Bayesian Student-*t* Stochastic Volatility Models via a Two-stage Scale Mixtures Representation

S.T. Boris Choy[1]*and C.M. Chan[2]

[1] Department of Mathematical Sciences,
University of Technology, Sydney, Australia

[2] Department of Statistics,
The Chinese University of Hong Kong, Shatin, Hong Kong

Abstract

In this paper, we provide a statistical analysis of the Stochastic Volatility (SV) models using full Bayesian approach. Student-t distribution is chosen as an alternative to the normal distribution for modelling white noise. Bayesian computation of the SV models completely relies on the Markov chain Monte Carlo methods. In particular, to speed up the efficiency of the Gibbs sampling scheme, we propose a two-stage scale mixtures representation for the Student-t

*Correspondence address: Department of Mathematical Sciences, University of Technology, Sydney, P.O. Box 123, Broadway, NSW 2007, Australia. Email: boris.choy@uts.edu.au

density function. In addition, our proposed scale mixtures scheme allows the identification of possible outlying observations.

Keywords: Scale Mixtures of Normal (SMN) Distributions, Scale Mixtures of Uniform (SMU) Distributions, Student-t Distribution, Markov chain Monte Carlo, Gibbs Sampler, Metropolis-Hastings Algorithm, Ratio-of-uniforms, t-normal Stochastic Volatility Model.

1 Introduction

Volatility of an asset or the return of an asset plays an important role in managing risks in financial world. To measure the volatility of an asset, different time series models have been proposed and studied. Twenty years ago, Eagle (1982) introduced the Autoregressive Conditional Heteroskedasticity (ARCH) models for modelling volatility of economic and financial time series. These models are superior to many conventional times series models because they allow the conditional variances of the time series data to be a deterministic function of the past observations. Thereafter, a number of variations of ARCH models, such as the Generalized ARCH (GARCH), Exponential GARCH (EGARCH), Integrated GARCH (IGARCH), Fractionally Integrated GARCH (FIGARCH), Factor ARCH (FACTOR-ARCH), Threshold GARCH (TGARCH), etc. were proposed to deal with a variety of data such as interest rates, exchange rates, equity returns, Treasury Bills, option pricing, etc.. See Bollerslev (1987), Nelson (1991), Engle and Bollerslev (1986), Engle *et al.* (1990) and Tong (1990). In addition, readers are referred to Eagle (1995) for a comprehensive review of the ARCH models.

An alternative class of models for modelling volatility is the stochastic volatility (SV) models. Unlike the ARCH-type models where the volatility depends on past realizations, the SV models formulate the volatility by an unobservable process that allows the volatilities to vary stochastically. See Taylor (1986, 1994) for details. Parameter estimation and forecasting for the SV-type models are, however, more difficult as the evaluation of the likelihood function usually involves high dimensional integration which is computationally expensive. This makes the likelihood-based analysis difficult and impractical. Therefore, attention has been shifted to the method of moments and quasi maximum likelihood approach (Harvey *et al.* 1994), etc.. Recent advancement in Bayesian computational techniques provides an efficient way for analysis of the SV models. Jacquier *et al.* (1994) adopts the Gibbs sampling scheme while Shephard and Pitt (1997) employ the Metropolis-Hastings scheme. Though the Markov chain Monte Carlo algorithms provide a flexible way for modelling times-dependent data, they suffer the serious drawback of high correlation amongst volatilities. To tackle this problem, Carter and Kohn (1994) and Shephard and Pitt (1997) use the blocking sampling techniques while De Jong and Shephard (1995) simulate the log-volatility using the Gaussian simulation smoother.

Although a number of improvements have been proposed to make statistical inference possible and simple, a deviation from the normality assumption for the time series data does impose an increasing computational burden to the analysis. In fact, many financial data exhibit a thick tail behaviour. This tempts statisticians and econometricians to model asset returns using heavy-tailed distributions such as Student-t distribution, symmetric stable distribution or exponential-power distribution. Harvey *et al.* (1994) and Jacquier *et al.* (1994) consider the Student-t distribution while Nelson (1988, 1994) deals with exponential-power distribution. For asymmetric choice of distribution, see Harvey and Shephard (1996) and Fernández and Steel (1998). However, the extension from normal to heavy-tailed distributional assumption increases the computational load dramatically. Therefore, with the expression of the complicated distributions into the scale mixtures forms, Bayesian computation can be easily performed. See Jacquier *et al.* (1994), who analyze the modified SV model using Markov chain Monte Carlo methods. For illustration of Markov chain Monte Carlo algorithms, see, for example, Gelfand and Smith (1990), Smith and Roberts (1993) and Tierney (1994).

This paper aims to provide a full Bayesian analysis for SV models with robustifying Student-t distribution for the asset return. A new representation for the Student-t density function is proposed which further simplifies the Bayesian computation of Jacquier *et al.* (1994) for the analysis of SV models from the Bayesian point of view. The structure of the paper is as follows. We first introduce a two-stage scale mixtures representation for the Student-t density function and then incorporate it into a basic SV model in a Bayesian framework for asset return to develop the basic formulae for the Gibbs sampling algorithm. For illustrative purpose, an exchange rate data set is analysed in detail. Finally, a discussion is given.

2 Two-stage Scale Mixtures Representation for Student-t Distribution

Andrews and Mallows (1974) characterized the univariate class of scale mixtures of normal (SMN) distributions using the Laplace transformation approach. For the standard random variable X having this normal scale mixtures representation, it can be expressed in the form of $X = Z \times \lambda$ where Z is the standard normal random variate and λ is a positive random variate known as the mixing variable, having a mixing distribution g, which can be either continuous or discrete. The Student-t, symmetric stable and exponential-power distributions are some of the well-known SMN distributions. See West (1987) and Choy and Smith (1997a) for the robustness properties of these distributions and Wakefield *et al.* (1994), Choy and Smith (1997b) and Fernandez and Steel (2000) for applications in Bayesian hierarchical models.

Let θ and σ be the location and scale parameters of the scale mixtures of normal random variable X. Then the probability density function of X has the following mixture form

$$f(x) = \int_{R^+} N(x|\theta, \lambda\sigma^2)\, g(\lambda) d\lambda \tag{2.1}$$

where $N(\cdot|\cdot)$ denotes the normal density, $g(\cdot)$ is a probability density function defined on $\Re^+ = (0, \infty)$ and λ is referred to as a mixing parameter which is commonly used as a global diagnostic check for outliers. See Choy and Smith (1997b) for details. In Bayesian framework, the mixture density in (2.1) can be expressed into a two-stage hierarchy of the form

$$\begin{aligned} X|\theta, \sigma^2, \lambda &\sim N(\theta, \lambda\sigma^2) \\ \lambda &\sim g(\lambda). \end{aligned}$$

For suitably chosen mixing density function g, a wide class of symmetric and unimodal SMN distributions can be obtained. In particular, the Student-t distribution with degrees of freedom α corresponds to an inverse gamma mixing distribution, i.e.

$$\lambda \sim IG\left(\frac{\alpha}{2}, \frac{\alpha}{2}\right)$$

where $IG(a, b)$ is the inverse gamma distribution with density

$$g(\lambda) = \frac{b^a}{\Gamma(a)\lambda^{a+1}}\, e^{-b/\lambda} \qquad \lambda > 0,\ a, b > 0.$$

To facilitate an efficient computation for the SV models, we make use of the class of scale mixtures of uniform (SMU) representation for the normal density. If X is a normal random variable with mean θ and variance σ^2, it can be easily shown that its density function can be rewritten into

$$N(x|\theta, \sigma^2) = \int_{\theta-\sigma\sqrt{u}}^{\theta+\sigma\sqrt{u}} \frac{1}{2\sigma\sqrt{u}} \, Ga\left(u \left| \frac{3}{2}, \frac{1}{2}\right.\right) du$$

where $Ga(u|a, b)$ is the gamma density function with parameters a and b. Details about the SMU distributions can be found in Walker and Gutiérrez-Peña (1999), Damien *et al.* (1999) and Choy and Walker (1999).

Theoretically, all SMN distributions can be expressed into the form of SMU distributions. Therefore, we can express the Student-t distribution with degrees of freedom α into the following hierarchy

$$\begin{aligned} X|\theta, \sigma^2, \lambda, u &\sim U(\theta - \sigma\lambda^{1/2} u^{1/2}, \theta + \sigma\lambda^{1/2} u^{1/2}) \\ \lambda &\sim IG\left(\frac{\alpha}{2}, \frac{\alpha}{2}\right) \\ u &\sim Ga\left(\frac{3}{2}, \frac{1}{2}\right) \end{aligned}$$

where $U(a, b)$ is a uniform distribution defined on the interval (a, b). Now, we shall show the advantages of using this two-stage scale mixtures representation for the Student-t distribution in the SV models in Section 3.

3 Bayesian Student-t SV Models

For modelling financial data, more and more attention has been focused on using heavy-tailed distributions such as Student-t, symmetric stable and exponential power distributions etc. Student-t distribution is probably the most popular among them and is therefore used in this paper.

Let r_t be the asset value of an equity at time $t = 0, 1, 2, \ldots, n$. The mean adjusted asset return y_t at time t is defined as

$$y_t = \ln\left(\frac{r_t}{r_{t-1}}\right) - \frac{1}{n}\sum_{i=1}^{n} \ln\left(\frac{r_i}{r_{i-1}}\right), \qquad t = 1, 2, \ldots, n.$$

Let H_t and h_t be the volatilities and log-volatilities, respectively. The standard (traditional) SV model for the asset return, y_t is defined as

$$\begin{aligned} y_t &= \beta\, H_t^{1/2}\, \epsilon_t, \qquad t = 1, 2, \ldots, n \\ h_t &= \begin{cases} \sigma\eta_1/\sqrt{1-\phi^2} & t = 1 \\ \phi\, h_{t-1} + \sigma\eta_t & t > 1 \end{cases} \end{aligned}$$

where $\{\epsilon_t\}$ and $\{\eta_t\}$ are independent standard Gaussian processes. β is a constant factor that represents the model instantaneous volatility which is usually set to one in many literatures. σ is the variance of the log-volatilities and $\phi \in (-1, 1)$ is the persistence of the volatility.

In this section, the mean adjusted returns are modelled by a Student-t distribution while the log-volatility is assumed to follow a normal distribution. In fact, we can also use heavy-tailed distribution for modelling the log-volatility to achieve a higher degree of robustness. However, our emphasis is on the development of an efficient simulation algorithm for the Gibbs sampler scheme. The extension to robustifying the log-volatility can be done without substantial increase in computational effort.

3.1 Bayesian t-N SV models

In the current setup, we write

$$y_t|h_t \overset{i.i.d.}{\sim} t_\alpha\left(0, \beta^2 H_t\right) \qquad t = 1, 2, \ldots, n.$$

The conditional distribution of the log-volatility h_t has a normal distribution of the form

$$h_t|h_{t-1}, \phi, \sigma^2 \overset{i.i.d.}{\sim} N\left(\phi h_{t-1}, \sigma^2\right)$$

while the marginal distribution is

$$h_t|\phi, \sigma^2 \overset{i.i.d.}{\sim} N\left(0, \frac{\sigma^2}{1-\phi^2}\right)$$

where $\phi \in (-1, 1)$ is the instantaneous persistence parameter and σ is a scale parameter of the log-volatility. As in many literatures, we may assume that β is fixed. For simplicity, we shall call this a t-N SV model. To complete a Bayesian framework, we assign independent priors for ϕ and σ^2 which are the shifted beta and inverse gamma distributions, respectively, i.e.

$$\frac{\phi+1}{2} \sim Be(a_\phi, b_\phi) \qquad \text{and} \qquad \sigma^2 \sim IG(a_\sigma, b_\sigma)$$

Now we can rewrite the t-N SV model hierarchically as

$$\begin{aligned}
y_t|h_t, \lambda_t, u_t &\sim U\left(-\beta H_t^{1/2}\lambda_t^{1/2}u_t^{1/2}, \beta H_t^{1/2}\lambda_t^{1/2}u_t^{1/2}\right)\\
\lambda_t &\sim IG\left(\alpha/2, \alpha/2\right)\\
u_t &\sim Ga\left(3/2, 1/2\right)\\
h_t|h_{t-1}, \phi, \sigma^2 &\sim N\left(\phi h_{t-1}, \sigma^2\right)\\
\frac{\phi+1}{2} &\sim Be\left(\alpha_\phi, \beta_\phi\right)\\
\sigma^2 &\sim IG\left(a_\sigma, b_\sigma\right)
\end{aligned}$$

for $t = 1, 2, \ldots, n$.

3.2 Gibbs Sampler for the t-N SV model

For statistical analysis, we implement the SV model using simulation-based Gibbs sampling approach. Let $\boldsymbol{y} = (y_1, \ldots, y_n)$, $\boldsymbol{h} = (h_1, \ldots, h_n)$, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)$, $\boldsymbol{u} = (u_1, \ldots, u_n)$ and, for $i = 1, \ldots, n$, $\boldsymbol{h}_{-t} = (h_1, \ldots, h_{t-1}, h_{t+1}, \ldots, h_n)$, $\boldsymbol{\lambda}_{-t} = (\lambda_1, \ldots, \lambda_{t-1}, \lambda_{t+1}, \ldots, \lambda_n)$ and $\boldsymbol{u}_{-t} = (u_1, \ldots, u_{t-1}, u_{t+1}, \ldots, u_n)$. With arbitrarily chosen starting values for $\boldsymbol{h}, \boldsymbol{\lambda}, \boldsymbol{u}, \sigma^2$ and ϕ, the Gibbs sampler iteratively samples random variates from a system of full conditional distributions and the resulting simulations are used to mimic a random sample from the target joint posterior distribution. Now the system of full conditionals is given below:-

1. Full conditional distribution for $h_t, t = 1, \ldots, n$:-

$$h_t | \boldsymbol{h}_{-t}, \boldsymbol{\lambda}, \boldsymbol{u}, \sigma^2, \phi, \boldsymbol{y} \sim N(a_t, b_t\sigma^2) \qquad h_t > \ln y_t^2 - 2\ln\beta \ln\lambda_t - \ln u_t$$

where

$$a_t = \begin{cases} \phi\, h_{t+1} - \sigma^2/2 & t = 1 \\ (1+\phi^2)^{-1}\big(\phi(h_{t-1} + h_{t+1}) - \sigma^2/2 & 2 \le t \le n-1 \\ \phi\, h_{t-1} - \sigma^2/2 & t = n \end{cases}$$

and

$$b_t = \begin{cases} 1 & t = 1, n \\ (1+\phi^2)^{-1} & 2 \le t \le n-1 \end{cases}$$

2. Full conditional distribution for σ^2:-

$$\sigma^2 | \boldsymbol{h}, \boldsymbol{\lambda}, \boldsymbol{u}, \phi, \boldsymbol{y} \sim IG\bigg(a_\sigma + \frac{n}{2}, b_\sigma + \frac{1}{2}\bigg[(1-\phi^2)h_1^2 + \sum_{t=2}^{n}(h_t - \phi h_{t-1})^2\bigg]\bigg)$$

3. Full conditional distribution for $\lambda_t, t = 1, \ldots, n$:-

$$\lambda_t | \boldsymbol{h}, \boldsymbol{\lambda}_{-t}, \boldsymbol{u}, \phi, \sigma^2, \boldsymbol{y} \sim IG\left(\frac{\alpha+1}{2}, \frac{\alpha}{2}\right) \qquad \lambda_t > \frac{y_t^2}{\beta^2 H_t u_t}$$

4. Full conditional distribution for $u_t, t = 1, \ldots, n$:-

$$u_t | \boldsymbol{h}, \boldsymbol{\lambda}, \boldsymbol{u}_{-t}, \phi, \sigma^2, \boldsymbol{y} \sim Exp\left(\frac{1}{2}\right) \qquad u_t > \frac{y_t^2}{\beta^2 H_t \lambda_t}$$

5. Full conditional distribution for ϕ:-

$$\phi|\boldsymbol{h},\boldsymbol{\lambda},\boldsymbol{u},\sigma^2,\boldsymbol{y} \sim N\left(\phi\left|\frac{\sum_{t=2}^{n} h_{t-1}h_t}{\sum_{t=2}^{n} h_t^2}, \frac{\sigma^2}{\sum_{t=2}^{n} h_t^2}\right.\right)(1+\phi)^{a_\alpha-1/2}(1-\phi)^{b_\alpha-1/2} \quad |\phi| \leq 1.$$

In the above expressions, conditional distributions for h_t, λ_t, u_t and σ^2 are the truncated normal, truncated inverse gamma, truncated exponential and inverse gamma distributions while that for ϕ is a product of a normal and a shifted beta distributions. Simulation of random variates from truncated exponential distribution can be straightforwardly done by the inversion method while simulation from truncated normal and gamma distributions can be done by applying the algorithms proposed by Robert (1995) and Philippe (1997), respectively.

The Appendix presents the corresponding system of full conditional distributions for all parameters without using the uniform mixtures representation. It can been seen that the full conditional distributions for $h_t, t = 1, 2, \ldots, n$, have non-standard densities. Therefore, our proposed two-stage scale mixtures representation for the Student-t density has the advantage of simplifying the full conditional distributions for the log-volatilities h_t and hence makes the Bayesian computation more efficient for the SV models. However, to simulate the non-standard density of ϕ, we still need to use the rejection method, or other simulation methods whatever the scale mixtures representations used.

4 Illustrative Example

To demonstrate our proposed t-N SV model, we analyse the exchange rates of US dollars to Sterling pounds. Fig. 1 presents 1000 mean adjusted daily closing exchange rate returns from January 2, 1981. Obviously, some trading days produce extreme returns. In the following simulation study, we set $\beta = 1$ for simplicity. To reflect the non-informative prior knowledge about the variance of the log-volatility σ^2, we choose $a_\sigma = b_\sigma = 0.001$. However, we shall adopt a beta $Be(20, 1)$ informative prior for $(\phi + 1)/2$.

In the examples, we ran the Gibbs sampler for a single series of 35000 iterations. The first 5000 iterations are discarded as the 'burn-in' period. To avoid high correlation between successive drawings of the financial data, we pick up simulated values at every 30th value to mimic a random sample of 1000 drawings from the target joint posterior distribution. Fig. 2 displays the ACF of $h_1, \lambda_1, u_1, \sigma$ and ϕ based on these 1000 values. The sampled values are rather uncorrelated except for σ.

====================================

Fig.1, Fig.2 and Table 1 about here

====================================

4.1 Parameter estimation

For various degrees of freedom α, posterior estimates of σ and ϕ, together with the standard errors and 95% posterior intervals, are presented in Table 1 and Table 2, respectively, the corresponding boxplots are presented in Fig. 3. We notice that σ increases gradually with α. It can be seen that a heavier tailed distribution for the returns results in a smaller variation of the log-volatility and a higher degree of persistence. Fig. 4 plots the volatilities, H_t, over the study period for a Cauchy and a normal sampling distributions, respectively and shows that small H_t values (and h_t values) are associated with small α values. This phenomenon can be explained by the robustifying property of the Student-t distribution that it protects inference of the volatility as its tails are heavy enough to capture the outlying observations.

====================================

Fig.3 and Fig.4 about here

====================================

4.2 Outlier diagnosis

In this paper, we propose the SMN-SMU scale mixtures presentation for the Student-t density in SV models in which computational issues and outlier identification are our primary concern. Though this is not compulsory as many other different simulation methods can be employed, our approach does simplify the system of full conditional distributions and hence provides a very efficient Gibbs sampling scheme. In addition, mixing parameters λ_t and u_t enable us to perform a global diagnosis of possible outliers. The posterior means of $\lambda_t \, u_t$ under a Cauchy-N SV model is plotted in Fig. 5. Large values of $\lambda_t \, u_t$ correspond to possible outlying daily returns. With reference to Fig. 1, extreme daily returns are clearly identified in Fig. 5. The five most influential daily returns are on Days 106, 697, 840, 34 and 858, respectively. Table 3 presents the extreme daily returns and their volatilities for Cauchy-N and N-N SV models. The volatilities of these days for the Cauchy-N and N-N models are very different - the former's volatilities are smaller than those of the latter's.

It is well-known that normal conjugate models do not provide a robustness analysis, the advantage of using heavy-tailed distribution is obvious when we compare the results for using the Cauchy-N and N-N models. The extreme daily returns are automatically downweighed in the analysis and inference of H_t and hence h_t are protected. We can notice from Table 1 that the posterior estimates of σ drops from 0.3796 to 0.3293 when the sampling distribution changes from normal to Cauchy.

====================================

Table 3 and Fig.5 about here

====================================

4.3 Model selection

To choose the most suitable value of α for the $t-N$ SV model, we adopt the model selection criterion suggested by San Martini and Spezzaferri (1984). Let M_α be the t-N SV model with α degrees of freedom. The posterior expected utility $U(\alpha)$ of this model is defined by

$$U(\alpha) = \frac{1}{n} \sum_{t=1}^{n} \ln\, p\,(y_t|M_\alpha)$$

and is evaluated from the Gibbs sampling outputs. The best model corresponds to the one that gives the largest value of $U(\alpha)$. Fig. 6 shows the pattern of the posterior expected utility $U(\alpha)$ against α. Obviously, the Cauchy-N model is the best choice in this comparative study.

====================================

Fig. 6 about here

====================================

Alternatively, we can treat the degrees of freedom α as random and assign, probably, a gamma $G(a_\alpha, b_\alpha)$ prior distribution to it. In that case the full conditional density of α has the form

$$p(\alpha|\boldsymbol{h}, \boldsymbol{\lambda}, \boldsymbol{u}, \sigma^2, \phi) \quad \propto \quad p(\boldsymbol{\lambda}|\alpha)\, p(\alpha)$$

$$\propto \left(\frac{(\alpha/2)^{\alpha/2}}{\Gamma(\alpha/2)} \right)^n Ga \left(\alpha \middle| a_\alpha, b_\alpha + \frac{1}{2} \sum_{t=1}^{n} (\lambda_t - \ln \lambda_t) \right)$$

where random variates can be simulated using, for example, the Metropolis-Hastings, ratio-of-uniforms (see Wakefield *et al.*, 1992) algorithms or using the adaptive rejection sampling (see Gilks and Wild, 1992) algorithm as the above conditional density can easily be shown to be log-concave.

4.4 Weekly data analysis

Now we consider the weekly data of the exchange rate returns over the same time period. The dataset contains about 200 observations. The Bayes estimates of σ, ϕ and the expected posterior utility for different α values are given in Table 4. When compared with the daily return analysis, we obtain a completely different picture. First of all, the log-volatility of the weekly data are more volatile than that of the daily data because the former has a larger σ value. Secondly, it seems that the estimates of σ and ϕ for the weekly data are quite insensitive to α except for small α values. This means that the SV model is quite robust to the choice of degrees of freedom α of the Student-t sampling distribution. This brings out an important issue on the robustness property of the Student-t distribution. For the weekly data, if we assume a Student-t distribution for modelling the returns but the true sampling distribution is normal, our results will not be very different from the results using the normal model. On the contrary, if the Student-t is in fact the true sampling distribution for the daily data, assuming a normal sampling distribution can result in very different parameter estimates and related analysis will be misleading. See Tables 1 and 2 for the comparison. Therefore, from robustness point of view, the advantage of using heavy-tailed distribution as an alternative to the normal distribution in statistical modelling is clearly illustrated.

==

Table 4 about here

==

5 Discussion

Student-t distribution has been widely used for modelling heavy-tailed events and for robustness consideration. For Bayesian inference, statisticians always express the distribution into a scale mixtures of normals form so as to simplify the Bayesian calculations, in particular when Gibbs sampling scheme is used. See Wakefield *et al.* (1994) and Choy and Smith (1997b) for details. For some complicated models where data are dependent such as the GARCH and SV models, the SMN representation may not provide a substantial improvement in computational efficiency. The significance of this paper is on the introduction of a two-stage scale mixture representation for the Student-t density. This new representation enables us to simplify the Bayesian computation for the SV models. Similar to the SMN representation, the product of the mixing parameters, $\lambda_i u_i$, provides a means for outlier diagnosis.

In this paper, we do not attempt to find the most suitable model for modelling the returns of the exchange rates. In fact, we can assume the degrees of freedom α to be an unknown quantity and assign a suitable prior distribution to it. In addition, for robustification consideration, it is an advantage to adopt heavy-tailed distribution as an alternative to the normal distribution in statistical modelling. If the normal component is correct, our analysis will not be erratic. The use of a heavy-tailed distribution can however, protect our inference from model misspecification when the normal assumption is not correct. In addition, another heavy-tailed distribution can be used to model the log-volatility and we expect that this additional modification can further protect inference from unusual volatilities of the time series.

Regarding the Gibbs sampling algorithm, we adopt a single-move sampling scheme although some researchers suggest using multiple-move to reduce the high dependence of the simulated values and to speed up the convergence rate. Multiple-move is worthy to be considered but a technical difficulty is that most of the full conditionals are truncated distributions. However, our single-move scheme performs satisfactorily in the t-N SV models and the power of the two-stage scale mixtures representation of the Student-t is is also proved.

Appendix

Without using the uniform mixtures representation for the normal density, the SV model considered in Section 3.1 is formulated hierarchically as

$$\begin{aligned} y_t|h_t, \lambda_t, &\sim N\left(0, \lambda_t \beta^2 H_t\right) \\ \lambda_t &\sim IG\left(\alpha/2, \alpha/2\right) \\ h_t|h_{t-1}, \phi, \sigma^2 &\sim N\left(\phi h_{t-1}, \sigma^2\right) \\ \frac{\phi+1}{2} &\sim Be\left(\alpha_\phi, \beta_\phi\right) \\ \sigma^2 &\sim IG\left(a_\sigma, b_\sigma\right) \end{aligned}$$

for $t = 1, 2, \ldots, n$ and the system of full conditionals is given below:-

1. Full conditional distribution for $h_t, t = 1, \ldots, n$:-

$$h_t|\boldsymbol{h}_{-t}, \boldsymbol{\lambda}, \sigma^2, \phi, \boldsymbol{y} \propto \begin{cases} N\left(y_t|0, \lambda_t\beta^2 H_t\right) N\left(h_t|\frac{h_{t+1}}{\phi}, \frac{\sigma^2}{\phi^2}\right) & t = 1 \\ N\left(y_t|0, \lambda_t\beta^2 H_t\right) N\left(h_t|\frac{\phi(h_{t-1}+h_{t+1})}{1+\phi^2}, \frac{\sigma^2}{1+\phi^2}\right) & 2 \le t \le n-1 \\ N\left(y_t|0, \lambda_t\beta^2 H_t\right) N\left(h_t|\phi h_{t-1}, \sigma^2\right) & t = n \end{cases}$$

2. Full conditional distribution for σ^2:-

$$\sigma^2|\boldsymbol{h}, \boldsymbol{\lambda}, \phi, \boldsymbol{y} \sim IG\left(a_\sigma + \frac{n}{2}, b_\sigma + \frac{1}{2}\left[(1-\phi^2)h_1^2 + \sum_{t=2}^{n}(h_t - \phi h_{t-1})^2\right]\right)$$

3. Full conditional distribution for $\lambda_t, t = 1, \ldots, n$:-

$$\lambda_t|\boldsymbol{h}, \boldsymbol{\lambda}_{-t}, \phi, \sigma^2, \boldsymbol{y} \sim IG\left(\frac{\alpha+1}{2}, \frac{\alpha}{2} + \frac{y_t^2}{2\beta^2 H_t}\right)$$

4. Full conditional distribution for ϕ:-

$$\phi|\boldsymbol{h}, \boldsymbol{\lambda}, \sigma^2, \boldsymbol{y} \sim N\left(\phi\left|\frac{\sum_{t=2}^{n} h_{t-1}h_t}{\sum_{t=2}^{n} h_t^2}, \frac{\sigma^2}{\sum_{t=2}^{n} h_t^2}\right.\right)(1+\phi)^{a_\alpha - 1/2}(1-\phi)^{b_\alpha - 1/2} \quad |\phi| \le 1.$$

Onviously, the full conditional densities for h_t and ϕ are non-standard.

Acknowledgements

The first author was partially supported by a grant from the Research Grants Council of HKSAR (Project No. HKU 7257/99H).

References

[1] Andrews, D.F., Mallows, C.L., 1974. Scale mixtures of normal distribution. Journal of the Royal Statistics Society Series B 36, 99-102.

[2] Bollerslev, T., 1987. A conditionally heteroskedastic time series model for speculative prices and rates of returns. Review of Economics and Statistics 69, 542-547.

[3] Carter C.K., Kohn, R., 1996. Markov chain Monte Carlo in conditionally Gaussian state space models. Biometrika 83, 589-601.

[4] Choy, S.T.B., Smith, A.F.M., 1997a. On robust analysis of a normal location parameter. Journal of Royal Statistical Society, Series B 59, 463-474.

[5] Choy, S.T.B., Smith, A.F.M., 1997b. Hierarchical models with scale mixtures of normal distribution. TEST 6, 205-211.

[6] Choy, S.T.B., Walker, S.G., 2003. The extended exponential power distribution and Bayesian robustness. To appear in Statistics and Probability Letters.

[7] Damien, P., Walkefield, J.C., Walker, S.G., 1999. Gibbs sampling for Bayesian non-conjugate models using auxiliary variables. Journal of Royal Statistical Society, Series B 61, 331-344.

[8] De Jong, P., Shephard, N., 1995. The Simulation Smoother for Time-Series Models. Biometrika 82, 339-850.

[9] Engle, R.F., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation. Econometrica 50, 987-1006.

[10] Engle, R.F., 1995. ARCH Selected Readings. Oxford University Press, New York.

[11] Engle, R.F., Bollerslev, T.. 1986. Modelling the persistence of conditional variance. Econometric Review 5, 1-50.

[12] Engle, R.F., Ng, V.K., Rothschild, M., 1990. Asset pricing with a FACTOR-ARCH covariance structure: Empirical estimates for Treasury Bills. Journal of Econometrics 45, 213-237.

[13] Fernandez, C., Steel, M.F.J., 1998. On Bayesian modelling of fat tails and skewness. Journal of the American Statistical Association 93, 359-371.

[14] Fernandez, C., Steel, M.F.J., 2000. Bayesian regression analysis with scale mixtures of normals. Econometric Theory 16, 80-101.

[15] Gelfand, A.E., Smith, A.F.M., 1990. Sampling-based approaches to calculating marginal densities. Journal of American Statistical Association 85, 398-409.

[16] Gilks, W.R., Wild, P., 1992. Adaptive rejection sampling for Gibbs sampling. Applied Statistics 41, 337-348.

[17] Harvey, A.C., Ruiz, E., Shephard, N., 1994. Multivariate stochastic variance models. Rev. Economic Studies 61, 247-264. Reprinted as 256-276.

[18] Harvey, A.C., Shephard, N., 1996. Estimation of an Asymmetric Stochastic Volatility Model for Asset Returns. Journal of Business and Economic Statistics 14, 429-434.

[19] Jacquier, E., Polson, N.G., Rossi, P.E., 1994. Bayesian analysis of stochastic volatility models (with discussion). Journal of Business and Economics Statistics 12, 371-417.

[20] Nelson, D.B., 1991. Conditional heteroskedasticity in asset returns: A new approach. Econometrica 59, 347-370.

[21] Philippe, A., 1997. Simulation of right and left truncated gamma distribution by mixtures. Statistics and Computing 7, 173-181.

[22] Robert, C.P., 1995. Simulation of truncated normal variables. Statistics and Computing 5, 121-125.

[23] Shephard, N., Pitt, M.K., 1997. Likelihood Analysis of Non-Gaussian measurement time series. Biometrika 84, 653-667.

[24] Smith, A.F.M., Roberts, G.O., 1993. Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo Methods. Journal of the Royal Statistical Society, Series B 55, 3-23.

[25] San Martini, A., Spezzaferri, F., 1984. A predictive model selection criterion. Journal of the Royal Statistical Society, Series B 46, 296-303.

[26] Taylor, S.J., 1986. Modelling Financial Time Series. John Wiley, New York.

[27] Taylor, S.J., 1994. Modelling stochastic volatility. Mathematical Finance 4, 183-204.

[28] Tierney, L., 1994. Markov chain for exploring posterior distributions (with discussion). The Annals of Statistics 22, 1701-1762.

[29] Tong, H., 1990. Non-Linear Time Series: A Dynamical System Approach. Oxford University Press, Oxford.

[30] Wakefield, J.C., Gelfand, A.E., Smith, A.F.M., 1992. Efficient generation of random variates via the ratio-of-uniforms method. Statistics and Computing 1, 129-133.

[31] Wakefield, J.C., Smith, A.F.M., Racine-Poon, A.E., Gelfand, A.E., 1994. Bayesian analysis of linear and non-linear population models by using the Gibbs sampler. Applied Statistics 43, 201-221.

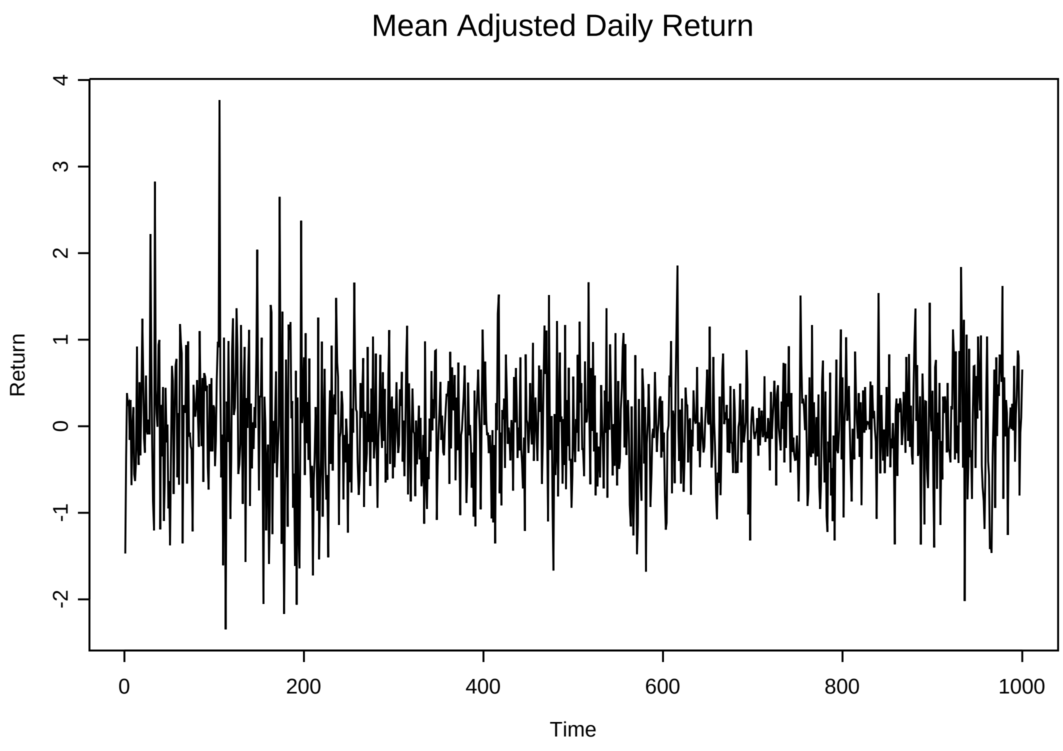
[32] Walker, S.G., Gutiérrez-Peña, E., 1999. Robustifying Bayesian Procedures (with discussion). In Bayesian Statistics 6, Bernardo JM, Berger JO, Dawid AP, Smith AFM (eds). Oxford University Press, New York, 685-710.

[33] West, M., 1984. Outlier models and posterior distributions in Bayesian linear regression. Journal of the Royal Statistical Society, Series B 46, 431-439.

[34] West, M., 1987. On scale mixtures of normal distributions. Biometrika 74, 646-648.

α	$E[\sigma\|\boldsymbol{y}]$	$SE[\sigma\|\boldsymbol{y}]$	95% posterior interval
1	0.3293	0.0411	(0.2536, 0.4238)
3	0.3304	0.0434	(0.2530, 0.4209)
5	0.3384	0.0456	(0.2563, 0.4404)
10	0.3524	0.0473	(0.2652, 0.4519)
15	0.3611	0.0484	(0.2772, 0.4652)
20	0.3623	0.0478	(0.2791, 0.4638)
∞	0.3796	0.0507	(0.2863, 0.4906)

Table 1: Bayes estimates, standard errors and 95% confidence intervals of σ for various α values.



α	$E[\phi\|\boldsymbol{y}]$	$SE[\phi\|\boldsymbol{y}]$	95% posterior interval
1	0.9903	0.0053	(0.9780,0.9988)
3	0.9839	0.0079	(0.9658,0.9967)
5	0.9802	0.0089	(0.9604,0.9960)
10	0.9752	0.0111	(0.9508,0.9953)
15	0.9725	0.0122	(0.9443,0.9930)
20	0.9713	0.0121	(0.9442,0.9923)
∞	0.9656	0.0140	(0.9370, 0.9902)

Table 2: Bayes estimates, standard errors and 95% confidence intervals of ϕ for various α values.

Time t	$y(t)$	N-N	Cauchy-N
34	2.8239	1.4833	0.2250
106	3.7680	1.9490	0.2851
697	-1.3220	0.3100	0.0387
840	1.5361	2.2450	0.0542
858	-1.3656	0.3231	0.0647

Table 3: Observed values of the extreme daily returns and their Bayes estimates of H_t under a Normal-Normal and Cauchy-Normal models.

α	$E[\sigma\|\boldsymbol{y}]$	$E[\phi\|\boldsymbol{y}]$	$U(\alpha)$
1	0.5572	0.9875	-0.3278
3	0.5163	0.9864	-0.1616
5	0.5107	0.9857	-0.1283
10	0.5038	0.9852	-0.1026
15	0.4987	0.9851	-0.0958
20	0.5009	0.9849	-0.0899
∞	0.4992	0.9843	-0.0782

Table 4: Bayes estimates of σ and ϕ and expected posterior utility of different t-N SV models for weekly return data.

Fig.1. *Time series plot of mean adjusted daily exchange rate returns.*

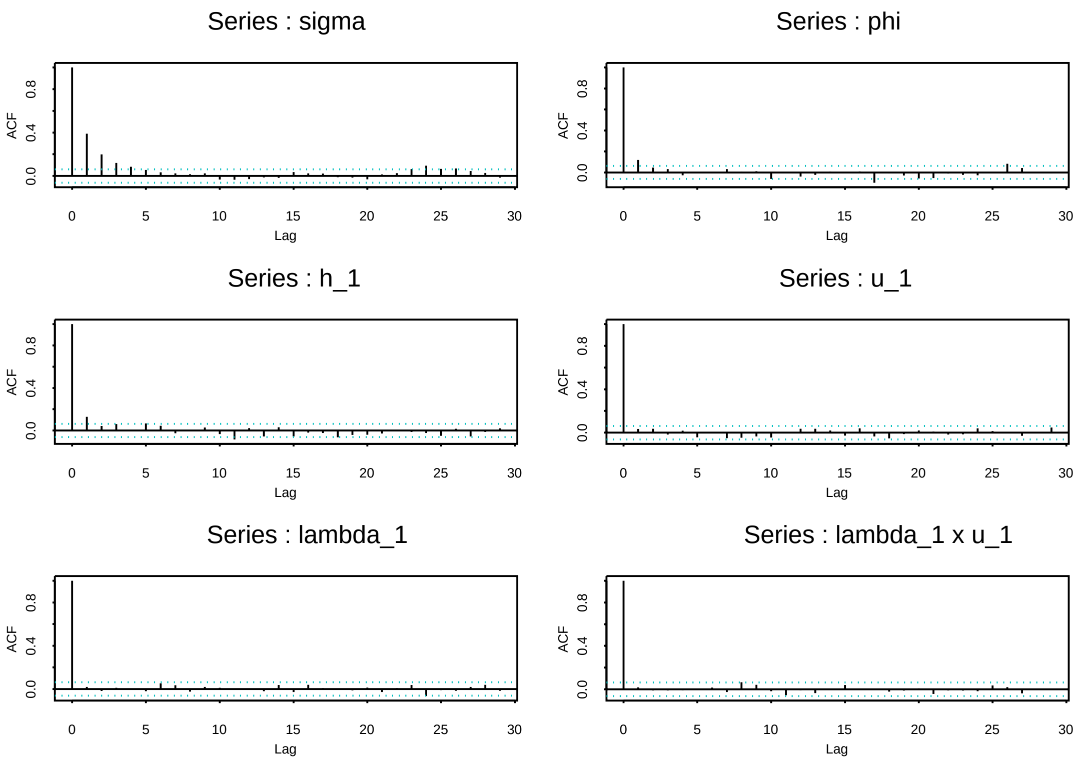


Fig.2. *Autocorrelation functions of 1000 simulated values of* $\sigma, \phi, h_1, \lambda_1, u_1, \lambda_i$ *and* $\lambda_1 u_1$.

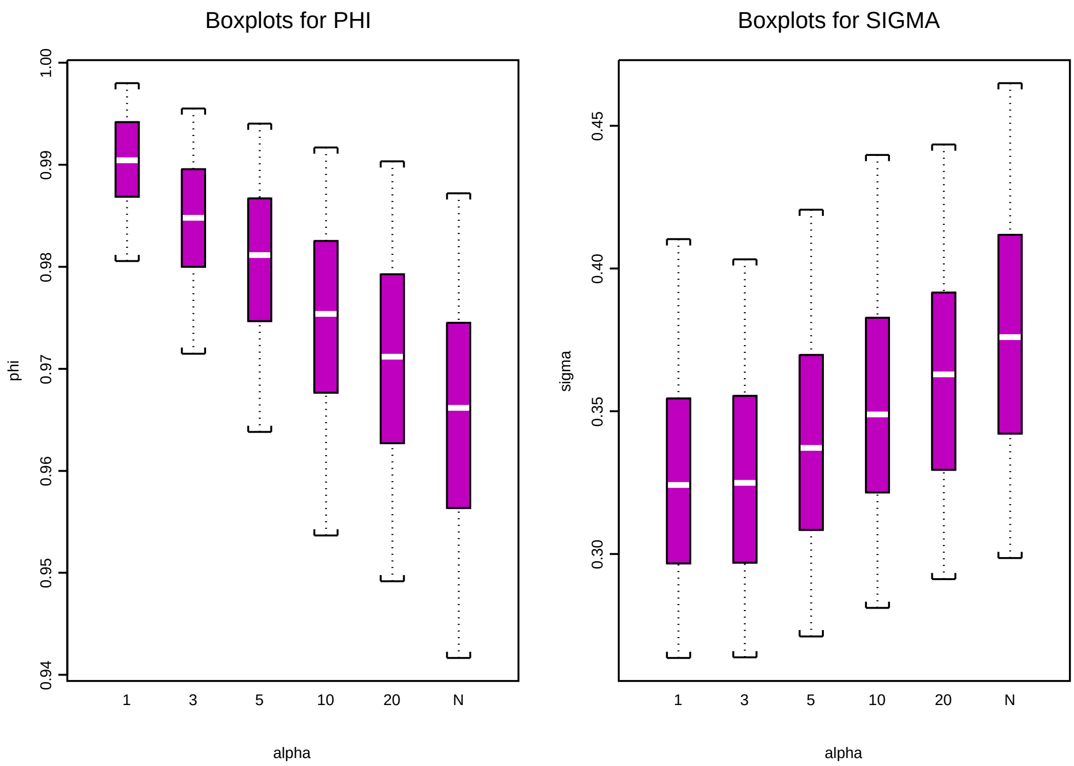


Fig.3. *Boxplots of σ and ϕ for different degrees of freedom α. The whiskers correspond to 5%, 25%, 50%, 75% and 95% of the values.*

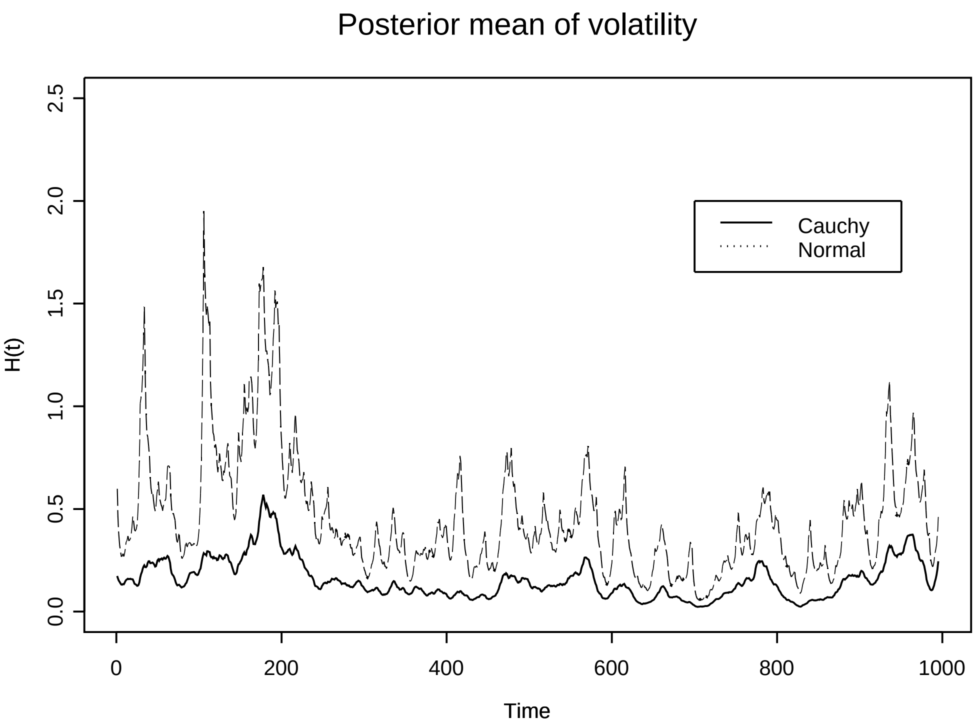


Fig.4. *Bayes estimates of volatility H_t for Cauchy-Normal and Normal-Normal SV models, respectively.*

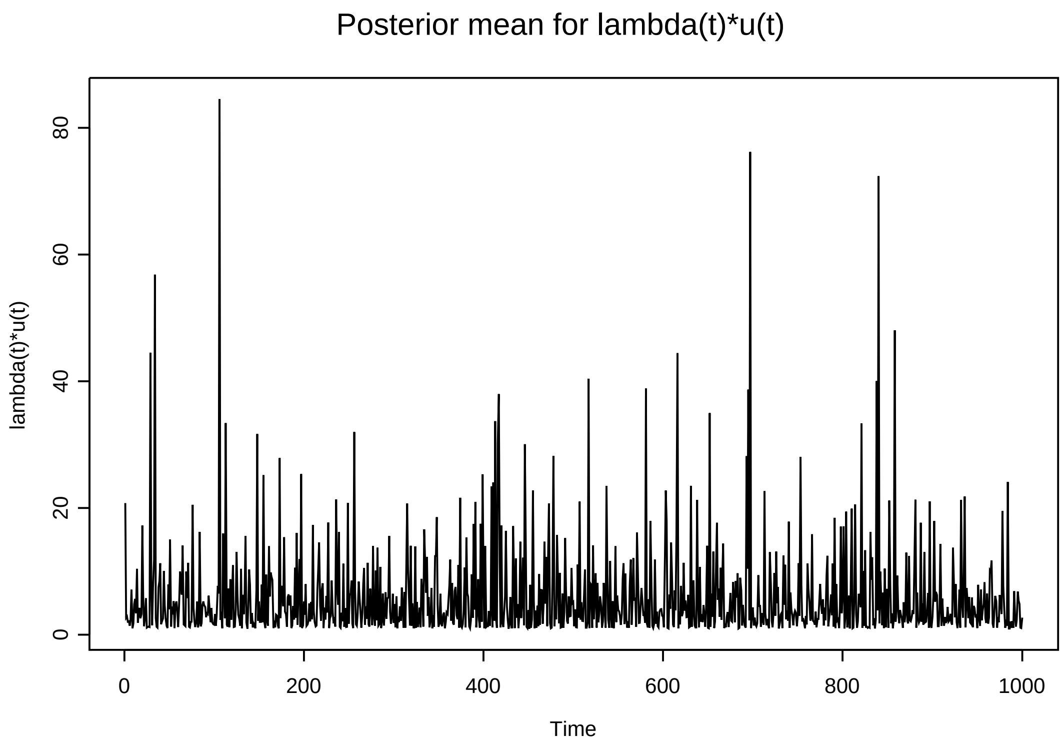


Fig.5. *Bayes estimates of the products of mixing parameters,* $\lambda_i\, u_i$ *for the Cauchy-Normal SV model. Large values correspond to extreme daily returns.*

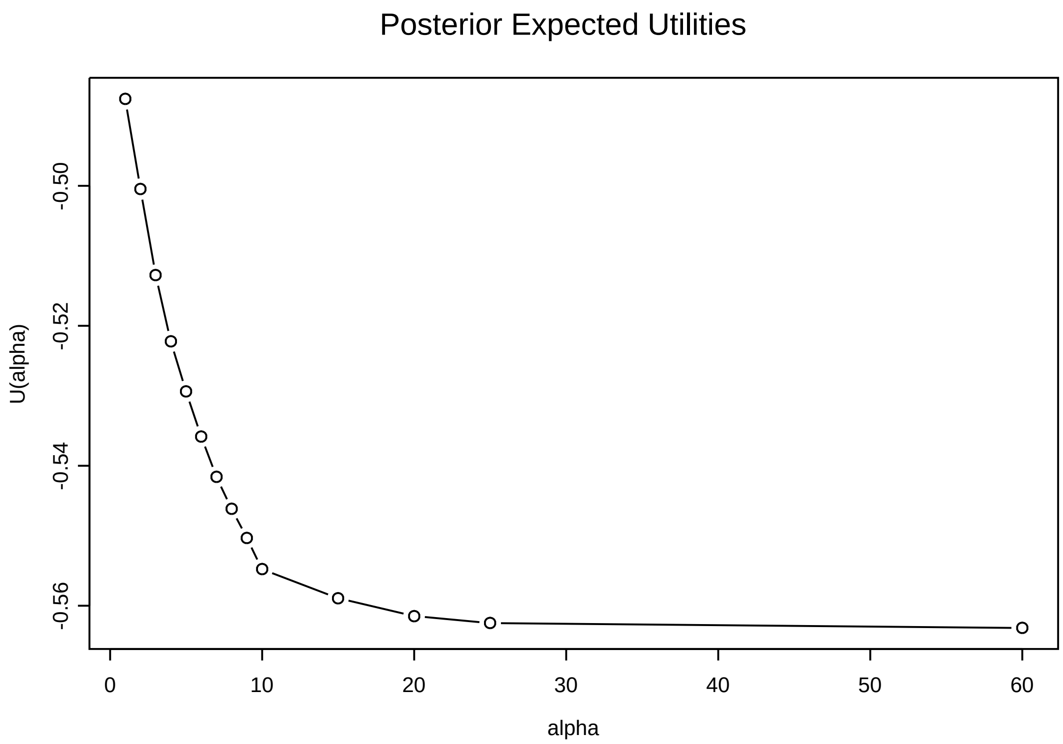


Fig.6. *Posterior expected utility* $U(\alpha)$ *of* $t - N$ *SV models with different degrees of freedom* α*.*