MCMC Bayesian Estimation of a Skew-GED Stochastic Volatility Model

Nunzio Cappuccio
Department of Economics
University of Padova
via del Santo 33, 35123 Padova
email: nunzio.cappuccio@unipd.it

Diego Lubian
Department of Economics
University of Verona
via dell'Università, 37129 Verona
email: diego.lubian@univr.it

Davide Raggi
Department of Statistics
University of Padova
via Cesare Battisti 241, 35121Padova
email:daviragg@stat.unipd.it

Abstract

In this paper we present a stochastic volatility model assuming that the return shock has a Skew-GED distribution. This allows a parsimonious yet flexible treatment of asymmetry and heavy tails in the conditional distribution of returns. The Skew-GED distribution nests both the GED, the Skew-normal and the normal densities as special cases so that specification tests are easily performed. Inference is conducted under a Bayesian framework using Markov Chain MonteCarlo methods for computing the posterior distributions of the parameters. More precisely, our Gibbs-MH updating scheme makes use of the Delayed Rejection Metropolis-Hastings methodology as proposed by Tierney and Mira (1999), and of Adaptive-Rejection Metropolis sampling. We apply this methodology to a data set of daily and weekly exchange rates. Our results suggest that daily returns are mostly symmetric with fat-tailed distributions while weekly returns exhibit both significant asymmetry and fat tails.

Keywords: Stochastic volatility, Markov Chain MonteCarlo, Skewness, Heavy tails, Bayesian inference, Metropolis-Hastings sampling.
JEL Classifications: C11, C15, G1.

1 Introduction

The time series econometrics literature on the modeling of time varying conditional variances of asset returns has grown enormously since the seminal paper on ARCH models by Engle (1982). A large literature has also grown up on modeling financial time series using stochastic volatility models (see Taylor, 1994; Ghysels et al., 1996 for a review). Several variants of ARCH and SV models have been proposed so far to account for the empirical regularities of financial time series. Amongst these regularities two are tackled in this paper within a stochastic volatility model, namely, the heavy tails and the asymmetry in the distribution of returns.

Fat tails have been documented since the work by Mandelbrot (1963) and Fama (1965) and several studies have been concerned with modeling of asset returns with stable distributions. In the stochastic volatility literature, Jacquier et al. (1999) and Liesenfeld and Jung (2000), amongst others, have provided consistent evidence that leptokurtic distributions, such as the Student's t or the GED ones, are more adequate to capture this empirical regularity. As for asymmetry, Corrado and Su (1997) suggests that fat tails and asymmetry jointly determine the so-called "volatility smile" in option pricing using the Black-Scholes approach and that explicit account of them improve accuracy in option pricing, Chunhachinda et al. (1997) show that the introduction of skewness affects significantly the construction of the optimal portfolio, Mittnik and Paolella (2000) argue that skewness and heavy tails should be taken into account explicitly in Value-at-Risk forecasts, Peiró (1999) provides further evidence of asymmetry in returns, both from stock market indices and from individual assets.

Heavy tails are usually accounted for by relaxing the normality assumption and assuming a distribution with fatter tails, such as the Student's t or the GED. In this paper we suggest that a flexible and parsimoniuos treatment of both asymmetry and heavy tails in the distribution of returns can be achieved by a skewed distribution built upon a fat tailed distribution. We provide such a direct treatment of the asymmetry in returns by exploiting a result by Azzalini (1985) which allows the construction of an asymmetric distribution which nests symmetric ones and whose asymmetry is characterized by as single parameter. In practice, this is accomplished by specifying a stochastic volatility model where returns shocks are modeled according to the Skew-GED distribution and volatility is modeled as an AR(1) process with Gaussian errors, independent on the returns shock. We are aware that as a consequences of the absence of correlation between the returns and the volatility shock we are not able to take into account the so-called "leverage" effect: episodes of high volatility induce expectations of lower future returns, hence the negative correlation between these shocks. However, in our empirical application of the model we consider exchange rates data where the "leverage" effect is usually not effective. The extension to correlated shocks is currently under development.

Inference on the Skew-GED stochastic volatility model is performed in a Bayesian framework via Markov Chain MonteCarlo methods (MCMC), as in Jacquier et al. (1994, 1999). MCMC permits to obtain the posterior distributions of the parameters by simulation rather than analytical methods. Our updating scheme for the transition kernel involves both standard Gibbs sampling steps and the use of the Delayed-Rejection Metropolis-Hastings algorithm, introduced by Tierney and Mira (1999), for the sampling of the volatility process. Direct sampling of volatilities is not feasible in our setup because their full conditional distributions have not a standard form. Further, the asymmetry and kurtosis parameter of the Skew-GED distribution are sampled by Adaptive-Rejection Metropolis Sampling. We apply our methodology to a data set of three exchange rates over the 1990s. Our findings indicate strong evidence in favor of the SGED specification for the ¥/US$, where a GED or a Skew-Normal model might be more appropriate for the DM/US$ and US$/£rates.

The paper is organized as follow. In section 2 we outline a standard stochastic volatility model with autoregressive volatility and introduce the Skew-GED distribution, characterizing both the marginal moments of the return process and the correlation structure of squared returns. In section 3 we briefly review the MCMC method and the most popular updating schemes, while in section 4 we discuss our Gibbs-MH updating scheme based upon Delayed-Rejection MH and Adaptive-Rejection Metropolis Sampling. Section 5 is devoted to the estimation of the model with a data set of exchange rates and to the posterior distributions analysis.

2 A Skew-GED stochastic volatility model

A stochastic volatility model for the observable return process y_t is usually specified as $y_t = \beta \exp\{h_t/2\}\epsilon_t$, where β is a scale factor, ϵ_t is a random return shock with some known distribution and h_t is the unobserved stochastic volatility process with some conditional (on past volatilities) distribution (see Ghysels et al. (1996) for a comprehensive survey). The most popular specification of this conditional distribution is the first-order autoregressive process for $\log(h_t)$ such as

$$y_t = \beta \exp\{h_t/2\}\epsilon_t \quad (1)$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \sigma_\eta \eta_t, \qquad t = 1, \ldots, T \quad (2)$$

where the scale factor β must be set equal to unity for identifiability reasons (see Kim et al. (1998)) and h_1 is drawn from some known distribution. Several contributions have considered different specification for the distribution of the return shock and its correlation with the volatility shock. When ϵ_t is Gaussian with zero mean and unit variance and is independent on η_t, we have the lagged autoregressive random variance model of Taylor (1994). Jacquier et al. (1994) and Kim et al. (1998) also assume that ϵ_t is Gaussian, while Jacquier et al. (1999) and Chib et al. (1998) consider a return shock with a Student's t distribution and some correlation between return and volatility shocks, Steel (1998) considers a Skew Exponential Power distribution for $\log(\epsilon_t^2)$ which implies a fat-tailed distribution for the return shock, and finally, Andersen (1996) suggests the use of the Generalized Exponential distribution.

To account for fat tails and asymmetry, we introduce the Skew-GED distribution, SGED in short, for the return shock ϵ_t. This distribution is completely characterized by two parameters: λ, which is related to the asymmetry and ν which measures how heavy are the tails. As we shall show below, this distributional assumption is very convenient since the SGED nests most previously used distributions.

The SGED density can be obtained via an ingenious Lemma by Azzalini (1985):

LEMMA 2.1 *(Azzalini (1985)) Let $f(\cdot)$ be a density function symmetric about* 0, *and $G(\cdot)$ an absolutely continuous distribution function such that $G'(\cdot)$ is symmetric about* 0. *Then $2G(\lambda x)f(x), -\infty < x < \infty$, is a density function for any real λ.*

In our application of Lemma 2.1, the density $f(\cdot)$ and the distribution function $G(\cdot)$ are those of a random variable with a Generalized Error Distribution (GED) because we believe that the GED distribution is a simple, yet effective, way of modeling the tail behavior of the distribution of return shocks. Letting Z be a GED random variable with zero mean, its probability density function is given by

$$f(z) = \frac{\nu \exp\{-0.5|z|^\nu\}}{2^{1+1/\nu}\Gamma(1/\nu)} \qquad -\infty < z < \infty \quad \nu > 0$$

The crucial parameter of the GED is ν which controls the thickness of the tails. In fact, for $\nu = 2$ the GED reduces to the normal density while for $\nu < 2$ we have a leptokurtic density (with heavier tails than the normal one) and for $\nu > 2$ we have a platykurtic density (with thinner tails than the normal one). Even moments of a GED random variable are given by

$$E[z^r] = 2^{r/\nu}\Gamma\left(\frac{r+1}{\nu}\right)/\Gamma(1/\nu)$$

so that, for instance, the variance is given by $\sigma^2 = 2^{2/\nu}\Gamma(3/\nu)/\Gamma(1/\nu)$ while odd moments are zero. Finally, this density is symmetric about 0 satisfying the conditions of Lemma 2.1.

The absolutely continuous distribution function of a GED random variable with zero mean can be written as

$$G(a) = \int_{-\infty}^{a} f(z)\mathrm{d}z = \frac{1}{2}\left\{1 + \frac{\mathrm{sign}(a)\gamma\left(\left(\frac{a}{2}\right)^{\nu}; \frac{1}{\nu}\right)}{\Gamma(1/\nu)}\right\}$$

where $\gamma((a/2)^{\nu}; 1/\nu)$ is the incomplete gamma function $\gamma(b; w) = \int_0^b t^{w-1}e^{-t}\mathrm{d}t$.

Following Lemma 2.1, for any real λ we are able to build a random variable X with probability density function given by[1]

$$f(x; \lambda, \nu) = \left\{1 + \frac{\mathrm{sign}(\lambda x)\gamma\left(0.5\,(\lambda x)^{\nu}; \frac{1}{\nu}\right)}{\Gamma(1/\nu)}\right\} \frac{\nu \exp(-0.5|x|^{\nu})}{2^{1+1/\nu}\Gamma(1/\nu)} \tag{3}$$

and we say that X has a SGED distribution. Several well-know densities can be obtained as special cases of the SGED random variable. The parameters λ and ν control the asymmetry and the fat tails of the distribution of returns, respectively. It is not difficult to see that when $\nu = 2$ and $\lambda = 0$ we have the standard Normal, for $\nu = 2$ and $\lambda \neq 0$ we obtain the Skew Normal of Azzalini (1985), for $\lambda = 0$ we have the GED distribution. Thus, the SGED nests several distributions which have been used to model the return shock in stochastic volatility models. In Figure 1 we graph the density of the SGED random variable for different values of λ and ν. For $\nu < 2\,(\nu > 2)$, this distribution exhibits fatter (thinner) tails then the benchmark standard normal case of $\lambda = 0$ and $\nu = 2$ (dotted line). Skewness is introduced when $\lambda \neq 0$: positive (negative) values of λ induce a longer and fatter right (left) tail.

The even moments of X are given by[2]

$$E[X^r] = 2^{r/2}\frac{\Gamma((r+1)/\nu)}{\Gamma(1/\nu)} \tag{4}$$

[1]The Skew-GED distribution is also defined in Azzalini (1986) as Distribution Type I (his formula (12)) with another parametrization. See also Goria (1998).

[2]For r even the moments of the skew distribution are equal to those of the symmetric one, see Proposition 1 in Azzalini (1986).

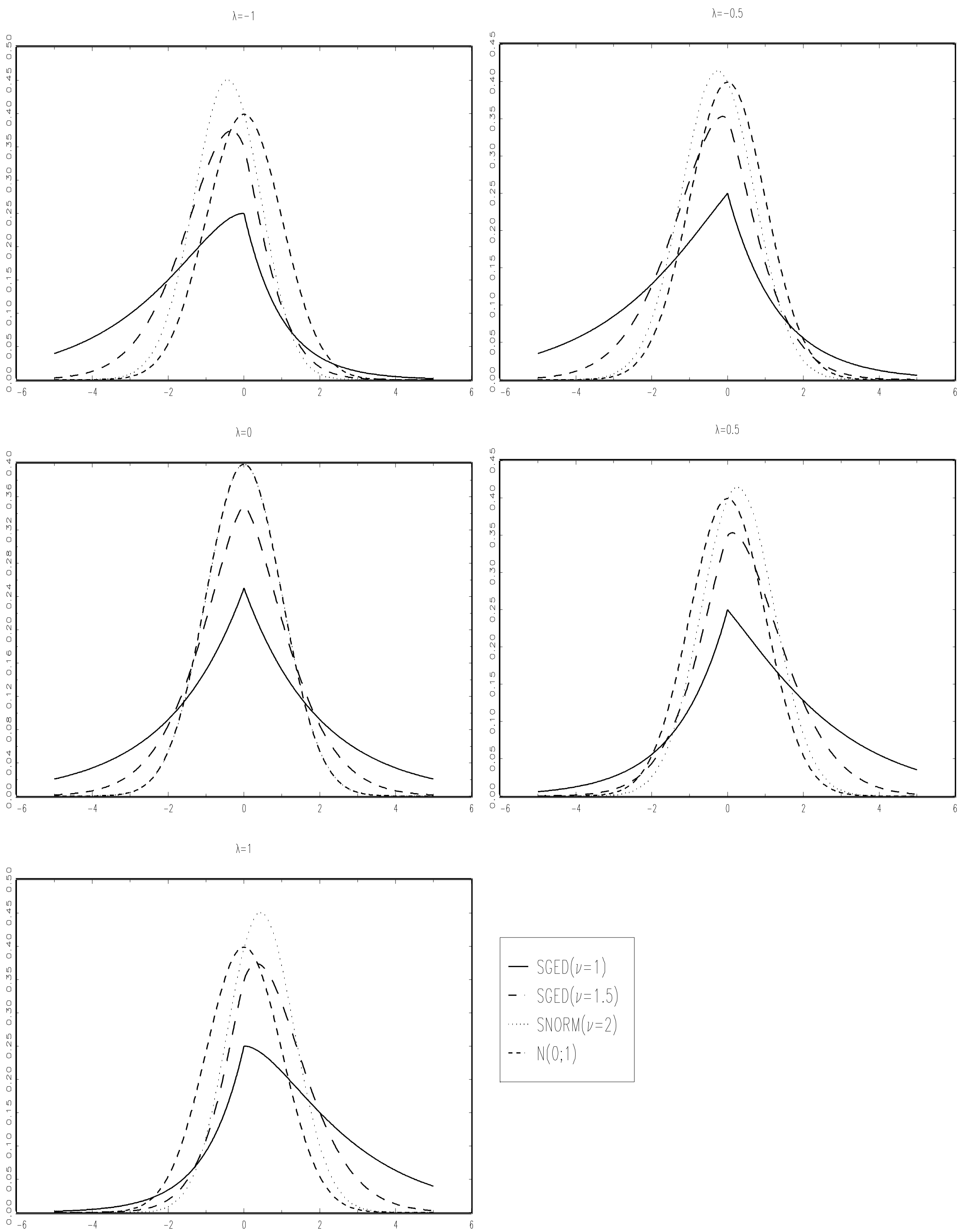

Figure 1: The *S-GED* distribution for different values of λ and ν.

while the odd moments are

$$E[X^r] = 2^{r/2}\frac{\text{sign}(\lambda)\Gamma((r+1)/\nu)}{\Gamma(1/\nu)}B\left(\frac{|\lambda|^\nu}{1+|\lambda|^\nu};\frac{1}{\nu},\frac{r+1}{\nu}\right) \tag{5}$$

where $B(z;p,q)$ is the distribution function of a Beta random variable.

In this paper, we assume that the return shock ϵ_t in the stochastic volatility model (1)-(2) has a SGED distribution with zero mean and is normalized to have unit variance. As for the initial condition for the volatility process, we assume that $h_1 \sim N(\mu, \sigma_\eta^2/(1-\phi^2))$. With this specification of the stochastic volatility model, the marginal moments of the return process y_t are given by

$$E[y^r] = \exp\left\{\frac{r}{2}\mu + \frac{r^2}{8}\frac{\sigma_\eta^2}{1-\phi^2}\right\}E[\epsilon_t^r]$$

where $E[\epsilon_t^r]$ is obtained from (4) and (5) according to the value taken by r. The dynamic properties of the model are summarized by the covariance structure of squared returns

$$\text{Cov}(y_t^2, y_{t-s}^2) = \exp\left\{2\mu + \frac{\sigma_\eta^2}{1-\phi^2}\right\}\left(\exp\left\{\frac{\sigma_\eta^2}{1-\phi^2}\phi^s\right\} - 1\right)$$

which is independent on the assumptions on the distribution of the return shock ϵ_t. These moment conditions suggest that a GMM estimator could be implemented; however we do not follow this avenue in the paper but a Bayesian approach as in Jacquier et al. (1994).

3 Markov Chain MonteCarlo

Letting $\boldsymbol{\theta} = (\mu, \phi, \sigma_\eta, \nu, \lambda)$ be the parameter vector, a consequence of our specification of the stochastic volatility model is that, conditional on $(h_t, \boldsymbol{\theta})$, the return y_t and the volatility h_{t+1} are stochastically independent. This allows the factorization of the joint density for a single observation as

$$p(y_t, h_{t+1}|h_t, \boldsymbol{\theta}) = p(y_t|h_t, \boldsymbol{\theta})p(h_{t+1}|h_t, \boldsymbol{\theta})$$

so that the likelihood can be written as

$$\begin{aligned} p(\boldsymbol{y}, \boldsymbol{h}|\boldsymbol{\theta}) &= p(h_1|\boldsymbol{\theta})\left[\prod_{t=1}^{T-1} p(y_t, h_{t+1}|h_t, \boldsymbol{\theta})\right] p(y_T|h_T, \boldsymbol{\theta}) \\ &= p(h_1|\boldsymbol{\theta})\left[\prod_{t=1}^{T-1} p(h_{t+1}|h_t, \boldsymbol{\theta})\right]\left[\prod_{t=1}^{T-1} p(y_t|h_t, \boldsymbol{\theta})\right] p(y_T|h_T, \boldsymbol{\theta}) \end{aligned}$$

where

$$p(y_T|h_T, \boldsymbol{\theta}) = \int p(y_T, h_{T+1}|h_T, \boldsymbol{\theta})\mathrm{d}h_{T+1}.$$

Analytic expressions are available both for the above likelihood and for $p(y_t|h_t, \boldsymbol{\theta})$ and $p(h_{t+1}|h_t, \boldsymbol{\theta})$. Once we specify some prior distribution of the parameters $p(\boldsymbol{\theta})$, assumed mutually independent throughout the paper, the joint density of returns, volatilities, and the parameters is available in closed form as

$$p(\boldsymbol{y}, \boldsymbol{h}, \boldsymbol{\theta}) = p(\boldsymbol{y}, \boldsymbol{h}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \tag{6}$$

Our task is to obtain the posterior distribution of the parameter vector given the data, namely, $p(\boldsymbol{\theta}|\boldsymbol{y})$.

Whenever both $\boldsymbol{y}$ and $\boldsymbol{h}$ are observable, the posterior distribution for the parameters $\boldsymbol{\theta}$, say $p(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{h})$, can be calculated. In the presence of unobserved data, i.e. the volatilities, Tanner and Wong (1987) argue that, the posterior distribution of interest is $p(\boldsymbol{\theta}|\boldsymbol{y})$ which may be difficult to calculate and they suggest that integration of (6) with respect to the latent data could be performed by simulation methods. If one could generate realizations of the latent process from its predictive density given the observed data, namely $p(\boldsymbol{h}|\boldsymbol{y})$, then one could evaluate the posterior density of interest as the expected value $p(\boldsymbol{\theta}|\boldsymbol{y}) = \int p(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{h})p(\boldsymbol{h}|\boldsymbol{y})\mathrm{d}\boldsymbol{h}$. In practice, suppose we can generate a sequence of simulated parameter vectors $\{\boldsymbol{\theta}^{(i)}\}_{i=1}^{M}$ and unobserved data $\{\boldsymbol{h}^{(i)}\}_{i=1}^{M}$ from $p(\boldsymbol{\theta}, \boldsymbol{h}|\boldsymbol{y})$, then by MonteCarlo we can integrate out $\boldsymbol{h}$. It follows that the sequence of parameter vectors is implicitly a sample from the posterior distribution of $\boldsymbol{\theta}$ given the data $\boldsymbol{y}$. This approach is called "data augmentation". In our context, the number of unobservables is quite large and simulation from $p(\boldsymbol{\theta}, \boldsymbol{h}|\boldsymbol{y})$ directly is not possible. In order to sample from such high-dimensional densities, we resort to Markov Chain MonteCarlo (MCMC) methods.

The basic idea behind MCMC is to build a Markov chain transition kernel

$$\mathbb{P}(x, A) = \mathrm{Prob}\{(\boldsymbol{\theta}^{(m+1)}, \boldsymbol{h}^{(m+1)}) \in A|(\boldsymbol{\theta}^{(m)}, \boldsymbol{h}^{(m)}) \in x\},$$

starting from some initial state $(\boldsymbol{\theta}^{(0)}, \boldsymbol{h}^{(0)})$, with limiting invariant distribution equal to the posterior distribution of the parameters given the data $p(\boldsymbol{\theta}|\boldsymbol{y})$. Under suitable conditions (Tierney, 1994; Chib and Greenberg, 1996), we can build such a transition kernel generating a Markov chain $\{\boldsymbol{\theta}^{(m+1)}, \boldsymbol{h}^{(m+1)}|\boldsymbol{\theta}^{(m)}, \boldsymbol{h}^{(m)}, \boldsymbol{y}\}_{m=1}^{M}$ whose elements (draws) converge in distribution to the (target) posterior density $p(\boldsymbol{\theta}|\boldsymbol{y})$. Once convergence is achieved, we obtain a sample of serially dependent simulated "observations" on the parameter vector $\boldsymbol{\theta}$ (and on the volatilities), which can be used to perform MonteCarlo inference[3].

Much effort has been devoted to the design of updating schemes able to generate a convergent transition kernel. The Metropolis-Hastings (MH) algorithm is one of them and it can be very effective in building the above mentioned Markov chain transition kernel (Metropolis et al., 1953; Hastings, 1970). This algorithm is very popular because it is possible to show that, under suitable (mild) conditions (see Robert and Casella 1999), the Markov chain converges to correct invariant distribution. The MH updating scheme works as follows: letting $\pi(x)$ be the target density and $q(x, y)$ be a proposal density function from which we generate the transition from state x to state y, the Markov chain is updated according to the following steps:

1. let x be the starting state (or the initial condition),
2. sample y from a proposal density $q(x, y)$,
3. sample u from the uniform density $U(0, 1)$.
4. compute $\alpha(x, y) = \begin{cases} \min\left(\frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}, 1\right) & \text{if } \pi(x)q(x, y) > 0 \\ 1 & \text{otherwise} \end{cases}$
5. new state of the chain $= \begin{cases} y & \text{if } u \leq \alpha(x, y) \\ x & \text{otherwise} \end{cases}$
6. go to step 2.

Several special cases of the MH scheme are of particular interest: if the proposal does not depend on the present state of the chain, that is $q(x, y) = f(y)$, MH generates a so called *Independence Chain*; if, on the other hand, $y = x + z$ with $z \sim f(z)$ implying $q(x, y) = f(y - x)$ we have a *Random Walk Chain*, see Chib and Greenberg (1995) for details.

An important special case of the MH algorithm arises when x is a vector. In this case it is possible to apply a "divide and conquer" strategy in which the vector is updated one component at a time. If the proposal for each component is the full conditional distribution, i.e. the distribution of each component conditional on all other components, the algorithm is known as Gibbs sampler. In practice, the acceptance probability $\alpha(x, y)$ is by definition equal to 1, so all suggestions are

[3]The period from the start of the chain until convergence to the stationary distribution is achieved is called the "burn-in" period. Since "draws" in this period do not come from the posterior distribution of interest, they are discarded when making inference on the parameter vector (see Gilks, Richardson and Spiegelhalter (1995) for a discussion on the appropriate choice of the "burn-in" period).

accepted. This updating scheme is particularly straightforward to implement since there is no need to evaluate $\alpha(x, y)$.

In practical works, the choice of the proposal is somehow arbitrary but subject to the condition that the chain has the stationary distribution $\pi(x)$, is π-irreducible and aperiodic[4]. From Chib and Greenberg (1996), if the Markov chain has an invariant distribution $\pi(x)$, it is π-irreducible and aperiodic then an ergodic distribution exists and that, irrespective of the starting point, the Markov chain with transition kernel $\mathbb{P}(x, A)$ will converge to the invariant distribution. A further important computational issue concerns the selection of a proposal distribution $q(x, y)$ so that "observations" may be generated easily.

4 A Gibbs-MH updating scheme

Many variants of the basic MH algorithm and Gibbs sampler have been proposed so far. The most relevant in our setup, where the state is a vector of parameter and volatilities, are the so-called "blocking schemes". Under these schemes one divides the parameter set, say $\mathcal{S}$, into subset S_i such that $\mathcal{S} = \{S_i | i = 1, \ldots, n\}$, and then samples each block S_i individually, conditional on the most recent value of the remaining blocks $\boldsymbol{S}_{-i}^{(j+1)} = \{S_1^{(j+1)}, \ldots, S_{i-1}^{(j+1)}, S_{i+1}^{(j)}, \ldots, S_n^{(j)}\}$. The distribution of S_i conditional on all other blocks and the data is called the full conditional distribution, and is denoted by $p(S_i | \boldsymbol{S}_{-i}, \boldsymbol{y})$. Similar techniques were applied to state space models in Kim et al. (1998), Steel (1998), Jacquier et al. (1994, 1999), and Carlin et al. (1992).

Alternative updating schemes may be implemented for the different blocks of the parameters, according to the difficulties originated by the problem at hand, giving rise to hybrid MH updating schemes (see Robert and Casella 1999 and Tierney 1994) For example, if we are not able to extract some drawings from a full conditional distribution inside a Gibbs sampler, it is possible to solve the problem introducing a Metropolis-Hastings step. This procedure is known (even if somehow improperly) as MH within Gibbs sampler. Other examples about this kind of strategies are in Tierney (1994).

Since the main problem we deal with is the high dimensionality of the latent

[4]Loosely speaking, a Markov chain with invariant distribution, π, is said to be π-irreducible if from any starting point x_0, the state A will be reached with positive probability. When the chain is π-irreducible then it said to be aperiodic if any point x_0 in the support of π can not be visited with probability 1 when starting from some specific set of states $\mathcal{A}$.

process, it is very important to look for a proposal distribution for each volatility that reduce the risk of an high probability of rejection and at the same time be computationally easy to sample. We update the volatility vector and each parameter one at a time beginning with the volatility vector, so our blocking scheme is the simplest one.

4.1 Sampling the volatility process

Given our specification of the prior densities, it is a matter of tedious calculation to obtain the full conditional density of h_t for $t = 2, \ldots, T-1$ as proportional to

$$\begin{aligned} p(h_t|\boldsymbol{y}, \boldsymbol{h}_{-t}, \boldsymbol{\theta}) \quad \propto \quad & \exp\left\{-\frac{h_t}{2}\right\} \exp\left\{-\frac{1}{2}\left|\frac{y_t \exp\{-h_t/2\} - \lambda_1}{\lambda_2}\right|^{\nu}\right\} \times \\ & \times \left[1 + sgn\left(\lambda\frac{\exp\{-h_t/2\}y_t - \lambda_1}{\lambda_2}\right) \times \right. \\ & \left. \times\, \gamma\left(\frac{1}{2}\left|\lambda\frac{\exp\{-h_t/2\}y_t - \lambda_1}{\lambda_2}\right|^{\nu}; \frac{1}{\nu}\right) / \Gamma(1/\nu)\right] \times \\ & \times \exp\left\{-\frac{1}{2}\left[\frac{(h_{t+1} - \mu) - \phi(h_t - \mu)}{\sigma_\eta}\right]^2\right\} \times \\ & \times \exp\left\{-\frac{1}{2}\left[\frac{(h_t - \mu) - \phi(h_{t-1} - \mu)}{\sigma_\eta}\right]^2\right\} \end{aligned}$$

where $\boldsymbol{h}_{-t} = (h_1, \ldots, h_{t-1}, h_{t+1}, \ldots, h_T)$. Further, for $t = 1$ we have

$$p(h_1|\boldsymbol{y}, \boldsymbol{h}_{-1}, \boldsymbol{\theta}) \propto p(y_1|h_1, \boldsymbol{\theta}) p(h_2|h_1, \boldsymbol{\theta}) p(h_1|\boldsymbol{\theta}) \tag{7}$$

while for $t = T$,

$$p(h_T|\boldsymbol{y}, \boldsymbol{h}_{-T}, \boldsymbol{\theta}) \propto p(h_T|h_{T-1}, \boldsymbol{\theta}) p(y_T|h_T, \boldsymbol{\theta}). \tag{8}$$

Some difficulties arise in this step because of the complexity of the full conditional distributions. Therefore, we make use of the Delayed Rejection MH algorithm proposed in Tierney and Mira (1999) who suggest that, in case of rejection of a draw from a proposal density, one should re-sample the new state of the chain from a different proposal, exploiting the information (of rejection) contained in the previous step[5]. This Delayed Rejection MH is used whenever we refuse a candidate for the volatility, appending a further MH step according to a new proposal. Mira (2001) and Tierney and Mira (1999) prove that, in order to maintain the reversibility of the chain, for each sub-step of the MH algorithm the acceptance probability

[5]See Cappuccio et al. (2001) for a similar stochastic volatility model with a different approach to sampling of the volatility process.

$\alpha(x, y_1, \cdots, y_n)$ has the form

$$\alpha_i(x, y_1, \ldots, y_i) = 1 \wedge \left\{ \frac{\pi(y_i)q_1(y_i, y_{i-1})q_2(y_i, y_{i-1}, y_{i-2}) \cdots q_i(y_i, y_{i-1}, \ldots, x)}{\pi(x)q_1(x, y_1)q_2(x, y_1, y_2) \cdots q_i(x, y_1, y_i)} \times \right.$$

$$\left. \frac{[1 - \alpha_1(y_i, y_{i-1})][1 - \alpha_2(y_i, y_{i-1}, y_{i-2})] \cdots [1 - \alpha_{i-1}(y_i, \ldots, y_1)]}{[1 - \alpha_1(x, y_1)][1 - \alpha_2(x, y_1, y_2)] \cdots [1 - \alpha_{i-1}(x, y_1, y_{i-1})]} \right\}$$

In the first step the proposal is based on an Independence Chain as proposed in Kim et al. (1998)

$$q(h_t \mid h_{t-1}, h_{t+1}, \boldsymbol{\theta}) \sim N(\mu_t, \sigma_t^2) \tag{9}$$

where $\mu_t = h_t^* + 0.5\sigma_t^2/[y_t^2 \exp\{-h_t^*\} - 1]$, $\sigma_t^2 = \sigma_\eta^2/(1 + \phi^2)$ and

$$h_t^* = \mu + \frac{\phi[(h_{t-1} - \mu) + (h_{t+1} - \mu)]}{1 + \phi^2}.$$

Even though this proposal was suggested for a stochastic volatility model with Gaussian errors, it turns out to work well with a SGED error term also, approximating in a precise way the full conditional distribution. In case of rejection we consider a random walk proposal with the same variance as the previous step. Combining these proposals allows to exploit the advantages of both: when the independence proposal is a good approximation of the invariant target distribution the number of rejections will be small whereas a rejection implies a poor approximation and a random walk proposal provides some control on this undesirable behavior[6].

4.2 Sampling the SGED parameters

Then, we move to sampling each component of the parameter vector $\boldsymbol{\theta}$ beginning from the parameters characterizing the SGED distribution. For these two parameter we specify uniform priors such as $\nu \sim U(0, \nu_H)$ and $\lambda \sim U(\lambda_L, \lambda_H)$ which cover a wide range of parameter values for both the tail thickness and the skewness measures.

[6]Sufficient conditions for the convergence of this chain are stated in Tierney (1994). Letting $f(\cdot)$ be the proposal density we have that if $f > 0$ almost everywhere on $\mathbb{R}$ then the random walk Metropolis kernel is irreducible and aperiodic. An Independence Metropolis kernel is irreducible and aperiodic if and only if $f > 0$ almost everywhere in $E^+ = \{x : \pi(x) > 0\}$. In practice, the sufficient conditions require that the proposal density be positive in the sets of positive probability of the target distribution π.

Then, the full conditional distribution for ν and λ are given by

$$\begin{aligned}p\left(\nu|\boldsymbol{y},\boldsymbol{h},\sigma_\eta^2,\phi,\mu,\lambda\right) \quad &\propto \quad p(\nu)\prod_{t=1}^{n}\frac{\nu}{2^{1+1/\nu}\Gamma(1/\nu)\lambda_2}\times \\ &\times\exp\left\{-\frac{1}{2}\left|\frac{\exp\{-h_t/2\}y_t-\lambda_1}{\lambda_2}\right|^\nu\right\}\times \\ &\times\left[1+sgn\left(\lambda\frac{\exp\{-h_t/2\}y_t-\lambda_1}{\lambda_2}\right)\times\right. \\ &\left.\times\,\gamma\left(\frac{1}{2}\left|\lambda\frac{\exp\{-h_t/2\}y_t-\lambda_1}{\lambda_2}\right|^\nu;\frac{1}{\nu}\right)/\Gamma(1/\nu)\right]\end{aligned}$$

and

$$\begin{aligned}p\left(\lambda|\boldsymbol{y},\boldsymbol{h},\sigma_\eta^2,\phi,\mu,\nu\right) \quad &\propto \quad p(\lambda)\prod_{t=1}^{n}\frac{1}{\lambda_2}\exp\left\{-\frac{1}{2}\left|\frac{\exp\{-h_t/2\}y_t-\lambda_1}{\lambda_2}\right|^\nu\right\}\times \\ &\times\left[1+sgn\left(\lambda\frac{\exp\{-h_t/2\}y_t-\lambda_1}{\lambda_2}\right)\times\right. \\ &\left.\times\,\gamma\left(\frac{1}{2}\left|\lambda\frac{\exp\{-h_t/2\}y_t-\lambda_1}{\lambda_2}\right|^\nu;\frac{1}{\nu}\right)/\Gamma(1/\nu)\right]\end{aligned}$$

respectively, where $\lambda_1 = -E[X]/\sqrt{\mathrm{Var}(X)}$ and $\lambda_2 = 1/\sqrt{\mathrm{Var}(X)}$ are introduced in order to standardize the SGED distribution.

Sampling ν and λ is accomplished via the Adaptive Rejection Metropolis Sampling (ARMS) proposed by Gilks, Best and Tan (1995). The rationale behind this sampling method is that the Adaptive-Rejection sampling method of Gilks and Wild (1992) for log-concave full conditional distributions cannot be used in the present context as the full conditional distribution is not log-concave. They argue that even though a MH algorithm could be used it is likely that a high probability of rejection will result, hence a slower convergence. To avoid this drawback they suggest to adapt the proposal density to the shape of the full conditional distribution. Since Adaptive-Rejection sampling is a way to accomplish this adapting, it can be effectively used to build a good proposal density. Thus, by supplementing the Adaptive-Rejection sampling with a MH step, an ARMS scheme is devised which preserves the stationary distribution of the Gibbs sampler.

4.3 Sampling the AR parameters

Following Kim et al. (1998), we assume a conjugate prior for the variance of the log-volatility $\sigma_\eta^2 \sim IG(n/2;\delta/2)$, so that the full conditional distribution follows directly

$$p(\sigma_\eta^2|\boldsymbol{y},\boldsymbol{h},\nu,\phi,\mu,\lambda) \propto IG\left(\frac{T+n}{2}; \frac{\delta+(h_1-\mu)^2(1-\phi^2)+\sum_{t=1}^{T-1}[(h_{t+1}-\mu)-\phi(h_t-\mu)]^2}{2}\right)$$

where IG stands for the inverse-gamma distribution. As for the autoregressive parameter, letting $\phi = 2\phi^* - 1$ where $\phi^* \sim \text{Beta}(b_1, b_2)$, our prior distribution is given by[7]

$$p(\phi) \propto \left(\frac{1+\phi}{2}\right)^{b_1-1}\left(\frac{1-\phi}{2}\right)^{b_2-1}, \text{with} \quad b_1, b_2 > 1/2$$

with support in the interval $(-1, 1)$ and prior mean of $2b_1/((b_1+b_2)-1$. The full conditional distribution for ϕ becomes

$$p(\phi|\boldsymbol{y},\boldsymbol{h},\sigma_\eta^2,\nu,\mu,\lambda) \propto p(\phi)p(h_1|\sigma_\eta^2,\nu,\mu,\lambda)\prod_{t=1}^{n-1} p(h_{t+1}|h_t,\sigma_\eta^2,\nu,\mu,\lambda)$$

where the full conditional for h_t with $t = 1,\ldots,T$ are given in subsection 4.1. Last, for the drift in the volatility process we assume a diffuse prior leading to the following full conditional distribution

$$p(\mu|\boldsymbol{y},\boldsymbol{h},\sigma_\eta^2,\phi,\nu,\lambda) \propto N(\tau,\sigma_\mu^2)$$

where

$$\begin{aligned}\tau &= \sigma_\mu^2\left[\frac{1-\phi^2}{\sigma_\eta^2}h_1 + \frac{1-\phi}{\sigma_\eta^2}\sum_{t=1}^{T-1}(h_{t+1}-\phi h_t)\right]\\ \sigma_\mu^2 &= \sigma_\eta^2/[T(1-\phi)^2]\end{aligned}$$

from which we sample directly. Therefore, in this step we have a standard Gibbs sampling update. Notice that because of the lack of identification of β and μ we find it easier to sample μ but reporting the value $\beta = \exp(\mu/2)$.

4.4 Summary

In short, our MCMC updating scheme can be summarized as follows. We begin with initialization of the volatilities and the parameter vector at some value $\boldsymbol{h}^{(0)}$ and $\boldsymbol{\theta}^{(0)}$, respectively. Then, for $i = 1,\ldots,M$

1. simulate the volatility vector $\boldsymbol{h}^{(i)}$ from the full conditional distribution

$$p(h_t|h_1^{(i)},\ldots,h_{t-1}^{(i)},h_{t+1}^{(i-1)},\ldots,h_T^{(i-1)},\boldsymbol{\theta}^{(i-1)},\boldsymbol{y})$$

via Delayed-Rejection MH sampling,

[7] See Kim et al. (1998) for a discussion on different prior densities.

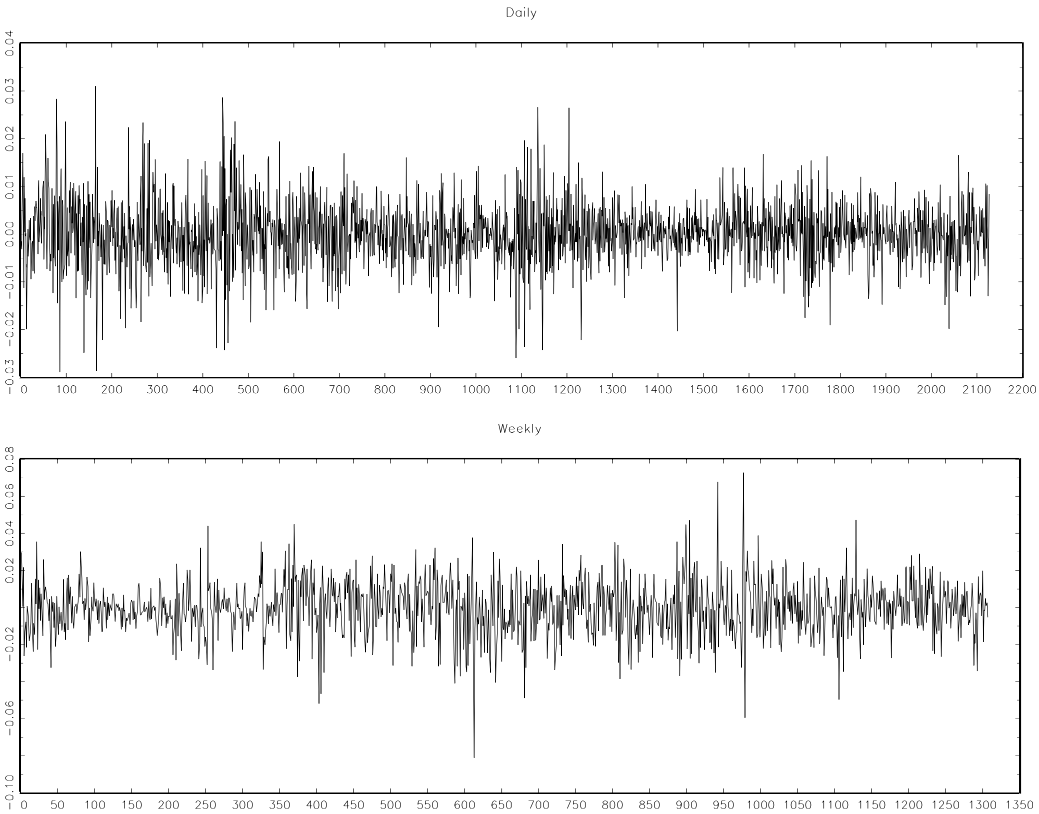

Figure 2: Returns DM/US$ - December 26,1990 - January 15,1999

2. sample one at a time the parameters from their full conditional distributions, namely
   (a) $\sigma_\eta^{2(i)}$ from $p(\sigma_\eta^2|\boldsymbol{h}^{(i)}, \phi^{(i-1)}, \mu^{(i-1)}, \nu^{(i-1)}, \lambda^{(i-1)})$ via full conditional distribution,
   (b) $\phi^{(i)}$ from $p(\phi|\boldsymbol{h}^{(i)}, \sigma_\eta^{2(i)}, \mu^{(i-1)}, \nu^{(i-1)}, \lambda^{(i-1)})$ via full conditional distribution,
   (c) $\mu^{(i)}$ from $p(\mu|\boldsymbol{h}^{(i)}, \sigma_\eta^{2(i)}, \phi^{(i)}, \nu^{(i-1)}, \lambda^{(i-1)})$ via full conditional distribution,
   (d) $\nu^{(i)}$ from $p(\nu|\boldsymbol{h}^{(i)}, \sigma_\eta^{2(i)}, \phi^{(i)}, \mu^{(i)}, \lambda^{(i-1)})$ via Adaptive-Rejection Metropolis sampling,
   (e) $\lambda^{(i)}$ from $p(\lambda|\boldsymbol{h}^{(i)}, \sigma_\eta^{2(i)}, \phi^{(i)}, \nu^{(i)})$ via Adaptive-Rejection Metropolis sampling.

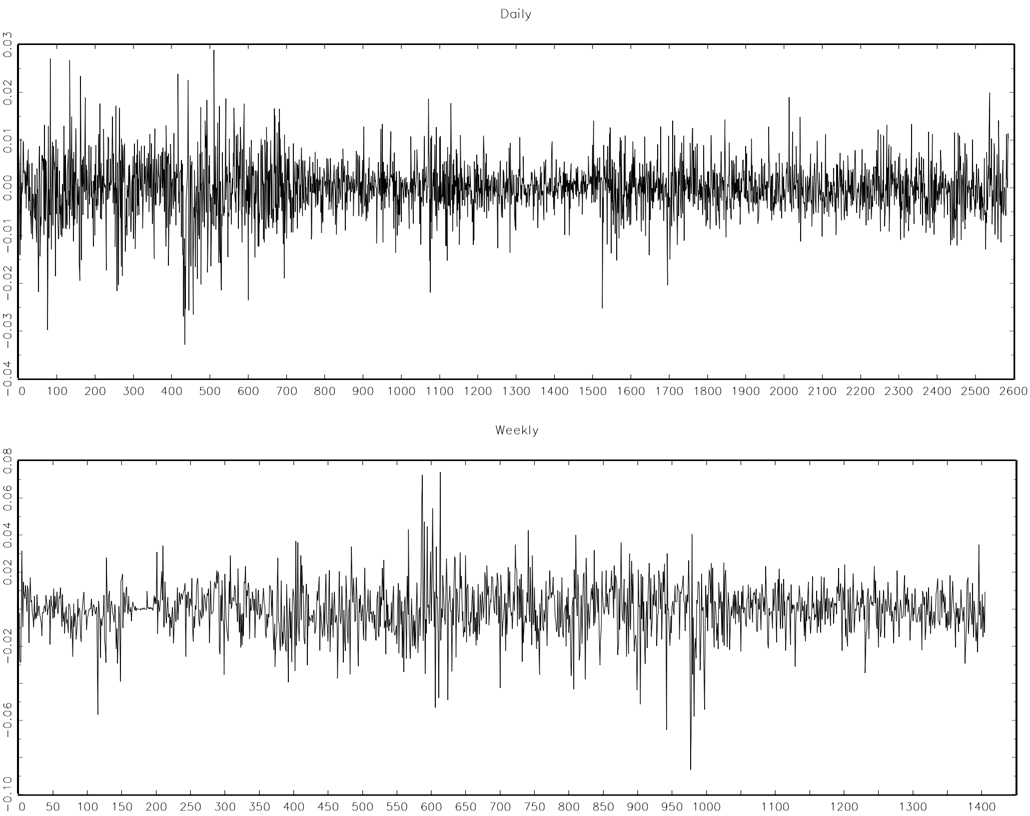

Figure 3: Returns US$/£- December 31, 1990 - December 4,2000

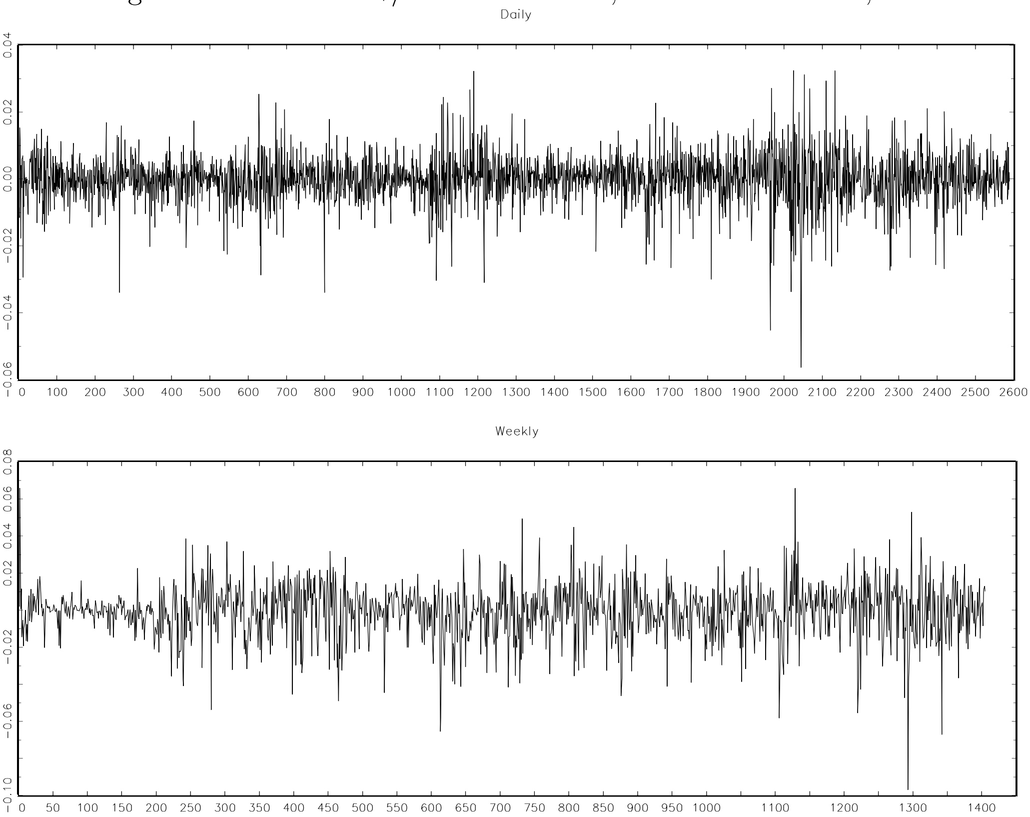

Figure 4: Returns ¥/US$ - December 31, 1990 - December 4,2000

Table 1: Exchange rates, daily and weekly series

Name	Symbol	Sample Period	
		Daily	Weekly
Deutsche Mark vs. US Dollar	DM/US$	12/31/1990- 1/15/1999	12/26/1973 - 1/13/1999
US Dollar vs. English Pound	US$/£	12/31/1990 -11/29/2000	12/26/1973 - 11/29/2000
Yen vs. US Dollar	¥/US$	12/31/1990 - 12/4/2000	12/26/1973 - 11/29/2000

5 An application to daily and weekly exchange rates

5.1 The data

Our empirical application concerns three daily and weekly (Wednesday quote) exchange rates over the 1990s as detailed in Table 1. Each returns series is regressed on daily and month dummies (the weekly series only on these dummies) to account for "day" and "month" effects. Moreover, an autoregressive filter is applied to remove the (weak) evidence of serial correlation in returns , with maximum lag variable with the series (detailed results are available from the authors upon request). The descriptive statistics in Table 2 are the per cent annualized sample mean, median and standard deviation computed multiplying the usual daily and weekly sample statistics by 256 and by 52, respectively. As for the standard test statistics for skewness and excess of kurtosis, we report the t statistics and their p-values. Under the null hypothesis of normality, these two test statistics are normally distributed with standard errors given by $SE(Skewness) = \sqrt{6/T}$ and $SE(ExcessKurtosis) = \sqrt{24/T}$, respectively. While the DM/US$ exchange rate does not exhibit significant skewness, both the US$/£and ¥/US$ display negative skewness. As for the excess kurtosis, the stylized fact of fat tails in the marginal distribution is confirmed for all exchange rates series. Last, the Jarque-Bera test is a test of the joint null hypothesis of no skewness and zero excess kurtosis, asymptotically distributed as a χ^2 with 2 degrees of freedom. This test reject the null hypothesis very soundly but, from previous results on testing for skewness and excess kurtosis separately, it is clear that for the DM/US$ exchange rate this is due to the strong evidence of thick tails.

Letting x_t be the original series, returns y_t are computed as $y_t = 100 \times [\ln(x_t) - \ln(x_{t-1})]$. Daily and weekly returns are displayed in Figures 2 to 4. To provide more evidence of asymmetry in the marginal distribution of returns we consider several nonparametric tests. Returns have been split in two sub-samples: $y_t^+ = (y_t - \bar{y})$ if $y_t > \bar{y}$, say positive "excess" returns, and $y_t^- = (\bar{y} - y_t)$ if $y_t < \bar{y}$, say negative "excess" returns, where $\bar{y}$ is the annualized sample mean of returns.

Table 2: Descriptive statistics for exchange rates returns

	DM/US$		US$/£		¥/US$	
	Daily	Weekly	Daily	Weekly	Daily	Weekly
Mean	1.441	-1.872	-2.816	-1.799	-1.984	-3.416
Median	0.000	-1.138	0.000	0.000	0.000	1.440
Maximum	3.103	7.274	2.889	7.397	3.239	6.586
Minimum	-2.896	-8.113	-3.28	-8.668	-5.630	-9.694
Std. Dev.	1.063	1.049	0.929	1.011	1.162	1.054
Skewness	0.040	-0.095	-0.23	-0.256	-0.600	-0.519
$t_{Skewness}$	0.771	-1.409	-4.893	-3.928	-12.478	-7.956
P-value	(0.440)	(0.158)	(0.000)	(0.000)	(0.000)	(0.000)
Kurtosis	4.919	4.860	5.981	6.572	7.751	6.185
$t_{ExcessKurtosis}$	18.068	13.727	30.953	27.337	49.345	24.375
P-value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Jarque-Bera	327.06	190.43	982.04	762.768	2590.67	657.46
P-value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	2127	1307	2586	1405	2589	1405

Table 3: Asimmetry Tests

	Wilcoxon		Siegel-Tukey		Kolmogorov-Smirnov	
	Daily	Weekly	Daily	Weekly	Daily	Weekly
DMUS$	0.183 (0.854)	-0.227 (0.819)	0.106 (0.914)	0.052 (0.958)	0.651 (0.788)	0.717 (0.681)
US$/£	0.169 (0.865659)	-0.038 (0.969)	0.159 (0.873)	2.270 (0.023)	0.568 (0.902)	0.864 (0.443)
¥US$	0.567 (0.570)	-2.030 (0.042)	1.474 (0.140)	2.956 (0.003)	0.795 (0.551)	1.503 (0.021)

The Wilcoxon, Siegel-Tukey and Kolmogorov-Smirnov statistics test the null hypothesis that the empirical distribution of positive and negative excess returns are identical.

Under symmetry both sub-samples should have the same empirical distribution. As in Peiró (1999), in Table 3 we report the Wilcoxon test and the Siegel-Tukey test based on rank and the Kolmogorov-Smirnov test based on the empirical distribution (see Hollander and Wolfe (1999) for a thorough treatment). Tests results indicate that there is strong evidence of asymmetry in the weekly exchange rate returns for all exchange rates series and evidence of asymmetry in the daily ¥/US$ exchange rate. Summarizing the evidence form descriptive statistics, we conclude that our findings of asymmetry and fat tails testifies a departure from normality in the marginal distribution of returns. Thus, our idea of joint parsimonious modeling of skewness and tails thickness by means of the SGED distribution may be useful in estimating a stochastic volatility model.

Bayesian estimation of the parameters of the stochastic volatility model (1)-(2) has been carried out via the MCMC algorithm presented in section 4. The hybrid

Gibbs-MH updating scheme has been implemented with the following specification of the prior distributions

1. $\phi = 2\phi^* - 1$ where $\phi^* \sim \text{Beta}(20, 1.5)$,
2. $\sigma_\eta^2|\phi, \nu \sim IG(2.5, 0.025)$,
3. $\mu \sim N(0, 20)$,
4. $\nu \sim U(0, 4)$,
5. $\lambda \sim U(-5, 5)$

The burn-in period has been set to 25000 and $M = 50000$. All calculations have been performed with the package Ox® v. 3.0.

Besides the SGED specification we also consider the Gaussian, the Skew-Normal (see Azzalini 1985) and GED specifications for the distribution of the returns shocks. To estimate these models we apply the algorithms described in section 4 for the relevant parameters[8].

5.2 Posterior analyisis

Analysis of the posterior distributions for the parameters of the stochastic volatility models is presented in Tables 5-7. We also report results for the square of the coefficient of variation of the volatility, i.e. $\text{Var}(\exp\{h_t\})/[E(\exp\{h_t\})]^2$, which provides a measure of relative dispersion since, partly, the mean and standard deviation tend to change together in many experiments, its knowledge is of some value in evaluating these experiments (Jacquier et al., 1994). For daily and weekly exchange rate series and for each parameter of the different models, we report the mean of the posterior distribution, the standard error (MCSE) of this mean, the standard deviation of the posterior distribution and a 95% confidence interval. Since draws from the posterior distributions are not independent, the reported MCSEs are an estimate of 2π times the spectral density matrix at frequency zero computed by standard time series method. In particular, our estimator is based on a VAR(1) prewhitening, than 2π times the spectral density matrix at frequency zero of VAR residuals is estimated by smoothing methods using the Parzen kernel and automatic bandwidth selection. Recolouring provides an estimate of 2π times the spectral density matrix at frequency zero of interest.

As expected from the descriptive statistics, we are not able to find significant asymmetry in the DM/US$ series, both daily and weekly. In fact, confidence in-

[8] For the Skew-Normal errors we impose the restriction $\nu = 2$ in the SGED case

tervals for the asymmetry parameter λ in the SGED and Skew-Normal models are almost symmetric around zero. However, we find strong evidence of fat tails in the conditional distribution of daily returns with posterior means for ν of about 1.5 and confidence intervals not including the threshold value of 2. Further, these findings remain unchanged under the SGED and GED specifications. Finally, there is some improvement in the coefficient of variation when we move from the standard normal model to a model with heavier tails. Thus, the standard normal and both skewed specifications are rejected for the daily data set in favor of the GED model. As for the weekly series, both the asymmetry and fat tails hypotheses appear to be inconsistent with the data, suggesting a model where returns shocks are normally distributed.

Figures 6 and 7 report empirical estimates (by kernel smoothing) of the posterior distributions for the parameters and of the coefficient of variation. In each panel, the solid line refers to the SGED model, the dash-and-dotted line to the Skew-Normal model, the short dashed line to the GED model and the dashed line to the Normal model, when they are available. It is noticeable the dramatic effect on the posterior distributions of ϕ and σ_η for daily data when the assumption of normality is relaxed. As for the Skew-Normal specification, in the daily data we observe a leftward shift of the posterior density of ϕ and a rightward one for σ_η with respect to the SGED or GED models. However, the most relevant result concerns the posterior distribution of the skewness parameter λ which is clearly uninformative being flat over the parameter space.

Results for the US$/£ exchange rate are summarized in Table 6. As far as we are concerned with asymmetry, we find strong evidence in the weekly data (negative skewness) but not in the daily data. In the former case we notice that explicit modeling of the asymmetry parameter has consequences on the tails thickness parameter ν, whose 95% confidence interval in the SGED model does include the value of 2 suggesting that a Skew-Normal model for the conditional distribution of returns may be more appropriate. In fact, under the Skew-normal specification we observe an increase in the skewness parameter and a more precise confidence interval. Since when under a GED specification ν is well below 2 and a 95% confidence interval does not contain it, we argue that failure of correct asymmetry modeling may result in spurious heavy tails. On the other hand, for daily data, the GED specification appears to be appropriate. Once again, by looking at the posterior densities for daily US$/£ in Figure 8, we notice that the posterior density of ϕ is very concentrated in the vicinity of unity and that it changes dramatically as we relax the hypothesis of normality, the same comment applies to σ_η. Once again, in the daily data the posterior density for λ in the Skew-Normal model is uninformative.

Figure 9 reports estimates of the posterior densities for weekly exchange rates. It is noticeable the higher precision of the posterior density for λ when moving from the SGED model to the Skew-Normal one.

Last, we consider the ¥/US$ exchange rate. Here, we find significant evidence of both asymmetry and fat tails in daily and weekly exchange rates since the 95% confidence intervals for λ and ν do not include the values of zero and 2, respectively. In particular, the posterior means of λ are -0.12 and -0.567 for the daily and weekly series respectively indicating significant negative skewness while the posterior means for ν are 1.336 and 1.323 for the daily and weekly series respectively signaling significant departures from the normal model in favor of a distribution with fatter tails. Failure of proper treatment of the tails of the distribution is reflected in higher estimates of the skewness parameter and of σ_η. Thus, neglecting the importance of heavy tails may results in higher estimates of asymmetry. This fact is even more evident as we consider the estimated posterior distributions for λ in the Skew-Normal and SGED models where we observe a significant leftward shift when the role of the tails is neglected. As usual, the posterior mean of the autoregressive parameter ϕ is close to unity increasing significantly as we move from the Normal and Skew-Normal specifications to the GED and SGED ones. Thus for this data set the SGED specification seems the correct one. As for the estimates of the posterior densities in Figures 10 and 11, we notice the usual shift towards unity of the density of ϕ as we relax the normality assumption. The posterior for σ is also affect in dramatic way by the normality assumption. The posteriors for the parameter ν in the daily exchange rates are almost identical while for the weekly data the SGED model entails a leftward shift with respect to the GED model, even though both models imply significant departure from normality.

5.3 Model ranking

The SGED stochastic volatility model, say $\mathcal{M}_1$, nests three models according to alternative restrictions on the parameters: for $\lambda = 0$ we have the GED model, say $\mathcal{M}_2$, for $\nu = 2$ we have the Skew-Normal model, say $\mathcal{M}_3$, and for $\lambda = 0$ and $\nu = 2$ we have the Gaussian model, $\mathcal{M}_4$. Both models $\mathcal{M}_2$ and $\mathcal{M}_3$ also nest model $\mathcal{M}_4$ but they are not nested between them. This entails a model reduction scheme such as in Figure 5 where we move from model $\mathcal{M}_1$ with skewed and heavy tailed errors to either model $\mathcal{M}_2$ with no skewness but fat tails or to model $\mathcal{M}_3$ with skewed but Gaussian errors and then to model $\mathcal{M}_4$ with symmetric Gaussian errors. The parameter vector in the four models can be summarized as

1. $\mathcal{M}_1 : \boldsymbol{\theta}_1 = (\phi, \sigma_\eta, \mu, \lambda, \nu)'$,
2. $\mathcal{M}_2 : \boldsymbol{\theta}_2 = (\phi, \sigma_\eta, \mu, \nu)'$,
3. $\mathcal{M}_3 : \boldsymbol{\theta}_3 = (\phi, \sigma_\eta, \mu, \lambda)'$,
4. $\mathcal{M}_4 : \boldsymbol{\theta}_4 = (\phi, \sigma_\eta, \mu)'$

from which is evident that $\mathcal{M}_2$ and $\mathcal{M}_3$ are not nested.

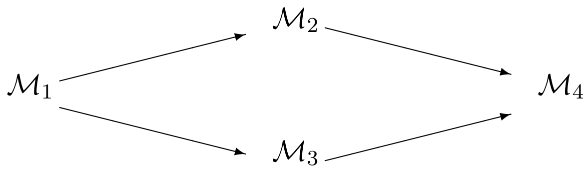

Figure 5: Model reduction

In our framework, it is natural to make use of the Bayes factor to compare the four different models. In general, the Bayes factor for comparing model $\mathcal{M}_j$ and model $\mathcal{M}_1$ is given by

$$BF_{21} = \frac{p(\boldsymbol{y}|\mathcal{M}_j)}{p(\boldsymbol{y}|\mathcal{M}_1)}, \qquad j = 2, 3, 4$$

where $p(\boldsymbol{y}|\mathcal{M}_j)$ is the predictive density of the data under model $\mathcal{M}_j$, namely $p(\boldsymbol{y}|\mathcal{M}_j) = \int_{\boldsymbol{h}} \int_{\boldsymbol{\theta}_j} p(\boldsymbol{y}, \boldsymbol{h}|\boldsymbol{\theta}_j, \mathcal{M}_j) p(\boldsymbol{\theta}_j|\mathcal{M}_j) \mathrm{d}\boldsymbol{\theta}_j \mathrm{d}\boldsymbol{h}$.

In the context of nested hypotheses on the parameters and because of the a priori independence amongst the parameters, the Bayes factors is referred to as the Savage-Dickey (SD) ratio and it simplifies dramatically [9]. The SD ratio for model $\mathcal{M}_2$ versus model $\mathcal{M}_1$ is given by

$$SD_{21} = \frac{p(\lambda = 0|\boldsymbol{y}, \mathcal{M}_1)}{p(\lambda = 0|\mathcal{M}_1)} \tag{10}$$

where $p(\lambda = 0|\boldsymbol{y}, \mathcal{M}_1)$ and $p(\lambda = 0|\mathcal{M}_1)$ stand for the marginal posterior and the marginal a priori density for λ, respectively, both under the SGED model $\mathcal{M}_1$ and

[9]See Verdinelli and Wasserman (1995) for a methodological treatment and Forbes et al. (2002) for an application to a stochastic volatility model.

evaluated at $\lambda = 0$. Analogously, the SD ratio for model $\mathcal{M}_3$ versus model $\mathcal{M}_1$ is given by

$$SD_{31} = \frac{p(\nu = 2|\boldsymbol{y}, \mathcal{M}_1)}{p(\nu = 2|\mathcal{M}_1)} \tag{11}$$

and a SD ratio for model $\mathcal{M}_4$ versus model $\mathcal{M}_1$ is then given by

$$SD_{41} = \frac{p(\lambda = 0, \nu = 2|\boldsymbol{y}, \mathcal{M}_1)}{p(\lambda = 0, \nu = 2|\mathcal{M}_1)} \tag{12}$$

In a similar fashion we also have SD ratios for model $\mathcal{M}_4$ versus model $\mathcal{M}_2$ and for model $\mathcal{M}_4$ versus model $\mathcal{M}_3$. These SD ratios are given

$$SD_{42} = \frac{p(\nu = 2|\boldsymbol{y}, \mathcal{M}_2)}{p(\nu = 2|\mathcal{M}_2)} \tag{13}$$

$$SD_{43} = \frac{p(\lambda = 0|\boldsymbol{y}, \mathcal{M}_3)}{p(\lambda = 0|\mathcal{M}_3)} \tag{14}$$

where both the posterior and the prior densities are those under the GED model $\mathcal{M}_2$ and the Skew-Normal model $\mathcal{M}_3$, respectively. As for the interpretation of the Bayes factor, we follow Kass and Raftery (1995) who suggest that the evidence for $\mathcal{M}_j$ is "negative" when $SD_{j1} < 1$, when $1 < SD_{j1} < 3.2$ the evidence is "not worth more than a bare mention", when $3.2 < SD_{j1} < 10$ the evidence is "substantial", when $10 < SD_{j1} < 100$ the evidence is "strong" and, finally, when $SD_{j1} > 100$ the evidence is "decisive" (see Jeffreys 1961). Intuitively, when the SD ratio is small the "probability mass" associated with the parameter restrictions implied by model $\mathcal{M}_j$ is unimportant relative to the a priori "probability mass" associated to those parameter values under the more general model $\mathcal{M}_1$. Hence, the evidence provided by the data does not support the restricted model. A similar reasoning applies for large values of the SD ratio, which provides increasing support to the restricted model $\mathcal{M}_j$.

Computation of the SD ratios is straightforward since the denominator is directly available from the a priori distributions and the numerator can be calculated using the MCMC simulation output by kernel smoothing estimation of the relevant marginal posterior densities at the point of interest. Results are reported in Table 4. This evidence complements our findings on 95% confidence interval for the skewness and tails parameters. For the daily data, we have strong evidence in favor of the GED model for the DM/US$ and £/US$ exchange rates while the evidence is substantial for the ¥/US$ exchange rate. For weekly data, we observe strong evidence in favor of the Gaussian model for the DM/US$ rate while for the £/US$ there is some evidence in favor of the Skew-Normal model and in the ¥/US$ case the SD ratios favors the Skew-GED model.

Table 4: Bayes Factors

	Daily				Weekly			
	SGED	GED	SNORM	NORM	SGED	GED	SNORM	NORM
	DM/US$							
SGED	1	33.39	0.00	0.00	1	4.57	5.09	22.60
GED		1	*	0.00		1	*	4.84
SNORM			1	4.90			1	4.12
NORM				1				1
	£/US$							
SGED	1	52.95	0.00	0.00	1	1.05	3.45	0.81
GED		1	*	0.00		1	*	0.71
SNORM			1	4.10			1	0.15
NORM				1				1
	¥/US$							
SGED	1	9.52	0.00	0.00	1	0.00	0.00	0.00
GED		1	*	0.00		1	*	0.00
SNORM			1	0.01			1	0.00
NORM				1				1

Entry (i, j) indicates the Bayes factor in favour of model j versus model i.

6 Concluding remarks

In this paper we have proposed a stochastic volatility model with an explicit modeling of asymmetry and fat tails. This is accomplished by assuming that the returns shock be distributed as Skew-GED distribution. We restricted our attention to a model with no correlation between the return and the volatility shock, this interesting extension is left to ongoing research. Inference on the model has been conducted in a Bayesian framework via Markov Chain MonteCarlo. Difficulties in the calculation of posterior distributions arising because the volatilities are not observed are overcome by the design of a Gibbs-MH updating scheme which allows to simulate these posterior distributions. We make use of both a Delayed-Rejection MH and an ARM sampling which permit us to simulate the volatilities and the parameters of the model in a fast and effective fashion. Finally, we have considered an application to daily and weekly exchange rates and found some significant evidence in favor of our Skew-GED model.

Summarizing the findings from our empirical application we conclude that the normal model may be relevant for the weekly DM/US$ case and the GED model for the daily DM/US$ returns. As for the US$/£ exchange rate we do find evidence of asymmetry and heavy tails in the daily data but the Skew-normal seems to be preferred for weekly data. Finally, for the ¥/US$ data we find consistent evidence in favor of the SGED specification.

Table 5: Posterior analysis for DM/US$

	Daily				Weekly			
	Normal	GED	SGED	SNormal	Normal	GED	SGED	SNormal
	ϕ							
$\widehat{E}(\phi\|data)$	0.976	0.983	0.984	0.974	0.962	0.967	0.963	0.960
$MCSE(\widehat{E}(\phi\|data))$	(0.0004)	(0.0003)	(0.0003)	(0.0005)	(0.0008)	(0.0011)	(0.0015)	(0.0012)
$\widehat{SD}(\phi\|data)$	0.008	0.006	0.006	0.0085	0.014	0.015	0.018	0.016
95% Conf. Inter.	[0.958, 0.990]	[0.969, 0.993]	[0.971, 0.993]	[0.955, 0.988]	[0.929, 0.986]	[0.931, 0.990]	[0.919, 0.989]	[0.920, 0.985]
	σ							
$\widehat{E}(\sigma\|data)$	0.142	0.115	0.109	0.150	0.185	0.170	0.183	0.189
$MCSE(\widehat{E}(\sigma\|data))$	(0.0017)	(0.0012)	(0.0014)	(0.001)	(0.0025)	(0.0036)	(0.0045)	(0.0036)
$\widehat{SD}(\sigma\|data)$	0.022	0.017	0.018	0.023	0.035	0.041	0.048	0.042
95% Conf. Inter.	[0.105, 0.192]	[0.085, 0.153]	[0.080, 0.150]	[0.107, 0.198]	[0.122, 0.262]	[0.112, 0.259]	[0.112, 0.292]	[0.117, 0.286]
	β							
$\widehat{E}(\beta\|data)$	0.601	0.611	0.614	0.599	1.311	1.319	1.314	1.309
$MCSE(\widehat{E}(\beta\|data))$	(0.0003)	(0.0005)	(0.0005)	(0.0003)	(0.0011)	(0.0015)	(0.0014)	(0.0015)
$\widehat{SD}(\beta\|data)$	0.044	0.052	0.054	0.043	0.107	0.120	0.113	0.103
95% Conf. Inter.	[0.520, 0.694]	[0.516, 0.724]	[0.517, 0.730]	[0.520, 0.691]	[1.117, 1.536]	[1.101, 1.572]	[1.110, 1.552]	[1.119, 1.532]
	coeff-var							
$\widehat{E}(coeff-var\|data)$	0.597	0.547	0.529	0.610	0.668	0.665	0.678	0.671
$MCSE(\widehat{E}(coeff-var\|data))$	(0.0030)	(0.0031)	(0.0037)	(0.0029)	(0.0040)	(0.0051)	(0.0068)	(0.0066)
$\widehat{SD}(coeff-var\|data)$	0.223	0.229	0.218	0.212	0.290	0.338	0.346	0.288
95% Conf. Inter.	[0.337, 1.143]	[0.284, 1.129]	[0.273, 1.091]	[0.350, 1.116]	[0.345, 1.333]	[0.314, 1.443]	[0.327, 1.390]	[0.329, 1.311]
	ν							
$\widehat{E}(\nu\|data)$		1.553	1.553			1.825	1.997	
$MCSE(\widehat{E}(\nu\|data))$		(0.0012)	(0.0015)			(0.0073)	(0.0206)	
$\widehat{SD}(\nu\|data)$		0.075	0.076			0.153	0.345	
95% Conf. Inter.		[1.413, 1.707]	[1.412, 1.709]			[1.560, 2.159]	[1.597, 2.969]	
	λ							
$\widehat{E}(\lambda\|data)$			0.026	-0.175			-0.379	-0.424
$MCSE(\widehat{E}(\lambda\|data))$			(0.0009)	(0.0051)			(0.0086)	(0.0049)
$\widehat{SD}(\lambda\|data)$			0.124	0.5338			0.723	0.561
95% Conf. Inter.			[-0.239, 0.260]	[-0.994, 0.807,]			[-1.820, 1.275]	[-1.274, 0.689]

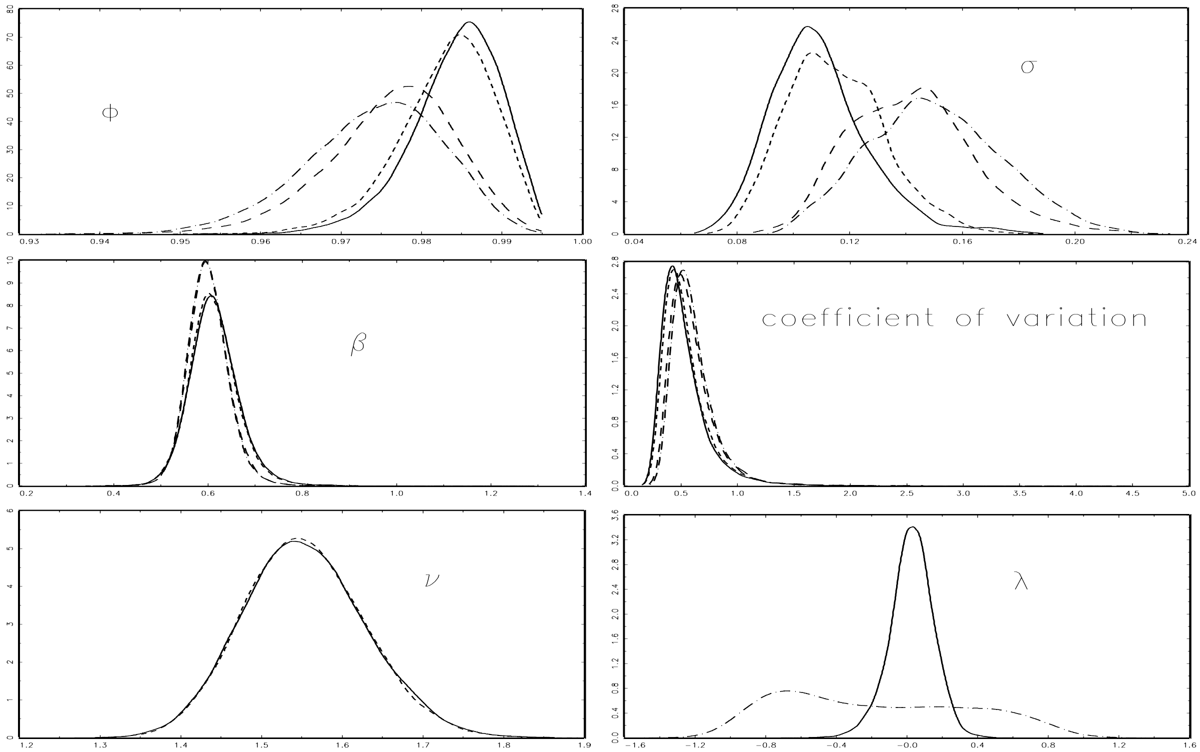

Figure 6: Posterior densities DM/US$daily

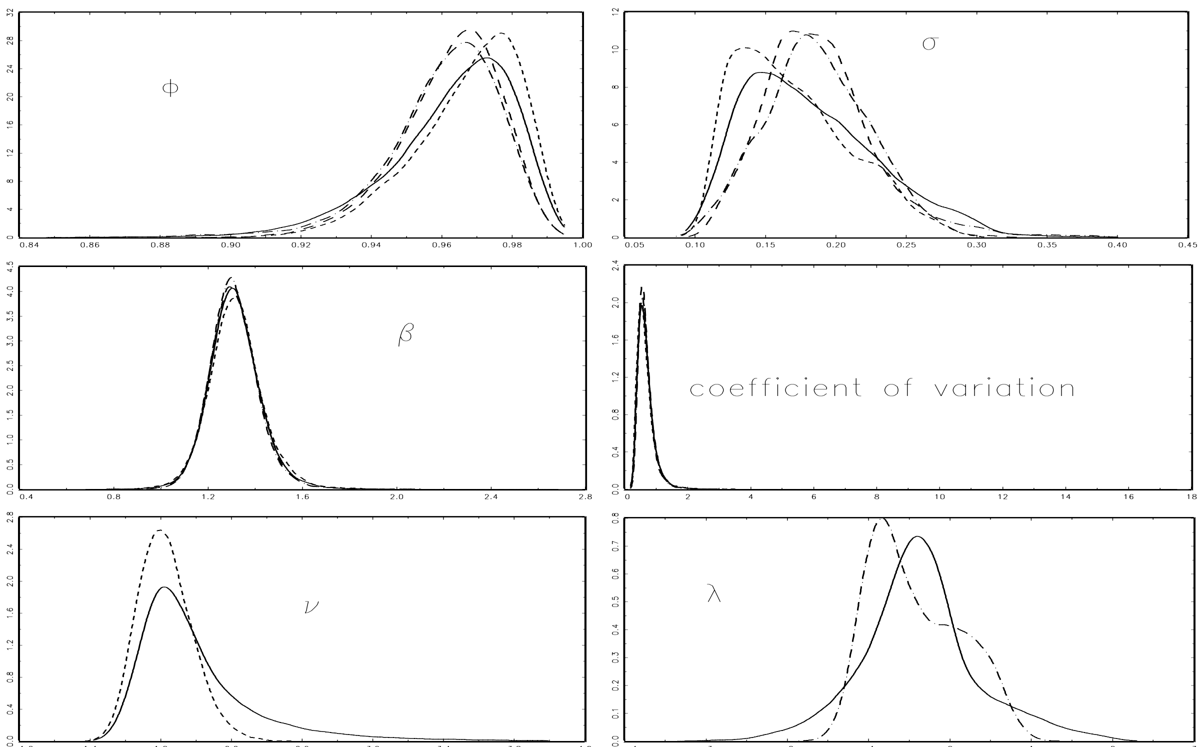

Figure 7: Posterior densities DM/US$ weekly

Solid line: SGED errors; Dash-and-dotted line: S-Normal errors; Short dashed line: GED errors; Dashed line: Normal errors.

Table 6: Posterior analysis for US$/£

	Daily				Weekly			
	Normal	GED	SGED	SNormal	Normal	GED	SGED	SNormal
	ϕ							
$\widehat{E}(\phi\|data)$	0.978	0.990	0.988	0.975	0.957	0.963	0.962	0.959
$MCSE(\widehat{E}(\phi\|data))$	(0.0005)	(0.0001)	(0.0002)	(0.0004)	(0.0005)	(0.0004)	(0.0005)	(0.0004)
$\widehat{SD}(\phi\|data)$	0.008	0.003	0.004	0.0075	0.013	0.011	0.012	0.0115
95% Conf. Inter.	[0.961, 0.990]	[0.982, 0.995]	[0.978, 0.994]	[0.958, 0.988]	[0.930, 0.979]	[0.939, 0.982]	[0.936, 0.982]	[0.933, 0.979]
	σ							
$\widehat{E}(\sigma\|data)$	0.151	0.094	0.106	0.161	0.262	0.235	0.242	0.253
$MCSE(\widehat{E}(\sigma\|data))$	(0.0022)	(0.0007)	(0.0014)	(0.0016)	(0.0022)	(0.0024)	(0.0023)	(0.0019)
$\widehat{SD}(\sigma\|data)$	0.025	0.012	0.017	0.023	0.038	0.038	0.038	0.033
95% Conf. Inter.	[0.106, 0.200]	[0.075, 0.121]	[0.077, 0.141]	[0.124, 0.212]	[0.195, 0.339]	[0.165, 0.309]	[0.166, 0.316]	[0.188, 0.324]
	β							
$\widehat{E}(\beta\|data)$	0.510	0.527	0.521	0.507	1.191	1.211	1.206	1.197
$MCSE(\widehat{E}(\beta\|data))$	(0.0004)	(0.0004)	(0.0005)	(0.0003)	(0.0011)	(0.0009)	(0.0010)	0.0008
$\widehat{SD}(\beta\|data)$	0.039	0.052	0.049	0.036	0.110	0.117	0.116	0.110
95% Conf. Inter.	[0.438, 0.594]	[0.432, 0.642]	[0.433, 0.630]	[0.439, 0.584]	[0.994, 1.423]	[0.999, 1.459]	[0.997, 1.457]	[0.995, 1.431]
	coeff-var							
$\widehat{E}(coeff-var\|data)$	0.765	0.607	0.645	0.784	1.438	1.320	1.361	1.392
$MCSE(\widehat{E}(coeff-var\|data))$	(0.0041)	(0.0047)	(0.0060)	(0.0033)	(0.0102)	(0.0147)	(0.0103)	(0.0108)
$\widehat{SD}(coeff-var\|data)$	0.269	0.238	0.265	0.255	0.663	0.740	0.669	0.815
95% Conf. Inter.	[0.438, 1.402]	[0.306, 1.218]	[0.334, 1.295]	[0.467, 1.412]	[0.724, 2.976]	[0.592, 2.945]	[0.662, 2.861]	[0.688, 2.872]
	ν							
$\widehat{E}(\nu\|data)$		1.432	1.457			1.687	1.806	
$MCSE(\widehat{E}(\nu\|data))$		(0.0008)	(0.0015)			(0.0028)	(0.0064)	
$\widehat{SD}(\nu\|data)$		0.060	0.063			0.122	0.218	
95% Conf. Inter.		[1.317, 1.553]	[1.338, 1.587]			[1.471, 1.947]	[1.483, 2.334]	
	λ							
$\widehat{E}(\lambda\|data)$			0.022	0.312			-0.876	-1.243
$MCSE(\widehat{E}(\lambda\|data))$			(0.0006)	(0.0053)			(0.0116)	(0.0028)
$\widehat{SD}(\lambda\|data)$			0.080	0.531			0.467	0.296
95% Conf. Inter.			[-0.133, 0.185]	[-0.741, 1.047]			[-1.818, -0.096]	[-1.715, -0.587]

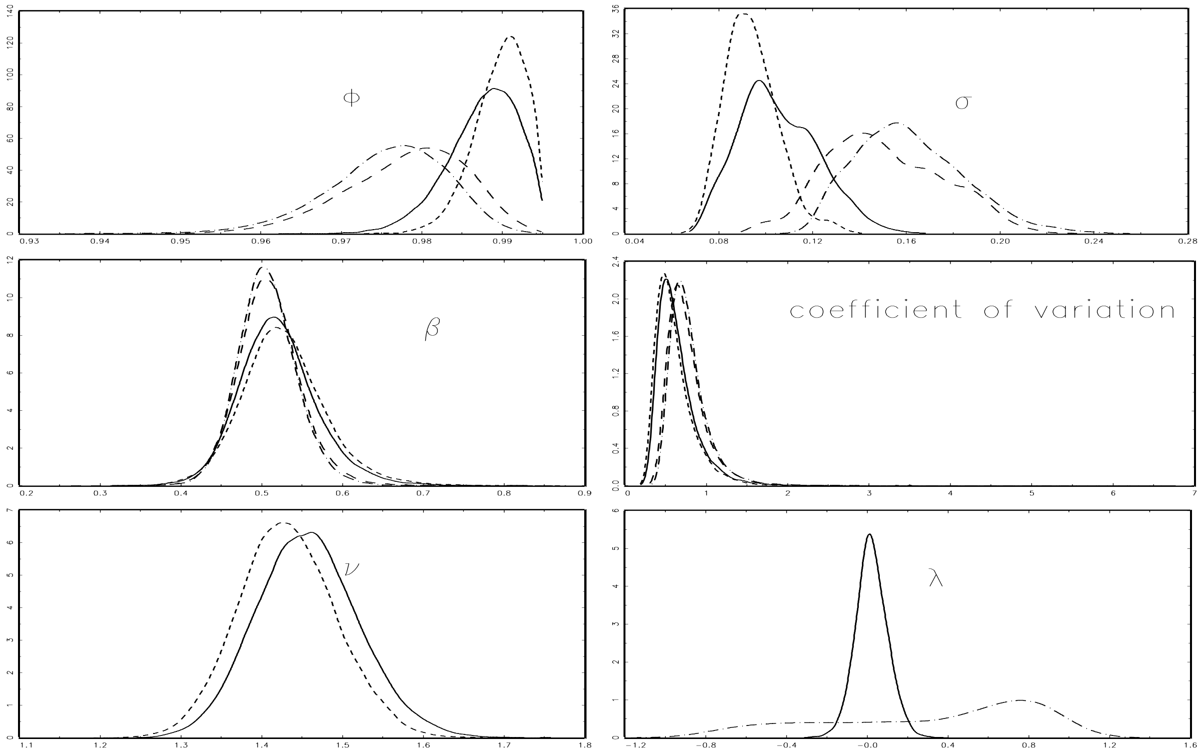

Figure 8: Posterior densities US$/£ daily

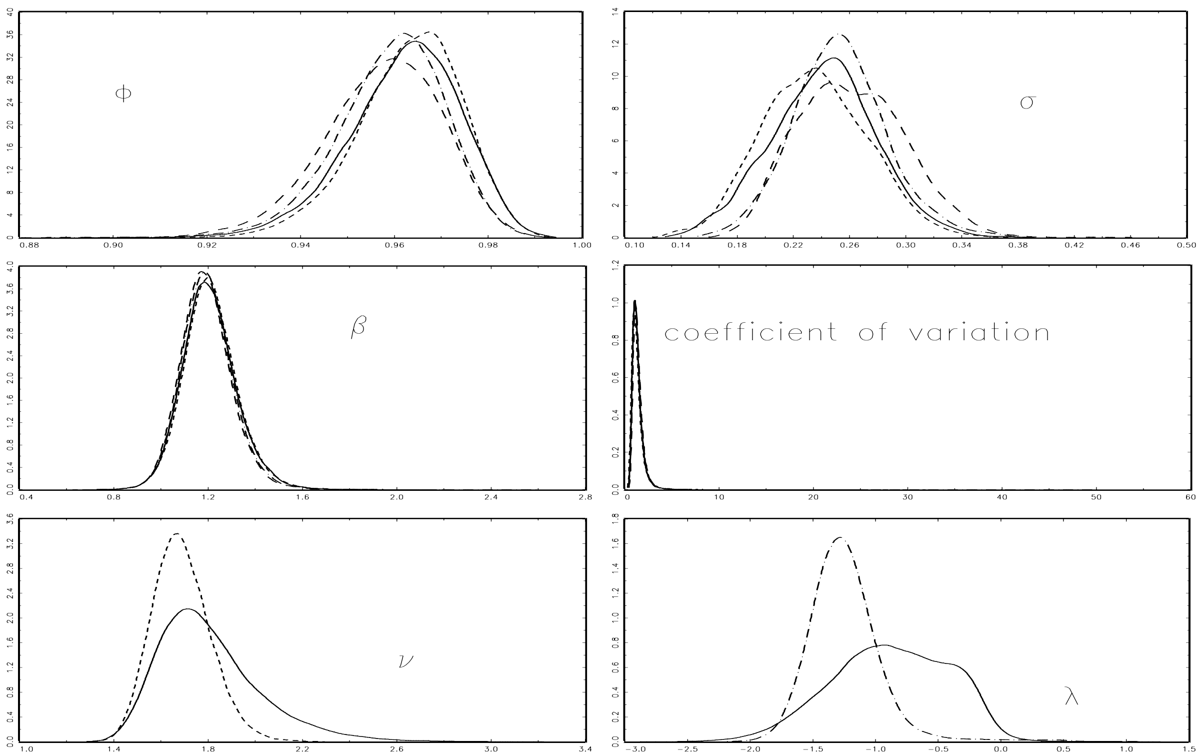

Figure 9: Posterior densities US$/£ weekly

Solid line: SGED errors; Dash-and-dotted line: S-Normal errors; Short dashed line: GED errors; Dashed line: Normal errors.

Table 7: Posterior analysis for ¥/US$

	Daily				Weekly			
	Normal	GED	SGED	SNormal	Normal	GED	SGED	SNormal
	ϕ							
$\widehat{E}(\phi\|data)$	0.927	0.984	0.985	0.944	0.953	0.983	0.985	0.962
$MCSE(\widehat{E}(\phi\|data))$	(0.0017)	(0.0002)	(0.0003)	(0.0017)	(0.0009)	(0.0004)	(0.0003)	(0.0007)
$\widehat{SD}(\phi\|data)$	0.021	0.005	0.005	0.018	0.016	0.007	0.006	0.013
95% Conf. Inter.	[0.878, 0.961]	[0.973, 0.993]	[0.974, 0.994]	[0.902, 0.973]	[0.917, 0.979]	[0.966, 0.994]	[0.971, 0.994]	[0.931, 0.984]
	σ							
$\widehat{E}(\sigma\|data)$	0.287	0.109	0.104	0.238	0.283	0.153	0.141	0.247
$MCSE(\widehat{E}(\sigma\|data))$	(0.0042)	(0.0012)	(0.0014)	(0.0044)	(0.0036)	(0.0022)	(0.0018)	(0.0031)
$\widehat{SD}(\sigma\|data)$	0.047	0.016	0.017	0.0447	0.050	0.029	0.026	0.0437
95% Conf. Inter.	[0.210, 0.388]	[0.081, 0.141]	[0.075, 0.139]	[0.162, 0.328]	[0.193, 0.393]	[0.102, 0.214]	[0.096, 0.195]	[0.171, 0.338]
	β							
$\widehat{E}(\beta\|data)$	0.615	0.652	0.652	0.623	1.206	1.261	1.260	1.219
$MCSE(\widehat{E}(\beta\|data))$	(0.0005)	(0.0003)	(0.0005)	(0.0006)	(0.0010)	(0.0013)	(0.0012)	(0.0009)
$\widehat{SD}(\beta\|data)$	0.027	0.051	0.053	0.0309	0.109	0.178	0.185	0.120
95% Conf. Inter.	[0.565, 0.672]	[0.559, 0.761]	[0.555, 0.765]	[0.567, 0.689]	[1.005, 1.440]	[0.939, 1.652]	[0.927, 1.665]	[0.996, 1.474]
	coeff-var							
$\widehat{E}(coeff-var\|data)$	0.831	0.527	0.510	0.736	1.580	1.279	1.215	1.476
$MCSE(\widehat{E}(coeff-var\|data))$	(0.0069)	(0.0026)	(0.0032)	(0.0063)	(0.0139)	(0.0115)	(0.0130)	(0.0092)
$\widehat{SD}(coeff-var\|data)$	0.165	0.199	0.194	0.159	0.789	0.922	0.836	0.719
95% Conf. Inter.	[0.560, 1.203]	[0.286, 1.031]	[0.271, 1.005]	[0.478, 1.096]	[0.788, 3.178]	[0.519, 3.440]	[0.477, 3.288]	[0.728, 3.149]
	ν							
$\widehat{E}(\nu\|data)$		1.338	1.336			1.409	1.323	
$MCSE(\widehat{E}(\nu\|data))$		(0.0010)	(0.0014)			(0.0026)	(0.0015)	
$\widehat{SD}(\nu\|data)$		0.055	0.055			0.088	0.080	
95% Conf. Inter.		[1.233, 1.450]	[1.233, 1.449]			[1.244, 1.591]	[1.177, 1.493]	
	λ							
$\widehat{E}(\lambda\|data)$			-0.112	-1.168			-0.567	-1.715
$MCSE(\widehat{E}(\lambda\|data))$			(0.0006)	(0.0030)			(0.0037)	(0.0029)
$\widehat{SD}(\lambda\|data)$			0.065	0.169			0.196	0.214
95% Conf. Inter.			[-0.263, -0.008]	[-1.473, -0.820]			[-0.981, -0.227]	[-2.147, -1.305]

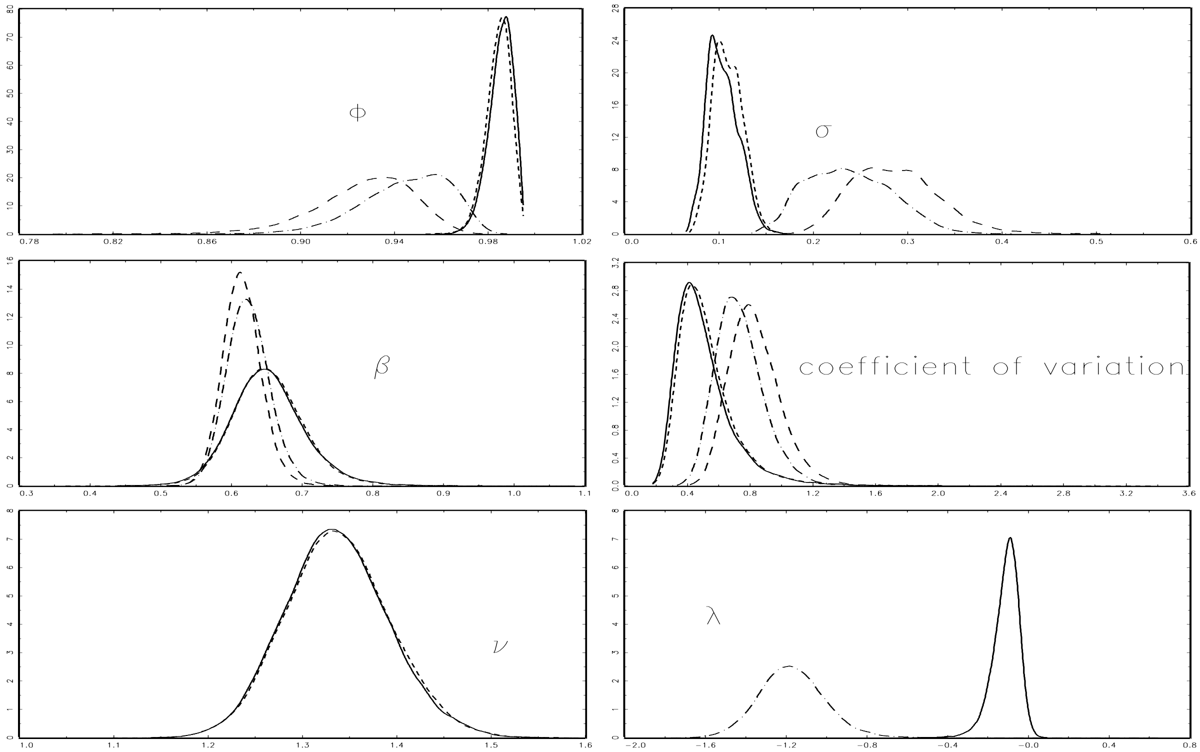

Figure 10: Posterior densities ¥/ US$ daily

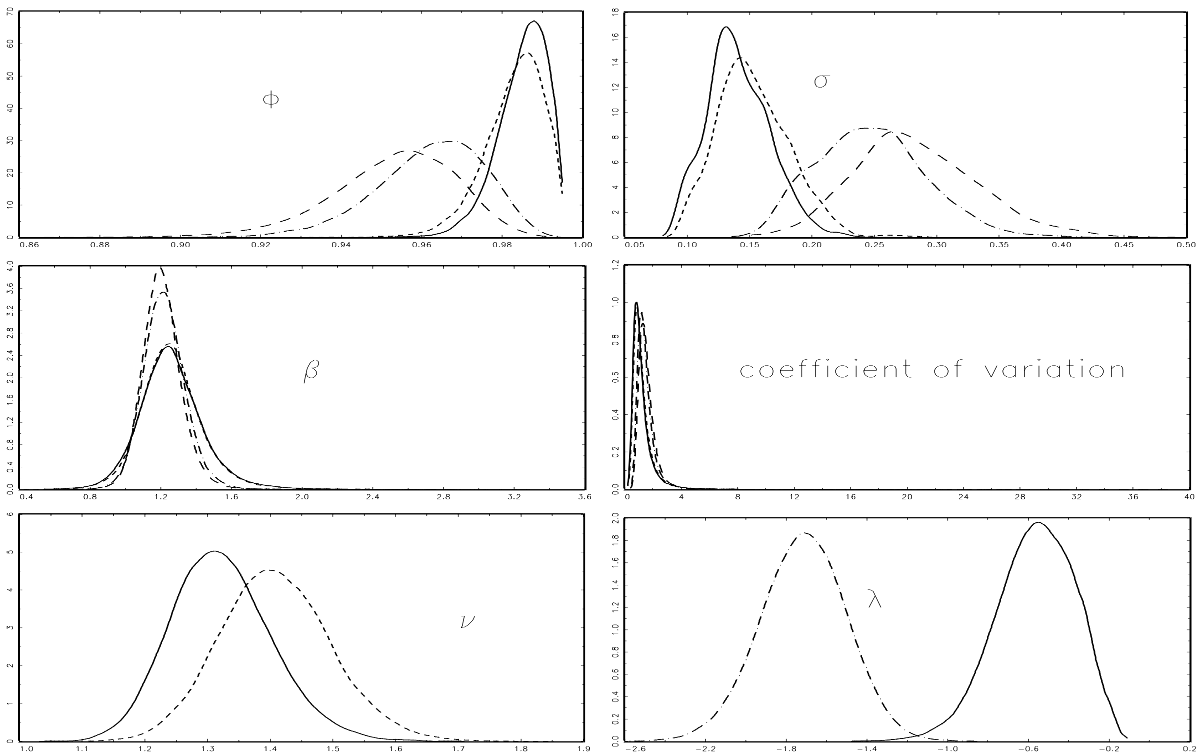

Figure 11: Posterior densities ¥/ US$ weekly

Solid line: SGED errors; Dash-and-dotted line: S-Normal errors; Short dashed line: GED errors; Dashed line: Normal errors.

Acknowledgements

Previous versions of this paper have been presented at a workshop on skewed distributions, held at the Department of Statistics, University of Padova, Padova (Italy) on September 19 2001, and at the Conference "Linear and non-linear dynamics in time series", Bressanone (Italy), June 6-7, 2003. Comments from the discussant Giuseppe Cavaliere (Bressanone) and participants to the workshops are acknowledged. We are grateful to Adelchi Azzalini for the precise and incisive comments. Financial support from MIUR PRIN 2000 is gratefully acknowledged.

References

Andersen, T. G. (1996), 'Return volatility and trading volume: an information flow interpretation of stochastic volatility', *The Journal of Finance* **51**, 169–204.

Azzalini, A. (1985), 'A class of distributions which includes the normal ones', *Scandinavian Journal of Statistics* **12**, 171–178.

Azzalini, Adelchi (1986), 'Further results on a class of distributions which includes the normal ones', *Statistica* **46**, 199–208.

Cappuccio, N., D. Lubian and D. Raggi (2001), Bayesian estimation of a stochastic volatility model with Skew-GED errors, *in* C.Provasi, ed., 'Metodi complessi e metodi computazionali intensivi per la stima e la previsione', CLEUP, pp. 313–318.

Carlin, B.P., N.G. Polson and D.S. Stoffer (1992), 'A Monte Carlo approach to nonnormal and nonlinear state-space modeling', *Journal of the American Statistical Association* **87**, 493–500.

Chib, S. and E. Greenberg (1995), 'Understanding the Metropolis-Hastings algorithm', *The American Statistician* **49**, 327–335.

Chib, S. and E. Greenberg (1996), 'Markov Chain Monte Carlo simulation methods in econometrics', *Econometric Theory* **12**, 409–431.

Chib, S., F. Nardari and N. Shephard (1998), Markov Chain Monte Carlo methods for generalized stochastic volatility models, Working paper, Oxford University, Nuffield College.

Chunhachinda, P., K. Dandapani, S. Hamid and A. J. Prakash (1997), 'Portfolio selection and skewness: evidence from international stock markets', *Journal of Banking and Finance* **21**, 143–167.

Corrado, C. J. and T. Su (1997), Implied volatility skews and stock index skewness and kurtosis implied by S&P500 index option prices, Working paper, University of Missouri - Columbia.

Engle, R. F. (1982), 'Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation', *Econometrica* **50**, 987–1008.

Fama, E. (1965), 'Portfolio analysis in a stable paretian market', *Management Science* **11**, 404–419.

Forbes, C. S., G. M. Martin and J. Wright (2002), Bayesian estimation of a stochastic volatility model using option and spot prices, Working paper, Monash University - Department of Econometrics and Business Statistics.

Ghysels, E., A. C. Harvey and E. Renault (1996), Stochastic volatility, *in* C. R.Rao and G. S.Maddala, eds, 'Statistical Methods in Finance', North-Holland, pp. 119–191.

Gilks, W., S. Richardson and D. J. Spiegelhalter, eds (1995), *Markon Chain Monte Carlo in practice*, Chapman & Hall.

Gilks, W.R., N.G. Best and K.K.C. Tan (1995), 'Adaptive rejection Metropolis sampling within Gibbs sampling', *Applied Statistic* **44**, 455–472.

Gilks, W.R. and P. Wild (1992), 'Adaptive rejection sampling for Gibbs sampling', *Applied Statistic* **41**, 337–448.

Goria, Sarah (1998), Uso di generalizzazioni della distribuzione normale nei modelli per serie storiche, PhD thesis, Facoltà di Sc. Statistiche, Università di Padova.

Hastings, W.K. (1970), 'Monte Carlo sampling methods using Markov Chains and their applications', *Biometrika* **57**, 97–109.

Hollander, M. and D. A. Wolfe (1999), *Nonparametric Statistical Methods*, Probability and Statistics, 2nd edn, Wiley.

Jacquier, E., N. G. Polson and P. Rossi (1994), 'Bayesian analysis of stochastic volatility models', *Journal of Business & Economic Statistics* **12**, 371–417.

Jacquier, E., N. G. Polson and P. Rossi (1999), Models and priors for multivariate stochastic volatility, Working paper, CIRANO, forthcoming in *Journal of Econometrics.*

Jeffreys, H. (1961), *Theory of probability*, Oxford University Press, Oxford.

Kass, R.E. and A. E. Raftery (1995), 'Bayes factors', *Journal of the American Statistical Association* **90**, 773–795.

Kim, S., N. G. Shephard and S. Chib (1998), 'Stochastic volatility: likelihood inference and comparison with ARCH models', *Review of Economic Studies* **65**, 361–393.

Liesenfeld, R. and R.C. Jung (2000), 'Stochastic volatility models: conditional normality versus heavy-tailed distributions', *Journal of Applied Economietrics* **15**, 137–160.

Mandelbrot, B. (1963), 'The variation of certain speculative prices', *Journal of Business* **36**, 314–419.

Metropolis, N., A. W. Rosenbluth, M.N. Rosenbluth, A. H. Teller and E. Teller (1953), 'Equations of state calculations by fast computing machines', *Journal of Chemical Physics* **21**, 1087–1091.

Mira, A. (2001), 'On Metropolis-Hastings algorithms with delayed rejections', *Metron* **59**, 231–241.

Mittnik, S. and M. S. Paolella (2000), 'Conditional density and value-at-risk prediction of asian currency exchange rates', *Journal of Forecasting* **19**, 313–333.

Peiró, A. (1999), 'Skewness in financial returns', *Journal of Banking and Finance* **23**, 847–862.

Robert, R. P. and G. Casella (1999), *Monte Carlo Statistical Methods*, Springer.

Steel, M. F. J. (1998), 'Bayesian analysis of stochastic volatility models with flexible tails', *Econometric Reviews* **17**, 109–143.

Tanner, M.A. and W.H. Wong (1987), 'The calculation of posterior distributions by data augmentation', *Journal of the American Statistical Association* **82**, 528–524.

Taylor, S. J. (1994), 'Modelling stochastic volatility: a review and comparative study', *Mathematical Finance* **4**, 183–204.

Tierney, L. (1994), 'Markov Chains for exploring posterior distributions', *Annals of Statistics* **22**, 1701–1728.

Tierney, L. and A. Mira (1999), 'Some adaptive Monte Carlo methods for bayesian inference', *Statistics in Medicine* **18**, 2507–2515.

Verdinelli, I and L. Wasserman (1995), 'Computing Bayes factors using a generalization of the Savage-Dickey density ratio', *Journal of the American Statistical Association* **90**, 614–618.