Bayesian Inference in the Non-central Student-t Model

Efthymios G. Tsionas

Department of Economics, Athens University of Economics and Business,
76 Patission Street, 104 34 Athens, Greece.
Tel. (++301) 820 3392, Fax: (++301) 823 8249 & 820 3301, Internet: tsionas@aueb.gr

ABSTRACT

The paper takes up Bayesian inference in linear models with disturbances from a non-central Student-t distribution. The distribution is useful when both long tails and asymmetry are features of the data. The distribution can be expressed as a location-scale mixture of normals with inverse weights distributed according to a chi square distribution. The computations are performed using Gibbs sampling with data augmentation. An empirical application to Standard and Poor's stock returns indicates that posterior odds strongly favor a non-central Student-t specification over its symmetric counterpart.

KEYWORDS**:** Markov Chain Monte Carlo; Gibbs sampling; skewness; fat tails.

1. INTRODUCTION

The symmetric Student-t distribution has a long history in statistics as a device for modeling data with outliers, see for example Lange, Little and Taylor (1989). Geweke (1993) uses Bayesian methods to perform inference in linear models with symmetric Student-t disturbances, and concludes that in possibly non-stationary macroeconomic time series the Student-t leads to a clear advantage over traditional normal specifications. The degree of long tails of the distribution is controlled by a single parameter, the degrees of freedom. In many cases, however, the data is not only heavy tailed but skewed as well. For example, stock market data containing severe crashes necessitate modeling both leptokurtosis and skewness. A straightforward way to accomplish this is to use a non-central t-distribution, where the non-centrality parameter controls the degree of skewness. An alternative, suggested by Fernandez and Steel (1998) and applicable to any symmetric distribution, involves assuming different scale parameters for positive and negative disturbances. Surprisingly enough, however, the idea of using *non-central* Student-t distributions to model asymmetric and leptokurtic behavior appears to be novel.

On the empirical side, applications of the non-central Student t-distribution have been limited by the fact that the probability density function is not expressible in closed form. It involves computation of complicated functions or infinite sums of incomplete beta functions. This makes maximum likelihood estimation a difficult task. Bayesian analysis organized around importance sampling techniques encounters more difficulties, since it requires fine-tuning an importance density to fit the posterior density. As a matter of fact, to the author's knowledge, Bayesian analysis of the non-central t-distribution has not been taken up so far in published research. Among the few empirical applications of the non-central t-distribution, Lahiri and Teigland (1987) and Dasgupta and Lahiri (1992) found the distribution useful in analyzing survey data and forecasting record data..

This paper takes up Bayesian analysis of a linear model with disturbances arising from a non-central t-distribution, with unknown degrees of freedom and unknown non-centrality parameter. It exploits the fact that the distribution can be expressed as a location-scale mixture of normals with inverse weights distributed according to the chi-square distribution. This establishes some similarity between the symmetric and non-central t-distributions, but Bayesian inference is more involved in the non-central case, because the distribution cannot be expressed simply as a *scale* mixture as in Geweke (1993). However, this does not pose a significant overhead because it is shown that conditional posterior distributions are in a form that facilitates synthetic random number generation. This motivates performing the computations using a Gibbs sampler with data augmentation.

Recent contributions in classical estimation of the non-centrality parameter appear in Chowdhury and Stedinger (1991) and Rukhin (1992). Contributions to computing the density function of the non-central t-distribution include Guenther (1975), Kuhlmeyer (1970), Lenth (1989), and Voit and Rust (1990). Approximations have been considered by Deutler (1984), Harley (1957), and Kraemer (1978). Detailed tables of percentiles appear in Bagui (1993). For computer programs regarding the computation of the cdf, see Cooper (1968) whose algorithm is included in Griffiths and Hill (1985) and Lenth (1989). See also Singh, Relya and Bartolucci (1992), and Posten (1993). Chattamvelli and Shanmugam (1994) have presented an algorithm that avoids the computation of incomplete beta functions.

The remainder of the paper is organized as follows. The model is presented in section 2. Conditional distributions necessary for Gibbs sampling are outlined in section 3. The stock market application and empirical results are presented in section 4. The final section concludes the paper.

2. THE MODEL

Let

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \sigma e_i, \quad i = 1,\ldots,n \tag{1}$$

be the linear model with n observations and k explanatory variables, and let e_i be a disturbance which arises from a non-central Student-t distribution. Here, σ is a positive scale parameter. The definition of a non-central Student-t distribution is as follows. If Z and w are independent random variables such that Z is standard normal, $\nu / w \sim \chi^2(\nu)$ and δ is a constant, then $u = w^{1/2}(Z+\delta)$ has the non-central Student-t distribution with ν degrees of freedom and non-centrality parameter δ. It should be mentioned that the "skewed Student-t" distribution is different than the one described in this paper. The former is presented in Fernandez and Steel (1998). The non-central Student-t distribution in this paper is the distribution of the classical one-sample t-statistic when the null hypothesis mean if off by an amount δ.

The probability density function of e can be expressed as

$$f_e(e) = \frac{\nu!}{2^{(\nu-1)/2}(\pi\nu)^{1/2}\Gamma(\nu/2)}\exp(-\frac{\nu\delta^2}{\nu+e^2})(\frac{\nu}{\nu+e^2})^{(\nu-1)/2}Hh_\nu(-\frac{\delta u}{(\nu+e^2)^{1/2}}) \tag{2}$$

where Airey's Hh function is defined by

$$Hh_\nu(x) = (\nu!)^{-1}\int_0^\infty u^\nu \exp[-(u+x^2)/2]du \tag{3}$$

Alternatively, the density can be expressed in terms of incomplete beta function ratios as follows:

$$\begin{aligned} f_e(e) &= \exp(-\delta^2/2)\Gamma((\nu+1)/2)(\pi\nu)^{-1/2}\Gamma(\nu/2)^{-1}(\nu/(\nu+2))^{-(\nu+1)/2} \\ &\sum_{j=0}^\infty \frac{\Gamma((\nu+j+1)/2)}{j!\Gamma((\nu+1)/2)}[e\delta\sqrt{2/(\nu+e^2}]^j \end{aligned} \tag{4}$$

Several approximations to the non-central Student-t cdf have been proposed in the literature, see for example Chattamvelli and Shanmugam (1994), Singh, Relyea and Bartolucci (1992) and Posten (1993) for recent contributions. For a review of older methods, see Johnson, Kotz and Balakrishnan (1995, volume 2, pp. 514-533. The non-centrality parameter of the t-distribution affects all central moments and the r*th* central moment can be expressed as a polynomial in δ^2. For large ν, the third moment is approximately $\mu_3 = \delta\nu^{-1}(3+1.25\delta^2\nu^{-1})$ and, therefore, skewness has the same sign as δ.

Let $\mathbf{y} = [y_1\, y_2\, \ldots\, y_n]'$ be the $n \times 1$ vector of observations for the dependent variable, $\mathbf{X}$ be the $n \times k$ matrix of observations for the explanatory variables. An equivalent specification of the linear model with non-central Student-t distributed disturbances, is

$$y_i \mid \mathbf{X}, \boldsymbol{\beta}, \sigma, \delta, \nu, \mathbf{w} \sim N(\mathbf{x}_i'\boldsymbol{\beta} + \delta\sigma w_i^{1/2}, \sigma^2 w_i)\,,\; i = 1,\ldots,n \tag{5}$$

This is because the disturbance can be expressed as $u_i = (\sigma^2 w_i)^{1/2}(Z_i + \delta)$ (conditional on w_i) where Z_i is standard normal, independent of w_i. The importance of non-centrality is that the model is no longer a scale mixture of normals as in the central Student-t case. Instead, the model is a location-scale mixture of normals, with the same weights w_i being used in the conditional mean and variance of the process.

The following prior is assumed:

$$\pi(\boldsymbol{\beta}, \sigma, \delta, \nu) \propto \sigma^{-1} \exp(-\omega\nu) \exp\{-\tfrac{1}{2}(\delta - \underline{\delta})^2 / \underline{\sigma}_\delta^2\}\,,\;\; \omega, \underline{\sigma}_\delta > 0\,,\; -\infty < \underline{\delta} < \infty \tag{6}$$

The prior for $\boldsymbol{\beta}$ is flat and improper, the prior for $\log(\sigma)$ is non-informative, and the prior on ν is exponential with mean ω^{-1}. As ω approaches zero, the prior imposes the assumption of normality on the error term of (1), as shown by a simple argument in Geweke (1993). The main reason for adopting this prior for the degrees of freedom parameter is that it is a conditionally conjugate prior, and facilitates computations while allowing for some flexibility in describing prior notions about this parameter. Geweke (1993) uses the same prior, so it is useful when conducting model comparisons. Finally, the prior on δ is normal with prior mean $\underline{\delta}$ and prior standard deviation $\underline{\sigma}_\delta$. As $\underline{\sigma}_\delta$ approaches zero, this prior becomes flat.

Using Bayes' theorem the kernel of the posterior distribution is

$$\pi(\boldsymbol{\beta},\sigma,\delta,\nu,\mathbf{w}\mid\mathbf{y},\mathbf{X}) \propto (\nu/2)^{n\nu/2}\Gamma(\nu/2)^{-n}\exp(-\omega\nu)\sigma^{-(n+1)}$$
$$\exp[-\tfrac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\delta\sigma\mathbf{w}^{1/2})'\mathbf{W}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\delta\sigma\mathbf{w}^{1/2})/\sigma^2] \tag{7}$$
$$\prod_{i=1}^{n} w_i^{-(\nu+3)/2}\exp(-\nu/2w_i)\exp\{-\tfrac{1}{2}(\delta-\underline{\delta})^2/\underline{\sigma}_\delta^2\}$$

where $\mathbf{W}=diag[w_i; i=1,..,n]$. Computations will be performed using Gibbs sampling, see Gelfand and Smith (1990), Smith and Roberts (1993) and Tierney (1994). Gibbs sampling requires drawing from the conditional distributions of parameters. These distributions and efficient ways to generate synthetic random samples from them are presented in the next section.

3. CONDITIONAL DISTRIBUTIONS

The conditional distribution of regression parameters is

$$\boldsymbol{\beta}\mid\sigma,\delta,\nu,\mathbf{w},\mathbf{y},\mathbf{X}\sim N_k([\mathbf{X}'\mathbf{W}^{-1}\mathbf{X}]^{-1}\mathbf{X}'\mathbf{W}^{-1}(\mathbf{y}-\delta\sigma\mathbf{w}^{1/2}),\sigma^2[\mathbf{X}'\mathbf{W}^{-1}\mathbf{X}]^{-1}) \tag{8}$$

where N_k denotes the k-variate normal distribution.

The kernel conditional distribution of the degrees of freedom is

$$\pi(\nu\mid\boldsymbol{\beta},\sigma,\delta,\mathbf{w},\mathbf{y},\mathbf{X}) \propto (\nu/2)^{n\nu/2}\Gamma(\nu/2)^{-n}\exp(-S\nu) \tag{9}$$

where $S=(1/2)\sum_{i=1}^{n}(\log w_i + w_i^{-1})+\omega$. This distribution is exactly the same as the posterior conditional kernel of degrees of freedom in the symmetric t-distribution model. Geweke (1993) has shown how efficient random number generation can be performed using optimal acceptance techniques.

The conditional posterior distribution of δ is in a particularly simple form. If we define $z_i=(y_i-\mathbf{x}_i'\boldsymbol{\beta})/(\sigma w_i^{1/2})$ and $\bar{z}=n^{-1}\sum_{i=1}^{n}z_i$ it follows that

$$\delta \mid \boldsymbol{\beta}, \sigma, \nu, \mathbf{w}, \mathbf{y}, \mathbf{X} \sim N_1\left(\frac{n\bar{z}\underline{\sigma}_\delta^2 + \underline{\delta}}{n\underline{\sigma}_\delta^2 + 1}, \frac{\underline{\sigma}_\delta^2}{n\underline{\sigma}_\delta^2 + 1}\right) \tag{10}$$

The conditional posterior distribution of heteroscedastic weights is given by

$$\pi(w_i \mid \boldsymbol{\beta}, \sigma, \delta, \nu, \mathbf{y}, \mathbf{X}) \propto w_i^{-(\nu+3)/2} \exp[-\frac{u_i^2/\sigma^2 + \nu}{2w_i} + \delta(u_i/\sigma)w_i^{-1/2}] \tag{11}$$

where $u_i = y_i - \mathbf{x}_i'\beta$. The conditional posterior distribution of σ satisfies

$$\pi(\sigma \mid \boldsymbol{\beta}, \delta, \nu, \mathbf{w}, \mathbf{y}, \mathbf{X}) \propto \sigma^{-(n+1)} \exp[-\tfrac{1}{2\sigma^2}\sum_{i=1}^{n} w_i^{-1}u_i^2 + \delta\sum_{i=1}^{n} w_i^{-1/2}u_i\sigma^{-1}] \tag{12}$$

These distributions are non-standard when the non-centrality parameter is different from zero. When $\delta = 0$ both w_i and σ^2 are conditionally inverted gamma, which leads to an extremely efficient Gibbs sampler for the symmetric Student-t distribution model as shown in Geweke (1993). Efficiency is preserved in the non-central case as well, but sampling methods are different because the conditional distributions are different. The sampling method is based on acceptance sampling from appropriate gamma distributions is detailed in Appendix 1.

4. EMPIRICAL APPLICATION

To illustrate the new methods, the model is applied to stock market data of daily returns for the Standard and Poor's index over the period 1/1/1995-7/30/1999, for a total of 1,156 observations. A time series plot of the data is provided in Figure 1, and some basic summary statistics are presented in Table 1. The skewness coefficient indicates that asymmetry is an important feature of this data set. Leptokurtosis is also a highly important feature since the coefficient of kurtosis is 9.214. These aspects of the data offer an initial indication that a non-central Student-t might provide a better description of the data compared to a symmetric Student-t specification.

If y_t denotes the logarithm of stock prices of date t, the following linear model is adopted:

$$\Delta y_t = \beta_0 + \sum_{i=1}^{m} \beta_i \Delta y_{t-i} + \beta_{m+1} y_{t-1} + \sigma\, e_t \tag{13}$$

where Δy_t is (100 times) the stock market return. This model specification is motivated by the literature on nonstationary time series, especially the formulation of the augmented Dickey and Fuller statistic and the literature on cointegration. Notice that if β_{m+1}<1 then stock returns are stationary. If $\beta_{m+1} \geq 1$ returns are non-stationary, and if, in particular, $\beta_{m+1} = 1$ they contain a unit root. The main focus of this section is not on time series behavior of returns but on properties of the disturbance term, u_t, which is *i.i.d* distributed according to a non-central Student t-distribution with degrees of freedom ν and non-centrality parameter δ. It must be mentioned, however, that based on daily returns for short time periods it would be extremely surprising if β_{m+1} turned out to be non-zero. For that reason, this parameter is set to zero.

Special attention is devoted to sensitivity of posterior inferences with respect to prior notions about the degrees of freedom. Four values of the prior parameter are considered, namely ω=1, 1/5, 1/10 and 1/50. Therefore, prior means of the degrees of freedom are 1, 5, 10 and 50. Notice that $\omega = 1/50$ almost imposes the assumption of normality. Regarding the non-centrality parameter, δ, a normal prior with mean zero and variance one is used. This implies that the parameter is between -2 and 2 with prior probability 95%. In view of the fact that the dependent variable is differences of stock returns, these values are quite reasonable, since we would expect the non-centrality parameter somewhere between -1 and 1 almost certainly. Posterior results with a flat prior on δ (in other words $\underline{\sigma}_\delta = \infty$) were very similar, and are not reported.

Gibbs sampling with data augmentation is based on 5,000 iterations with 1,000 iterations used in the burn-in period to mitigate start up effects. All computations were performed using GAUSS 386-i, version 3.01. The present model satisfies the simple conditions of Roberts and Smith (1993) which guarantee convergence to the posterior. Standard tests indicated that the samplers converge very quickly in less than 100 iterations. Graphs of the autocorrelation functions (acf) of basic parameters, ν and δ, in Figure 2 indicate that autocorrelation is not destructive for the Gibbs samplers. These samplers reach the equilibrium distribution quickly.

Posterior statistics of parameters are presented in Table 2. Posterior medians of the non-centrality parameter are close to -0.20 with small posterior semi-interquartile ranges. Almost the entire range of posterior variation in δ consists of negative values, since $\Pr(\delta \le 0 \mid \mathbf{y}, \mathbf{X})$ exceeds 0.90. Posterior medians of degrees of freedom are close to 4. The posterior probability $\Pr(\nu \le 2 \mid \mathbf{y}, \mathbf{X})$ was zero for all values of ω. These findings imply leptokurtic distributions for stock returns, for which moments higher than the fourth do not exist, and also imply significant negative asymmetry.

It also follows that posterior medians are not overly sensitive to values of ω. This fact emerges also from an inspection of Figures 3-5 that present posterior distributions of degrees of freedom (see Figure 3), and posterior distributions of the non-centrality parameter (see Figure 4). Predictive distributions were computed by utilizing the infinite sum representation of the non-central Student-t distribution in (4), and averaging with respect to the parameters in standard Rao-Blackwell fashion. The non-central Student-t densities were computed to precision at least 10^{-7}. The posterior predictive distribution for $\omega = 1$ and a kernel density estimate from the data are reported in Figure 5. Only the results for $\omega = 1$ are reported, since for other values of ω these distributions were virtually identical.

These posteriors and predictive distributions differ very little even though ω values differ significantly. This finding is at odds with previous results in Bayesian analysis of central Student-t models: For example, Geweke (1993) finds that marginal posteriors of degrees of freedom are sensitive to the value of ω, in US quarterly macroeconomic data. In other words, the data do not seem to update the degrees of freedom to a significant extent. Tsionas (1999) in a comparison of symmetric stable and Student-t models finds that the same holds true in monthly exchange rates. If this absence of sensitivity is not an idiosyncratic feature of the particular data set, it would seem that the non-central Student-t model has a desirable property that may facilitate its wider use. Part of the explanation is, of course, that posterior correlation between degrees of freedom and non-centrality parameter is not destructively large.

Inspection of Figure 5 indicates that predictive distributions are consistent with significant negative asymmetry in stock returns as mentioned before. A formal comparison between the non-central and central Student-t models can be facilitated by the use of Bayes factors. Since the second model is nested to the first, the Bayes factor in favor of central Student t-distribution is the Bayes factor in favor of the hypothesis $H_0 : \delta = 0$. The method of Verdinelli and Wasserman (1995) can be used to compute the Bayes factor. Since the conditional distribution of δ is normal, computation of the Bayes factor by this method is straightforward and involves computing

$$BF = \frac{\int \pi(\delta = 0 | \beta, \sigma, \nu, w, y, X) d\beta d\sigma d\nu dw}{\pi(\delta = 0)} \tag{14}$$

The numerator can be computed by averaging the normal conditional distribution of δ over Gibbs draws, in standard Rao-Blackwell manner. The denominator is simply the prior evaluated at zero. Bayes factors in favor of H_0 are reported in Table 2. It is apparent that the hypothesis receives very little support from the data, since Bayes factors range from 0.091 to 0.129. This finding implies that the non-central Student-t model is strongly favored over the symmetric Student-t counterpart. The Bayes factors were not sensitive to the adopted value of ω.

Although the hypothesis of symmetric Student-t receives almost no support from the Bayes factor, it is instructive to compare the symmetric and non-central models in terms of their predictive power. To that end, one needs to obtain the posterior predictive distribution of the data under a symmetric Student-t specification. This necessitates conducting formal inference for this model. Bayesian inference for the symmetric Student-t model can be accomplished by using either the Gibbs algorithm proposed in this study, subject to the constraint $\delta = 0$, or by using the more efficient Gibbs sampling algorithm presented in Geweke (1993). Bayesian inference has been conducted using the first alternative with 5,000 iterations, 1,000 of which are discarded to mitigate the impact of start up effects. The prior is exactly the same as in (6) with $\omega = 1$.

The posterior predictive distributions for the symmetric and non-central Student-t specifications are provided in Figure 6. Although the two distributions are fairly close, the symmetric Student-t predictive has a thinner negative tail and a fatter positive tail: The non-central Student-t specification implies higher probabilities of negative returns and captures the stock market crashes better. A more formal comparison can be based on the measures $K_S = \max_{r \in R} | f_S(r) - \hat{f}(r) |$ and $K_N = \max_{r \in R} | f_N(r) - \hat{f}(r) |$, where r denotes stock returns, R is a set of values for returns over which the posterior predictive distributions are computed, $f_S(r)$ is the posterior predictive distribution of the symmetric Student-t model, $f_N(r)$ denotes the posterior predictive distribution of the non-central Student-t model, and $\hat{f}(r)$ is the kernel density estimate. Since the data is given, and parameter uncertainty has been fully accounted for in the computation of posterior predictive distributions, from the Bayesian viewpoint these measures are fixed numbers, not statistics. So we do not have to examine their sampling behavior, or compare them with critical values. The ratio K_S / K_N was about 1.5 for all values of ω indicating a superior posterior predictive fit of the non-central Student-t model.

From the financial point of view, the ability of a model to explain tails and asymmetric behavior better, is an important property: Market funds are often interested not in predicting average returns *per se* but rather in predicting extreme returns in the positive or negative part of the return distribution. Therefore, a model that provides a better description of tails and asymmetric behavior would be a significant addition to the arsenal of market funds or stock market forecasters. The evidence provided in this paper points to the direction that the non-central Student-t model might be a prime candidate for such a model.

5. CONCLUSIONS AND DIRECTIONS FOR FUTURE RESEARCH

The paper presented Bayesian analysis of linear models with disturbances arising from a non-central Student t-distribution. The distribution can be expressed as a location-scale mixture of normal distributions, with inverse weights distributed according to a chi-square distribution. Computations are as straightforward as in the symmetric Student-t case, and efficient implementation of Gibbs sampling with data augmentation is feasible. The new method was illustrated using Standard and Poor's daily return data over the 1987-1988 period. The empirical results suggest that Gibbs sampling is robust, mixes well and posterior inferences are not sensitive to prior assumptions about the degrees of freedom parameter. Computed Bayes factors suggest that the symmetric Student-t model is strongly rejected in favor of the non-central counterpart, which allows for significant negative asymmetry in stock returns.

An important avenue of future research would be the empirical application of non-central Student-t distributions in other stock return or exchange rate data. Posterior predictive distributions may be used to conduct formal forecasting exercises, and compare the out-of sample forecasting ability of non-central versus symmetric Student-t distributions. Useful extensions of the model would include GARCH models with non-central Student-t innovations: Such models are capable of capturing three features of financial time series, namely conditional heteroscedasticity of autoregressive type, leptokurtic innovations (beyond leptokurtosis captured by GARCH) and asymmetric behavior. Evaluating the overall performance of such models would be an important empirical matter.

APPENDIX: GENERATING RANDOM NUMBERS FROM THE CONDITIONAL DISTRIBUTIONS OF w_i AND σ

The conditional distributions of $t_i = w_i^{-1/2}$ and $\phi = \sigma^{-1}$ are log-concave. Instead of using a general log-concave sampler, one can use gamma source densities. The conditional distribution of each t_i is

$$\pi_{t_i}(t \mid .) \propto t^{\nu} \exp(-\tfrac{q}{2} t^2 + rt)$$

where

$$q = u_i^2 / \sigma^2 \text{ and } r = \delta u_i / \sigma .$$

The conditional distribution of ϕ satisfies

$$\pi(\phi \,|\, .) \propto \phi^{n-1} \exp(-a / 2\phi^2 + b\phi)$$

where

$$a = \sum_{i=1}^{n} w_i^{-1} u_i^2 \text{ and } b = \delta \sum_{i=1}^{n} w_i^{-1/2} u_i .$$

The two distributions are members of the same family whose kernel is given by

$$f(x) \propto x^{N-1} \exp[-(A/2)x^2 + Bx].$$

Suppose the proposal density is $gamma(N, \theta)$, *i.e.*

$$g(x) \propto \theta^N x^{N-1} \exp(-\theta x) .$$

The problem of optimal selection of parameter θ is $\min_\theta \max_x : f(x) / g(x)$. It can be shown that the parameter is the unique positive root of the quadratic $\theta^2 + B\theta - AN = 0$. The optimizing value of x is $x^* = N / \theta^*$. Then, the sampling strategy is to draw a candidate x from a $gamma(N, \theta)$, and accept if $\dfrac{f(x) / g(x)}{f(x^*) / g(x^*)} \geq U$, where U is a standard uniform variate. Practical experience with the algorithm suggests that it is extremely efficient. Acceptance rates were typically near 90%.

Comparisons indicated that the Wild and Gilks (1993) sampler for log-concave densities is quite inefficient relative to the present algorithm. The reason is that the Wild and Gilks (1993) sampler is efficient provided a large number of draws is to be made from a single log-concave distribution. This is because there are set up costs involved in the algorithm, like computing the mode of the distribution from which random numbers are sought, computing suitable bounds which utilize the first derivatives of the logarithm of the distribution *etc*. In the non-central t-distribution linear model, we have to draw n values for the w_i's and one for σ. For large n, the setup costs accumulate and become prohibitive relative to the simpler gamma proposal density sampling scheme proposed here.

Since the exponential term of the target distribution $f(x)$ is the kernel of a normal distribution with mean BA^{-1} and variance A^{-1}, I have tried a normal proposal density as well. For $B > 0$ the optimized acceptance sampler works well, but it does not work for $B < 0$ because the second order conditions for solving the minimax problem are not satisfied. Even when it works, however, the normal proposal density scheme is not efficient because it requires solving a nonlinear equation to obtain the optimal normal distribution mean for each heteroscedastic weight. The nonlinear equation arises because the proposal has to be truncated below at zero and, therefore, the normal cdf is involved in the minimax problem.

Finally, following the suggestion of a reviewer, it should be mentioned that the entire sampling scheme is a kind of Metropolis-Hastings algorithm, see for example Tierney (1994).

REFERENCES

Chattamvelli, R. and Shanmugam, R. (1994), “An enhanced algorithm for non-central t-distribution”, *Journal of Statistical Computation and Simulation*, 49, 77-83.

Chowdhury, J.V. and Stedinger, J. R. (1991), “Confidence intervals for design floods with estimated skew coefficients”, *Journal of Hydraulic Engineering,* 177, 811-831.

Cooper, B.E. (1968), “Algorithm AS-5. The normal integral; the integral of Student’s t distribution; an auxiliary function for distribution integrals; the integral of the non-central t distribution”, *Applied Statistics*, 17, 186-194.

Dasgupta, S. and Lahiri, K. (1992), “A comparative study of alternative methods of quantifying qualitative survey responses using NAPM data”, *Journal of Business and Economic Statistics,* 10, 391-400.

Deutler, T. (1984), “A series expansion for the cumulants of the x-distribution and a Cornish-Fisher expansion for the non-centrality parameter of the non-central t-distribution”, *Communications in Statistics-Simulation and Computation,* 13, 507-513.

Fernandez, C., and Steel, M. F. J. (1998), “On Bayesian modelling of fat tails and skewness”, *Journal of the American Statistical Association,* 93, 359-371.

Gelfand, A.E., and Smith, A. F. M. (1990), “Sampling based approaches to calculating marginal densities”, *Journal of the American Statistical Association,* 85, 398-409.

Geweke, J. (1993), “Bayesian treatment of the independent Student-t linear model, *Journal of Applied Econometrics*, 8, S19-S40.

Griffiths, P. and Hill, I. D. (eds.) (1985), *Applied statistics algorithms*, Chichester: Ellis Horwood.

Guenther, W. C. (1975), “Evaluation of non-central distribution integrals with a desk calculator, Research paper 80 S-1975-538”, College of Commerce and Industry, University of Wyoming, Laramie.

Harley, B. I. (1957), “Relation between the distributions of non-central t and a transformed correlation coefficient”, *Biometrika*, 44, 219-224.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994), *Continuous univariate distributions*, volume 2, second edition, New York: Wiley.

Kraemer, H. C. (1978), “A central t-approximation to the non-central t-distribution”, *Communications in Statistics-Simulation and Computation*, 20, 23-31.

Kuhlmeyer, M. (1970), *Die Nichtzentrale t-verteilung*, Lecture Notes in Operations Research and Mathematical Systems, New York: Springer-Verlag.

Lahiri, K. and Teigland, C. (1987), "On the normality of probability distributions of inflation and GNP forecasts", *International Journal of Forecasting*, 3, 269-279.

Lange, K. L., Little, R. J. A., and Taylor, J. M. G. (1989), "Robust statistical modeling using the t distribution", *Journal of the American Statistical Association*, 84, 881-896.

Lenth, R. V. (1989), "Algorithm AS 243. Cumulative distribution function of the non-central t distribution", *Applied Statistics*, 38,185-189.

Posten, H. O. (1986), "Algorithms for the beta distribution function, in *Proceedings in Computational Statistics*, in F. de Antoni, N. Lauro and A. Rizzi (eds), Vienna: Physica, 309-319.

Roberts, C. O., and Smith, A. F. M. (1994), "Simple conditions for the convergence of the Gibbs sampler and Hastings-Metropolis algorithms", *Stochastic Processes and their Applications*, 49, 207-216.

Rukhin, A. L. (1992), "Estimating the non-centrality parameter of a t-distribution", *Systems Science and Mathematical Sciences*, 5, 1-8.

Singh, K.P., Relyea, G. E., and Bartolucci, A. A. (1992), "On the tail probabilities of the non-central t-distribution", *Computational Statistics*, 7, 67-80.

Smith, A. F. M., and Roberts, C. O. (1993), "Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo methods", *Journal of the Royal Statistical Society*, B55, 3-23.

Tierney, L. (1994), "Markov chains for exploring posterior distributions (with discussion)", *Annals of Statistics*, 22, 1701-1762.

Tsionas, E.G. (1999), "Monte Carlo inference in econometric models with symmetric stable disturbances", *Journal of Econometrics*, 88, 365-401.

Wild, P. and Gilks, W. R. (1995), "Algorithm AS 287. Adaptive rejection sampling from log- concave density functions", *Applied Statistics*, 42, 701-709.

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

Voit, E. O. and Rust, P. F. (1990), "Evaluation of the non-central t-distribution in S-systems", *Biometrical Journal*, 32, 681-695.

Table 1. Summary statistics for Standard and Poor's data

Mean	0.092
Standard deviation	0.994
Minimum	-7.112
Maximum	4.989
Skewness	-.579
Kurtosis	9.214

Notes: Returns are computed as 100 times the first difference of log stock prices.

Table 2. Posterior statistics

	Prior specification of degrees of freedom $(E(\nu) = \omega^{-1})$			
Parameter	$\omega = 1$	$\omega = 1/5$	$\omega = 1/10$	$\omega = 1/50$
β_0	.285 (.071)	.305 (.071)	.310 (.076)	.308 (.076)
β_1	-.0038 (.019)	-.0040 (.020)	-.0027 (.020)	-.0040 (.021)
β_2	-.054 (.019)	-.055 (.0109)	-.056 (.019)	-.056 (.020)
β_3	-.047 (.019)	-.049 (.020)	-.048 (.020)	-.048 (.019)
β_4	-.024 (.020)	-.023 (.019)	-.022 (.020)	-.022 (.020)
σ^2	.486 (.024)	.495 (.025)	.495 (.025)	.494 (.025)
δ	-.195 (.087)	-.219 (.089)	-.232 (.095)	-.227 (.093)
ν	3.766 (.307)	3.921 (.325)	3.922 (.330)	3.920 (.338)
$\Pr(\delta \le 0 \mid \mathbf{y}, \mathbf{X})$	.935	.962	.953	.938
BF	.129	.091	.103	.126

Notes: The table reports posterior medians for parameters. Posterior semi-interquartile ranges appear in parentheses. *BF* is the Bayes factor in favor of the symmetric Student-t specification, *i.e.* in favor of the hypothesis $\delta = 0$. $\Pr(\delta \le 0 \mid \mathbf{y}, \mathbf{X})$ denotes the posterior probability that parameter δ is non-positive.

Figure 1. Plot of Standard & Poor's returns (1995-1999)

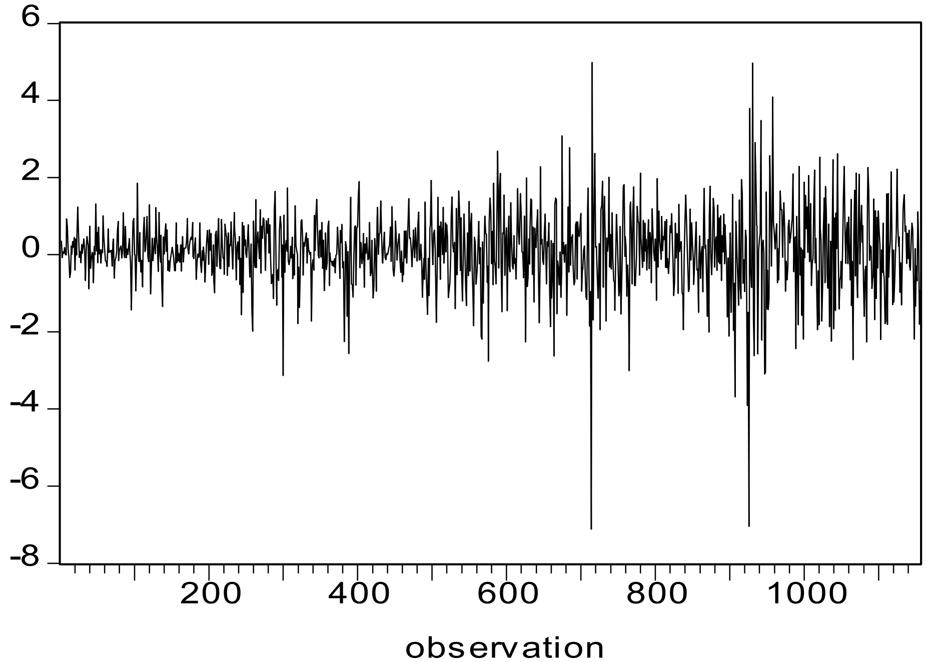

Figure 2. Autocorrelation functions

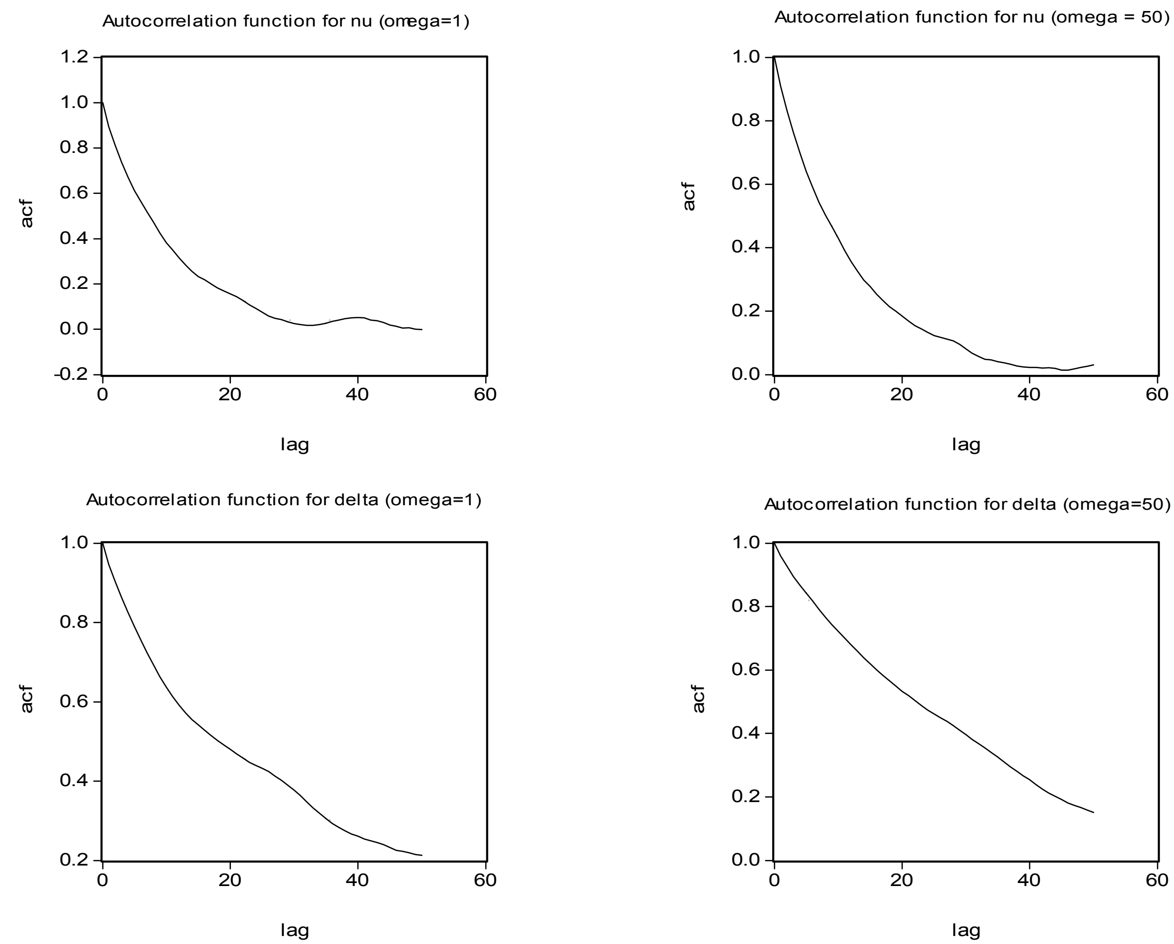

Figure 3. Posterior distibutions of degrees of freedom

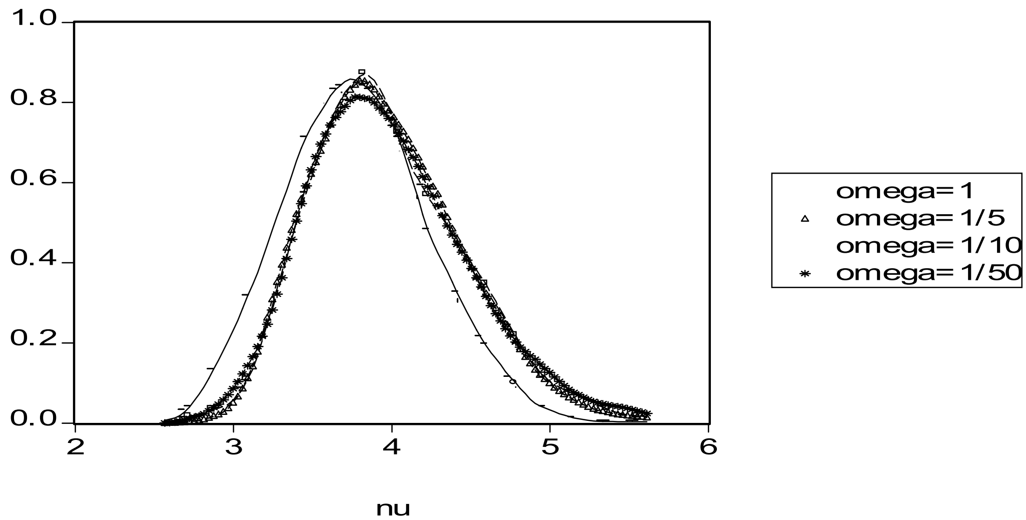

Figure 4. Posterior distributions of non-centrality parameter

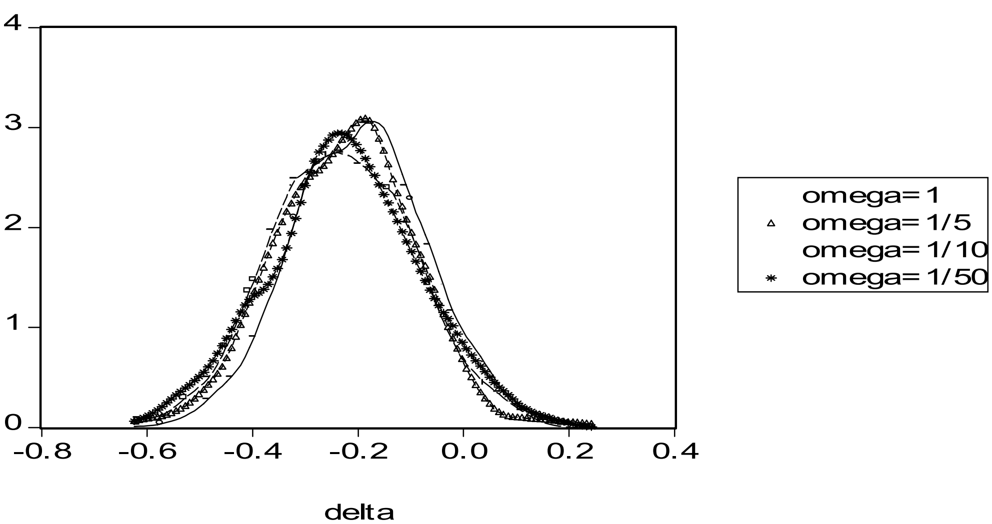

Figure 5. Posterior predictive distribution

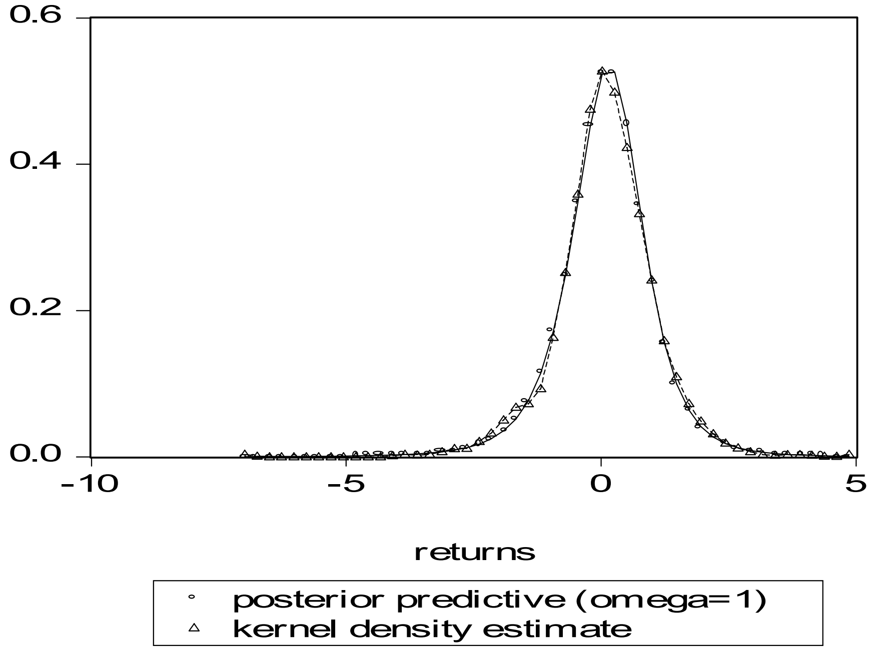

Figure 6. Log-predictive distributions of non-central and symmetric model