The Statistician 32 (1983)

Monte Carlo Analysis of Skew Posterior Distributions: an Illustrative Econometric Example†

H. K. van DIJK and T. KLOEK

Erasmus University Rotterdam, PO Box 1738, 3000 DR Rotterdam, Netherlands

Abstract: The posterior distribution of a small-scale illustrative econometric model is used to compare symmetric simple importance sampling with asymmetric simple importance sampling. The numerical results include posterior first and second order moments, numerical error estimates of the first order moments, posterior modes, univariate marginal posterior densities and bivariate marginal posterior densities plotted in three-dimensional figures.

1 Preliminaries

Our research is directed towards finding finite sample approximations for posterior moments, functions of posterior moments and marginal posterior densities of parameters of econometric models in multidimensional cases. For this purpose we make use of Monte Carlo integration methods. The problem may be stated briefly as follows. Let θ be an s vector of interesting parameters and $g(\theta)$ some function of θ, then

$$Eg(\theta)=\frac{\int g(\theta)p'(\theta|\text{data})\,d\theta}{\int p'(\theta|\text{data})\,d\theta} \tag{1}$$

where $p'(\theta|\text{data})$ is a kernel of the posterior density. We are interested in the efficient computation of the right hand side of equation (1).

So far our approach was to generate a random sample $\theta_1, \ldots, \theta_N$ from a density $I(\theta)$ and compute the posterior expected value of $g(\theta)$ by

$$\bar{g}=\frac{\sum_{i=1}^{N} g(\theta_i)\,w(\theta_i)}{\sum_{i=1}^{N} w(\theta_i)} \tag{2}$$

with $w(\theta_i)=p'(\theta_i|\text{data})/I(\theta_i)$. The density $I(\theta)$ is called the importance function. For details see e.g. Hammersley and Handscomb (1964) and Rubinstein (1981). In two earlier papers (Kloek and Van Dijk, 1978, hereafter KVD, and Van Dijk and Kloek, 1980, hereafter VDK), we applied importance sampling in some Monte Carlo integration problems. In these papers we emphasized as a condition for the feasibility of this approach that an importance function can be found which is a reasonable approximation to the posterior

† This paper is a companion to a forthcoming paper entitled: "Some Alternatives for Simple Importance Sampling in Monte Carlo Integration". In the present paper we emphasize a particular application. In the other paper we emphasize the methodology of alternative Monte Carlo techniques. The authors wish to thank G. den Broeder and E. Gerritsen for assistance with the necessary computer programs.

density. Since this condition is not always satisfied in econometric applications we started to investigate alternative approaches. In the present paper we discuss a particular econometric application. We revisit the Johnston model, studied in KVD, but we consider a different prior density. Our prior is again uniform, but on the interval $(-2, +2)$ rather than $(0, +1)$ for each of the three parameters. The advantage of this choice is that we get more insight in the integration problems of very skew distributions. As mentioned in KVD this skewness is due to the contribution of the Jacobian to the likelihood, but in that paper we eliminated much of the skewness by means of truncation. Apart from skewness the posterior has some interesting features which are described in more detail in section 2. Finally we wish to emphasize that this is a preliminary report (see the introductory footnote).

2 Some results

We take as an example the three-dimensional marginal posterior density of the structural parameters β_1, β_2 and γ_2 of the Johnston model (see KVD, section 4). The prior for β_1, β_2 and γ_2 is uniform on the interval $(-2, +2)$ for all three parameters in the present paper. Hence it is considerably less informative than in the KVD case. The prior for the covariance matrix of the structural disturbances is one proposed by Malinvaud (for details see Malinvaud, 1970, pp. 248–9). The prior for the constant terms is uniform on a large region. In this particular case of two stochastic equations the marginal posterior density of $(\beta_1, \beta_2, \gamma_2)$ is equivalent to the concentrated likelihood function. For some technical details we refer to Van Dijk and Kloek (1977).

We consider two families of importance functions: the multivariate Student density and a product of a univariate Student and log Student densities. The log transformation is rather obvious as a tool to introduce skewness. The problem is to find the proper direction(s) of skewness. This is done in a rather *ad hoc* manner in the present case. More mechanical procedures are a topic of current research.

A multivariate Student density of the s vector θ may be written as

$$I(\theta \mid \mu, V, \lambda) = c\{\lambda + (\theta-\mu)'V^{-1}(\theta-\mu)\}^{-1/2(s+\lambda)} \tag{3}$$

with

$$c = \frac{\lambda^{1/2\lambda}\Gamma\{\frac{1}{2}(s+\lambda)\}}{\pi^{1/2s}\Gamma(\frac{1}{2}\lambda)\,|V|^{1/2}}$$

where μ is the centre of the distribution, V a positive definite symmetric matrix and λ the degrees of freedom parameter. We consider two ways of assigning the parameters of (3). In both cases the degrees of freedom parameter and a common scale parameter are fixed at unity for the sake of convenience. The two cases differ in the following respects. Case I consists of taking the posterior mode for μ and minus the inverse of the Hessian of the log posterior for V. This we shall name the *local* approximation case. Case II consists of taking the posterior mode for μ in asymmetric importance sampling and the posterior mean as estimate for μ in symmetric importance sampling. The posterior covariance matrix is the estimate for V. These posterior estimates for μ and V are obtained after a first round of Monte Carlo. We name case II the *global* approximation case.

The parameter estimates are presented in Table 1. It is seen that the posterior densities for β_1 and β_2 are skew because the modes and means differ considerably. Also the local approximation of V fails to hold globally in at least two respects. First, the posterior standard deviations for β_1 and β_2 are for the global case roughly eight times as large as their local approximations. Second, the posterior correlation between β_2 and γ_2 is positive in the global case, while the local approximation indicates that it is negative.

Table 1. Estimates of importance function parameters[a] μ and V

	μ_1	μ_2	μ_3
Posterior mode (=FIML)	0·46	0·09	0·36
Posterior mean	−0·57	−0·31	0·30
	v_1	v_2	v_3
Local approximation in mode	0·10	0·04	0·11
Posterior standard deviations	0·78	0·33	0·14
	r_{12}	r_{13}	r_{23}
Local approximation in mode	0·88	0·17	−0·16
Posterior correlations	0·93	0·25	0·35

[a] The square roots of the diagonal elements of V are denoted by v (lower case). Note that the local approximation of v_3 reported here differs slightly from the value presented in KVD (Table 1). The present values are computed on a DEC 2050 computer in double precision. The correlation coefficients of the parameters are denoted by r.

Table 2. Numerical error estimates of posterior means of structural parameters for alternative Monte Carlo methods[a]

	SSIS	ASIS	Best method
Case I (local approximation)			
β_1	4·44	13·28	SSIS
β_2	5·99	7·10	SSIS
γ_2	3·34	5·44	SSIS
Case II (global approximation)			
β_1	0·67	0·55	ASIS
β_2	0·60	0·63	SSIS
γ_2	0·58	0·44	ASIS

[a] SSIS = symmetric simple importance sampling; ASIS = asymmetric simple importance sampling.

Table 2 presents results for the numerical error of the posterior means of β_1, β_2 and γ_2 for the Monte Carlo methods. As a measure of numerical error we take the ratio ($\times$ 100) of the standard deviation of the Monte Carlo estimate of the posterior mean (see KVD, section 6) and the posterior standard deviation given in Table 1. This relative measure of numerical inaccuracy has been chosen, since we are more concerned to estimate a posterior mean accurately if the posterior variance is small, than if it is large.

We consider two methods. The first is symmetric simple importance sampling (SSIS), as described in our earlier papers. The second method is asymmetric simple importance sampling (ASIS). We generate random drawings as follows. Standard Student random drawings (s_1, s_2, s_3) are generated. Next we use the transformation $s_1^* = 1 - e^{s_1}$, $s_2^* = 1 - e^{s_2}$, $s_3^* = s_3$. Finally the values of s_1^*, s_2^* and s_3^* are rotated in the usual way with estimates of the posterior mode and the local approximation (case I) and the global approximation (case II). This introduction of skewness is rather *ad hoc*. We are currently investigating more mechanical procedures.

The results of Table 2 clearly indicate that the local approximation of V is a poor starting point for computational efficiency. Comparing the Monte Carlo methods it is seen that SSIS gives the lowest error when the local approximation is used, but that asymmetric importance sampling is better in two out of three cases for global approximation. We want to stress that these numerical results are rather preliminary. We have taken the results for the different methods after roughly 15 minutes CPU-time in order to avoid large sampling errors in the estimates of these numerical errors. However, a careful comparison consists of recording the number of function evaluations; the numbers of accepted and rejected drawings using identical random number sequences, etc. This has still to be performed.

Next we present the univariate marginal posterior densities for β_1, β_2 and γ_2 and bivariate marginal posterior densities for (β_1, β_2), (β_1, γ_2) and (β_2, γ_2) in Figure 1 and Figures 2a, 2b and 2c. These have been computed by making use of the formulae given in KVD, section 7, but with asymmetric simple importance sampling.

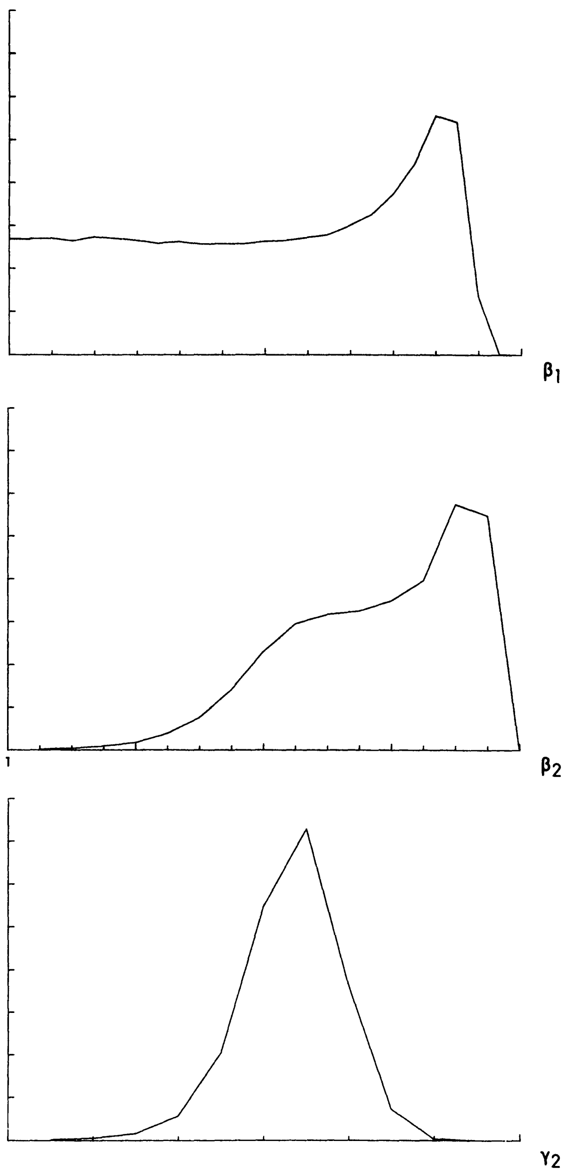

Fig. 1. Univariate marginal posterior densities for β_1, β_2 and γ_2.

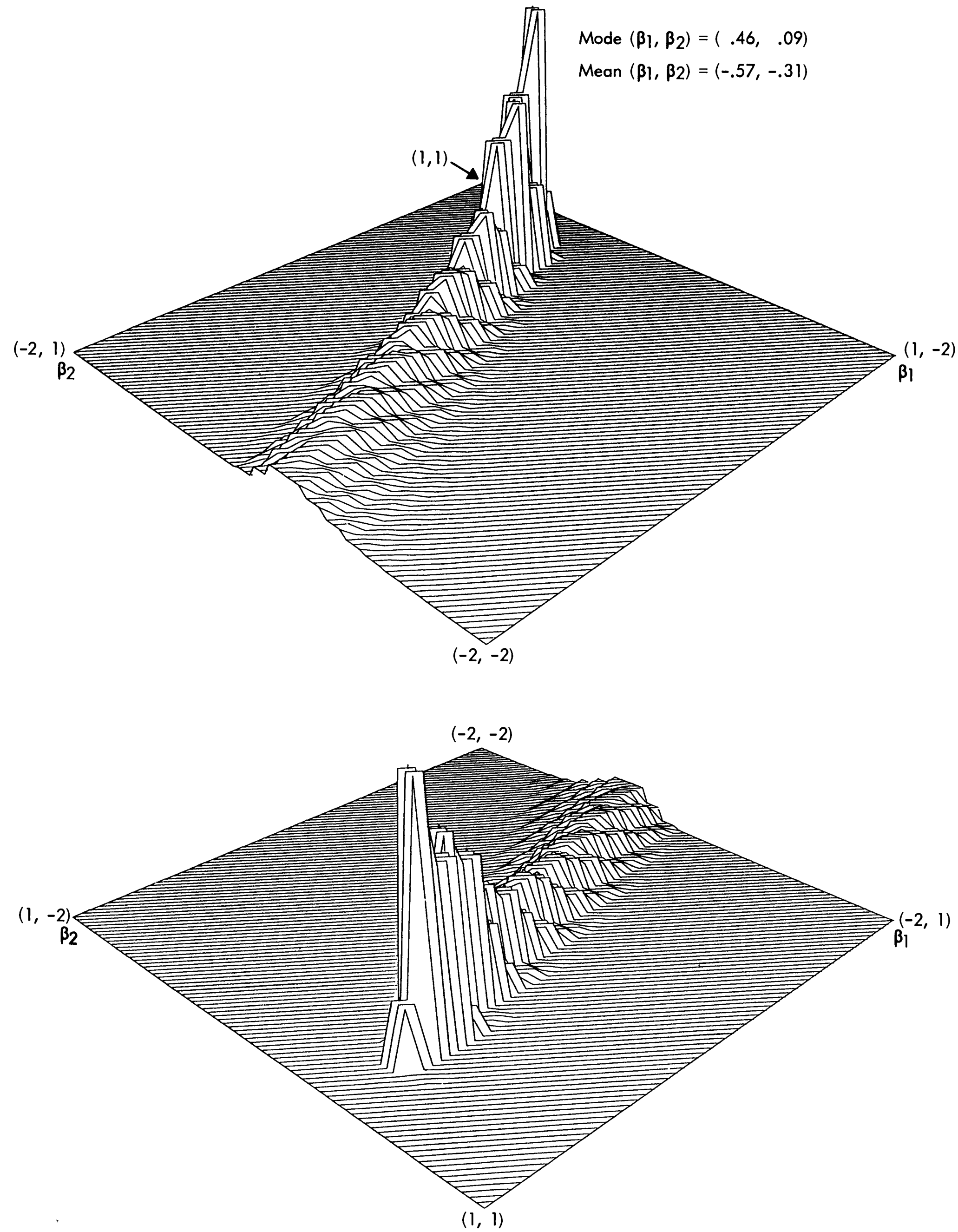

Fig. 2a. Bivariate marginal posterior densities for (β_1, β_2).

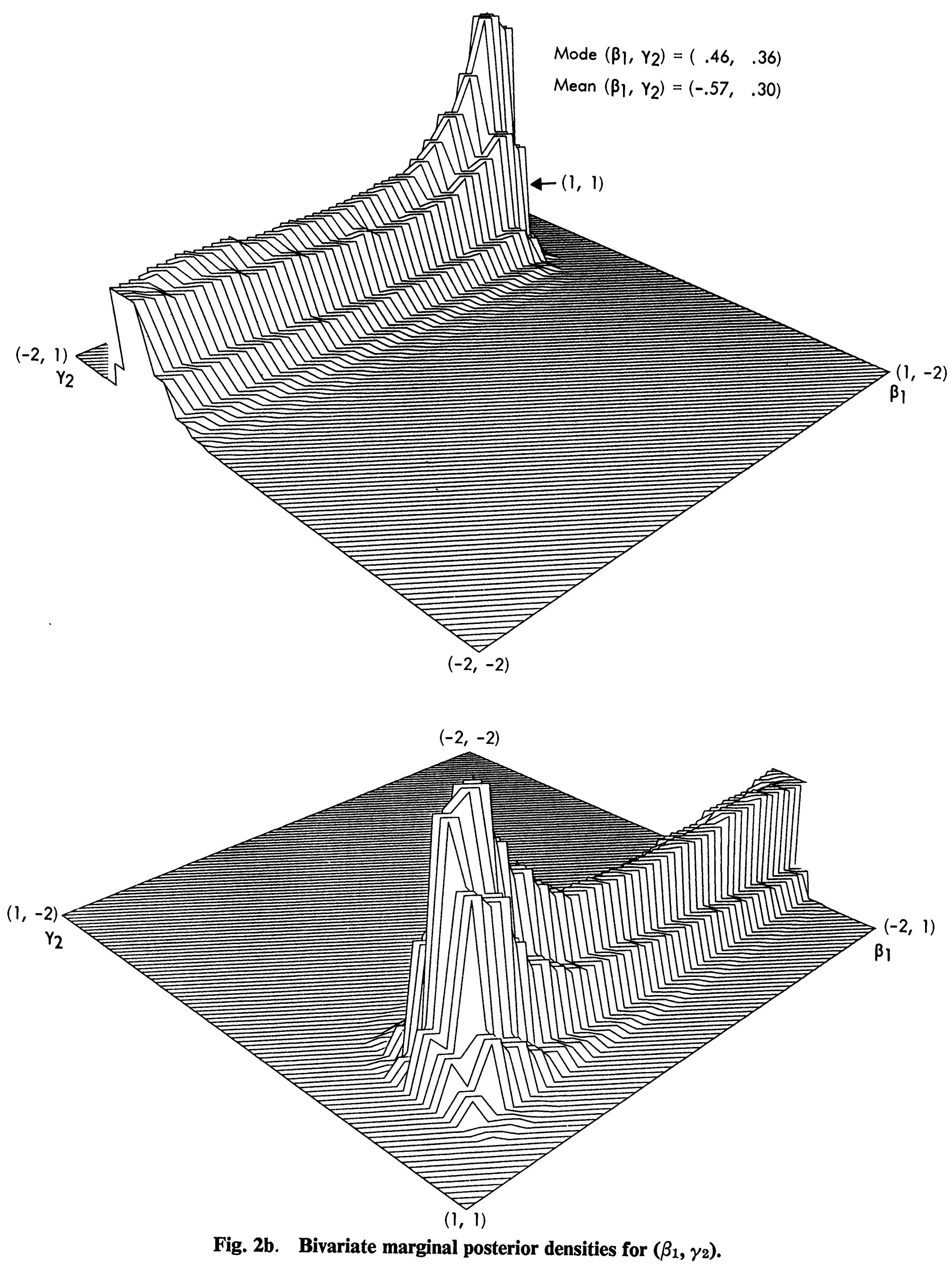

Fig. 2b. **Bivariate marginal posterior densities for (β_1, γ_2).**

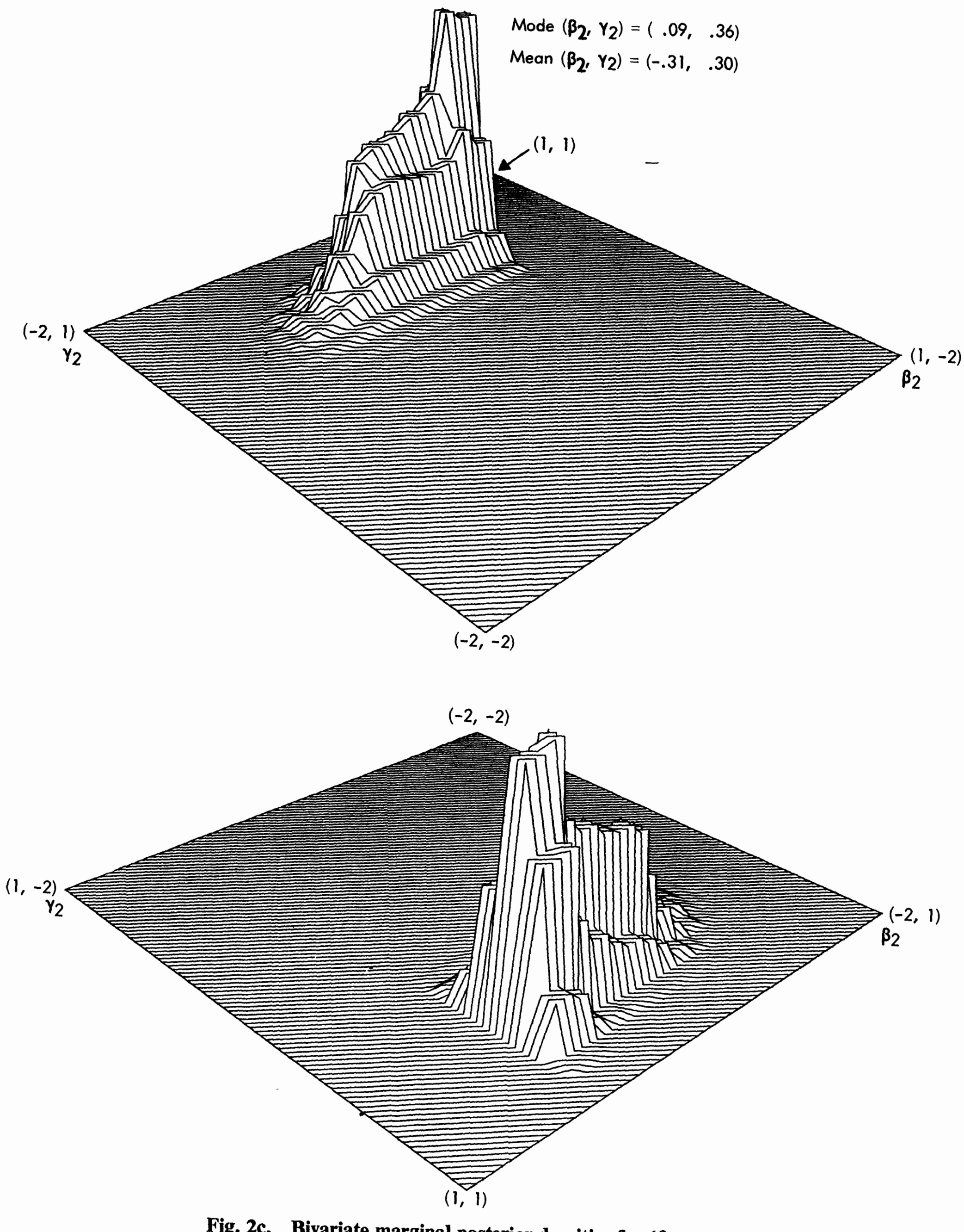

Fig. 2c. Bivariate marginal posterior densities for (β_2, γ_2).

Figure 2a clearly illustrates the effect of the Jacobian $|1-\beta_1-\beta_2|^n$ (compare KVD, sections 2 and 4). If β_1 and β_2 are both negative, the Jacobian is greater than unity. Figures 1, 2a and 2b indicate that the constraint $\beta_1 \geqslant -2$ truncates the posterior. Figure 2c has a more regular shape, though rather skew in the direction of β_2.

We conclude this section with two remarks.

1. For the local approximation case we performed a sensitivity analysis with respect to the common scale parameter of the covariance matrix. However, the search for an optimal value of such a parameter is computationally rather costly, when one has to run the same computer program for different values in a rather wide interval.

2. We generated structural disturbances from a multivariate normal process around the posterior mode and re-estimated the marginal posterior densities by means of Monte Carlo. Roughly the same results occur. Thus, specification errors, which are probably present in Johnston's model, are not the main cause of the problem.

References

Hammersley, J. M. and Handscomb, D. C. (1964). *Monte Carlo Methods.* Methuen, London.

Kloek, T. and van Dijk, H. K. (1978). Bayesian estimates of equation system parameters; an application of integration by Monte Carlo. *Econometrica* **46,** 1–19; reprinted in A. Zellner (ed.) (1980). *Bayesian Analysis in Econometrics and Statistics, Essays in Honour of Harold Jeffreys.* North-Holland, Amsterdam.

Malinvaud, E. (1970). *Statistical Methods of Econometrics,* 2nd edn. North-Holland, Amsterdam.

Rubinstein, R. Y. (1981). *Simulation and the Monte Carlo Method.* Wiley, New York.

Van Dijk, H. K. and Kloek, T. (1977). Predictive moments of simultaneous econometric models, A Bayesian approach. In *New Developments in the Applications of Bayesian Methods* (eds A. Aykaç and C. Brumat). North-Holland, Amsterdam.

Van Dijk, H. K. and Kloek, T. (1980). Further experience in Bayesian analysis using Monte Carlo integration. *Journal of Econometrics* **14,** 307–28.