon_t^4) \exp\left[\sigma_h^2\right], \quad (3)$$

where

$$\sigma_h^2 = \frac{\sigma_\eta^2}{1-\phi^2},$$

which represents the unconditional variance of h_t (see Appendix A in Liesenfeld and Jung (2000) for the derivation).

The standard normal distribution is usually assumed for the distribution of ϵ_t. If so, $E(\epsilon_t^4) = 3$ and hence the kurtosis of r_t is:

$$k = 3\exp\left[\sigma_h^2\right] \geq 3,$$

where $k = 3$ only if $\sigma_h^2 = 0$. This result indicates that, as long as the volatility changes over time, the unconditional distribution of r_t is leptokurtic even if ϵ_t follows the standard normal. However, it does not necessarily follow that the leptokurtosis of asset returns can fully be explained by changes in volatility. The distribution of ϵ_t itself may possibly be leptokurtic. In this article, we also fit leptokurtic distributions to ϵ_t. Specifically, we consider a Student-t distribution and a generalized error distribution (GED).

The probability density function (PDF) of the t-distribution with mean zero and variance normalized to one is given by

$$f(\epsilon_t) = [\pi(\nu-2)]^{-1/2}\,\frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)}\left[1+\frac{\epsilon_t^2}{\nu-2}\right]^{-(\nu+1)/2}, \quad \nu > 2, \qquad (4)$$

where the parameter ν represents the degree-of-freedom and $\Gamma(\cdot)$ denotes the gamma function. As long as $\nu > 4$, the kurtosis of the t-distribution is

$$E(\epsilon_t^4) = 3(\nu-2)/(\nu-4),$$

which is greater than three if $\nu < \infty$. Needless to say, the $t-$distribution approaches the standard normal distribution when $\nu \to \infty$.

The PDF of the GED with mean zero and variance one is given by

$$f(\epsilon_t) = \frac{\upsilon \exp\left[-\frac{1}{2}|\epsilon_t/\beta|^{\upsilon}\right]}{\beta\Gamma(1/\upsilon)2^{(1+1/\upsilon)}}, \quad 0 < \upsilon < \infty, \qquad (5)$$

where

$$\beta = \left[2^{-2/\upsilon}\frac{\Gamma(1/\upsilon)}{\Gamma(3/\upsilon)}\right]^{1/2}. \qquad (6)$$

The kurtosis of the GED is given by

$$E(\epsilon_t^4) = \Gamma(1/\upsilon)\Gamma(5/\upsilon)/\left[\Gamma(3/\upsilon)\right]^2,$$

which is greater than three if $\upsilon < 2$. The GED becomes the standard normal distribution when $\upsilon = 2$.

3. BAYESIAN ANALYSIS

In this section, we explain the method used in this article for the analysis of the SV models with a normal distribution, a Student-t distribution, and a GED. As is well known, it is difficult to estimate the parameters in the SV model using the maximum likelihood method. Several alternative methods have, however, been proposed. Among such methods, we extend the Bayesian method using Markov-chain Monte Carlo (MCMC) techniques introduced by Jacquier et al. (1994) and developed by Shephard and Pitt (1997).

3.1 Parameter Estimation

For the unknown parameters in the SV-normal model, following Kim et al. (1998), we work with the following prior distributions.

$$\mu \sim N(k_1, k_2),$$

$$2\phi - 1 \sim Beta(\phi_1, \phi_2),$$

$$\sigma_\eta^2 \sim IG(\sigma_r/2, S_\sigma/2),$$

where $N(\cdot,\cdot)$, $Beta(\cdot,\cdot)$, and $IG(\cdot,\cdot)$ represent the normal, beta, and inverse gamma distributions respectively. Specifically, we set $k_1 = 0$, $k_2 = 10$, $\phi_1 = 20$, $\phi_2 = 1.5$, $\sigma_r = 5$, and $S_\sigma = 0.01 \times \sigma_r$. We sample these parameters as well as the latent variable $\{h_t\}_{t=0}^T$ from their full conditional distributions using MCMC techniques. The prior distribution of h_0 is set to be the unconditional distribution of h_t, i.e. $h_0 \sim N(\mu, \sigma_\eta^2/(1-\phi^2))$. It is straightforward to obtain the full conditional distributions of the parameters and h_0 and sample from them (see Appendix A1 for their full conditional distributions). For sampling $\{h_t\}_{t=1}^T$, we use the multi-move sampler proposed by Shephard and Pitt (1997), where a block of disturbances $\{\eta_s\}_{s=t}^{t+k}$ in equation (2) are sampled from their conditional density

$$f(\{\eta_s\}_{s=t}^{t+k} | h_{t-1}, h_{t+k+1}, \{r_s\}_{s=t}^{t+k}, \theta). \tag{7}$$

where $\theta = (\mu, \phi, \sigma_\eta^2)$ (see Appendix A1 for details).

In the SV-t model, ϵ_t can be represented by:

$$\epsilon_t = \sqrt{\omega_t} z_t, \tag{8}$$

where $z_t \sim i.i.d.N(0,1)$ and $(\nu - 2)/\omega_t \sim i.i.d.\chi^2(\nu)$.

We sample $\{\omega_t\}_{t=1}^T$ and ν as well as the other parameters and $\{h_t\}_{t=0}^T$ from their full conditional distributions. For μ, ϕ, and σ_η^2, we adopt the same priors as those in the SV-normal model, so that their full conditional distributions remain the same. Following Geweke (1992b, 1993) and Fernández and Steel (1998), we set the prior distribution for ν as a truncated exponential with probability density function (PDF),

$$f(\nu) = \begin{cases} c\lambda \exp(-\lambda\nu), & \nu > 4, \\ 0, & \text{otherwise}, \end{cases} \tag{9}$$

where $c = \exp(4\lambda)$. Specifically, we set λ equal to 0.1. We sample ν and $\{\omega_t\}_{t=1}^T$ from their conditional distributions $f(\nu|\,\{\omega_t\}_{t=1}^T)$ and $f(\{\omega_t\}_{t=1}^T\,|\nu$, $\{\epsilon_t\}_{t=1}^T)$. Conditional on ν, the $(\epsilon_t^2 + \nu - 2)/\omega_t$ follow independent $\chi^2(\nu+1)$ distribution. Hence, it is straightforward to sample from $f(\{\omega_t\}_{t=1}^T|\nu,\{\epsilon_t\}_{t=1}^T)$. It is, however, more troublesome to sample from $f(\nu|\,\{\omega_t\}_{t=1}^T)$. We use the method proposed by Watanabe (2001), which is based on the acceptance-rejection/Metropolis-Hastings (A-R/M-H) algorithm proposed by Tierney (1994) (see Appendix A2 for details). We employ the multi-move sampler again to sample the latent variable $\{h_t\}_{t=1}^T$.

When ϵ_t follows the GED, it is not possible to represent it as equation (8). Hence, for the SV-GED model, we sample υ as well as the other parameters and $\{h_t\}_{t=0}^T$ from their full conditional distributions. The priors for μ, ϕ, and σ_η^2 remain the same. The prior distribution for υ is given by

$$\upsilon = I[1,2], \tag{10}$$

where $I[1,2]$ is the indicator function of the interval $[1,2]$. We sample υ from its conditional distributions $f(\upsilon|\{\epsilon_t\}_{t=1}^T)$ using the A-R/M-H algorithm (see Appendix A3 for details). We employ the multi-move sampler again to sample the latent variable $\{h_t\}_{t=1}^T$.

3.2 Bayes factors

Model comparison in a Bayesian framework can be performed using posterior odds ratio. Let $R = \{r_t\}_{t=1}^T$ denote the observation vector. Then, posterior odds ratio between model i, M_i, and model j, M_j, is given by

$$\text{POR} = \frac{f(M_i|R)}{f(M_j|R)} = \frac{f(R|M_i)}{f(R|M_j)} \frac{f(M_i)}{f(M_j)},$$

where $f(R|M_i)/f(R|M_j)$ and $f(M_i)/f(M_j)$ are called Bayes factor and prior odds ratio respectively.

As is the usual practice, we set the prior odds to be 1, so that the posterior odds ratio is equal to the Bayes factor. To evaluate the Bayes factor, which is the ratio of the marginal likelihoods, we follow the basic marginal likelihood identity in Chib (1995). The log (base 10) of the marginal likelihood of model M_i can be written as[1]

$$\begin{aligned} &\log f(R|M_i) \\ &= \log f(R|M_i, \theta_i) + \log f(\theta_i|M_i) - \log f(\theta_i|M_i, R), \end{aligned} \tag{11}$$

where θ_i is the set of unknown parameters for model M_i, $\log f(R|M_i, \theta_i)$ is the likelihood, $\log f(\theta_i|M_i)$ is the prior density, and $\log f(\theta_i|M_i, R)$ is the posterior density.

The above identity holds for any value of θ_i, but following Chib (1995), we set θ_i at its posterior mean calculated using the MCMC draws. The likelihood is evaluated using the Accelerated Gaussian Importance Sampling (AGIS) proposed by Danielsson and Richard (1993) and Danielsson (1994), and the posterior density is calculated using the method of Chib and Jeliazkov (2001), which is based on the Metropolis-Hastings algorithm (see Appendix B).

[1] Bayes factor is usually shown as its log (base 10) value. We denote log (base 10) as log and log (base e) as ln in this article.

4. EMPIRICAL APPLICATION

4.1 Data Description and Preliminary Results

We illustrate our method using daily data of the following financial series: the spot exchange rates for the Japanese Yen/U.S. Dollar exchange rate from January 4, 1990 to December 28, 1999 and the Tokyo stock price index (TOPIX) from January 4, 1990 to September 30, 1999. We define the both returns as

$$100 \times \{\ln(P_t) - \ln(P_{t-1})\}$$

where P_t is the closing price on day t.

The descriptive statistics are summarized in Table 1. The statistics reported are the mean, the standard deviation, the kurtosis, and the Ljung-Box (LB) statistics for 12 lags corrected for heteroskedasticity following Diebold (1988). The kurtosis of the returns for the both series is significantly above three, indicating leptokurtic return distributions. The LB statistics indicate that the return for the yen/dollar exchange rate is serially uncorrelated while the return for the TOPIX is serially correlated. Hence, as for $\{r_t\}$, we use for the former series returns with the mean subtracted and for the latter series the residuals from the AR(2) model, where the lag length 2 is selected based on the Schwartz (1978) Information Criterion (SIC).

4.2 Estimation Details

For parameter estimation, we conduct the MCMC simulation with 15000 iterations. The first 5000 draws are discarded and then the next 10000 are recorded. Following Kim et al. (1998), we record σ_η and $\exp(\mu/2)$ in place of σ_η^2 and μ. Using these 10000 draws for each of the parameters, we calculate the posterior means, the standard errors of the posterior means, the 95% intervals, and the convergence diagnostic (CD) statistics proposed by Geweke (1992a). The posterior means are computed by averaging the simulated draws. The standard errors of the posterior means are computed using a Parzen window with a bandwidth of 1000 (see Shephard and Pitt (1997) for details). The 95% intervals are calculated using the 2.5th and 97.5th percentiles of the simulated draws. Geweke (1992a) suggests assessing the convergence of the MCMC by comparing values early in the sequence with those late in the sequence. Let $\theta^{(i)}$ be the ith draw of a parameter in the recorded 10000 draws, and let $\bar{\theta}_A = \frac{1}{n_A}\sum_{i=1}^{n_A}\theta^{(i)}$ and $\bar{\theta}_B = \frac{1}{n_B}\sum_{i=10001-n_B}^{10000}\theta^{(i)}$. Using these values, Geweke (1992a) proposes the following statistic called *convergence diagnostics* (CD).

$$\text{CD} = \frac{\bar{\theta}_A - \bar{\theta}_B}{\sqrt{\hat{\sigma}_A^2/n_A + \hat{\sigma}_B^2/n_B}}, \tag{12}$$

where $\sqrt{\hat{\sigma}_A^2/n_A}$ and $\sqrt{\hat{\sigma}_B^2/n_B}$ are standard errors of $\bar{\theta}_A$ and $\bar{\theta}_B$. If the sequence of $\theta^{(i)}$ is stationary, it converges in distribution to the standard normal. We set $n_A = 1000$ and $n_B = 5000$ and compute $\hat{\sigma}_A^2$ and $\hat{\sigma}_B^2$ using Parzen windows with bandwidths of 100 and 500 respectively.

An advantage of using the MCMC Bayesian method is that we can also sample the parameters such as the kurtosis and the unconditional variance, which are functions of the model parameters, from their posterior distributions. For example, all we have to do to sample the kurtosis from its posterior distribution is to substitute the draws of the model parameters sampled from their posterior distribution into equation (3). Hence, we calculate the posterior means, the standard errors of the posterior means, the 95% intervals, and the CD statistics for the unconditional variance and the kurtosis as well as the model parameters.

The details of the marginal likelihood evaluation are provided in Appendix B.

4.3 Estimation Results

Table 2 shows the results for the mean-subtracted return for the Yen/Dollar exchange rate. Table 2(A) presents log (base 10) of Bayes factors. The log value of Bayes factor of the SV-normal model compared to the SV-t is -6.96 and that of the SV-t compared to the SV-GED is 3.14, indicating that the SV-t fits the exchange rate data better than the SV-normal and the SV-GED. This result is consistent with that of Liesenfeld and Jung (2000). Table 2(B) shows the posterior means, the 95% intervals, and the CD statistics, which has already been explained in the previous subsection, for the model parameters, the unconditional variance of volatility $\sigma_h^2(=\sigma_\eta^2/(1-\phi^2))$, and the kurtosis. According to the CD values, the null hypothesis that the sequence of 10000 draws is stationary is accepted at any standard level for all parameters in all models. The posterior mean and 95% interval of ϕ of the SV-t are placed on upward, compared to those of the SV-normal and the SV-GED, showing a higher persistence in volatility of the SV-t than those of the other two models. The posterior means and 95% intervals of the conditional standard deviation of volatility σ_η and the unconditional variance of volatility σ_h^2 are both smaller in the SV-t than those in the SV-normal and SV-GED, indicating that the volatility of the SV-t is less variable than those of the other models. These results are also consistent with those of Liesenfeld and Jung (2000). The posterior mean and 95% interval of the kurtosis of the SV-t are larger than those of the SV-normal and the SV-GED while the 95% intervals of the kurtosis in all models include the sample kurtosis of 6.5889.

Table 3 shows the results for the AR(2) residuals of the TOPIX return series, which are qualitatively unaltered, although the quantitative differences among the three models are smaller, compared to those for the Yen/Dollar exchange rate series. Log values of Bayes factors presented in Table 3(A) indicate that the SV-t fits the TOPIX data better than the SV-normal and the SV-GED. According to the estimates of ϕ, σ_η, and σ_h^2 in Table 3(B), the posterior mean and 95% interval of ϕ are larger and those of the conditional standard deviation and the unconditional variance of volatility are smaller in the SV-t than those in the SV-normal and the SV-GED. The posterior mean and 95% interval of the kurtosis of the SV-t are larger than those of the SV-normal and the SV-GED while the 95% intervals of the kurtosis in all models include the sample kurtosis of 7.2813.

The reason why the volatility of the SV-t is estimated to be more persistent and less variable can be understood by comparing the probability density function (PDF) of the t-distribution to those of the normal and GED. Figures 1 and 2 show these PDFs for the Yen/Dollar exchange rate and the TOPIX respectively, where the parameters ν and υ are set equal to the corresponding posterior means. Needless to say, the PDFs of the t-distribution and GED have fatter tails than that of the normal. The PDF of the GED puts emphasis on the sharpness around the mean rather than the tail fatness, so that the PDF of the t-distribution has a fatter tail than that of the GED. Therefore, the SV-t attributes a larger proportion of extreme return values to ϵ_t instead of η_t than the SV-normal and the SV-GED do, making the volatility of the SV-t less variable. It also increases the persistence in volatility of the SV-t if extreme returns are less persistent.

This interpretation is confirmed by comparing the volatility estimates. Figures 3 and 4 plot the posterior means of volatilities $\{\exp(h_t/2)\}$ jointly with the squared returns for the returns of the Yen/Dollar exchange rate and the squared residuals for the AR(2) residuals of the TOPIX returns respectively. For the exchange rate data, the posterior means of volatilities from the SV-t model exhibit smoother movements than those from the SV-normal and SV-GED models. Extreme returns such as the returns at the Asian currency crisis beginning in July 1997 make the difference among the three models clear. The volatilities associated with these extreme return values jump up more under the SV-normal and the SV-GED than under the SV-t model. Contrary to the exchange rate data, no major difference in the posterior means of volatilities among the three models can be seen in the TOPIX data because the differences in parameter estimates are small.

Next, we evaluate the three SV models by comparing the posterior distributions for the autocorrelation coefficients of squared returns to the corresponding sample autocorrelation coefficients. It is straightforward to sample the autocorrelation coefficients of squared returns from their conditional distributions. The τth order autocorrelation of r_t^2 is given by

$$\rho(\tau) = \frac{\exp(\sigma_h^2 \phi^\tau) - 1}{E[\epsilon_t^4] \exp(\sigma_h^2) - 1}. \tag{13}$$

All we have to do is to substitute the draws of the parameters sampled from their posterior distributions into the above equation. Figures 5 and 6 show the posterior means of autocorrelation coefficients up to the 100th order and the 95% confidence interval jointly with the sample autocorrelation coefficients for the two series. In the exchange rate data, the autocorrelation function of the SV-t is flatter than those of the SV-normal and SV-GED because the estimates of ϕ for the SV-t is closer to one than those for the other two models. The 95% intervals for the SV-t include the corresponding sample autocorrelation coefficients more than the other two models especially in the higher order area. The same is true for the TOPIX data but the differences are smaller.

From the viewpoint of Value at Risk (VaR), it is important to examine how the specification of conditional distribution influences the Bayesian confidence intervals of future returns. The Bayesian confidence interval of

r_{T+1} can be obtained by sampling from

$$f(r_{T+1}|R) = \int f(r_{T+1}|h_{T+1}, R) f(h_{T+1}|h_T) f(h_T|R) dh_{T+1} dh_T, \qquad (14)$$

where $R = \{r_t\}_{t=1}^T$. This sampling is straightforward. First, N draws $\left\{h_T^{(1)}, \ldots, h_T^{(N)}\right\}$ are sampled from $f(h_T|R)$ jointly with the model parameters using the MCMC method explained in Section 3. Then, given these N draws, sample N draws $\{h_{T+1}^{(1)}, \ldots, h_{T+1}^{(N)}\}$ from $f(h_{T+1}|h_T^{(1)}), \ldots, f(h_{T+1}|h_T^{(N)})$ using equation (2). Finally, given these N draws, sample N draws $\{r_{T+1}^{(1)}, \ldots, r_{T+1}^{(N)}\}$ from $f(r_{T+1}|h_{T+1}^{(1)}), \ldots, f(r_{T+1}|h_{T+1}^{(N)})$ using equation (1). N is set equal to 10000. Although methods for sampling from the standard normal and t-distribution are well known, it is not obvious for the GED. We propose a method for sampling from the GED by using the inverse incomplete gamma function and probability integral transformation (see Appendix C). For each model, we calculate the 90%, 95%, and 99% intervals with the posterior mean of volatility, $\exp(h_{T+1}/2)$. The confidence intervals of r_{T+1} would become narrower in ordering of the SV-t, the SV-GED, and the SV-normal if the volatility $\exp(h_{T+1}/2)$ were the same in all three models. The volatility estimates, however, differ among the three models, so that the effect of the specification of conditional distribution on the widths of confidence intervals is vague. For the exchange-rate data, the means of volatility decrease in ordering of the SV-t, the SV-GED, and the SV-normal, so that the confidence intervals become narrower in the same order. The TOPIX data, however, is not the case. Since the means of volatility decrease in ordering of the SV-normal, the SV-t, and the SV-GED, the confidence intervals in the SV-GED is narrower than those of the SV-normal. In any case, since the specification of the conditional distribution obviously affects the confidence intervals of r_{T+1}, it is important to specify the conditional distribution correctly in constructing confidence intervals.

All analyses so far are based on the SV model defined by equations (1) and (2). To analyze the effect of the volatility specification on the estimates of ν and υ, we also estimate the following GARCH model with different distributions for $\{\epsilon_t\}$.

$$r_t = \sigma_t \epsilon_t, \quad E(\epsilon_t) = 0, V(\epsilon_t) = 1,$$
$$\sigma_t^2 = a_0 + a_1 r_{t-1}^2 + b_1 \sigma_{t-1}^2.$$

Although it is possible to estimate the GARCH model using MCMC Bayesian methods (see Bauwens and Lubrano (1998), Nakatsuma (2000), and Mitsui and Watanabe (2000)), we simply apply the conventional maximum likelihood method. Table 5 shows the estimation results. For the both Yen/Dollar exchange rate and TOPIX, the estimates of ν in the GARCH-t model and υ in the GARCH-GED model are much smaller than the corresponding posterior means in the SV-t and SV-GED models. This result indicates that the GARCH model requires a more leptokurtic distribution for ϵ_t than the SV model. Volatility in the GARCH model is determined only by the past volatility σ_{t-1}^2 and the past squared return r_{t-1}^2, while that in the SV model depends also on the current shock η_t, which increases the leptokurtosis of r_t and the estimates of ν and υ. In the GARCH model, volatility persistence is measured by $a_1 + b_1$. In ordering of GARCH-normal, GARCH-GED, and GARCH-t, the estimates of $a_1 + b_1$ rise, which is consistent with the result based the SV model.

5. CONCLUSIONS AND EXTENSIONS

This article analyzes SV models with the Student-t distribution or generalized error distribution (GED). A Bayesian method via Markov-chain Monte Carlo (MCMC) techniques is used to estimate parameters and Bayes factors are calculated to compare the fit of distributions. The results, based on daily data from the yen/dollar exchange rate and the TOPIX, reveal that the t distribution fits the both data better than the normal and the GED. We also examine the effects of the choice of the conditional distribution on the shapes of autocorrelation functions and the confidence intervals of future returns.

This article has certain contributions, but several extensions are still possible.

1. We focus on leptokurtic distributions as distributions for ϵ_t, but it is also worthwhile fitting skewed distributions such as skewed t distribution (see Hansen (1994) and Fernández and Steel (1998)) and the generalized exponential beta distribution (GEB) (see Wang et al. (2001)).

2. We specify the log of volatility as a simple AR(1) process. According to Figures 5 and 6, the number of the sample autocorrelation coefficients that are not included in the 95% intervals is, however, not negligible even in the SV-t model. Hence, more elaborate models such as higher order ARMA models and long memory models may be required for the specification of volatility.

3. We neglect the well-known phenomenon in stock markets of a negative correlation between current returns and future volatility. A stochastic volatility model extended to accommodate this phenomenon (see Danielsson (1994), Harvey and Shephard (1996), and Watanabe (1997)) should also be applied to the TOPIX data.[2]

[2] The application of multimove sampler to such an asymmetric SV model is under study.

APPENDIX A: SAMPLING METHOD FOR THE SV MODELS

A.1 The SV-normal Model

As shown by Kim et al. (1998), the full conditional distributions of μ and σ_η^2 are given by

$$\mu|\cdot \sim N\left(\hat{\mu}, \sigma_\mu^2\right), \tag{A.1}$$

$$\sigma_\eta^2|\cdot \sim IG(A, B), \tag{A.2}$$

where

$$\begin{aligned}
\sigma_\mu^2 &= \frac{k_2 \sigma_\eta^2}{k_2\left\{T(1-\phi)^2 + 1 - \phi^2\right\} + \sigma_\eta^2}, \\
\hat{\mu} &= \sigma_\mu^2 \left\{ \frac{(1-\phi^2)}{\sigma_\eta^2} h_0 + \frac{(1-\phi)}{\sigma_\eta^2} \sum_{t=1}^{T} (h_t - h_{t-1}) + \frac{k_1}{k_2} \right\}, \\
A &= \frac{T + 1 + \sigma_r}{2}, \\
B &= \frac{1}{2} \left\{ S_\sigma + (h_0 - \mu)^2 (1-\phi^2) + \sum_{t=1}^{T} (h_t - \mu(1-\phi) - \phi h_{t-1})^2 \right\}.
\end{aligned}$$

It is straightforward to sample from these distributions.

The log of the full conditional distribution of ϕ is represented by:

$$\ln f(\phi|\cdot) = \text{const} + g(\phi) - \frac{\sum_{t=1}^{T} \left\{h_t - \mu(1-\phi) - \phi h_{t-1}\right\}^2}{2\sigma_\eta^2}, \tag{A.3}$$

where

$$g(\phi) = \ln f(\phi) - \frac{(h_1 - \mu)^2 (1-\phi^2)}{2\sigma_\eta^2} + \frac{1}{2} \ln(1-\phi^2).$$

To sample from this distribution, we use the method of Chib and Greenberg (1994), which is based on the Metropolis-Hastings algorithm. Specifically, given the current value $\phi^{(i-1)}$ at the $(i-1)$-st iteration, sample a proposal value ϕ^* from the truncated normal distribution $N(\hat{\phi}, V_\phi)I[-1, 1]$ where $\hat{\phi} = \sum_{t=1}^{T}(h_t - \mu)(h_{t-1} - \mu) / \sum_{t=1}^{T}(h_{t-1} - \mu)^2$ and $V_\phi = \sigma_\eta^2 / \left\{\sum_{t=0}^{T}(h_t - \mu)^2\right\}$.

Then, accept this proposal value as $\phi^{(i)}$ with probability $\exp\{g(\phi^*)-g(\phi^{(i-1)})\}$. If the proposal value is rejected, set $\phi^{(i)}$ to equal $\phi^{(i-1)}$.

The full conditional distribution of h_0 is:

$$h_0|\cdot \sim N(\mu(1-\phi)+\phi h_1, \sigma_\eta^2). \tag{A.4}$$

It is straightforward to sample from this distribution.

The log of the conditional distribution (7) is expressed as

$$\begin{aligned} &\ln f(\{\eta_s\}_{s=t}^{t+k}|\cdot) \\ &= \text{const} + \ln f(\{\eta_s\}_{s=t}^{t+k}) + \ln f(\{r_s\}_{s=t}^{t+k}|\{h_s\}_{s=t}^{t+k}) \\ &= \text{const} - \frac{1}{2\sigma_\eta^2}\sum_{s=t}^{t+k}\eta_s^2 + \sum_{s=t}^{t+k}\ln f(r_s|h_s). \end{aligned} \tag{A.5}$$

For the SV-normal model, we have

$$\ln f(r_s|h_s) = \text{const} - \frac{1}{2}h_s - \frac{r_s^2}{2}\exp(-h_s). \tag{A.6}$$

We denote (A.6) by $l(h_s)$ and write the derivative of this density with respect to h_s as l' and l'' respectively. Applying a Taylor series expansion to the log-density $\ln f(\{\eta_s\}_{s=t}^{t+k}|\cdot)$ around some preliminary estimate $\{\hat{\eta}_s\}_{s=t}^{t+k}$, we have

$$\begin{aligned} &\ln f(\{\eta_s\}_{s=t}^{t+k}|\cdot) \\ &\approx \text{const} - \frac{1}{2\sigma_\eta^2}\sum_{s=t}^{t+k}\eta_s^2 + \sum_{s=t}^{t+k}\left\{l(\hat{h}_s) + (h_s-\hat{h}_s)l'(\hat{h}_s) + \frac{1}{2}(h_s-\hat{h}_s)^2 l''(\hat{h}_s)\right\} \\ &= \ln g, \end{aligned} \tag{A.7}$$

where $\{\hat{h}_s\}_{s=t}^{t+k}$ are the estimate of $\{h_s\}_{s=t}^{t+k}$ corresponding to $\{\hat{\eta}_s\}_{s=t}^{t+k}$.

Define variables v_s and $\hat{y}_s$ as follows.[3] For $s = t, \ldots, t+k-1$,

$$v_s = -1/l''(\hat{h}_s) \tag{A.8}$$

$$\hat{y}_s = \hat{h}_s + v_s l'(\hat{h}_s). \tag{A.9}$$

[3] Shephard and Pitt (1997) define v_s and $\hat{y}_s$ for all s using equations (A.8) and (A.9). Watanabe and Omori (2001) show that this mistake may cause a significant bias in estimates for the both parameters and latent variables. When $t+k < T$, v_{t+k} and $\hat{y}_{t+k}$ must be defined using equations (A.10) and (A.11).

For $s = t + k$, if $t + k < T$,

$$v_{t+k} = \sigma_\eta^2 / \left\{ \phi - \sigma_\eta^2 l''(\hat{h}_{t+k}) \right\} \tag{A.10}$$

$$\hat{y}_{t+k} = \hat{h}_{t+k} + v_{t+k} \left[l'(\hat{h}_{t+k}) + \phi \left\{ h_{t+k+1} - \mu(1-\phi) - \phi \hat{h}_{t+k} \right\} \right], \tag{A.11}$$

and if $t + k = T$,

$$v_{t+k} = -1/l''(\hat{h}_{t+k}) \tag{A.12}$$

$$\hat{y}_{t+k} = \hat{h}_{t+k} + v_{t+k} l'(\hat{h}_{t+k}). \tag{A.13}$$

Then, the normalized version of g is a k-dimensional normal density, which is the exact density of $\{\eta_s\}_{s=t}^{t+k}$ conditional on $\{\hat{y}_s\}_{s=t}^{t+k}$ in the linear Gaussian state space model:

$$\hat{y}_s = h_s + \epsilon_s, \quad \epsilon_s \sim N(0, v_s), \tag{A.14}$$

$$h_s = \phi h_{s-1} + \eta_s, \quad \eta_s \sim N(0, \sigma_\eta^2), \tag{A.15}$$

Applying the de Jong and Shephard (1995) simulation smoother to this model with the artificial $\{\hat{y}_s\}_{s=t}^{t+k}$ enables us to sample $\{\eta_s\}_{s=t}^{t+k}$ from the density g.

Following Shephard and Pitt (1997), we select the expansion block $\{\hat{h}_s\}_{s=t}^{t+k}$ as follows. Once an initial expansion block $\{\hat{h}_s\}_{s=t}^{t+k}$ is selected, we can calculate the artificial $\{\hat{y}_s\}_{s=t}^{t+k}$. Then, applying the Kalman smoother to the linear Gaussian state space model that consists of equations (A.14) and (A.15) with the artificial $\{\hat{y}_s\}_{s=t}^{t+k}$ yields the mean of $\{h_s\}_{s=t}^{t+k}$ conditional on $\{\hat{y}_s\}_{s=t}^{t+k}$ in the linear Gaussian state space model, which is used as the next $\{\hat{h}_s\}_{s=t}^{t+k}$. We use five iterations of this procedure to obtain the expansion block $\{\hat{h}_s\}_{s=t}^{t+k}$.

Since g does not bound f, we cannot use the conventional acceptance-rejection sampling method to simulate $\{\eta_s\}_{s=t}^{t+k}$ from the true density f.

Instead, Shephard and Pitt (1997) suggest using the Acceptance-Rejection/ Metropolis-Hastings (A-R/M-H) algorithm proposed by Tierney (1994) (see also Chib and Greenberg (1995) for details). Let us denote the previously sampled value of $\{\eta_s\}_{s=t}^{t+k}$ by x. Suppose that the candidate y is produced from the acceptance-rejection algorithm. Then, the A-R/M-H algorithm proceeds as follows.

1. If $f(x) < g(x)$, then let $\alpha = 1$;
   If $f(x) \geq g(x)$ and $f(y) < g(y)$, then let $\alpha = g(x)/f(x)$;
   If $f(x) \geq g(x)$ and $f(y) \geq g(y)$, then let $\alpha = \min\left\{\frac{f(y)g(x)}{f(x)g(y)}, 1\right\}$.

2. Generate u from a standard uniform distribution.

3. If $u \leq \alpha$, return y.
   Else, return x.

To implement the multi-move sampler, we must select the knots. Following Shephard and Pitt (1997), we select the K knots, that is equivalently $K+1$ blocks, randomly with U_i being independent uniforms and

$$k_i = \text{int}[T \times \{(i + U_i)/(K+2)\}], \quad i = 1, \dots, K,$$

where $\text{int}[x]$ represents the operator that rounds x down to the nearest integer. The stochastic knots ensure that the method does not become stuck by an excessive amount of rejections. In all of the analyses in this paper, the number of knots K are set equal to 40 for the Yen/Dollar exchange rate and 50 for the TOPIX.

A.2 The SV-t Model

For the SV-t model that consists of equations (1), (2), and (8), we sample $\{\omega_t\}_{t=1}^{T}$ and ν as well as the other parameters and $\{h_t\}_{t=0}^{T}$ from their full conditional distributions. The full conditional distributions for μ, ϕ, σ_η^2, and

h_0 are the same as those in the SV-normal model. It is straightforward to sample $\{h_t\}_{t=1}^T$ from their full conditional distributions using the multi-move sampler. All we have to do is to replace equation (A.5) by

$$\ln f(r_s|h_s,\omega_s) = \text{const} - \frac{1}{2}h_s - \frac{r_s^2}{2\omega_s}\exp(-h_s). \tag{A.16}$$

Conditional on ν, the $(\epsilon_t^2+\nu-2)/\omega_t$ follow independent $\chi^2(\nu+1)$ distribution. Hence, it is straightforward to sample from $f(\{\omega_t\}_{t=1}^T|\nu,\{\epsilon_t\}_{t=1}^T)$.

Under the prior (9), the log of conditional distribution of ν is given by

$$\begin{aligned}&\ln f(\nu|\,\{\omega_t\}_{t=1}^T)\\ &= \text{const} + \frac{T\nu}{2}\ln\left(\frac{\nu-2}{2}\right) - T\ln\Gamma\left(\frac{\nu}{2}\right) - \eta\nu,\end{aligned} \tag{A.17}$$

where

$$\eta = \frac{1}{2}\sum_{t=1}^{T}\left\{\ln(\omega_t) + \frac{1}{\omega_t}\right\} + \lambda. \tag{A.18}$$

To sample from this distribution, we use the method proposed by Watanabe (2001), which is based on the A-R/M-H algorithm proposed by Tierney (1994). Specifically, we use a normal distribution as a proposal density in the A-R step. The mean and variance of this distribution are chosen as follows. We find the mode of ν^* by numerical optimization, which is used for the mean, and calculate $d^2\ln f(\nu|\cdot)/d\nu^2$ at $\nu=\nu^*$, which is used for the variance. It follows from the following theorem that the variance is always positive.

Theorem. *For the full conditional distribution of ν defined in equation (A.17), the second derivative is negative for all $\nu>0$, that is*

$$\frac{d^2\ln f(\nu|\cdot)}{d\nu^2} < 0. \tag{A.19}$$

Proof. Differentiating equation (A.17), we have

$$\frac{d^2 \ln f(\cdot)}{d\nu^2} = \frac{T}{2}\left\{\frac{\nu-4}{(\nu-2)^2} - \frac{1}{2}\psi'\left(\frac{\nu}{2}\right)\right\}, \qquad \text{(A.20)}$$

where $\psi(x)$ is a psi (digamma) function defined by $\psi(x) = d\ln\Gamma(x)/dx$ and $\psi'(x)$ is a trigamma function defined by $\psi'(x) = d\psi(x)/dx$.

$\ln\Gamma(\nu/2)$ is represented by

$$\ln\Gamma\left(\frac{\nu}{2}\right) = \frac{\ln(2\pi)}{2} + \frac{\nu-1}{2}\ln\left(\frac{\nu}{2}\right) - \frac{\nu}{2} + \frac{\theta}{6\nu}, \quad 0<\theta<1, \qquad \text{(A.21)}$$

(see equation 6.1.38 in Abramowitz and Stegun (1970)). Substituting the derivative of (A.21) into (A.20) yields

$$\frac{d^2 \ln f(\nu|\cdot)}{d\nu^2} = -\frac{T\theta}{3\nu^3} - \frac{T(\nu^2+4)}{2(\nu-2)^2\nu^2} < 0. \quad \square$$

A.3 The SV-GED Model

For the SV-GED model, we sample υ as well as the other parameters and $\{h_t\}_{t=0}^T$ from their full conditional distributions. The full conditional distributions for μ, ϕ, σ_η^2, and h_0 are the same as those in the SV-normal and SV-t models. We sample $\{h_t\}_{t=1}^T$ from their full conditional distributions using the multi-move sampler. All we have to do is to replace equation (A.5) by

$$\ln f(r_s|h_s) = \text{const} - \frac{h_s}{2} - \frac{1}{2}\left|\frac{r_s}{\beta}\right|^{\upsilon}\exp(-\upsilon\frac{h_s}{2}), \qquad \text{(A.22)}$$

where β is defined by equation (6). It is straightforward to prove that this is log-concave.

Under the prior (10), the log of conditional distribution of υ is given by

$$\begin{aligned}
&\ln f(\upsilon|\{r_s\}_{s=1}^T, \{h_s\}_{s=1}^T) \\
&= \text{const} + T\ln\upsilon - T\ln\beta - T\left(1+\frac{1}{\upsilon}\right)\ln 2 \\
&\quad -T\ln\Gamma\left(\frac{1}{\upsilon}\right) - \frac{1}{2}\sum_{s=1}^T\left|\frac{r_s}{\beta}\right|^{\upsilon}\exp\left(-\frac{h_s}{2}\upsilon\right). \qquad \text{(A.23)}
\end{aligned}$$

To sample from this density, we use the A-R/M-H algorithm similar to that used to sample ν in the SV-t model. The problem is that $\ln f(\upsilon|\cdot)$ is no longer concave and hence $d^2 \ln f(\upsilon|\cdot)/d\upsilon^2$ can be positive.

To see this, we use the second derivative of $\ln f(\upsilon|\cdot)$,

$$\begin{aligned}\frac{1}{T}\frac{d^2 \ln f(\upsilon|\cdot)}{d\upsilon^2} &= \left[-\frac{1}{\upsilon^2} - \frac{d^2 \ln \beta}{d\upsilon^2} - \frac{2}{\upsilon^3}\left[\ln 2 + \psi\left(\frac{1}{\upsilon}\right)\right] - \frac{1}{\upsilon^4}\psi'\left(\frac{1}{\upsilon}\right)\right] \\ &\quad - \frac{1}{2\beta^\upsilon}\frac{1}{T}\sum_{s=1}^{T}|\epsilon_s|^\upsilon\left[-\frac{d^2 \ln \beta}{d\upsilon^2} + \left(\frac{d \ln \beta}{d\upsilon} + \frac{h_s}{2}\right)^2\right]. \quad (A.24)\end{aligned}$$

Noting that

$$\psi(x) = \ln x - \frac{1}{2x} - \frac{\theta}{12x^2}, \quad 0 < \theta < 1,$$

which we used in Appendix A.2, we have

$$\begin{aligned}\frac{d \ln \beta}{d\upsilon} &= \frac{1}{\upsilon^2}(\ln 2 + 3\ln 3 - 2\ln \upsilon) + \frac{\theta}{18} > 0, \\ \frac{d^2 \ln \beta}{d\upsilon^2} &= -\frac{1}{\upsilon^3}(1 + \ln 2 + 3\ln 3 - 2\ln \upsilon) < 0,\end{aligned}$$

for $1 \leq \upsilon \leq 2$. We obtained these inequalities by $3 > e > \upsilon$. Hence, we have

$$-\frac{1}{2\beta^\upsilon}\frac{1}{T}\sum_{s=1}^{T}|\epsilon_s|^\upsilon\left[-\frac{d^2 \ln \beta}{d\upsilon^2} + \left(\frac{d \ln \beta}{d\upsilon} + \frac{h_s}{2}\right)^2\right] < 0.$$

Thus, the second term of equation (A.24) is always negative. But, the derivative of $\psi(x)$ is

$$\psi'(x) = \frac{1}{x} + \frac{1}{2x^2} + \frac{\theta}{6x^3},$$

which implies that, for the first term of (A.24),

$$\begin{aligned}&\left[-\frac{1}{\upsilon^2} - \frac{d^2 \ln \beta}{d\upsilon^2} - \frac{2}{\upsilon^3}\left[\ln 2 + \psi\left(\frac{1}{\upsilon}\right)\right] - \frac{1}{\upsilon^4}\psi'\left(\frac{1}{\upsilon}\right)\right] \\ &= \frac{1}{\upsilon^3}\left[\frac{5}{2}(2-\upsilon) + 2\ln\left(\frac{2}{\upsilon}\right) + \frac{1}{3}(4 - \theta\upsilon^2) + 6\ln 3 - 4 - 2\ln 2\right] > 0.\end{aligned}$$

Therefore, it depends on the values of $T^{-1}\sum_{s=1}^{T}|\epsilon_s|^\upsilon$, $T^{-1}\sum_{s=1}^{T}|\epsilon_s|^\upsilon h_s$ and $T^{-1}\sum_{s=1}^{T}|\epsilon_s|^\upsilon h_s^2$ that $d^2 \ln f(\upsilon|\cdot)/d\upsilon^2$ is positive or negative.

If it is positive, the variance of the proposal density will be negative. In such a case, we can implement an *ad hoc* adjustment; we set the variance of the proposal density as $-1/D_s(\hat{h}_s)$ where

$$D_s(\hat{h}_s) = \min \left[\left. \frac{d^2 \ln f(\upsilon|\cdot)}{d\upsilon^2} \right|_{\upsilon=\upsilon^*}, -0.0001 \right]. \tag{A.25}$$

We should note that $d^2 \ln f(\upsilon|\cdot)/d\upsilon^2$ is always negative in our empirical analysis.

APPENDIX B: EVALUATING BAYES FACTORS

B.1 Likelihood Estimation

Let $R = \{r_t\}_{t=1}^T$ and $H = \{h_t\}_{t=0}^T$. Then, the likelihood of the SV models is given by

$$f(R|\theta) = \int_{R^t} f(R, H|\theta) dH. \tag{B.1}$$

The joint density $f(R, H|\theta)$ can be factorized in an importance function (IF) $\psi(H|R)$ and a remainder function (RF) $\pi(H, R)$ such that

$$f(R, H|\theta) = \pi(H, R)\psi(H|R). \tag{B.2}$$

Thus, the likelihood $f(R|\theta)$ is the expectation $E_\psi [\pi(H, R)]$, which can be estimated by sampling $\left\{H^{(k)}\right\}_{k=1}^K$ from $\psi(H|R)$ and calculating the sample mean

$$\hat{f}_N(R|\theta) = \frac{1}{K} \sum_{k=1}^K \pi(H, R). \tag{B.3}$$

An initial factorization is obtained as follows:

$$\psi_0(H|R) = f(h_0) \prod_{t=2}^T f(h_t|h_{t-1}) \tag{B.4}$$

$$\pi_0(H, R) = \prod_{t=1}^T f(r_t|h_t), \tag{B.5}$$

where $f(h_t|h_{t-1})$ is the density of h_t conditional on h_{t-1}, which is, according to equation (2), the normal with mean $\mu(1-\phi)$ and variance σ_η^2. $f(h_0)$ is the unconditional density of h_t, which is the normal with mean μ and variance $\sigma_\eta^2/(1-\phi^2)$. $f(r_t|h_t)$ is the density of r_t conditional on h_t. For the SV-normal model, $f(r_t|h_t)$ is the normal with mean 0 and variance $\exp(h_t)$. For the SV-t and SV-GED models, $f(r_t|h_t)$ are respectively given by

$$f(r_t|h_t) = [\pi(\nu-2)\exp(h_t)]^{-1/2}\frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)}\left[1+\frac{r_t^2}{\exp(h_t)(\nu-2)}\right]^{-(\nu+1)/2}, \tag{B.6}$$

and

$$f(r_t|h_t) = \frac{\upsilon\exp\left[-\frac{1}{2}|r_t/\left\{\exp(h_t/2)\beta\right\}|^{\upsilon}\right]}{\exp(h_t/2)\beta\Gamma(1/\upsilon)2^{(1+1/\upsilon)}}, \tag{B.7}$$

where β is defined in equation (6).

This initial factorization is, however, inefficient in the sense that the resulting sampling variance of $\hat{f}_N(R|\theta)$ increases dramatically with the dimension of the integral T. To solve this inefficiency problem, Danielsson and Richard (1993) propose an acceleration method, called Accelerated Gaussian Importance Sampling (AGIS). The AGIS method searches for a Gaussian IF which minimizes the sampling variance of the corresponding RF.

Let Q_0 denote a 2×2 matrix and Q_t denote a 3×3 matrix for $t=1,\dots,T$. Also, define

$$\lambda_0' = (h_0, 1),$$

and

$$\lambda_t' = (h_t, h_{t-1}, 1), \quad t=1,\dots,T.$$

Then, a variance reduction function $\xi(H,Q)$ is defined as

$$\xi(H,Q) = \prod_{t=0}^{T}\xi(h_t, Q_t), \tag{B.8}$$

where

$$\xi(h_t, Q_t) = \exp\left(-\frac{1}{2}\lambda_t' Q_t \lambda_t\right). \tag{B.9}$$

This variance reduction function is used to construct a new pair of an IF and a RF as follows:

$$\psi(H|R) = \psi_0(H|R)\xi(H,Q)/k(Q), \tag{B.10}$$

$$\pi(H,R) = \pi_0(H,R)k(Q)/\xi(H,Q), \tag{B.11}$$

where $k(Q)$ represents the integration constant which ensures that the new IF integrates to one and is given by

$$k(Q) = \int_{R^T} \psi(H|R)\xi(H,Q)dH. \tag{B.12}$$

Starting with the initial IF $\psi_0(H|R)$, an initial simulated sample $\{H_{0,n}\}_{n=1}^N$ is drawn and used to run the regression for every time period $t = 1, \ldots, T$:

$$\ln \pi_0(h_{0,n,t}) = a_{1,t} + b_{1,t}h_{0,n,t} + c_{1,t}h_{0,n,t}^2 + \text{error term}, \quad n = 1, \ldots, N. \tag{B.13}$$

Then, the OLS-estimates of the coefficients are used to construct the following matrix for every period t:

$$\hat{Q}_{1,t} = \begin{pmatrix} -2\hat{c}_{1,t} & 0 & -\hat{b}_{1,t} \\ 0 & 0 & 0 \\ -\hat{b}_{1,t} & 0 & -2\hat{a}_{1,t} \end{pmatrix}, \quad t = 1, \ldots, T. \tag{B.14}$$

A first new IF is given by $\psi_1(H|R) = \psi_0(H|R)\xi(H,\hat{Q}_1)/k(\hat{Q}_1)$ where $\hat{Q}_1 = \left\{\hat{Q}_{1,t}\right\}_{t=1}^T$. A second-step IF $\psi_2(H|R)$ is constructed in the same fashion by drawing a random sample from $\psi_1(H|R)$ and regressing $\ln \pi_0(h_{1,n,t})$ on a constant, $h_{1,n,t}$, $h_{1,n,t}^2$. With the resulting sequence of matrices $\hat{Q}_2$, one can determine $\psi_2(H|R)$. This procedure is repeated until $\hat{Q}_j$ is sufficiently close to the one-step-ahead matrices $\hat{Q}_{j-1}$. Danielsson and Richard (1993) showed that the convergence is reached very quickly, typically after less than five iterations. Finally, the simulated $\{H_{j,n}\}_{n=1}^{N}$ from $\psi_j(H|R)$ is used to calculate the jth step AGIS estimate of the integral

$$\hat{f}_{N,j}(R|\theta) = \frac{1}{N}\sum_{n=1}^{N}\frac{\pi_0(H_{j,n},R)k(\hat{Q}_j)}{\xi(H_{j,n},\hat{Q}_j)}. \tag{B.15}$$

Here, a simulation sample size of $N = 5000$ and four iterations for the AGIS algorithm are used.

B.2 Posterior Density Estimation

We first consider the SV-normal model, whose posterior density at the posterior mean $(\sigma_\eta^{2*}, \mu^*, \phi^*)$ is represented by

$$\begin{aligned}&\log f(\sigma_\eta^{2*}, \mu^*, \phi^*|R)\\ =\ &\log f(\sigma_\eta^{2*}|R) + \log f(\phi^*|R, \sigma_\eta^{2*}) + \log f(\mu^*|R, \sigma_\eta^{2*}, \phi^*).\end{aligned} \tag{B.16}$$

To estimate $f(\sigma_\eta^{2*}|R)$, we do not need any additional MCMC runs. All we have to do is to substitute the MCMC draws $\left\{H^{(m)}, \mu^{(m)}, \phi^{(m)}\right\}_{m=1}^{M}$ obtained for the parameter estimation into the following equation.

$$\begin{aligned}&f\left(\sigma_\eta^{2*}|R\right)\\ =\ &\frac{1}{M}\sum_{m=1}^{M} f\left(\hat{\sigma}_\eta^{2}|R, H, \mu^{(m)}, \phi^{(m)}\right)\\ =\ &\frac{1}{M}\sum_{m=1}^{M}\exp\left(A\ln(B^{(m)}) - (A+1)\ln(\sigma_\eta^{2*}) - \ln\Gamma(A) - \frac{B^{(m)}}{\sigma_\eta^{2*}}\right),\end{aligned} \tag{B.17}$$

where

$$A = \frac{T+1+\sigma_r}{2},$$

$$\begin{aligned}B^{(m)} &= \frac{1}{2}\Big\{S_\sigma + (1-\phi^{(m)2})(h_0^{(m)} - \mu^{(m)})^2\\ &\quad + \sum_{t=1}^{T}(h_t^{(m)} - \mu^{(m)}(1-\phi^{(m)}) - \phi^{(m)}h_{t-1}^{(m)})^2\Big\}.\end{aligned}$$

It is more troublesome to estimate $f(\phi^*|R, \sigma_\eta^{2*})$ because its normalizing constant is unknown. To overcome this problem, we use the method of Chib and Jeliazkov (2001). Let $q(\phi, \phi'|R, H, \sigma_\eta^{2*}, \mu)$ denote the proposal density for the transition from ϕ to ϕ'. Specifically, $q(\phi, \phi'|\cdot)$ is given by the truncated normal distribution $N(\hat{\phi}, V_\phi)I[-1, 1]$ where $\hat{\phi} = \sum_{t=1}^{T}(h_t - \mu)(h_{t-1} - \mu)/\sum_{t=1}^{T}(h_{t-1} - \mu)^2$ and $V_\phi = \sigma_\eta^2/\left\{\sum_{t=0}^{T}(h_t - \mu)^2\right\}$. Also, let

$$\begin{aligned} &\alpha(\phi, \phi'|R, H, \sigma_\eta^{2*}, \mu) \\ &= \min\left[1, \frac{f(\phi\prime|R, H, \sigma_\eta^{2*}, \mu)q(\phi\prime, \phi|R, H, \sigma_\eta^{2*}, \mu)}{f(\phi|R, H, \sigma_\eta^{2*}, \mu)q(\phi, \phi\prime|R, H, \sigma_\eta^{2*}, \mu)}\right] \end{aligned} \tag{B.18}$$

denote the probability of move. Chib and Jeliazkov (2001) proved that the following equation holds.

$$\begin{aligned} &f(\phi^*|R, H, \sigma_\eta^{2*}) \\ &= \frac{E_1\left[\alpha(\phi, \phi^*|R, H, \sigma_\eta^{2*}, \mu)q(\phi, \phi^*|R, H, \sigma_\eta^{2*}, \mu)\right]}{E_2\left[\alpha(\phi^*, \phi|R, H, \sigma_\eta^{2*}, \mu)\right]}, \end{aligned} \tag{B.19}$$

where the numerator expectation E_1 is with respect to the distribution $f(\phi, \mu, H|R, \sigma_\eta^{2*})$ while the denominator expectation E_2 is with respect to the distribution $f(\mu, H|R, \phi^*)q(\phi^*, \phi|R, H, \sigma_\eta^{2*}, \mu)$.

To estimate the numerator, since the expectation is conditioned on σ_η^{2*}, we continue the MCMC simulation for an additional G iterations with σ_η^2 given at σ_η^{2*}. Specifically, given σ_η^{2*}, we sample ϕ and μ from their full conditional densities and H using the multi-move sampler. For the denominator, since the expectation is conditioned on σ_η^{2*} and ϕ^*, we continue the MCMC simulation for an additional J iterations. Specifically, given σ_η^{2*} and ϕ^*, we sample μ from its full conditional densities and H using the multi-move sampler. At each iteration of this reduced run, given the values $(\mu^{(j)}, H)$, we also generate a variate

$$\phi^{(j)} \sim q(\phi^*, \phi|R, \sigma_\eta^{2*}, \mu^{(j)}, H^{(j)}).$$

The resulting triple $(\mu^{(j)}, H^{(j)}, \phi^{(j)})$ is a draw from the distribution

$$f(\mu, H|R, \sigma_\eta^{2*}, \phi^*)q(\phi^*, \phi|R, \sigma_\eta^{2*}, \mu, H).$$

The marginal ordinate can now be estimated as

$$\begin{aligned} & f(\phi^*|R, \sigma_\eta^{2*}) \\ = & \quad \frac{\frac{1}{G}\sum_{g=1}^{G}\left[\alpha(\phi^{(g)}, \phi^*|R, \sigma_\eta^{2*}, \mu^{(g)}, H^{(g)})q(\phi^{(g)}, \phi^*|R, \sigma_\eta^{2*}, \mu^{(g)}, H^{(g)})\right]}{\frac{1}{J}\sum_{j=1}^{J}\left[\alpha(\phi^*, \phi^{(j)}|R, \sigma_\eta^{2*}, \mu^{(j)}, H^{(j)})\right]}. \end{aligned} \tag{B.20}$$

Next, to estimate the reduced conditional ordinate $f(\mu^*|R, \sigma_\eta^{2*}, \phi^*)$, we only need to use the values of $H^{(j)}$ from the above reduced run to form the average

$$f(\mu^*|R, \sigma_\eta^{2*}, \phi^*) = \frac{1}{J}\sum_{j=1}^{J} f(\mu^*|R, \sigma_\eta^{2*}, \phi^*, H^{(j)}).$$

A similar method can be applied to the SV-t and SV-GED models. We apply the method of Chib and Jeliazkov (2001) to the estimation of the conditional densities $f(\nu^*|\cdot)$ and $f(\upsilon^*|\cdot)$. Notice, however, that this method requires the normalized constant of the proposal densities $q(\nu, \nu'|\cdot)$ and $q(\upsilon, \upsilon'|\cdot)$. If we use the A-R/M-H algorithm to sample ν and υ, the normalizing constant of the proposal density is unknown. Hence, we use the A-R/M-H algorithm in the MCMC runs for the parameter estimation while we remove the A-R step and use the M-H algorithm without the A-R step in those for the posterior density estimation. Then, the proposal densities $q(\nu, \nu'|\cdot)$ and $q(\upsilon, \upsilon'|\cdot)$ in the latter MCMC runs are given by the truncated normal distributions $N(\nu^*, B)I[4, \infty]$ and $N(\upsilon^*, C)I[1, 2]$, where ν^* and υ^* are the mode of the conditional densities $f(\nu|\cdot)$ and $f(\upsilon|\cdot)$ and $B = d^2 \ln(\nu|\cdot)/d\nu^2|_{\nu=\nu^*}$ and $C = d^2 \ln(\upsilon|\cdot)/d\upsilon^2|_{\nu=\nu^*}$. In this article, we set $G = J = 5000$.

APPENDIX C: SAMPLING FROM GED

The density function of the GED with mean zero and variance one is given by equation (5), which can be written as

$$f(x) = K\beta^{-1} \exp\left(-\frac{1}{2}\left|\frac{x}{\beta}\right|^{v}\right), \quad 0 < v < \infty, \tag{C.1}$$

where

$$\beta = \left[2^{-2/v}\frac{\Gamma(1/v)}{\Gamma(3/v)}\right]^{1/2}, \quad K = v\left[\Gamma(1/v)2^{(1+1/v)}\right]^{-1}.$$

Since the cumulative distribution function (CDF) of the GED is expressed by using the incomplete gamma function, we propose a sampling method by the probability integral transformation.

When $x < 0$, the CDF of GED is given by

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} K\beta^{-1} \exp(-(-y/\beta)^{v}/2)dy \\ &= \int_{x^*}^{\infty} (K2^{1/v}/v)t^{1/v-1}e^{-t}dt \\ &= (K2^{1/v}/v)\int_{x^*}^{\infty} t^{1/v-1}e^{-t}dt \\ &= (K\Gamma(1/v)2^{1/v}/v)\{1 - P(1/v, x^*)\} \\ &= \frac{1}{2}\left\{1 - P\left(\frac{1}{v}, x^*\right)\right\}, \end{aligned}$$

where $x^* = (-x/\beta)^v/2$, and $P(\cdot,\cdot)$ is the incomplete gamma function defined in Abramowitz and Stegun (1970, p.260). We can easily verify that

$$F(0) = \frac{\Gamma(1/v, 0)}{2\Gamma(1/v)} = \frac{\Gamma(1/v)}{2\Gamma(1/v)} = 1/2.$$

Since GED is a symmetric distribution, we have, for $x \geq 0$,

$$F(x) = \frac{1}{2}\left\{1 - P\left(\frac{1}{v}, \frac{1}{2}\left(\frac{x}{\beta}\right)^{v}\right)\right\}.$$

Setting $u = F(x)$, we have the inverse function

$$x = F^{-1}(u) = \begin{cases} -\beta\left\{2P^{-1}\left(\frac{1}{v}, 1-2u\right)\right\}^{1/v} & (u < 0.5) \\ \beta\left\{2P^{-1}\left(\frac{1}{v}, 2u-1\right)\right\}^{1/v} & (u \geq 0.5). \end{cases} \tag{C.2}$$

We can rewrite the above equation by using the CDF of a χ^2-variant. Since the CDF of $z \sim \chi^2(k)$ is given by $Q(z \mid k) = P(k/2, z/2)$, the CDF of the GED becomes

$$F(x) = \begin{cases} \frac{1}{2}\{1 - Q((-x/\beta)^v \mid 2/v)\} & (x < 0) \\ \frac{1}{2}\{1 + Q((x/\beta)^v \mid 2/v)\} & (x \geq 0). \end{cases}$$

Thus, its inverse function is

$$x = F^{-1}(u) = \begin{cases} -\beta\left\{Q^{-1}\left(1 - 2u \middle| \frac{2}{v}\right)\right\}^{1/v} & (u < 0.5) \\ \beta\left\{Q^{-1}\left(2u - 1 \middle| \frac{2}{v}\right)\right\}^{1/v} & (u \geq 0.5). \end{cases} \tag{C.3}$$

We can, therefore, sample from GED by using equation (C.2) or (C.3) and the probability integral transformation.

References

Abramowitz, M. and Stegun, N. (1970), *Handbook of Mathematical Functions,* (Dover Publications Inc., New York).

Bauwens, L., and M. Lubrano (1998), "Bayesian Inference on GARCH Models using the Gibbs Sampler," *Econometrics Journal,* **1,** c23–c46.

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics,* **31,** 307–327.

Bollerslev, T., R.F. Engle, and D.B. Nelson (1994), "ARCH Models," in R. F. Engle and D. McFadden, eds., *The Handbook of Econometrics,* **4,** North-Holland: Amsterdam.

Chib, S. (1995), "Marginal Likelihood from the Gibbs Output," *Journal of the American Statistical Association,* **90,** 1313–1321.

Chib, S., and Greenberg, E. (1994), "Bayes Inference for Regression Models with ARMA(p,q) Errors," *Journal of Econometrics,* **64,** 183–206.

Chib, S., and Greenberg, E. (1995), "Understanding the Metropolis-Hasting Algorithm," *The American Statistician,* **49,** 327–335.

Chib, S., and Jeliazkov, I. (2001), "Marginal Likelihood from the Metropolis-Hastings Output," *Journal of the American Statistical Association,* **96,** 270–291.

Danielsson, J. (1994), "Stochastic Volatility in Asset Prices: Estimation with Simulated Maximum Likelihood," *Journal of Econometrics,* **64,** 375–400.

Danielsson, J., and Richard, J. F. (1993), "Accelerated Gaussian Importance Sampler With Applications to Dynamic Latent Variable Models," *Journal of Applied Econometrics,* **8,** 153–173.

de Jong, P., and Shephard, N. (1995), "The Simulation Smoother for Time Series Models," *Biometrika,* **82,** 339–350.

Diebold, F. X. (1988), *Empirical Modeling of Exchange Rate Dynamics,* New York: Springer-Verlag.

Fama, E.F. (1965), "The Behavior of Stock Market Prices," *Journal of Business,* **38**, 34–105.

Fernández, C. and Steel, M. F. J. (1998), "On Bayesian Modeling of Fat Tails and Skewness," *Journal of the American Statistical Association,* **93**, 359–371.

Geweke, J. (1992a), "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments," in *Bayesian Statistics 4,* eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Oxford, U.K.: Oxford University Press, pp.169–193.

Geweke, J. (1992b), "Priors for Macroeconomic Time Series and Their Application," Discussion Paper 64, Federal Reserve Bank of Minneapolis, Institute for Empirical Macroeconomics.

Geweke, J. (1993), "Bayesian Treatment of the Student-t Linear Model," *Journal of Applied Econometrics,* **8**, S19–S40.

Ghysels, E., Harvey, A.C., and Renault, E. (1996), "Stochastic Volatility," in *Handbook of Statistics 14,* eds. G. S. Maddala and C. R. Rao, Amsterdam, Netherlands: Elsevier Science B.V., 119–191.

Hansen, B.E. (1994), "Autoregressive Conditional Density Estimation," *International Economic Review,* **35,** 705–730.

Harvey, A.C., and Shephard, N. (1996), "Estimation of an Asymmetric Stochastic Volatility Model for Asset Returns," *Journal of Business & Economic Statistics,* **14,** 429–434.

Jacquier, E., N.G. Polson and P.E. Rossi (1994), "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business & Economic Statistics*, **12**, 371–389.

Kim, S., Shephard, N., and Chib, S. (1998), "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies,* **65,** 361–393.

Liesenfeld, R. and Jung, R.C. (2000), "Stochastic Volatility Models: Conditional Normality versus Heavy-Tailed Distributions," *Journal of Applied Econometrics,* **15,** 137–160.

Mandelbrot, B. (1963), "The Variation of Certain Speculative Prices," *Journal of Business,* **36**, 394–419.

Mitsui, H., and Watanabe, T. (2000), "Bayesian Analysis of GARCH Option Pricing Models," (in Japanese), Research Paper Series No.21, Faculty of Economics, Tokyo Metropolitan University.

Nakatsuma, T. (2000), "Bayesian Analysis of ARMA-GARCH Models: A Markov Chain Sampling Approach," *Journal of Econometrics,* **95,** 57–69.

Schwartz, G. (1978), "Estimating the Dimension of a Model," *Annals of Statistics,* **6,** 461–464.

Shephard, N. (1996), "Statistical Aspects of ARCH and Stochastic Volatility," in *Time Series Models in Econometrics, Finance and Other Fields,* eds. D. R. Cox, D. V. Hinkley, and O. E. Barndorff-Nielsen, London, U.K: Chapman & Hall, pp.1–67.

Shephard, N., and Pitt, M. K. (1997), "Likelihood Analysis of Non-Gaussian Measurement Time Series," *Biometrika,* **84,** 653–667.

Tierney, L. (1994), "Markov Chains for Exploring Posterior Distributions" (with discussion), *Annals of Statistics,* **21,** 1701–1762.

Wang, K.-L., Fawson, C., Barrett, C.B., and McDonald, J.B. (2001), "A Flexible Parametric GARCH Model with an Application to Exchange Rates," *Journal of Applied Econometrics,* **16,** 521–536.

Watanabe, T. (1999), "A Non-linear Filtering Approach to Stochastic Volatility Models with an Application to Daily Stock Returns," *Journal of Applied Econometrics,* **14,** 101–121.

Watanabe, T. (2000), "Excess Kurtosis of Conditional Distribution for Daily Stock Returns: The Case of Japan," *Applied Economics Letters,* **7**, 353–355.

Watanabe, T. (2001), "On Sampling the Degree-of-Freedom of Student-t Disturbances," *Statistics and Probability Letters,* **52**, 177–181.

Watanabe, T., and Y. Omori (2001), "A Note on Multi-move Sampler for Estimating Non-Gaussian Time Series Models: Comments on Shephard and Pitt (1997)," Research Paper Series No.25, Faculty of Economics, Tokyo Metropolitan University.

TABLE 1. Descriptive Statistics of Daily Returns (%) for the Yen/Dollar Exchange Rate and the TOPIX

Statistics	Yen/Dollar rate	TOPIX	
	Raw data	Raw data	Residual
Sample size	2467	2403	2401
Mean	-0.0143	-0.0268	0
St. dev.	0.7395	1.2753	1.2615
Kurtosis	6.5889	7.4684	7.2813
LB(12)	8.73	35.09	10.40

NOTE: For the Yen/Dollar exchange rate, statistics for the raw return series are calculated. For the TOPIX, statistics for the raw return series and the residual series from the AR(2) model are calculated. LB(12) is the heteroskedasticity-corrected Ljung-Box statistic including twelve lags for the return series. The corrected Ljung-Box statistic is calculated following Diebold (1988). The critical values for LB(12) are: 18.55 (10%), 21.03 (5%), 26.22 (1%).

TABLE 2. Estimation Results for the Yen/Dollar Exchange Rate

(A) Bayes Factors

	t	GED
Normal	-6.96	-3.82
t	–	3.14

NOTE: The numbers in the table are log (base 10) of Bayes factors for row model against column model.

(B) Parameter Estimates

Parameter	Mean	Standard Error	95% Interval	CD
Normal Distribution				
$\exp(\mu/2)$	0.6307	0.0008	[0.5620,0.7051]	-1.43
ϕ	0.9583	0.0013	[0.9317,0.9780]	-1.52
σ_η	0.2113	0.0039	[0.1595,0.2785]	1.48
σ_h^2	0.5678	0.0040	[0.4011,0.7995]	0.79
kurtosis	5.3224	0.0203	[4.4802,6.6734]	0.76
Student t Distribution				
$\exp(\mu/2)$	0.6503	0.0011	[0.5478,0.7621]	-0.84
ϕ	0.9827	0.0007	[0.9688,0.9930]	1.50
σ_η	0.1234	0.0032	[0.0897,0.1622]	-1.69
ν	8.1161	0.1851	[5.8650,11.5648]	0.34
σ_h^2	0.4830	0.0034	[0.2882,0.8514]	-1.23
kurtosis	7.7837	0.1612	[5.5983,11.8064]	-0.07
GED				
$\exp(\mu/2)$	0.6441	0.0004	[0.5593,0.7353]	-0.34
ϕ	0.9742	0.0010	[0.9532,0.9881]	1.50
σ_η	0.1542	0.0036	[0.1118,0.2174]	-1.40
υ	1.5715	0.0048	[1.4270,1.7357]	-1.15
σ_h^2	0.4896	0.0025	[0.3205,0.7695]	-0.61
kurtosis	6.0101	0.0266	[4.9353,7.8503]	-0.47

NOTE: The first 5000 draws are discarded and then the next 10000 are used for calculating the posterior means, the standard errors of the posterior means, 95% interval, and the convergence diagnostic (CD) statistics proposed by Geweke (1992). The posterior means are computed by averaging the simulated draws. The standard errors of the posterior means are computed using a Parzen window with a bandwidth of 1000. The 95% intervals are calculated using the 2.5th and 97.5th percentiles of the simulated draws. The CD is computed using equation (12), where we set $n_A = 1000$ and $n_B = 5000$ and compute $\hat{\sigma}_A^2$ and $\hat{\sigma}_B^2$ using a Parzen window with bandwidths of 100 and 500 respectively. σ_h^2 represents the unconditional variance of volatility $\sigma_\eta^2/(1-\phi^2)$

TABLE 3. Estimation Results for the TOPIX

(A) Bayes Factors

	t	GED
Normal	-2.67	-1.28
t	–	1.39

NOTE: The numbers in the table are log (base 10) of Bayes factors for row model against column model.

(B) Parameter Estimates

Parameter	Mean	Standard Error	95% Interval	CD
Normal Distribution				
$\exp(\mu/2)$	1.0449	0.0011	[0.9254,1.1772]	0.84
ϕ	0.9552	0.0008	[0.9334,0.9737]	1.19
σ_η	0.2478	0.0026	[0.1997,0.3020]	-1.26
σ_h^2	0.7221	0.0031	[0.5279,0.9890]	-0.79
kurtosis	6.2223	0.0195	[5.0859,8.0660]	-0.67
Student t Distribution				
$\exp(\mu/2)$	1.0699	0.0021	[0.9300,1.2268]	-0.68
ϕ	0.9676	0.0008	[0.9485,0.9823]	-0.67
σ_η	0.2002	0.0027	[0.1581,0.2497]	0.71
ν	13.2045	0.5077	[7.3395,21.4837]	-0.52
σ_h^2	0.6545	0.0033	[0.4531,0.9739]	0.21
kurtosis	7.3334	0.0997	[5.7225,10.5928]	0.42
GED				
$\exp(\mu/2)$	1.0617	0.0011	[0.9351,1.2110]	-0.05
ϕ	0.9627	0.0006	[0.9424,0.9792]	0.90
σ_η	0.2154	0.0024	[0.1640,0.2666]	-1.09
υ	1.6975	0.0037	[1.5281,1.8873]	-1.33
σ_h^2	0.6561	0.0036	[0.4640,0.9376]	-1.11
kurtosis	6.5877	0.0151	[5.3530,8.6671]	-0.68

NOTE: The first 5000 draws are discarded and then the next 10000 are used for calculating the posterior means, the standard errors of the posterior means, 95% interval, and the convergence diagnostic (CD) statistics proposed by Geweke (1992). The posterior means are computed by averaging the simulated draws. The standard errors of the posterior means are computed using a Parzen window with a bandwidth of 1000. The 95% intervals are calculated using the 2.5th and 97.5th percentiles of the simulated draws. The CD is computed using equation (12), where we set $n_A = 1000$ and $n_B = 5000$ and compute $\hat{\sigma}_A^2$ and $\hat{\sigma}_B^2$ using a Parzen window with bandwidths of 100 and 500 respectively. σ_h^2 represents the unconditional variance of volatility $\sigma_\eta^2/(1-\phi^2)$

TABLE 4. Confidence Intervals for Return at $T+1$

	Normal	t	GED
Exchange Rate			
90%	[-0.7632,0.7943]	[-0.9821,0.9967]	[-0.8200,0.8216]
95%	[-0.9419,0.9761]	[-1.2528,1.2406]	[-1.0263,1.0393]
99%	[-1.3385,1.4120]	[-1.9177,1.9182]	[-1.5398,1.4701]
mean of $\exp(h_{T+1}/2)$	0.4479	0.5171	0.4896
TOPIX			
90%	[-3.2686,3.1970]	[-3.3220,3.3391]	[-3.1261,3.1016]
95%	[-4.0301,3.9822]	[-4.3015,4.2163]	[-3.8508,3.8180]
99%	[-5.6504,5.4389]	[-6.4485,6.1505]	[-5.4893,5.6495]
mean of $\exp(h_{T+1}/2)$	1.9128	1.8544	1.8429

TABLE 5. Maximum Likelihood Estimates for GARCH Models

(A) Yen/Dollar Exchange Rate

	Normal	t	GED
a_0	0.0180	0.0076	0.0109
	(0.0049)	(0.0029)	(0.0040)
a_1	0.0871	0.0603	0.0702
	(0.0136)	(0.0120)	(0.0140)
b_1	0.8805	0.9279	0.9102
	(0.0199)	(0.0145)	(0.0188)
ν		5.5129	
		(0.5946)	
υ			1.3298
			(0.0495)
log-likelihood	-1.0434	-1.0127	-1.0177
$a_1 + b_1$	0.9676	0.9882	0.9804

(B) TOPIX-Residual

	Normal	t	GED
a_0	0.0703	0.0492	0.0574
	(0.0094)	(0.0125)	(0.0137)
a_1	0.1443	0.1187	0.1275
	(0.0127)	(0.0170)	(0.0179)
b_1	0.8166	0.8554	0.8394
	(0.0160)	(0.0197)	(0.0222)
ν		5.8686	
		(0.7142)	
υ			1.3408
			(0.0480)
log-likelihood	-1.5478	-1.5199	-1.5234
$a_1 + b_1$	0.9610	0.9741	0.9669

NOTE: QML standard errors are in parenthesis.

Figure 1: PDF of the Standard Normal, t, and GED
Yen/Dollar Exchange Rate
Center Area

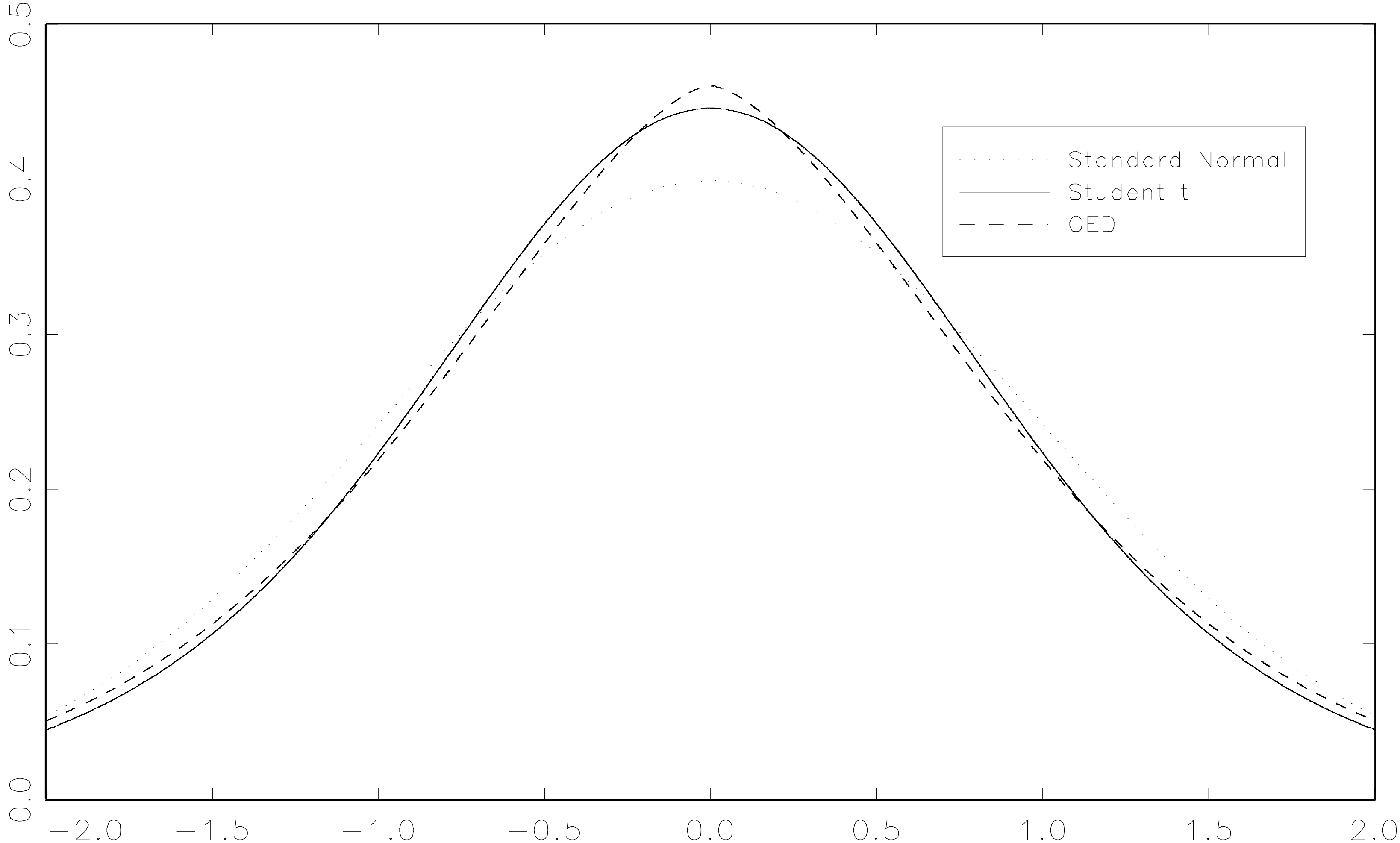

Figure 1: PDF of the Standard Normal, t, and GED
Yen/Dollar Exchange Rate
Tail Area

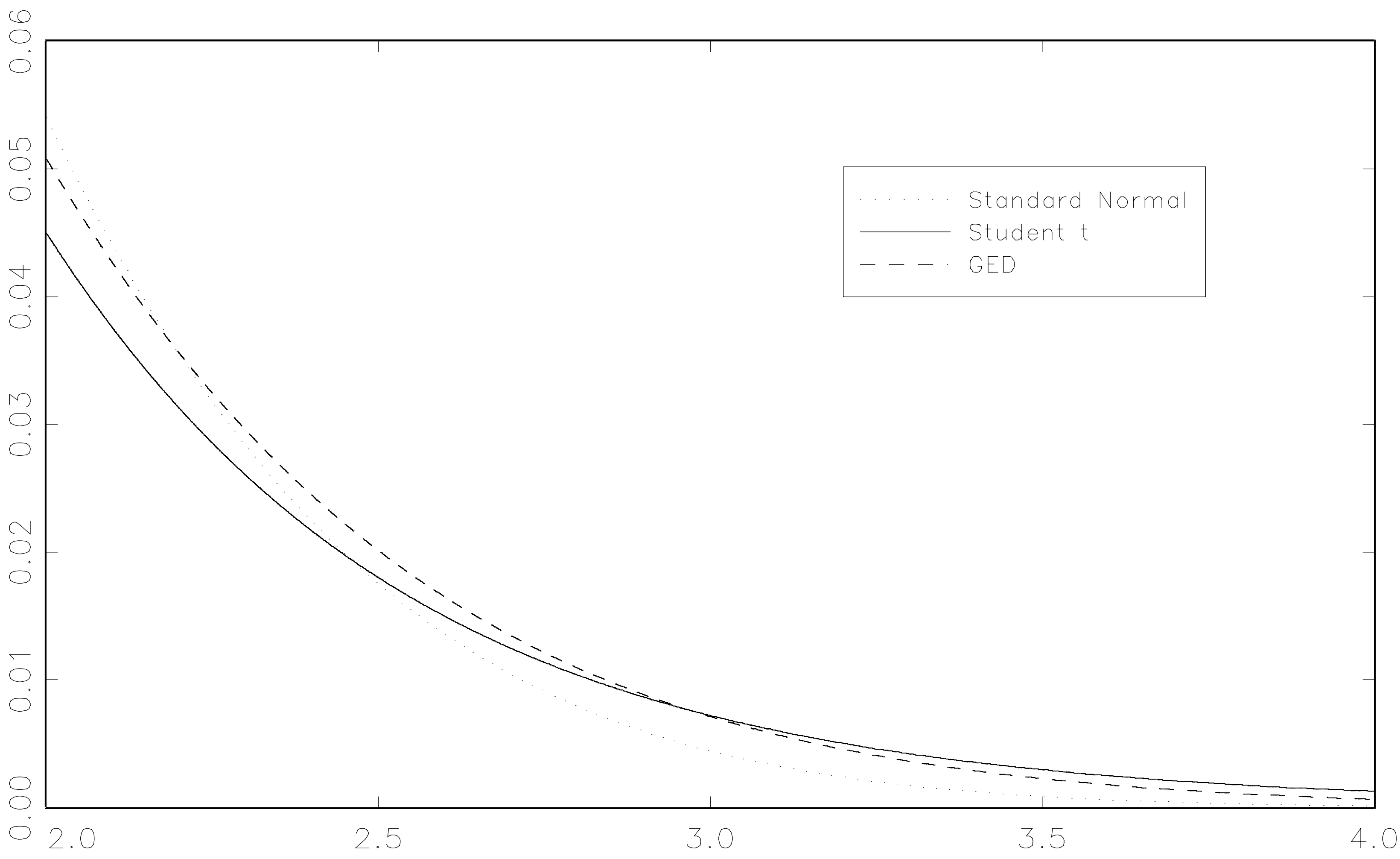

Figure 2: PDF of the Standard Normal, t, and GED
TOPIX
Center Area

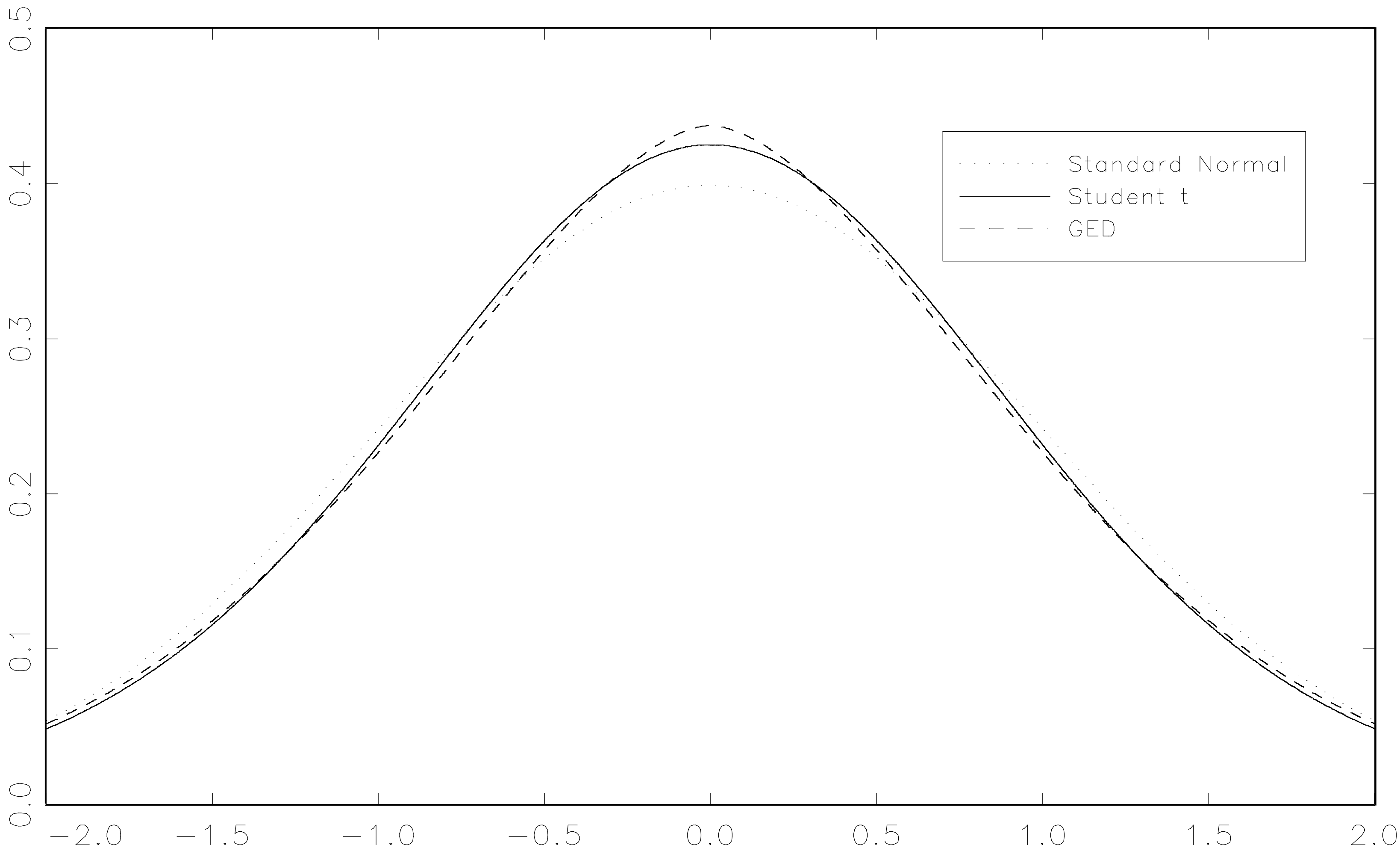

Figure 2: PDF of the Standard Normal, t, and GED
TOPIX
Tail Area

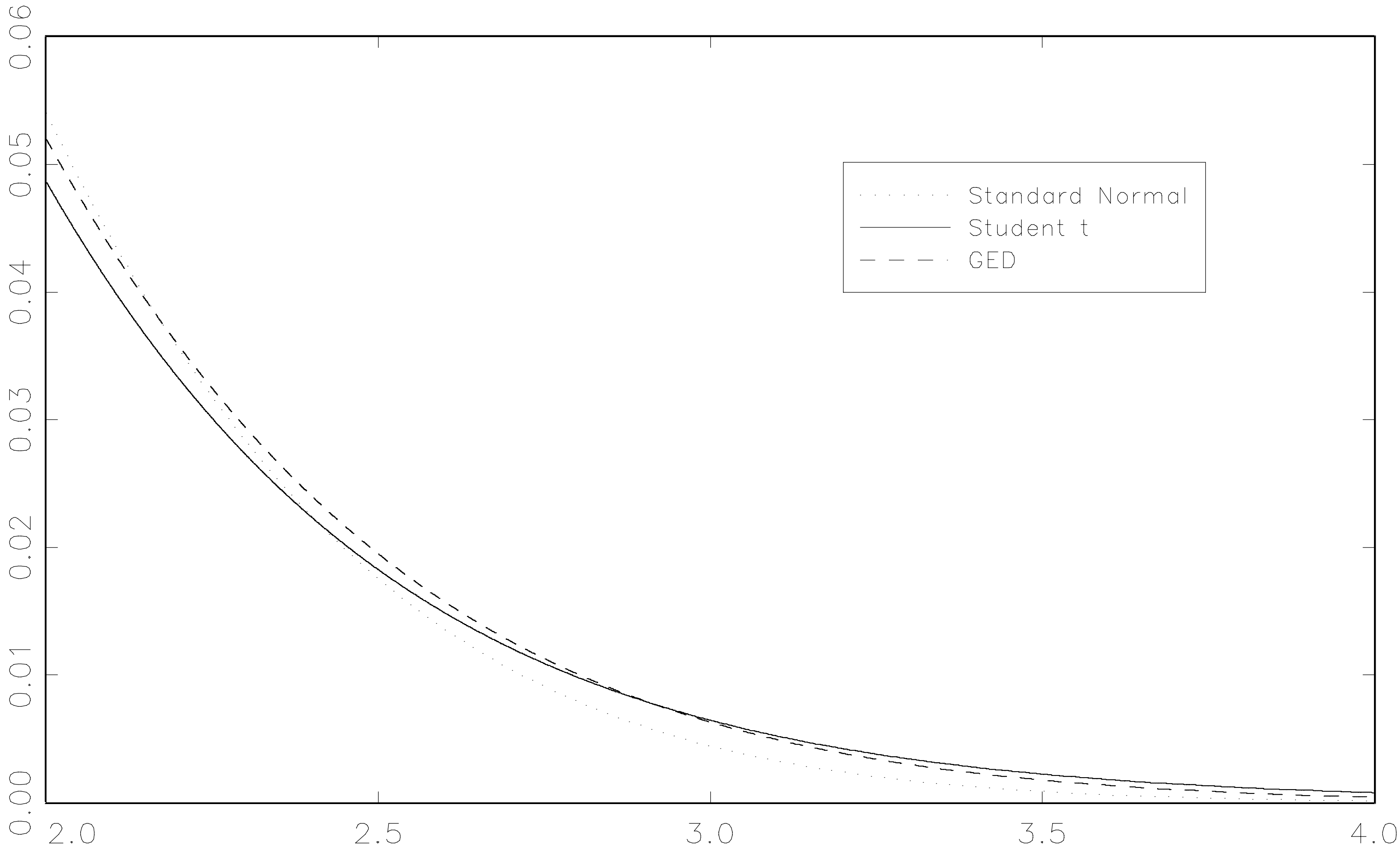

Figure 3: Volatility Estimates
Yen/Dollar Exchange Rate
Normal Distribution

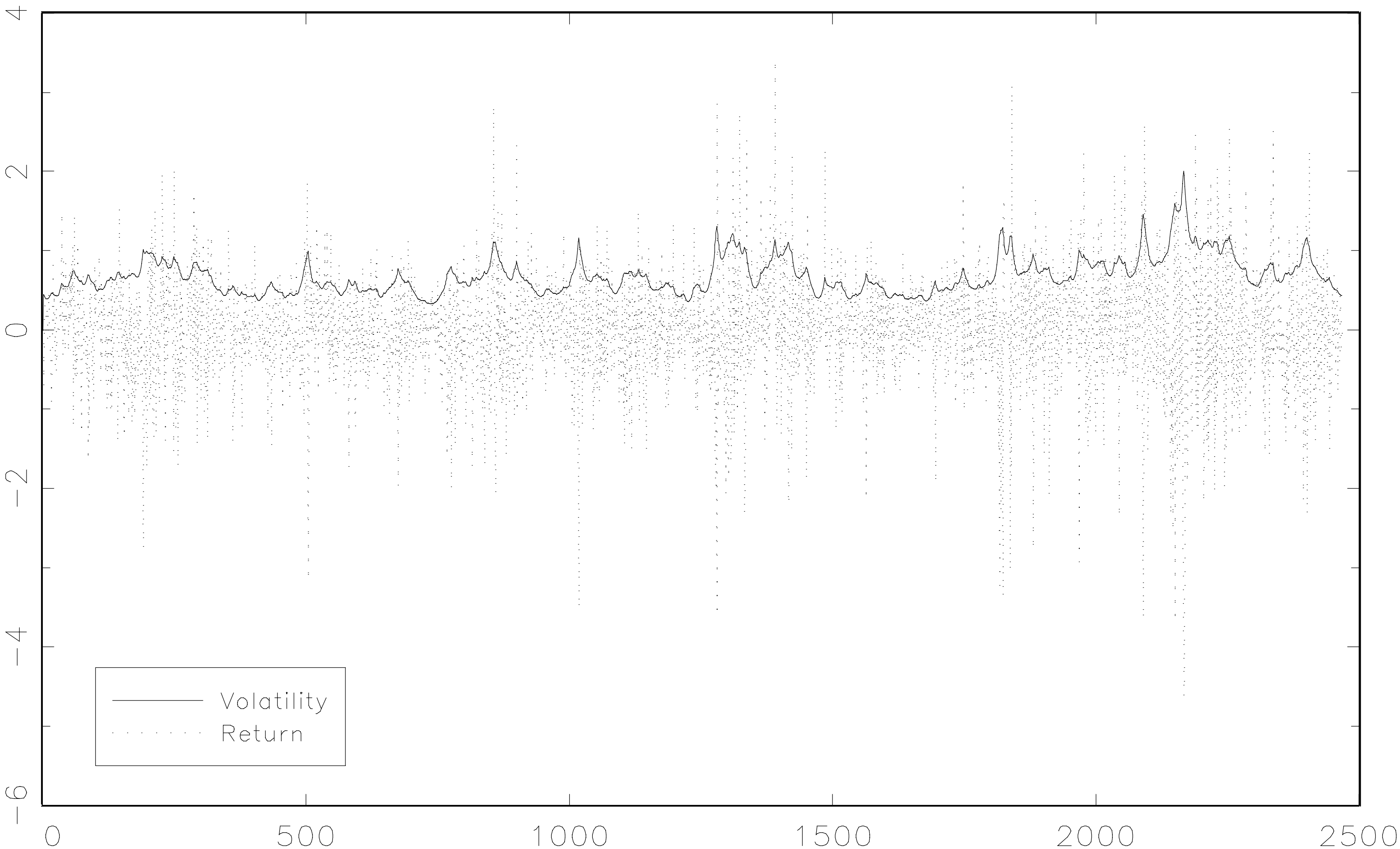

Figure 3: Volatility Estimates
Yen/Dollar Exchange Rate
Student t Distribution

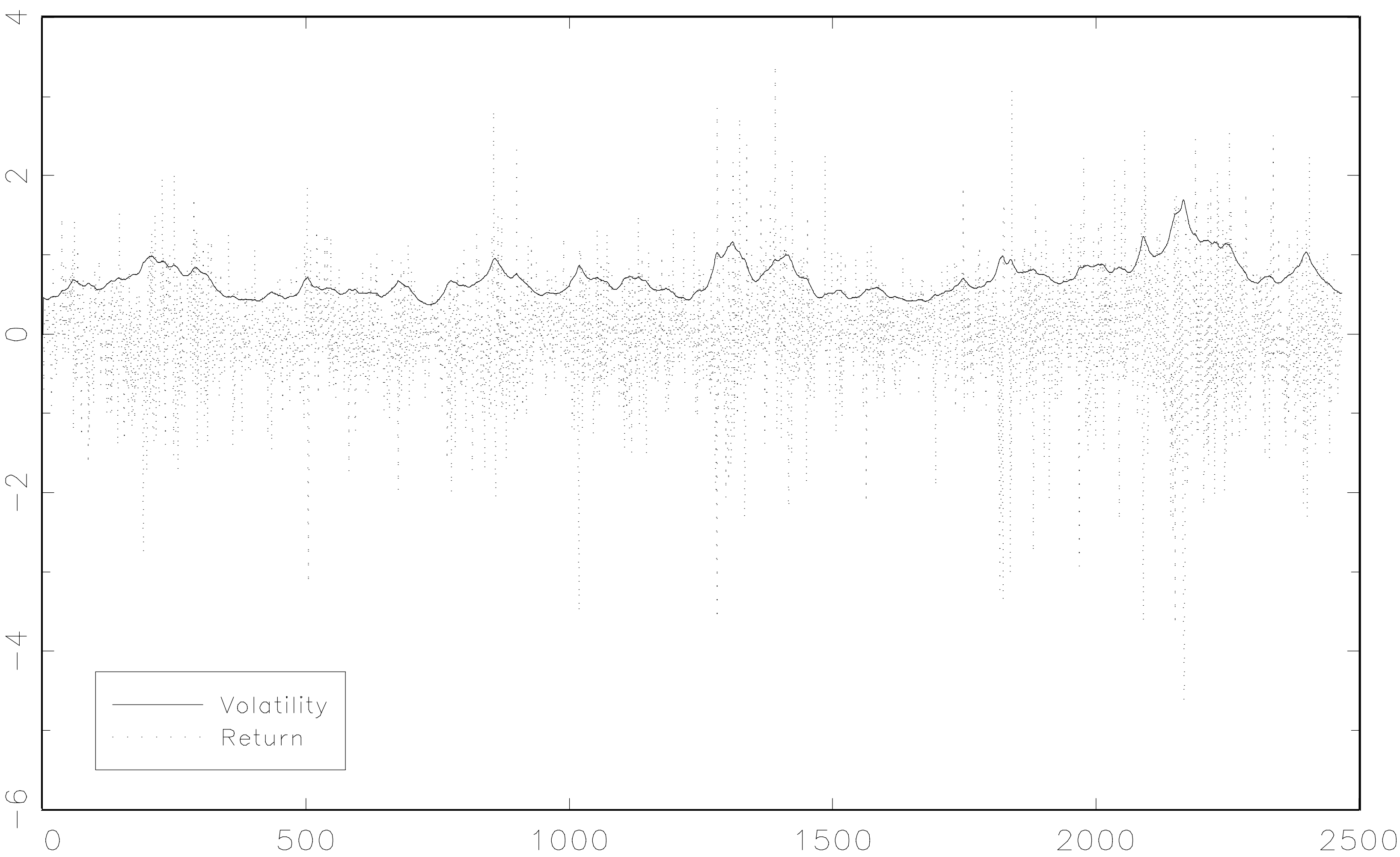

Figure 3: Volatility Estimates
Yen/Dollar Exchange Rate
GED

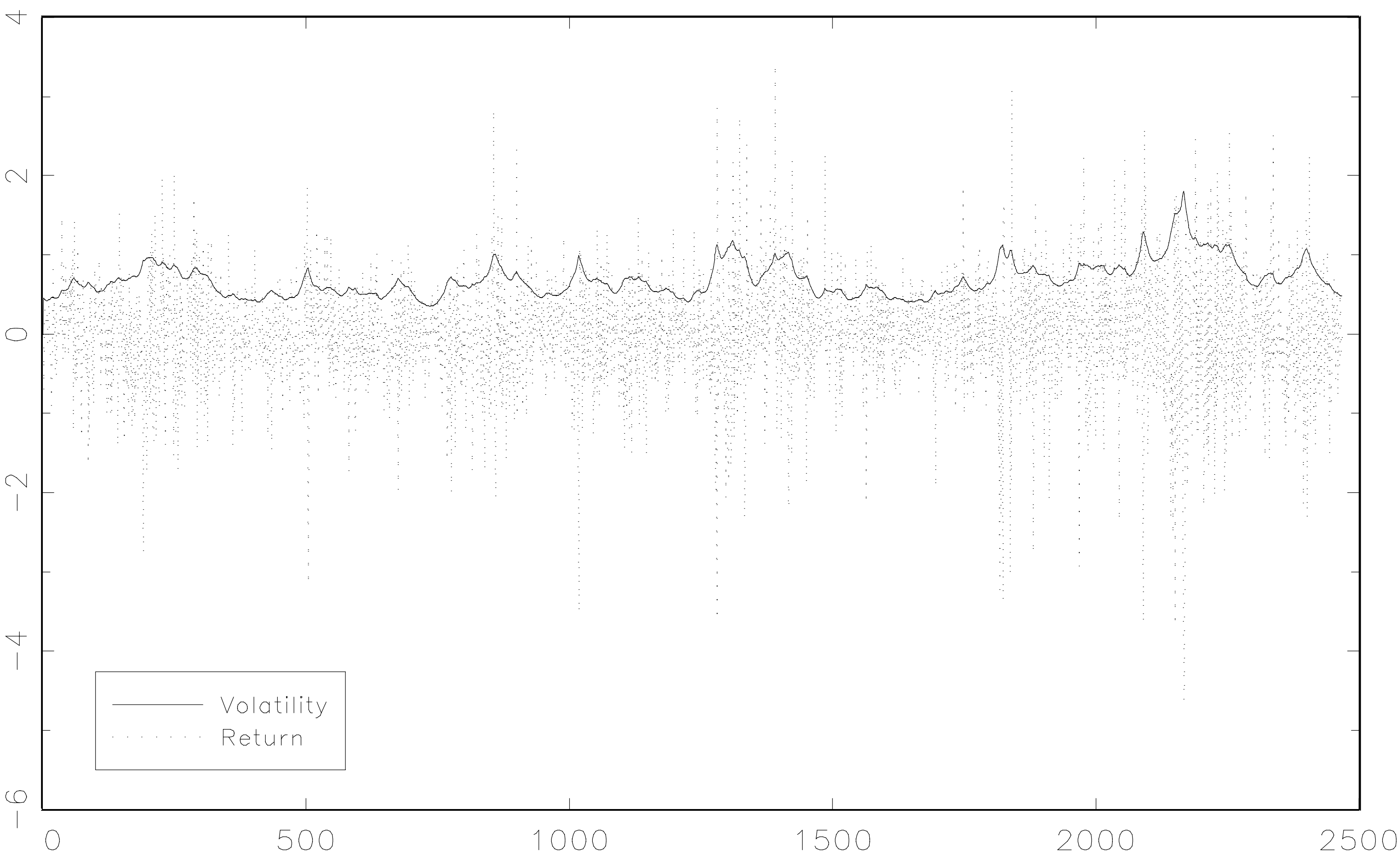

Figure 4: Volatility Estimates
TOPIX
Normal Distribution

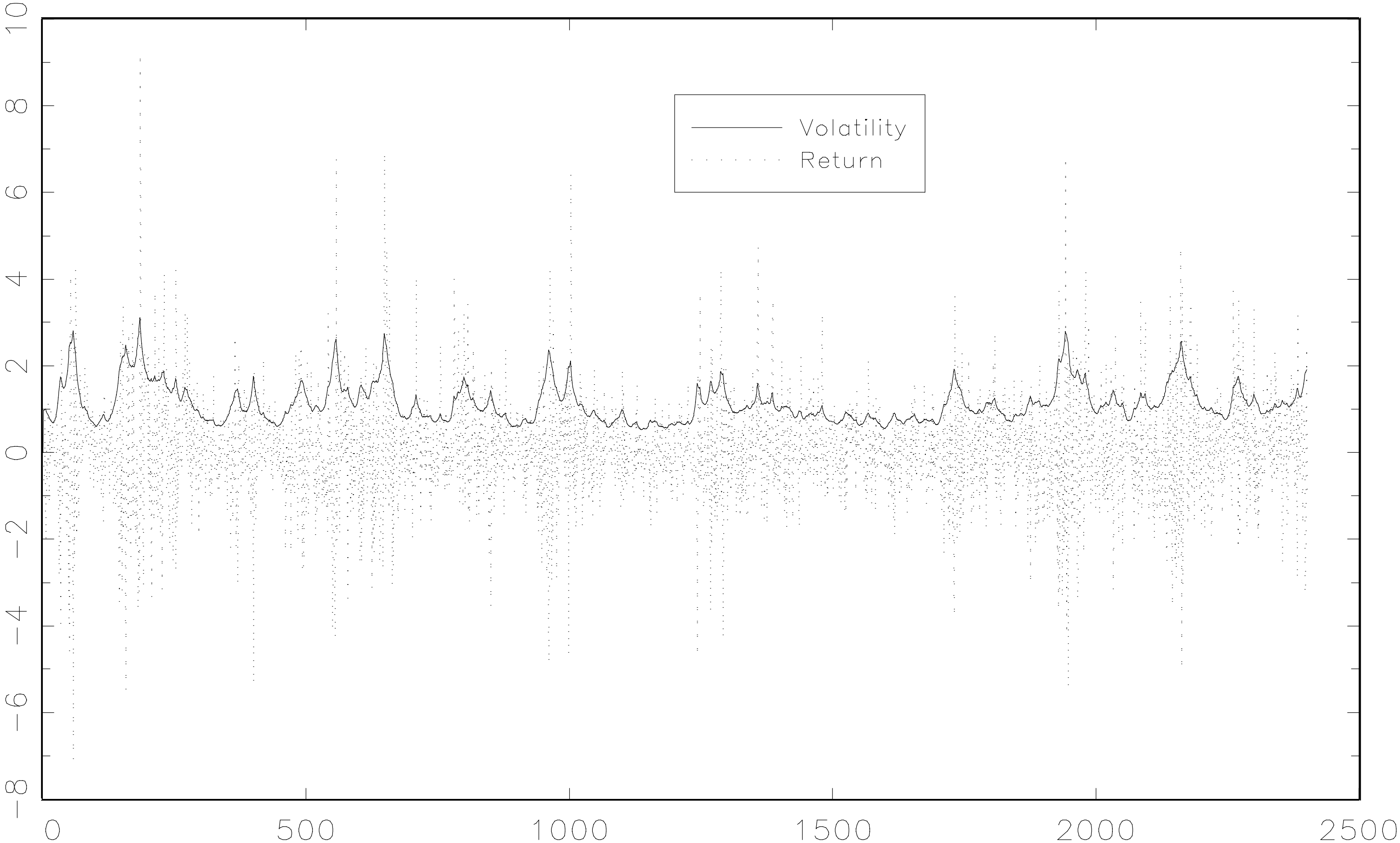

Figure 4: Volatility Estimates
TOPIX
Student t Distribution

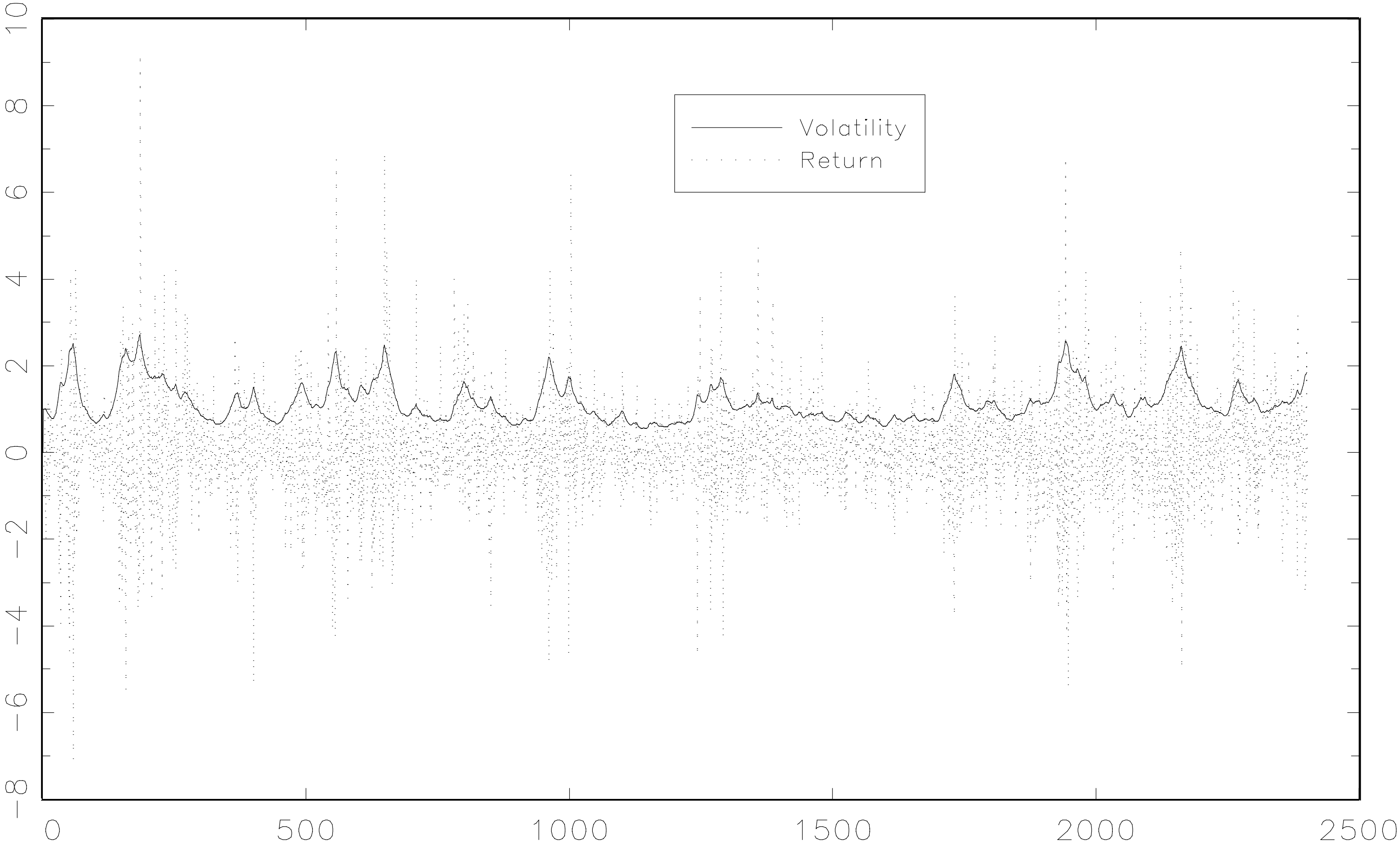

Figure 4: Volatility Estimates
TOPIX
GED

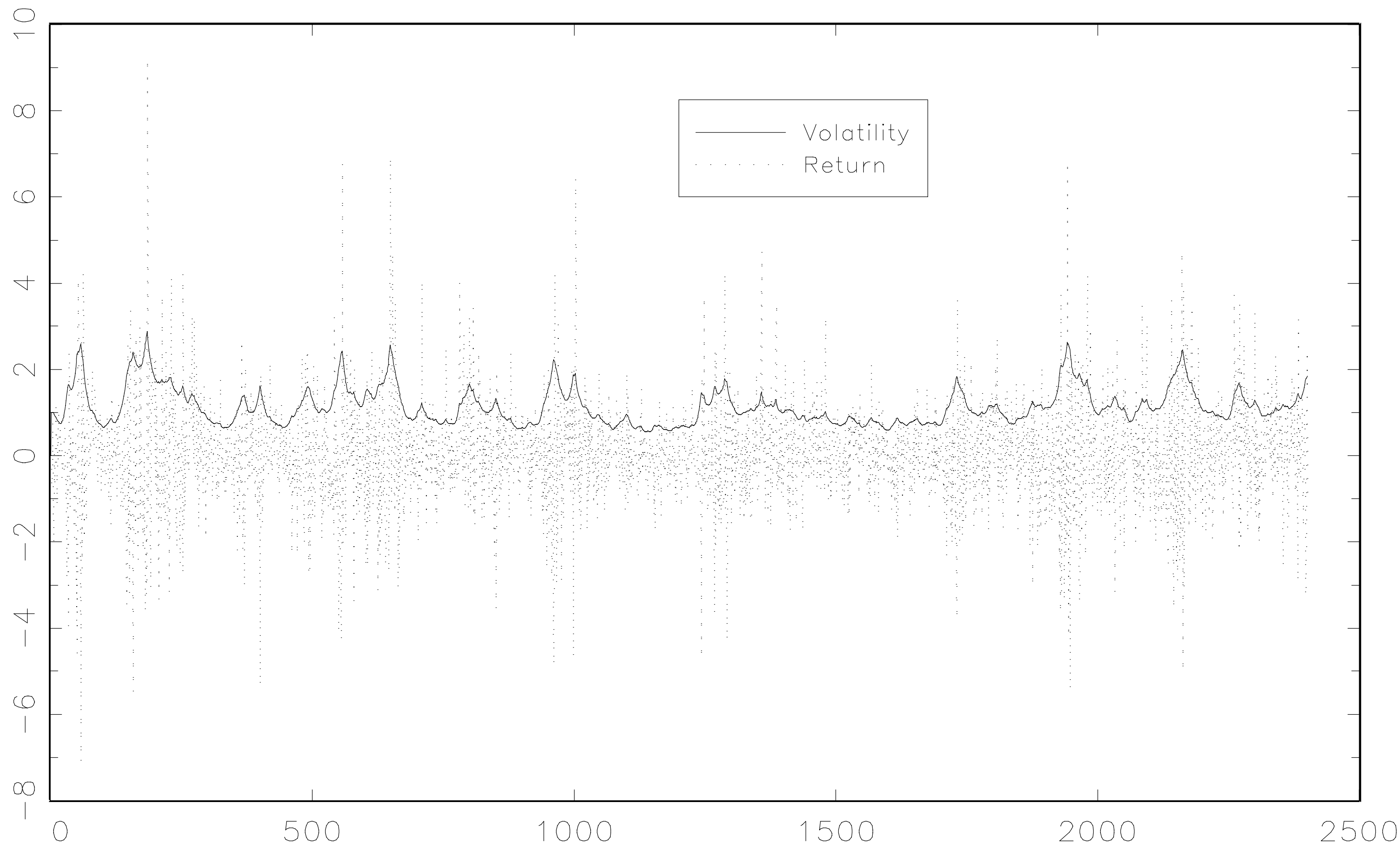

Figure 5: Confidence Interval for the Autocorrelation Coefficients of Squared Returns (Yen/Dollar Exchange Rate) Normal Distribution

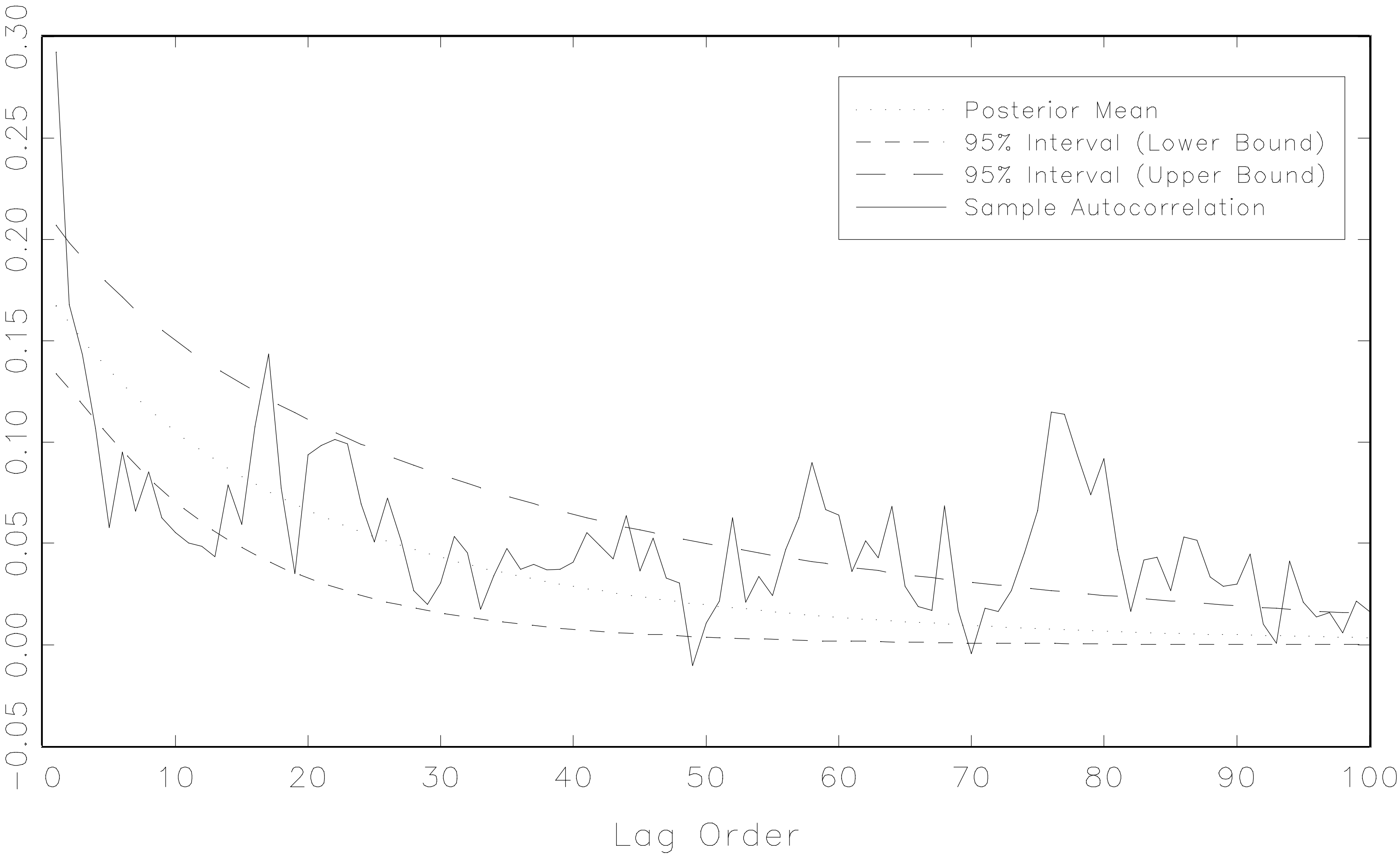

Figure 5: Confidence Interval for the Autocorrelation Coefficients of Squared Returns (Yen/Dollar Exchange Rate)
Student t Distribution

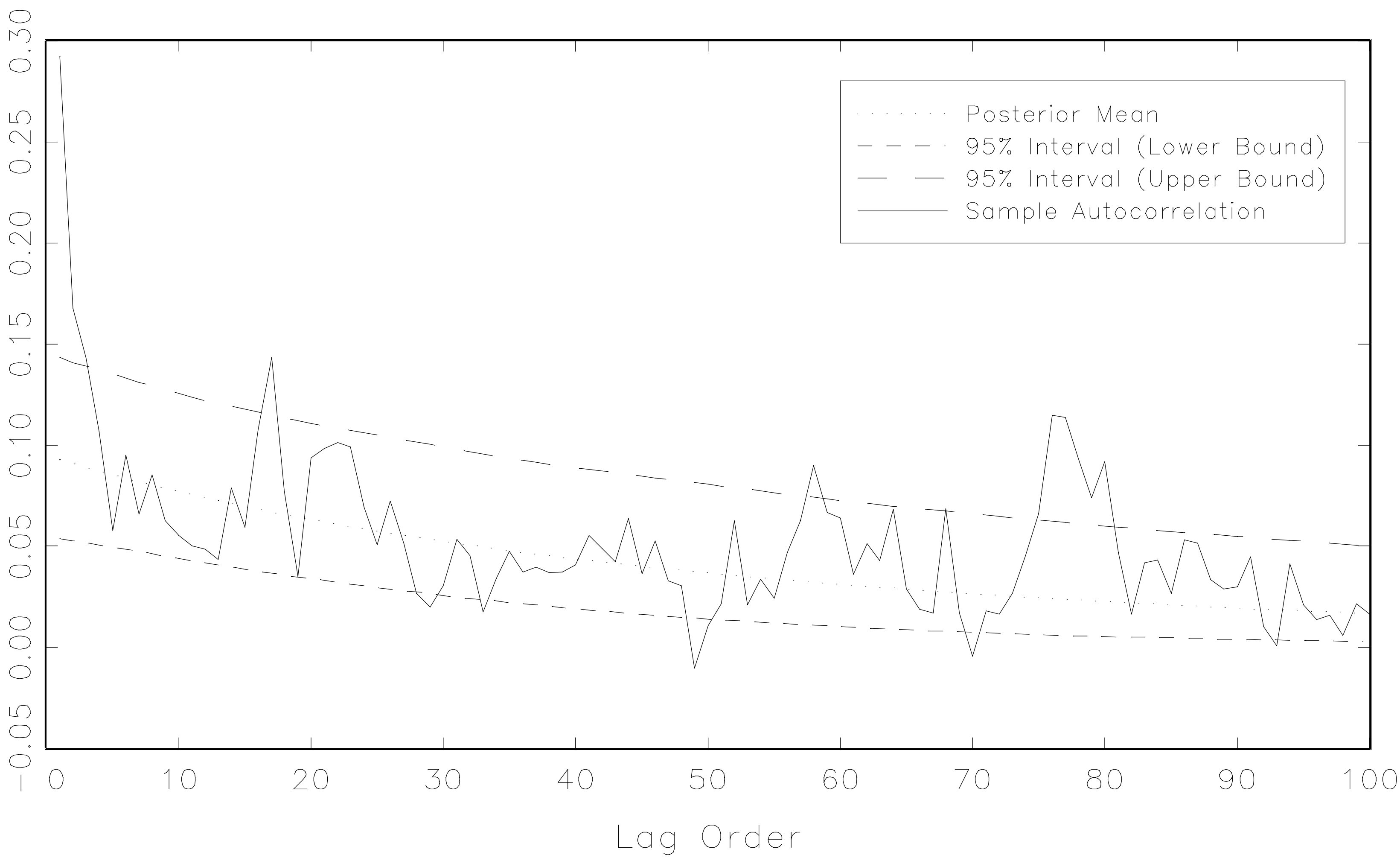

Figure 5: Confidence Interval for the Autocorrelation Coefficients of Squared Returns (Yen/Dollar Exchange Rate)
GED

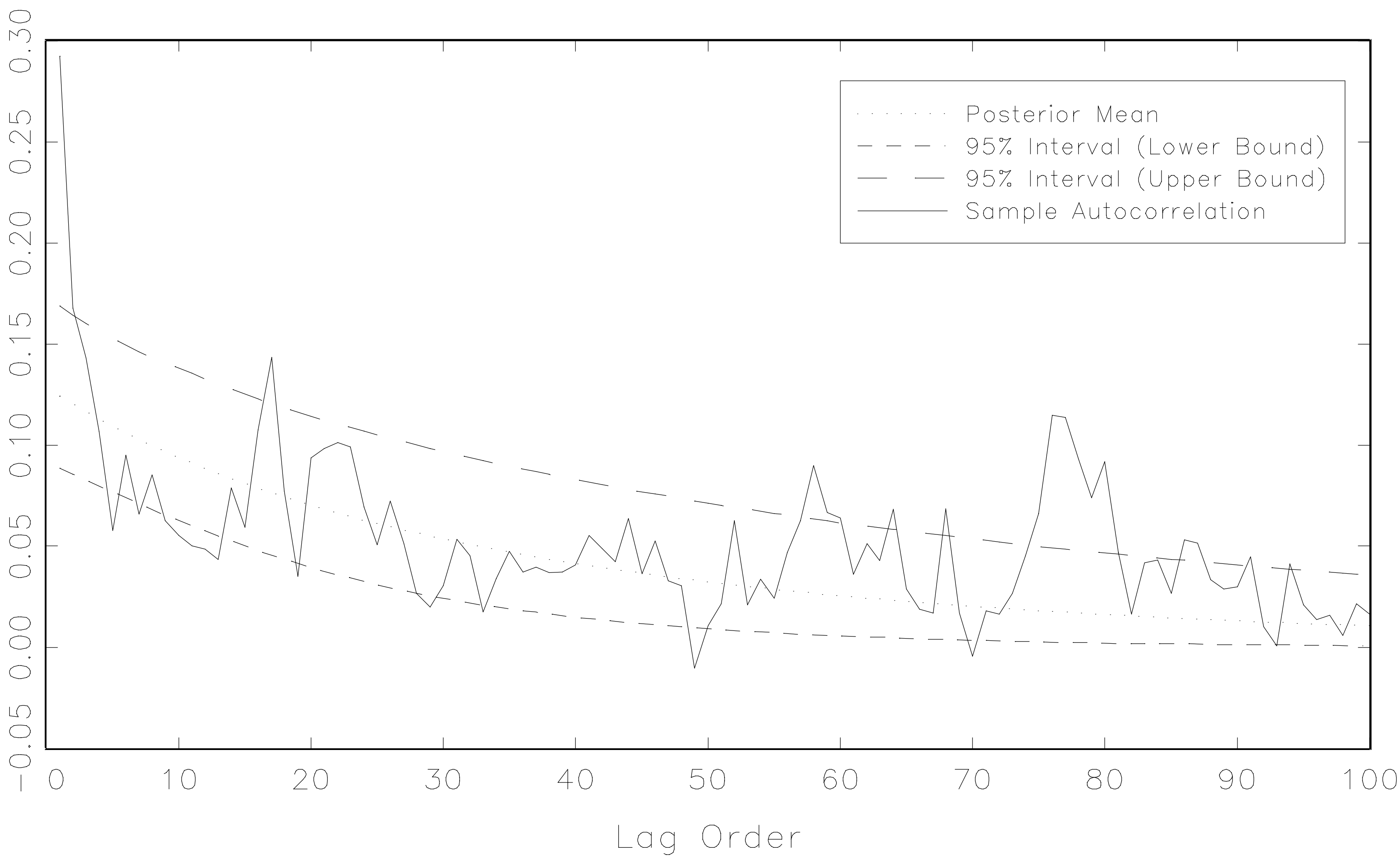

Figure 6: Confidence Interval for the Autocorrelation Coefficients of Squared Returns (TOPIX)
Normal Distribution

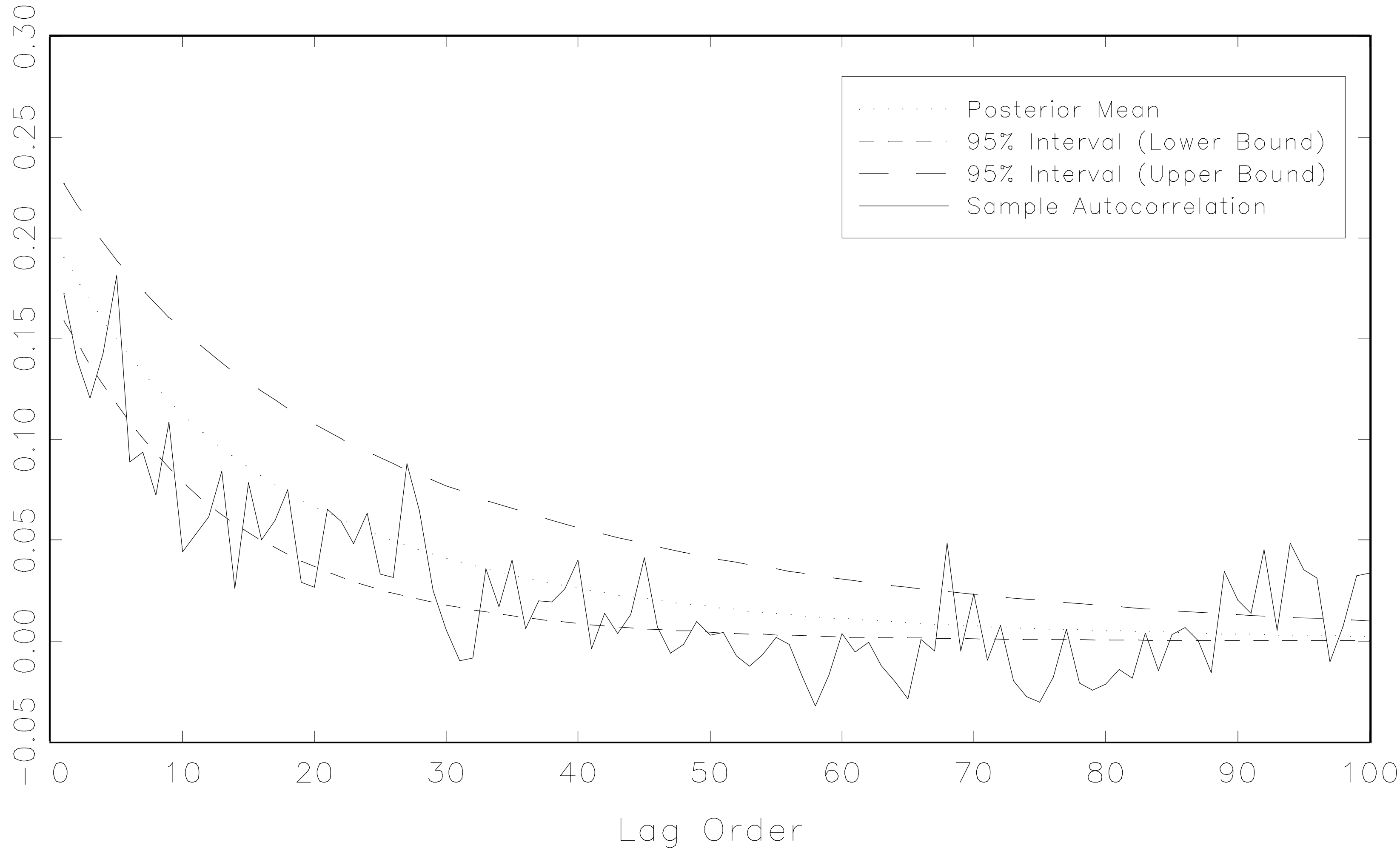

Figure 6: Confidence Interval for the Autocorrelation Coefficients of Squared Returns (TOPIX)
Student t Distribution

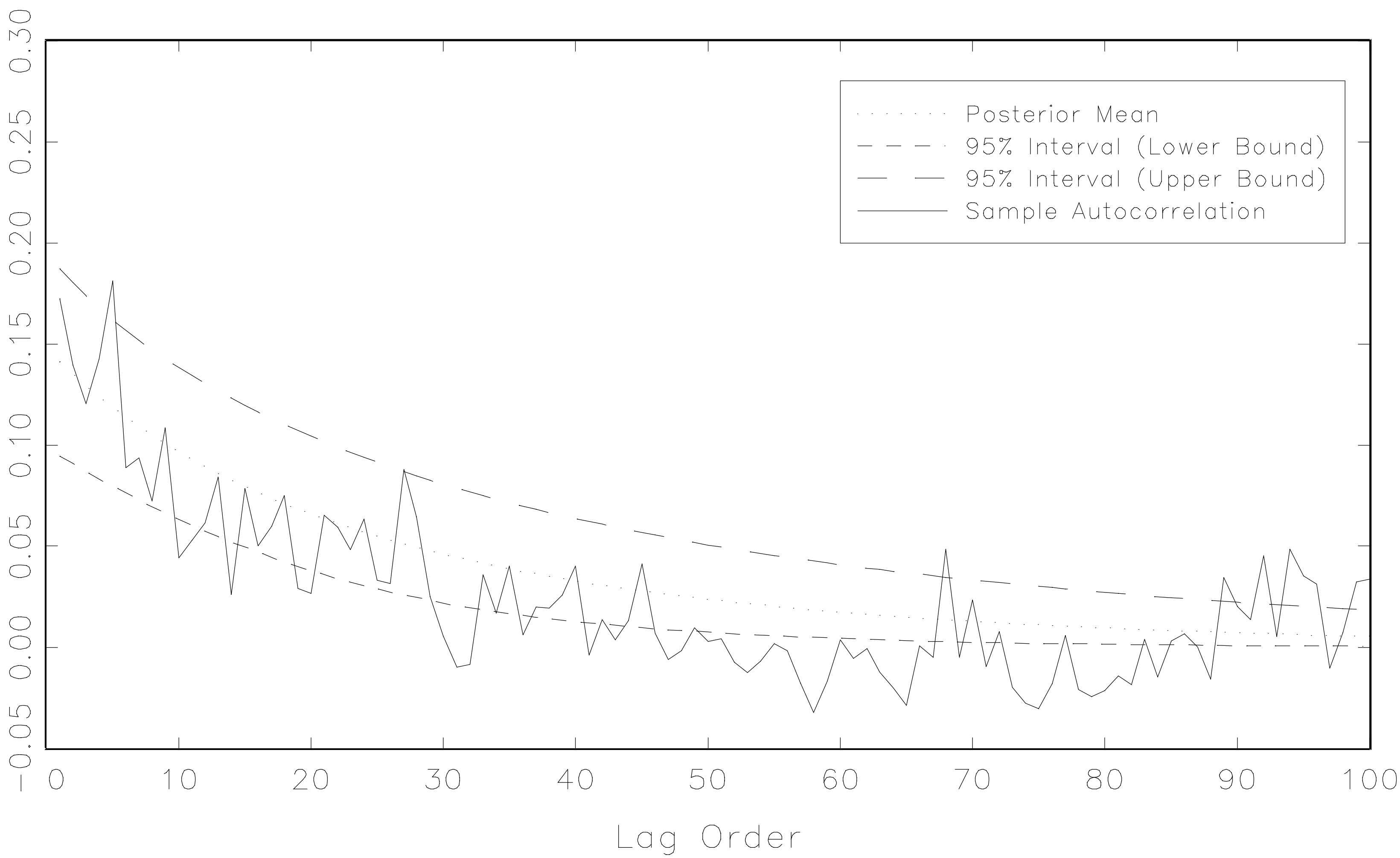

Figure 6: Confidence Interval for the Autocorrelation Coefficients of Squared Returns (TOPIX)
GED