Sampling Schemes for Asymmetric Models a comparative study

Luis A. T. Medrano
IM - UFRJ
Helio S. Migon
IM / COPPE - UFRJ

June 14, 2007

Sampling Schemes for Asymmetric Models a comparative study

Abstract

This paper discusses numerical aspects of the Bayesian estimation of the parameters in a special class of asymmetric models, the normal-gamma stochastic frontier model, without restriction in the shape parameter. As is well known, some of the full conditional distributions are not available in closed form and besides that in same cases the distribution is not log-concave. Alternative MCMC schemes are compared in both artificial and real data. Our main finding is that slice sampling produces more reliable estimates for the parameters related to the stochastic frontier model, being a more realistic alternative to current proposals, which fail to produce good estimates for the quantities of interest when the distribution of the inefficiencies concentrates a significant portion of its probability mass far from the origin.

Keywords: Bayesian stochastic frontier, Slice sampling, Optimized Metropolis sampler, Stochastic simulation - MCMC.

1 Introduction

Current research in statistics is strongly driven by the development of new models to analyze asymmetric data. Skew-normal regression models are closely related to the econometrics of the stochastic production frontier. This paper takes into account the relationship among those research areas and points out some difficulties with estimating its parameters.

The focus of this paper is to compare alternative sampling schemes recently proposed in the literature to deal with the Bayesian inference in asymmetric models. In particular, we are interested in the asymmetric model obtained after mixing a normal error component with a gamma distribution to model the inefficiency term in a stochastic production frontier. Our proposal is to do inference in this model without imposing a restriction on the shape parameter of the gamma component. For instance, we emphasize the cases where the method proposed by Tsionas (2000) fails to give sensible results. We also compare the performance of the slice sampling algorithm (Neal, 2003) with the former to deal with the inference in the stochastic frontier model.

The broad class of asymmetric distributions introduced by Azzalini (1985) is in evidence in the current statistical literature. Particular attention has been devoted to the skew-normal

(SN) and to the skew t-Student (ST) cases. The most recent and important results related to the skew-normal models are presented in Azzalini (2005) and Genton (2004). Some numerical difficulties in estimating the skewness parameter are well documented in the literature. The likelihood function is not well behaved, which means that with positive probability the maximum likelihood estimator is not finite (see Liseo and Loperfido (2006) for details). The ST introduced by Branco and Dey (2001) and Azzalini and Capitanio (2003) is a natural extension of the SN and involves an extra parameter that controls the tail thickness of the distribution. This class of models allows one to control simultaneously the skewness and kurtosis of the data. The estimation of the degrees of freedom also demands some caution (Fonseca, Ferreira and Migon, 2005).

A very natural context for the use of skew-normal or skew-t distributions arises in the so-called *stochastic frontier (SF) models*, used to evaluate the efficiency of some economic agents, such as firms or countries. The basic idea is that the observed production of a single unit cannot exceed the (latent) potential production, which is named the *frontier*; the difference between the frontier and the actual production can be used as a measure of inefficiency. Consider, then, the model

$$\mathbf{y} = \mathbf{f}(X, \boldsymbol{\beta}) + \mathbf{e} \quad \text{with} \quad \mathbf{e} = \mathbf{v} - \mathbf{u} \tag{1}$$

where $\boldsymbol{\beta}$ is a $k \times 1$ vector of parameters, X is an $n \times k$ full column rank matrix of non-stochastic inputs, $\mathbf{y}$ is an $n \times 1$ vector of outputs, $\mathbf{v}$ is an $n \times 1$ random vector of disturbances, and $\mathbf{u}$ is an $n \times 1$ vector of inefficiency terms, with $u_i > 0 \; \forall \; i = 1, \cdots, n$ and $\mathbf{f}$ is a smooth function. Aigner et al. (1977) show that, when $\mathbf{v} \sim N_n(\mathbf{0}, \Sigma_v)$, multivariate normal distributed, and $\mathbf{u} \sim N_n^0(\mathbf{0}, \Sigma_u)$, multivariate normal distributed truncated at zero, then the difference $\mathbf{e} = \mathbf{v} - \mathbf{u}$ is exactly the density of the multivariate skew-normal distribution (MSN) proposed by Azzalini and Dalla Valle (1996) and Azzalini and Capitanio (1999). The connection between the SF models and the multivariate skew-normal has been discussed, among others, in Domínguez-Molina et al. (2004).

Therefore, stochastic frontier models can be represented as standard regression models with multivariate skew-normal errors; the extension to multivariate skew-t errors, with ν degrees of freedom, can be easily accommodated by assuming that the difference vector $\boldsymbol{e}$ in (1), is replaced by $\omega^{-1/2}(\mathbf{v} - \mathbf{u})$, where $\omega \sim Ga(\frac{\nu}{2}, \frac{\nu}{2})$ (Azzalini and Capitaneo, 2002), with $Ga(a, b)$ denoting the gamma distribution with shape parameter a and scale parameter b. The inference for the parameters of this model, based on the likelihood function, is described in Tancredi (2003), and a recent paper by Dey and Tchumtchua (2007) introduces its Bayesian counterpart.

The skewed compound error vector $\boldsymbol{e}$ is obtained after many different choices of the components. Assuming that $\mathbf{v}$ is multivariate normal and u is the vector with independent gamma distributed components leads to the normal-gamma stochastic frontier proposed by Greene (1990) and Beckers and Hammond (1987) as a natural extension of the original proposal of Aigner, Lovell, and Schmidt (1977), which provides a richer and more flexible parameterization of the inefficiency distribution than either of the canonical forms: normal-half normal or normal-exponential. Nevertheless, several attempts to operationalize this model have had limited success, as the log likelihood is complex. Greene (1990) attempted a direct, but crude maximization procedure which, as documented by Ritter and Simar (1997), was not sufficiently accurate to produce satisfactory estimates. Stevenson (1980) restricted his attention to the Erlang case, a very restrictive model, avoiding some difficulties in the estimation. Previous Bayesian work involving the normal-gamma model was presented by van der Broeck et al. (1994). In particular, these authors concentrated in the Erlang case, integer values for the form parameter, and implemented the inference using importance sampling. Koop, Steel and Osiewalski (1995) were the first to use the Gibbs sampler to analyze this model but did so assuming a known shape parameter.

The unrestricted normal-gamma stochastic frontier model presents some difficulties in the MCMC implementation. The full conditional distributions for same parameters are not in a known closed form, and in same cases, they are not log-concave. Tsionas (2000) proposed a rejection method to draw values from these distributions. Unfortunately his method does not work properly for shape parameters greater than one and also when the ratio of the inefficiency variance and the total variance of the model is relatively small, as he has shown empirically. More details will be presented in the Section 4.1.

The rest of the paper is organized as follows. The class of skew models is reviewed in Section 2. The normal-gamma stochastic frontier model is presented in Section 3, including prior specification and the full conditional distributions. Alternative methods of stochastic simulation to sample from the full conditional distributions when they are not available in closed form are also introduced in this section. In Section 4, some numerical illustrations based on artificial data are presented with the objective of evaluating the performance of alternative MCMC methods. An application based on the Greene's (1990) electricity data set is presented in Section 5. Finally, concluding remarks and some extensions are discussed in Section 6.

2 The Class of Skew Distributions

A general class of skew distributions (Azzalini and Capitanio, 2003) is defined as follows

$$f(\mathbf{e}) = 2g(\mathbf{e})Q(W(\mathbf{e})) \tag{2}$$

where $g(\mathbf{e})$ is the density function of an n-dimensional continuous random variable which is centrally symmetric around 0, Q is a scalar distribution function such that $Q(-\mathbf{e}) = 1-Q(\mathbf{e})$ for all $\mathbf{e}$ and $W(\mathbf{e})$ is a function from $\Re^n$ to $\Re$ such that $W(-\mathbf{e}) = W(\mathbf{e})$ for all $\mathbf{e} \in \Re^n$. A special case is obtained setting $g = \phi_n$ and $Q = \Phi_n$, respectively the standard multivariate normal density and distribution function, and $W(\mathbf{e}) = \boldsymbol{\lambda}^T\mathbf{e}$, which corresponds exactly to the multivariate skew normal distribution, denoted as $SN_n(0, I, \boldsymbol{\lambda})$.

Although the class of skewed distributions introduced before is very broad, it is common practice to begin with the more well-known members of the family. So the first choices are, of course, the multivariate SN and the multivariate skew-t. An alternative characterization of the multivariate skew-normal, is as a linear combination of two independent random variables. Let $u \sim N(0,1)$ and $\mathbf{v} \sim N_n(\mathbf{0}, I_n)$ be independent variables, and $\Delta = diag(\delta_1, ..., \delta_n)$ where $\delta_i \in (-1, 1)$. Then

$$\mathbf{e} = (I_n - \Delta^2)^{1/2}\mathbf{v} + \Delta\mathbf{1}_n|u| \tag{3}$$

has distribution $SN_n(\mathbf{0}, I_n, \Lambda)$, with $\Lambda = \left(\frac{\delta_1}{\sqrt{1-\delta_1^2}}, \cdots, \frac{\delta_n}{\sqrt{1-\delta_n^2}}\right)$. For instance, if $\delta_i = -\frac{\sigma_u}{\sqrt{\sigma_v^2+\sigma_u^2}}, \forall i = 1, \cdots, n$, then $\mathbf{e}$, in (1), is $SN_n(\mathbf{0}, \sqrt{\sigma_v^2 + \sigma_u^2}I_n, -(\sigma_u/\sigma_v)\mathbf{1}^T)$.

The skew factor depends on the ratio of the two scale parameters: the larger σ_u^2 is compared to σ_v^2, then the larger the probability is that u_i assumes large values and the larger the amount of skewness induced in the law of e_i.

It is worth pointing out that the SN frontier models are related to the hierarchical formulation of O'Hagan and Leonard (1976), introduced in the context of estimating a normal location parameter subject to an inequality constraint indicated by theoretical considerations. Let

$$\begin{aligned} \mathbf{y}|\mathbf{u}, &\sim N_n(X\boldsymbol{\beta} - \mathbf{u}, \Sigma_v) \\ \mathbf{u} &\sim N_n^0(\mathbf{0}, \Sigma_u) \end{aligned}$$

where $\Sigma_v = diag(\sigma_v^2, ..., \sigma_v^2)$ and $\Sigma_u = diag(\sigma_u^2, ..., \sigma_u^2)$. After some standard algebra it follows that the marginal distribution of $\mathbf{y}$ is skew-normal, that is:

$$\mathbf{y} \sim SN_n(X\boldsymbol{\beta}, (\sigma_v^2 + \sigma_u^2)I_n, -(\sigma_u/\sigma_v)\mathbf{1}^T) \tag{4}$$

This is our preferred form, to be used in the implementation of many alternative stochastic frontier models, as for example the normal-gamma studied in this paper.

3 Normal-Gamma Stochastic Frontier

In the Bayesian approach, if we assume that the components of the vector $\mathbf{u}$'s are unknown, the model in (1) can be written as follows

$$\begin{aligned} (\mathbf{y}|\boldsymbol{\beta}, \Sigma_v, \mathbf{u}) &\sim N(X\boldsymbol{\beta} - \mathbf{u}, \Sigma_v) \\ (\mathbf{u}|P, \theta) &\sim GaI(P, \theta) \end{aligned} \tag{5}$$

where GaI denotes the product of the gamma density function obtained under the supposition of independence and identical distribution of the component of the vector $\mathbf{u}$, and $\Sigma_v = diag(\sigma_v^2, ..., \sigma_v^2)$.

The density of $\mathbf{e}$ follows easily by solving $\int_{\Re^n} p(\mathbf{e}+\mathbf{u})p(\mathbf{u})d\mathbf{u}$:

$$p(\mathbf{e}) = \frac{\theta^{nP}}{\Gamma(P)^n} \exp(\theta \sum_{i=1}^{n} e_i + n\sigma_v^2\theta^2/2) \prod_{i=1}^{n} Pr[Q > 0|e_i]h[P-1, e_i] \tag{6}$$

where $h(r, e_i) = E[Q^r|Q > 0, e_i]$ and $Q|e_i \sim N[-(e_i + \theta\sigma_v^2), \sigma_v^2]$, as presented in Greene (1990). Although this expression is not in the form of equation (2), it corresponds to a skew distribution.

The prior distribution for the parameters $(\boldsymbol{\beta}, \sigma_\nu^2, P, \theta)$ is assumed to be independent:

$$p(\boldsymbol{\beta}, \sigma_\nu^2, P, \theta) = p(\boldsymbol{\beta})p(\sigma_\nu^2)p(P)p(\theta) \tag{7}$$

where $\boldsymbol{\beta} \sim N_k(m, C)$, $\sigma^{-2} \sim Ga\,(n_0/2, n_0\sigma_0^2/2)$, $P \sim Ga\,(d_0, \varepsilon_0)$ and $\theta \sim Ga\,(\upsilon_0, \omega_0)$. The hyperparameters were chosen in such away that these distributions were vague. For instance, we set $m = 0$ and $c_{i,i} \to \infty$ in the prior for the elasticities β´s.

3.1 MCMC Methods in Normal-Gamma Frontier

The full conditional posterior distributions for the parameters $\boldsymbol{\beta}$, σ^2 and θ have known form, as shown in Appendix 1, then Gibbs sampler is used. The same is not true for the other parameters: P and $u_i, i = 1, \cdots, n$. As can easily be verified, the full conditional distribution of u_i is not log-concave, $\frac{\partial^2}{\partial^2 u_i} ln(p(u_i)) = \frac{1-P}{u_i^2} > 0$, if $P < 1$. To contemplate this case, an optimized acceptance/rejection method (optimized Metropolis) was, recently, introduced in

the literature, while for the log-concave case, $P > 1$, the ratio-of-uniform method (Ripley, 1988) is advocated. Finally, an exponential prior is defined for P, then the full conditional distribution of P is log-concave and the optimized Metropolis method is applied.

We propose in this paper the use of slice sampling independently of the nature of the full conditional distribution. In what follows we briefly discuss the methods we will be comparing.

- **Optimized Metropolis**

  Let $p(x)$ be the density we want to draw and let $g(x, \gamma)$ be a proposed distribution for x. Define $R(x, \gamma) = p(x)/g(x, \gamma)$ and $r(x, \gamma) = \log R(x, \gamma)$, where the parameter λ is chosen to maximize $R(x, \gamma)$, that is, to find the $max_x\ min_\gamma\{r(x, \gamma)\}$.

  Given the optimal value of γ, the algorithm follows two steps:

  1. Draw x^{new} from $g(x, \gamma)$.
  2. Accept x^{new} with probability $\alpha(x^{new}) = R(x^{new})/R(x^*)$ where x^* is the optimum value of x, that is, x^{new} will be accepted if $\alpha(x^{new}) \geq U$, where U is uniformly distributed in $(0, 1)$.

  The proposal distributions suggested in the literature are the exponential and a gamma distributions, respectively, to P and u_i. More details on these procedures can be found in the work previously cited.

- **Slice Sampling**

  Let us suppose that we want to draw from the density function $p(x) = g(x)/c$ where $x \in X \subseteq \Re$. The sliced sampling algorithm introduces an auxiliary variable y and defines the joint distribution of x and y as uniform over the region $U = \{(x, y) : 0 < y < g(x)\}$. Then the joint density of (x, y) is

  $$p(x, y) \quad = \quad \frac{1}{c} I_{(0, g(x))}(y) \tag{8}$$

  Consequently, the marginal density of x is $p(x) = \int_0^{g(x)} (1/c) dy = g(x)/c$.

  In order to draw x, we need to jointly sample from and ignore the second component - y. Sometimes it is not easy to draw the elements of (x, y), jointly sampling uniformly from U. An alternative scheme is Gibbs sampling, which demands the following conditional distributions:

- $(y|x) \sim U(0, g(x))$
- $(x|y) \sim U(S(y))$ where $S(y) = \{x : y < g(x)\}$

However, drawing uniformly one value of x in $S(y)$ could still be difficult. Alternatively, Neal (2003) proposes to sample one value of x following these steps:

(a) Draw a value of y from the uniform distribution in $(0, g(x_0))$, where x_0 is an initial value. Then a horizontal slice is defined as: $S(y) = \{x : y < g(x)\}$.

(b) Find an interval, $I = (L, R)$, around x_0 containing the biggest part of the slice.

This step can be implemented in many different ways. The scheme involves defining an interval of width w centered around x_0 and then increasing this interval with increments of size w until its extremes fall out of the slice.

(c) The final step consists of drawing a new point x_1, from the part of the slice included in the interval, that is $S \cap I$. So, a new point x_1 is sampled from an U(L,R) and this proposed value is accepted if $y < g(x_1)$, otherwise if $x_1 < x_0$ then $L = x_1$ or $R = x_1$.

The main advantages of the slice sampling procedure are:

- it demands less implementation effort; and in particular
- it takes less time to tune it to obtain a satisfactory performance than with other commonly used methods, like the oft-employed Metropolis-Hastings algorithm.

4 Numerical Study

In this section we present two exercises with the objective of evaluating the performance of alternative MCMC methods. The first is based on two artificial samples, one corresponding to $P < 1$ - not log concave and the other one to the log-concave case, that is $P \geq 1$. The second exercise shows a Monte Carlo study where several simulated datasets are generated with the objective of evaluating the inefficiency of the MCMC schemes.

4.1 Illustration with Artificial Data

Our aim is to evaluate the performance of the methods introduced in Section 3.1, via artificially generated data.

Two datasets were generated from the model in eqn. 5, using an S-plus routine, with $N = 100$ and $N = 500$ elements. We set $k = 3$ regressors with equal coefficients, that is: $\beta = (1, 1, 1)$. The design matrix X is obtained as: the first column is a vector of 1's corresponding to the intercept. The second and third columns were generated from $N(0, 1)$ for each of the N firms involved. With respect to P and θ, two cases will be considered. In the first, **case 1**, we set $(P, \theta) = (0.8, 1)$, corresponding to the case where the full conditional distribution of u is not log-concave. The second one, **case 2**, corresponds to set $(P, \theta) = (2, 1)$. In this case, the estimates obtained by Tsionas (2000) were not very precise. In both case cases, $\sigma^2 = 0.05$ was fixed. These values where chosen mimicking this reference.

We are mainly interested in three functions of the parameters: the mean inefficiency defined as $m = P/\theta$; the inefficiency standard deviation, given as $S = (P/\theta^2)^{1/2}$; and the frontier variance proportion, given by $\zeta = \frac{P/\theta^2}{P/\theta^2+\sigma^2}$.

In this empirical exercise, we have used 15,000 iterations for the MCMC algorithm, leaving out the first 5,000 iterations, corresponding to the burn-in period. All the results presented below are based on the samples after reaching convergence (Brooks and Gelman, 1998). It is worth pointing out that the optmized Metropolis or slice sampling are used to sample the parameters P and $u_i, i = 1, \cdots, n$ and, for the others, a Gibbs sampler is applied.

The main results for **case 1** using optimized Metropolis or slice sampling for a sample of 100 units are showed in Table 1, where we present the posterior mean, the standard deviation, the median and the 2.5 and 97.5 percentiles for all the parameters of the model. We observe that for both methods, the posterior means are reasonably close to the true value. All the true values belong to the 95% credibility interval.

*** Insert Table 1 here ***

For the data set with sample size $N = 500$, although the results are not reported here, both algorithms again performed quite well. The parameters posterior densities are more concentrated around the true parameter value than in the case of $N = 100$. This is not a surprising, since the posterior standard deviation must decreases with the sample size.

The marginal posterior densities of P and θ for both methods are presented in Figure 1 for the case 2 and $N = 100$. Clearly, the results obtained with the slice sampling method are quite impressive, since the densities for almost all the parameters are strongly concentrated around the true parameter value (the vertical solid line). Nevertheless, the same is not true with $P > 1$. Using the proposal of Tsionas (2000), the marginal posterior density for P is asymmetric to the right.

*** Insert Figure 1 here ***

Finally, for case 2 with $N = 500$, both algorithms produce reasonable results. The marginal posterior densities for the parameters P and θ are shown in Figure 2, where we see that the densities are centered around the true value of the parameters.

*** Insert Figure 2 here ***

4.2 Simulation Evaluation

The sampling schemes presented in Section 3.1 gave rise to chains with different convergence properties. Theoretical characteristics, like their autocorrelation or the convergence ratio can ideally be used to compare those chains.

The autocorrelation is the key ingredient to compare the efficiency of the Monte Carlo estimators, of any quantity of interest $h(\theta)$, obtained after the generated data. Its variance $\bar{h}_M = (1/M)\sum_{i=1}^{M} h(\theta^{(i)})$ is given by $Var[\bar{h}_M] = (\sigma_h^2/M)i$, where σ_h^2 is the posterior variance of $h(\theta)$ and i is the chain inefficiency factor, given as:

$$i = 1 + 2\sum_{k=1}^{M-1} \frac{(M-k)}{M}\rho_k \tag{9}$$

where ρ_i is the lag k autocorrelation of the sampled values of $h(\theta)$. Therefore, the i factor evaluates the sampling inefficiency of a MCMC scheme.

The simulation study developed in this section includes four alternative settings, differing in the sample size N (100 or 500) and on the value of the parameter P of the gamma component: 0.8 and 2. These values correspond just to the cases considered before in the literature and also used in the former section. The other parameter of the gamma component is fixed as $\theta = 1$.

For each alternative setting, 100 replications were generated and analyzed via the two alternative sampling schemes discussed in this paper. Based on these sampling replications, envelopes for the autocorrelation function and for the inefficiency factor were built up.

We run a total of $15,000$ iterations for each chain. For all the chains, the convergence for all the parameters of interest occurred after the first 5,000 iterations, although in some cases it happened almost immediately. We developed the computer code in Fortran and used an AMD Athlon XP 2300, 1.8GHz, 512MB RAM to run it. The computational times spent by each method are shown in Table 2.

Notice that for N small and any value of P, the optimized Metropolis method takes 1.5 times longer than the slice sampling method. For $N = 500$, the computational time is 2.2 times longer than for the optimized Metropolis method, this is because this method requests an optimization of each interaction.

*** Insert Table 2 here ***

Figures 3 and 4 show the autocorrelation functions for the chains of values of P, θ. For each of the 4 combinations of values N and P, 100 estimates of the autocorrelation functions are obtained, and thus, mean and envelopes can be constructed. For P and θ, the optimized Metropolis method shows a slow decay in the autocorrelation for the generated chains, reflecting the strong dependence between the parameters previously mentioned. The slice sampling method, in turn, presents a very fast decay in those autocorrelations. For example, for the case $P = 2$ and $N = 100$ (resp. $N = 500$), the autocorrelation for the chain generated with the slice method is about 0.50 (resp. 0.60) at lag 20, while the optimized Metropolis reaches the same value only at lag 100.

*** Insert Figure 3 here ***

*** Insert Figure 4 here ***

These results can be summarized by seeing them as inefficiencies aggregated into the figures referred to in the previous section. For each value of P and each scheme, 100 MCMC inefficiency values for the parameters θ, P and σ^2 are obtained, one for each replication. These can be summarized as in the box-plots of Figures 5 and 6. For all schemes considered, the inefficiencies becomes larger as the length of the sample gets larger and as P gets smaller. The figures show that the slice sampling method appears to be the most efficient scheme for P, θ, β_2 and σ_v^2. As can be seen the central boxes do not overlap for almost all the simulated alternatives and all the parameters examined.

*** Insert Figure 5 here ***

*** Insert Figure 6 here ***

For each MCMC replication, we save the values of $\boldsymbol{u}$ only for the best and the worst firm (DM unit). Remember that we known precisely who they are, since they were artificially generated. During the MCMC iterations, we evaluate the autocorrelation function for lags

up to 100 and summarize these results using the MCMC inefficiency statistic defined by eqn. (9). The above-mentioned procedure is replicated M times, providing a sample from the MCMC inefficiency measure. We denote by u_{min} and u_{max}, respectively, the MCMC inefficiency statistics calculated over the $\boldsymbol{u}$´s for each of the two firms mentioned above. These values are plotted in the box-plots presented in Fig. 7, for the cases $N = 100, 500$ and $P = 0.8, 2$. Again, for almost all the simulation alternatives, the median inefficiencies obtained using the slice method are slightly better.

*** Insert Figure 7 here ***

5 An Example using the Greene Electricity Data

In this section we present an illustration using a real data set previously explored in the literature. The dataset considered by Greene (1990) includes 123 electricity firms in the United States for 1970. These data were also analyzed by Van den Broeck et al. (1994).

Greene (1990) fitted a cost function derived from the Cobb-Douglas production function, but generalized it to include a quadratic term (log of the output Q), allowing the return of scale to vary with Q. In this data set, we have three input factors: labor, capital and fuel, with the respective prices denoted by P_l, P_k and P_f; The cost function is specified as:

$$\begin{aligned} y_i = & - \beta_0 - \beta_1 ln Q_i - \beta_2 ln^2 Q_i - \beta_3 ln(P_l/P_f) \\ & - \beta_4 ln(P_k/P_f) + \upsilon_i - u_i, \end{aligned} \tag{10}$$

where y_i= - ln(cost of the ith firm / P_f)

The aim of this application is to show how well the slice sampling algorithm performs on real data. Furthermore, we use this application to reinforce the advantages of the Bayesian approach, which allows obtaining the posterior density probability function of the inefficiencies for each firm $r_i = exp(-u_i)$.

All the results presented are based on the final 10,000 iterations out of the 20,000, in this way avoiding the presence of transients. In Table 3 the posterior mean, median, standard deviation and the 2.5 and 97.5 percentiles for all the parameters obtained using slice sampling are shown, together with the results previously obtained by Greene (1990) and Van den Broeck et al. (1994).

*** Insert Table 3 here ***

Note from Table 3 that the results are not that different when compared with those obtained by Greene (1990) and Van den Broeck et al. (1994). Figure 8 shows the posterior marginal densities for the parameters θ, P, σ^2, β_1 obtained via the slice sampling algorithm.

*** Insert Figure 8 here ***

The posterior moments for the individual efficiencies r_i $(i = 1, \ldots, 5)$ for the first five firms, are shown in Table 4, together with the point estimates of van den Broeck et al. (1994).

Comparing the figures in Table 4 with those obtained by Greene (1990) and van den Broeck et al. (1994), one notes some differences in the estimates for the parameters describing the efficiencies. For example, it is easy to note that the efficiencies for the first five firms in the dataset are smaller when obtained via slice sampling than those obtained by Greene (1990) and van den Broeck et al. (1994). We can also note that the ranking of these five firms obtained using the slice sampling method and that of van den Broeck et al. (1994) are almost the same.

*** Insert Table 4 here ***

6 Conclusions and Extensions

In this paper we discuss some numerical aspects involved with the normal-gamma stochastic frontier model and present two alternative strategies to deal with the stochastic simulation procedure.

Based on the results obtained with both artificial and real data sets, we conclude that the slice sampling algorithm performed better to draw samples from the non log-concave full conditional posterior distribution, in the normal-gamma model. Besides being more efficient, it demands less computational time.

The first exercise clearly demonstrate that the optimized Metropolis algorithm does not perform well for several cases.

Since the error vector component $\mathbf{u}$ is an N dimensional parameter, the implementation of the algorithm proposed by Tsionas (2000) needs to solve N optimization problems in each Gibbs sampling step, imposing a high computational cost for its implementation.

References

[1] Aigner, D., Lowell, C., Schmidt, P., 1977. Formulation and Estimation of Stochastic Frontier Production Function Models, Journal of Econometrics, 6, 21-37.

[2] Azzalini, A., 1985. A Class of distributions which includes the normal ones. Scandinavian Journal of Statistics 12, 171-178.

[3] Azzalini, A., Capitanio, A., 1999. Statistical applications of the multivariate skew normal distribution. J. Roy. Statist. Soc. B 61, 579-602.

[4] Azzalini, A., Capitanio, A., 2003. Distributions generated by perturbations of symmetry with emphasis on a multivariate skew t distribution. Journal of the Royal Statistical Society, Series B 65, 367-389.

[5] Azzalini, A., Dalla-Valle, A., 1996. The multivariate skew-normal distribution. Biometrika 83, 715-726.

[6] Azzalini, A., 2005. The Skew-normal Distribution and Related Multivariate Families. Scandinavian Journal of Statistics, 32 159-188.

[7] Beckers, D., Hammond, C., 1987. A Tractable Likelihood Function for the Normal-Gamma Stochastic Frontier Model. Economics Letters 24, 33-38.

[8] Branco, M. D., Dey, D. K., 2001. A general class of multivariate skew-elliptical distributions. Journal of Multivariate Analysis 79, 99-113.

[9] van den Broeck, J., Koop, G., Osiewalski, J., Steel, M.F.J., 1994. Stochastic frontier models: A Baysesian perspective. Journal of Econometrics 61, 273-303.

[10] Brooks, S.P., Gelman, A., 1998. Alternative methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics 7, 434-455.

[11] Dey, D. K., Tchumtchua, S., 2007. Bayesian Estimation of Stochastic Frontier Models with Multivariate Skew t Error Terms. Tech Report Department of Statistics, University of Connecticut.

[12] Domínguez-Molina, J. A., González-Farías, G., Ramos-Quiroga, R., 2004. Skew-normality in stochastic frontier analysis. In Skew-elliptical distributions and their applications (ed. M. G. Genton), chapter 13, 223-242. Chapman & Hall / CRC, London.

[13] Genton, M.G., 2004. Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality. CRC Boca Raton:: Chapman & Hall.

[14] Gilks W., 1992. Derivative-free adaptive rejection sampling for Gibbs sampling. Appl. Statist. 41, 337-348.

[15] Greene, W. H., 1990. A Gamma Distributed Stochastic Frontier Model. Journal of Econometrics 46, 141-164.

[16] Koop, G., Steel, M. F. J., Osiewalski, J., 1995. Posterior Analysis of Stochastic Frontier Models Using gibbs Sampling. Computational Statistics 10, 353-373.

[17] Liseo, B., Loperfido N., 2006. Default Bayesian analysis of the skew-normal distribution. Journal of Statistical Planning and Inference 136, 373-389.

[18] Neal, R. M., 2003. Slice sampling. Annals of Statistics 31, 705-767.

[19] O'Hagan, A., Leonard, T., 1976. Bayes estimation subject to uncertainty about parameter constraints. Biometrika 63, 201-202.

[20] Osiewalski, J., Steel, M. F. J., 1998. Numerical Tools for the Bayesian Analysis of Stochastic Frontier Models. Journal of Productivity Analysis 10, 103-117.

[21] Ritter, C., Simar, L., 1997. Pitfalls of Normal-Gamma Stochastoc Frontier Models. Journal of Productivity Analysis 8, 167-182.

[22] Stevenson, R., 1980. Likelihood Functions for Generalized Stochastic Frontier Estimation. Journal of Econometrics 13, 57-66.

[23] Tancredi, A., 2002. Accounting for Heavy Tails in Stochastic Frontier Models. Working paper. Dipartiment di Scienze Statistiche, Università di Padova, n. 2002.16.

[24] Tsionas, E.G., 2000. Full Likelihood Inference in Normal-Gamma Stochastic Models. Journal of Productivity Analysis 13, 183-205.

Appendix - Full Conditional Distribution of the Parameters

In order to implement the MCMC procedure, the full conditional distributions of the parameters must be evaluated. These distributions are given as:

- $\beta|y, X, \sigma^2, P, \theta, u \sim N_k\left((X'X)^{-1}[X'(y+u)], \sigma^2(X'X)^{-1}\right)$
- $\sigma^2|y, X, \beta, P, \theta, u \sim GI\left(\frac{(N+n_0)}{2}, \frac{((y+u-X\beta)'(y+u-X\beta)+a_0)}{2}\right)$
- $\theta|y, X, \beta, \sigma^2, P, u \sim G\left(NP + v_0, \sum_{i=1}^{N} u_i + \omega_0\right)$

  As can be seen, the above full conditional distributions are available for sampling, since they have known form. In contrast, the full conditional distribution for the parameters P and u_i's are not in closed form.

- $\pi(P|Y, X, \beta, \sigma^2, \theta) \propto P^{d_0-1}\Gamma(P)^{-N}\exp[(\sum_{i=1}^{N}\log u_i - \varepsilon + N\log\theta)P]$
- Para $i = 1, \ldots, N$, $\pi(u_i|y, X, \beta, \sigma^2, \theta, P) \propto u_i^{P-1}\exp\left[\frac{-(u_i+\theta\sigma^2+y_i-x_i'\beta)^2}{2\sigma^2}\right]$

  It is worth noting that the distribution of P is not log-concave for $d_0 < 1$, and for $P < 1$ the full conditional distribution of u_i is not log-concave.

  There is a natural hierarchy in the application of the following alternative sampling schemes. If the full conditionals are not available for sampling but if they are log-concave, we can apply the adaptive sampling procedure of Gilks et al. (1992). Moreover if log concavity fails, slice sampling (Neal, 2003) can be used because the full conditional distribution is defined in a limited interval. A third alternative consists of using Metropolis-Hastings. For instance, the strategy used in WinBugs is to generate u_i and P through the slice sampling method.

Table 1: Posteriori distribution summary for all the parameters of the model obtained via slice sampling and optimized Metropolis, considering case 1: $(P, \theta) = (0.8, 1)$ and N=100 (the true value of the parameters are inside the brackets)

Case 1	Methods									
N=100	Slice Sampling					Optimized Metropolis				
Parameters	Mean	s.d.	Median	2.5%	97.5%	Mean	s.d.	Median	2.5%	97.5%
$\beta_1[1]$	1.026	0.103	1.016	0.960	1.082	1.050	0.107	1.063	0.973	1.125
$\beta_2[1]$	0.972	0.046	0.972	0.940	1.001	0.959	0.045	0.959	0.929	0.988
$\beta_3[1]$	0.914	0.048	0.913	0.883	0.945	0.910	0.048	0.911	0.882	0.941
$\sigma[0.05]$	0.077	0.023	0.074	0.060	0.090	0.054	0.025	0.051	0.036	0.068
$\theta[1]$	0.862	0.255	0.834	0.681	1.003	0.861	0.234	0.857	0.695	1.011
$P[0.8]$	0.716	0.276	0.662	0.533	0.838	0.733	0.243	0.738	0.554	0.892
$m[0.8]$	0.831	0.181	0.816	0.702	0.945	0.847	0.148	0.847	0.747	0.944
$s[0.89]$	1.009	0.201	0.978	0.867	1.117	1.017	0.171	0.994	0.900	1.108
$\zeta[0.94]$	0.923	0.036	0.930	0.904	0.950	0.948	0.029	0.953	0.934	0.968

Table 2: Computing times (in seconds)

Sampling Scheme	N=100		N=500	
	$P = 0.8$	$P = 2$	$P = 0.8$	$P = 2$
Slice	329	320	713	720
Optimized Metropolis	442	454	1691	1565

Table 3: Posterior distribution summary for all models´ parameters using slice sampling approach and also the results obtained by van den Broeck et al. (1994) and Greene (1990) - Industrial Electricity Service - US

	Frontier de Normal-Gama								
	Slice Sampling					van den Broeck et al 1994		Greene 1990	
Parameters	Mean	s.d.	Median	2.5%	97.5%	Mean	s.d.	Mean	s.d
β_1	-7.471	0.362	-7.470	-7.713	-7.226	-7.442	0.343	-7.810	0.378
β_2	0.413	0.043	0.413	0.385	0.442	0.407	0.042	0.473	0.043
β_3	0.030	0.003	0.030	0.028	0.032	0.030	0.003	0.026	0.003
β_4	0.248	0.069	0.247	0.201	0.294	0.257	0.068	0.291	0.068
β_5	0.059	0.065	0.058	0.015	0.102	0.060	0.064	0.025	0.066
σ^2	0.015	0.003	0.015	0.013	0.017	0.016	0.004	0.018	0.002
θ	17.432	6.561	16.827	12.236	21.691	23.446	11.368	21.347	3.944
P	2.841	2.013	2.109	1.129	4.192	2	-	2.450	1.102
m	0.149	0.066	0.141	0.094	0.201	-	-	0.115	-
s	0.093	0.019	0.093	0.081	0.105	-	-	0.005	-
ζ	0.367	0.118	0.366	0.287	0.446	-	-	0.226	-

Table 4: Posterior distribution summary of the efficiencies for the first five firms using slice sampling and the approaches of van den Broeck et al. (1994) and Greene (1990) - Industrial Electricity Service - US

	Normal-gama Frontier						
	Slice Sampling			van den Broeck et al., 1994		Greene 1990	
Parameters	Mean	s.d.	Rank	Mean	Rank.	Mean	s.d
r_1	0.723	0.112	120	0.826	121	0.743	0.101
r_2	0.930	0.055	2	0.998	3	0.939	0.036
r_3	0.887	0.077	48	0.912	36	0.906	0.059
r_4	0.872	0.082	68	0.897	77	0.896	0.065
r_5	0.917	0.063	4	0.928	4	0.928	0.043

Figure 1: Posterior density for: P, θ using the: slice sampling (solid line) and optimized Metropolis methods (broken line), based on **case 2**: $(P, \theta) = (2, 1)$ and N=100

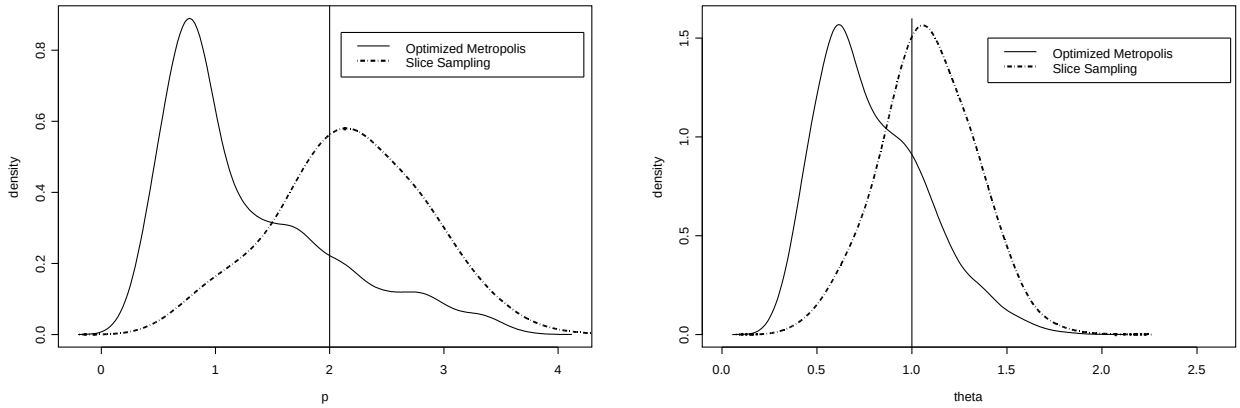
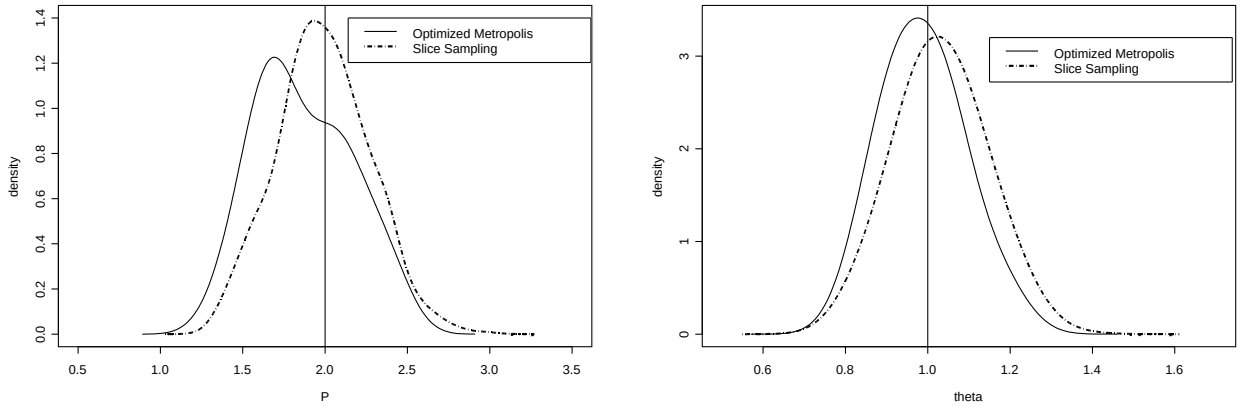
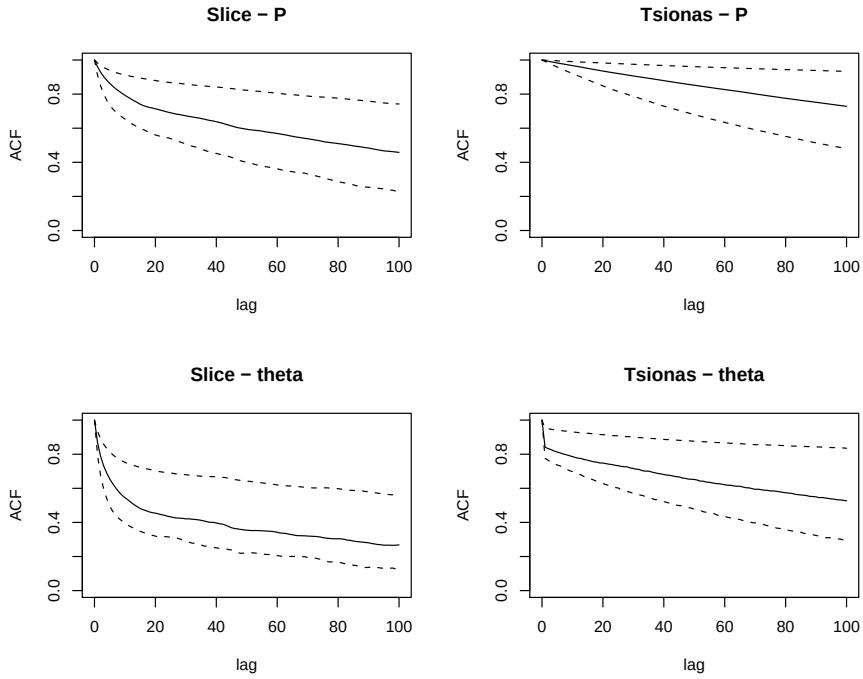
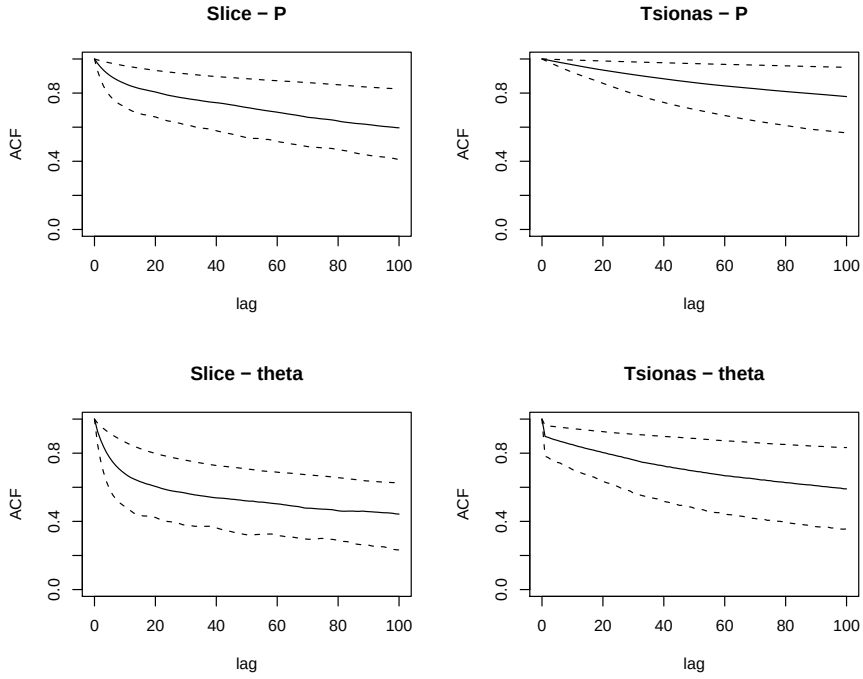


Figure 2: Posterior density for: P, θ using the: slice sampling (solid line) and optimized Metropolis (broken line), based on **case 2**: $(P, \theta) = (2, 1)$ and N=500



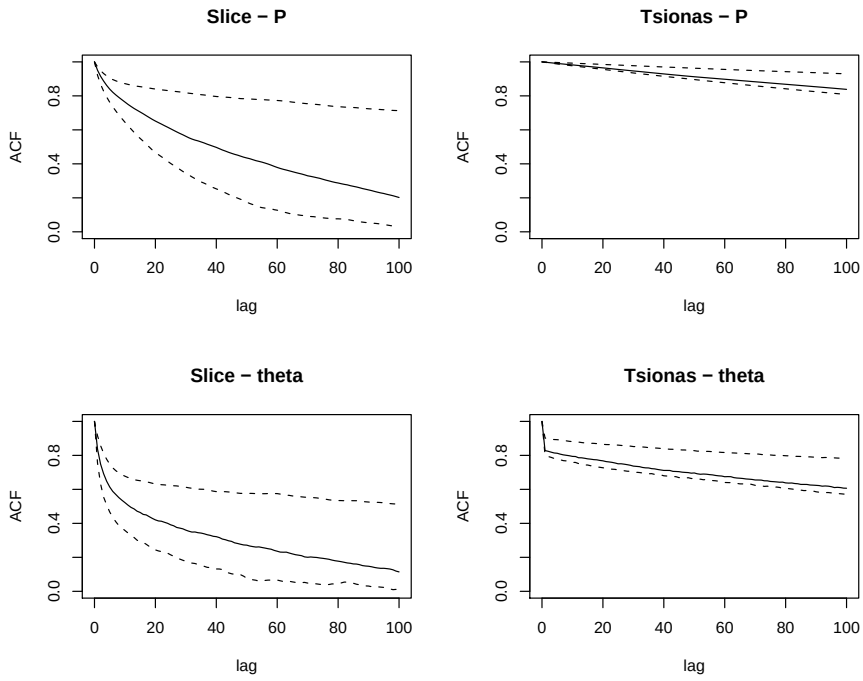


(a) P=0.8 and N=100

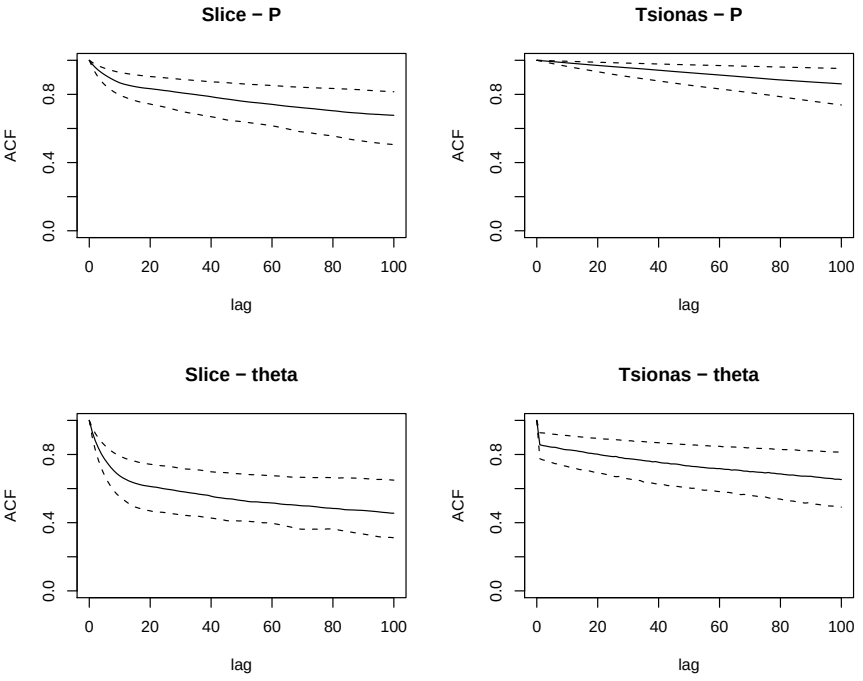


(b) P=2 and N=100

Figure 3: Autocorrelation function for the chain values of P and θ for two alternative sampling schemes and two different value of P: (a) $P = 0.8$ and $N = 100$; (b) $P = 2$ and $N = 100$. The solid lines are the means of 100 replications and the broken lines the 90% credibility interval.

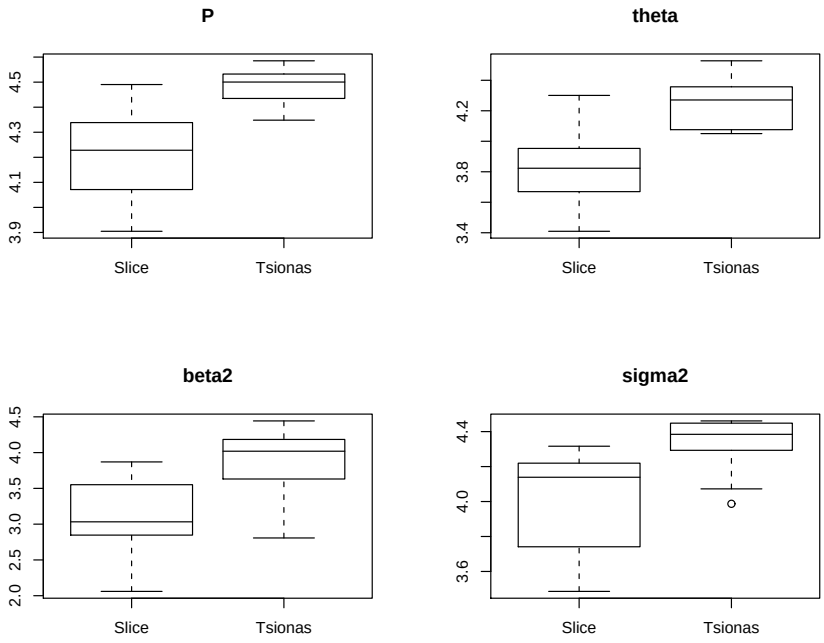


(a) P=0.8 and N=500

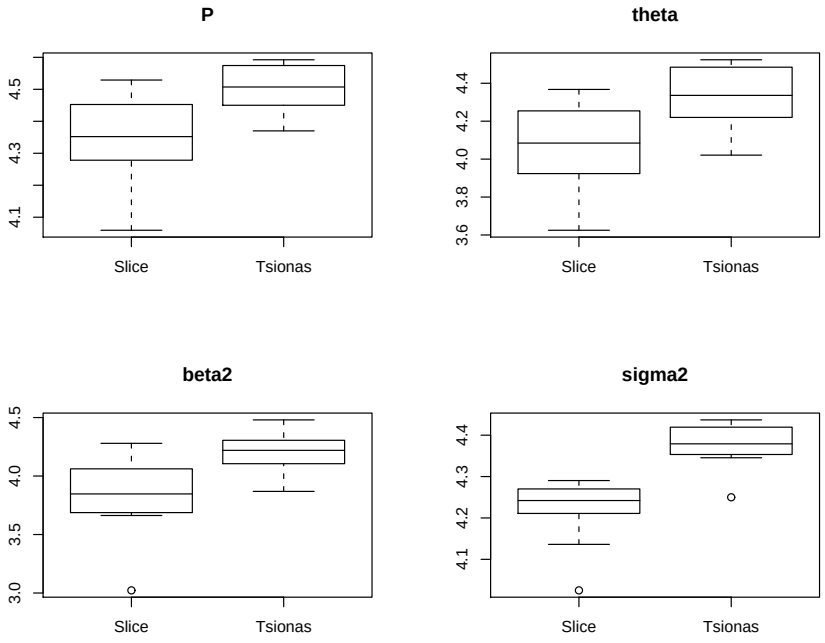


(b) P=2 and N=500

Figure 4: Autocorrelation function for the chain values of P and θ for alternative sampling schemes and two different value of P: (a) $P = 0.8$ and $N = 500$; (b) $P = 2$ and $N = 500$. The solid lines are the means of 100 replications and the broken lines the 90% credibility interval.

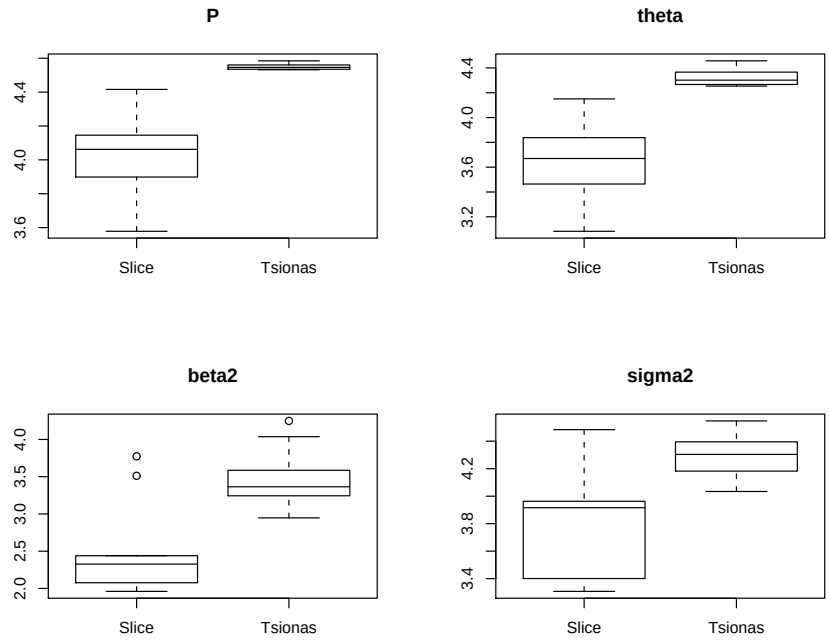


(a) P=0.8 and N=100

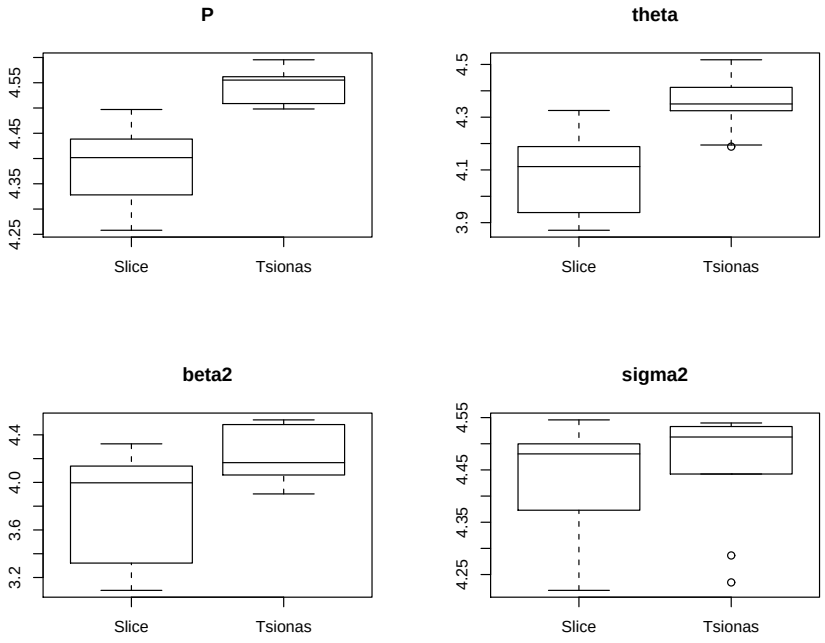


(b) P=2 and N=100

Figure 5: Box-plots of inefficiency factors for P, θ, β_2 and σ_v^2 for two alternative sampling schemes and two different value of P: (a) $P = 0.8$ and $N = 100$; (b) $P = 2$ and $N = 100$.

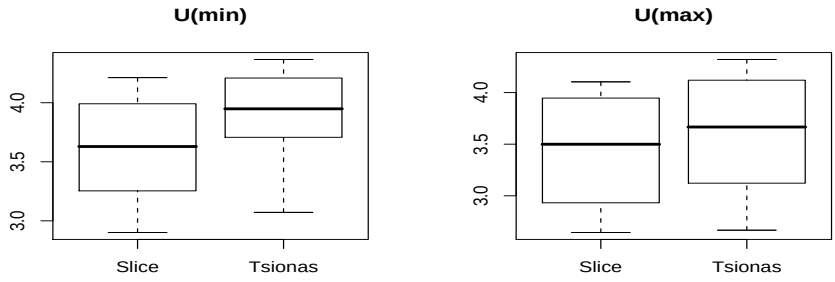


(a) P=0.8 and N=100

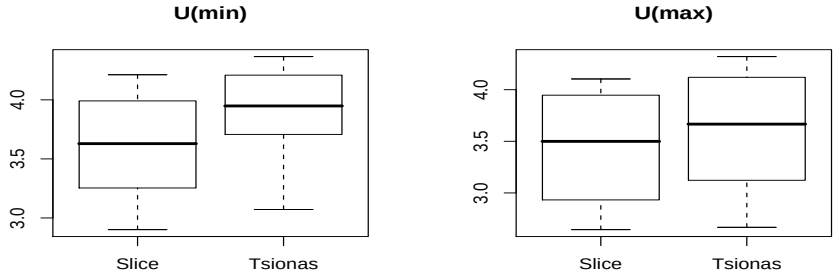


(b) P=2 and N=100

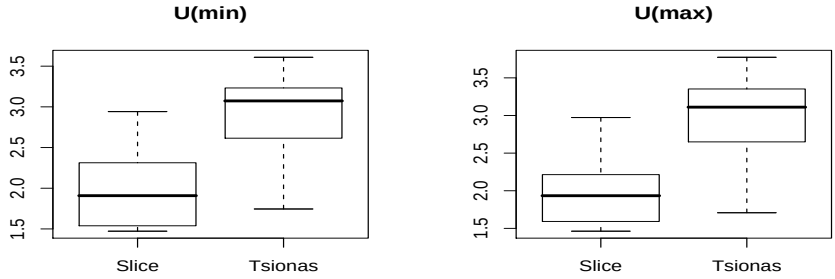
Figure 6: Box-plots of inefficiency factors for P, θ, β_2 and σ_v^2 for two alternative sampling schemes and two different value of P: (a) $P = 0.8$ and $N = 500$; (b) $P = 2$ and $N = 500$.



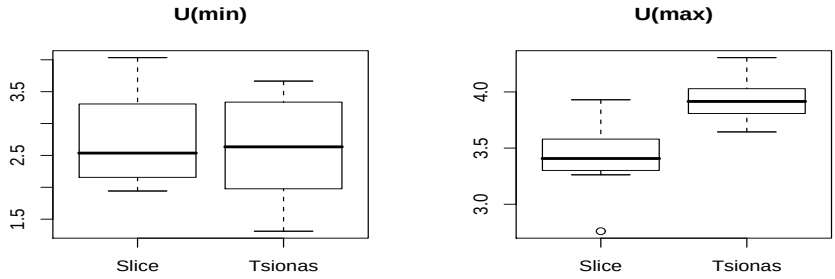
(a) P=0.8 and N=100



(b) P=2 and N=100



(c) P=0.8 and N=500



(d) P=2 and N=500

Figure 7: Box-plots of inefficiency factors of u_{min}, u_{med} and u_{max} for 2 sampling schemes and different values of P: (a) $P = 0.8$ and $N = 100$; (b) $P = 2$ and $N = 100$; (c) $P = 0.8$ and $N = 500$; (d) $P = 2$ and $N = 500$.

Figure 8: Posterior density for: θ, P, σ^2, β_1 using the slice sampling method