Journal of Econometrics 14 (1980) 307–328. North-Holland Publishing Company

FURTHER EXPERIENCE IN BAYESIAN ANALYSIS USING MONTE CARLO INTEGRATION*

H.K. van DIJK and T. KLOEK

Erasmus University, Rotterdam, The Netherlands

Received November 1978, final version received June 1980

An earlier paper [Kloek and Van Dijk (1978)] is extended in three ways. First, Monte Carlo integration is performed in a nine-dimensional parameter space of Klein's model I [Klein (1950)]. Second, Monte Carlo is used as a tool for the elicitation of a uniform prior on a finite region by making use of several types of prior information. Third, special attention is given to procedures for the construction of importance functions which make use of nonlinear optimization methods.

1. Introduction

In a previous paper [Kloek and Van Dijk (1978)] we proposed integration by Monte Carlo as a tool for finding posterior moments and posterior densities. In this paper we report about further experience in this area. In comparison to the previous paper, the analysis is extended in three ways.

First, there is a difference in dimension. The Monte Carlo approach is now applied to Klein's model I [Klein (1950)], which implies that we compute nine-dimensional integrals numerically; the dimensionality of the previous paper was only three.

Second, we demonstrate how Monte Carlo can play a helpful role in the specification of prior distributions. We specify uniform priors on finite regions in the space of structural parameters and make use of four types of information. These include:

(i) information on signs of structural parameters and short-run and long-run multipliers;
(ii) information on upper bounds of the absolute values of structural parameters, and of short-run and long-run multipliers;
(iii) information on the stability of the system;
(iv) information on the length of the period of oscillation.

*This paper started as a revision of Van Dijk and Kloek (1978). In the course of the work our ideas developed to such an extent that the final result is an almost completely new paper. We are indebted to a referee for a number of very useful suggestions. We also wish to thank A.S. Louter and G. den Broeder of the Econometric Institute for their help in preparing the necessary computer programs.

The order of the priors is chosen in such a way that *a priori* implausible implications of one prior are (partially) corrected in the next one. In a sense one may speak of elicitation of a prior of economically interesting structural parameters by looking at the implied prior multipliers and dynamic characteristics.

Third, we partially revise the suggestions for the specification of importance functions given in the previous paper. In particular, we emphasize that it is better to start with maximizing the log posterior density and to use (minus the inverse of) its Hessian evaluated at the optimum for describing the covariance structure, than with integration based on an importance function which is a very rough approximation of the posterior density. In this context it is a handicap that the posterior of our example (based on a uniform prior) is not twice continuously differentiable at the posterior mode. For that reason we experiment with two types of approximate priors which have the desired differentiability property. In addition, improvements in the accuracy of the posterior results are investigated by comparing the use of different members of the family of multivariate Student densities.

The plan of the paper is as follows. Section 2 deals with prerequisites such as model specification, prior information as far as this is a routine operation, and a few results of our previous paper. Section 3 discusses the specification of prior densities of the interesting parameters. In section 4 we treat the problem of constructing importance functions. Posterior results of Klein's model I are discussed in section 5. Our conclusions are given in section 6. The appendix gives some details on the use of information contract curves, which are used as a tool in the construction of importance functions.

2. Prerequisites

This section deals with such prerequisites as the model, the likelihood and the prior information with respect to the constant terms and the covariance matrix. The statistical model is the same as that in our previous paper. It can be summarized as follows. Our starting point is the well-known linear simultaneous equation model

$$Y\Gamma + XB = U, \tag{1}$$

where Y is an $n \times G$ matrix of observations on G current endogenous variables and X an $n \times K$ matrix of observations on predetermined variables. The rows of U are assumed to be independently normally distributed with a covariance matrix

$$\begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix},$$

where Σ (a non-singular $G_1 \times G_1$ matrix) corresponds to the stochastic equations. The prior density of the constant terms (β_0) is uniform and the prior density of Σ is proportional to $|\Sigma|^h$, where $h = -\frac{1}{2}(G_1 + 1)$ and G_1 is the number of stochastic equations [see, e.g., Zellner (1971)]. As a consequence, analytical integration with respect to β_0 and Σ is possible. Identification is treated in the traditional way; for an alternative approach, see Kiefer (1979). As a result we have a number of exactly known parameters (not only identifying zeros, but also normalizing unities) which are substituted in the likelihood function. The remaining parameters are summarized in a vector θ. As a consequence we can write

$$\Gamma = \Gamma(\theta), \qquad B = B(\beta_0, \theta).$$

Prior densities for θ (the so-called interesting parameters) will be specified in section 3 below. These are all uniform, but the regions differ.

The posterior density of $(\theta, \beta_0, \Sigma)$ is obtained by combining the likelihood function of the linear simultaneous equation system and the prior density by means of Bayes theorem. Using an analytical integration procedure with respect to Σ and β_0, one obtains the posterior density of θ, marginal with respect to Σ and β_0. It is denoted by[1]

$$p'(\theta \mid Y, X) \propto \kappa(\theta \mid Y, X) p(\theta), \tag{2}$$

where $p(\theta)$ is the prior density of θ. We note that $\|\Gamma\|^n$, the nth power of the absolute value of the determinant of Γ, is a factor in the κ-function; this factor is the cause of considerable skewness of the posterior density in one direction.

Next, we briefly introduce Klein's model I. This is important for the interpretation of the parameters to be estimated. Also, the exact part of the prior information is implied. The structural equations of Klein I read

$$\begin{aligned}
C &= \alpha_1 P + \alpha_2 P_{-1} + \alpha_3 W + \alpha_4 + u_1, \\
I &= \beta_1 P + \beta_2 P_{-1} - \beta_3 K_{-1} + \beta_4 + u_2, \\
W_1 &= \gamma_1 X + \gamma_2 X_{-1} + \gamma_3 t + \gamma_4 + u_3, \\
X &= C + I + G, \\
P &= X - W_1 - T, \\
K &= K_{-1} + I, \\
W &= W_1 + W_2.
\end{aligned}$$

[1]The precise formula for the κ-function can be found in Kloek and Van Dijk (1978, eq. (2.5)). There is a subtle difference with the concentrated likelihood, explained in section 7 of that paper. In our case this difference is numerically important. Details of the analytical integration steps have been spelled out in Van Dijk and Kloek (1977).

Consumption expenditure (C) is structurally dependent on profits (P), profits lagged one year (P_{-1}) and on total wages (W). Net investment expenditure (I) depends on profits, profits lagged and on the capital stock at the beginning of the year (K_{-1}); note the minus sign before β_3 in the investment equation. Finally, private wage income (W_1) depends on net private product at market prices (X), the same variable lagged (X_{-1}) and on a trend term (t). The model is closed by four identities, which provide links with three exogenous variables: the government wage bill (W_2), government nonwage expenditure, including the net foreign balance, (G) and business taxes (T). The model counts seven jointly dependent variables (C, I, W_1, X, P, K, W) and eight predetermined variables (1, P_{-1}, X_{-1}, K_{-1}, G, T, W_2, t). All variables (except 1 and t) are measured in constant dollars.

For a more detailed exposition of the model the reader is referred to Klein (1950). Note however that the use of the symbols Y for net national income and G for government nonwage expenditure is not uniform in the literature on Klein's model I. We shall use $Y(=X-T+W_2)$ for net national income and follow the notation of Theil, Boot and Kloek (1965). Klein (1950) uses G for government expenditure including wages ($=G+W_2$ in our notation). Other authors, e.g., Rothenberg (1973), use Y instead of X for net private product. This notational point is relevant for the interpretation of a number of reduced and final form multipliers. For details, see section 5.

3. Prior densities

In this section we shall specify a number of prior densities of θ and demonstrate how Monte Carlo may be used to investigate the implied prior information with respect to the reduced form parameters, the stability characteristics of the model and the final form parameters (if these exist). Our starting point is the vector θ' which equals (α_1, α_2, α_3, β_1, β_2, β_3, γ_1, γ_2, γ_3); compare the preceding section.

Our first and simplest prior for this vector is uniform on the nine-dimensional unit region[2] minus the region where $\|\Gamma\|<0.01$. The latter region has been subtracted in order to guarantee that the implied prior moments of the multipliers exist. The reason for choosing such a uniform prior is that it is easy to understand and to specify. Suppose, for example, that the most important theoretical information on some parameter α is that it is positive, but that one has no idea whether it will be close to zero or closer to 0.5, say. So one wants to specify a prior with the following properties: (i) the implied prior probability on the interval $(-0.5, 0)$ should be zero or small; (ii) the implied prior probability on (0, 0.5) should be substantial; (iii) the prior density on (0, 0.5) should be constant or at least approximately constant. It

[2] We use the term *unit interval* for the interval (0, 1) and the term *unit region* for a Cartesian product of unit intervals.

is impossible to specify a normal or Student prior which simultaneously has these three properties. The simplest way to avoid this problem is to specify a uniform prior on (0, b) , where b is given some appropriate value. In that interval the likelihood determines the posterior. It goes without saying that such a prior need not reflect in all detail the betting odds one might be willing to accept.

Next we investigate the implications of our prior information for the multipliers and dynamic characteristics of the model. We obtained the implied prior means and standard deviations of these functions of θ by drawing θ vectors from the nine-dimensional standard uniform distribution. Each θ vector was checked with respect to the condition $\|\Gamma\| > 0.01$. In case this condition was not satisfied, the vector was rejected and replaced by a new vector. Each experiment was stopped when 2,500 θ vectors satisfying the constraint were obtained. For each θ vector we computed the implied reduced form parameters or short-run multipliers (SRM) and some other characteristics, to be discussed below. These were the basis for the computation of implied prior means and second-order moments.

The reduced form equations form a system of linear difference equations. The roots of the characteristic polynomial of this system summarize the dynamic properties of the system [see Theil and Boot (1962)]. Dependent on these roots the system may be damped or explosive, and oscillating or monotone. For each of these four states we computed the prior probabilities implied by the specified structural prior density. In case the system is oscillating one may compute the period of oscillation, and in case the system is damped one may compute the final form parameters or long-run multipliers (LRM).

As a next step we modified our first prior in several ways by adding sets of extra constraints. The constraint of prior 1 was maintained in all stages. The sets of extra constraints, which were introduced partly one at a time and partly in various combinations, will now be described:

(1) The system is assumed to be stable. So we only accepted vectors θ satisfying $|DRT| < 1$, where DRT is the dominant root of the characteristic polynomial. In the present example this is of the third degree; for computational aspects, see Uspensky (1948).
(2) The long-run effects in the structural equations $\alpha_1 + \alpha_2$, $\beta_1 + \beta_2$, $\gamma_1 + \gamma_2$ are all assumed to be in the unit interval.
(3) The SRM's are assumed to be less than five in absolute value (with an exception for the reduced form equation of K, where an upper bound of ten was adopted) and to have the correct sign (positive for effects of W_2 and G, negative for effects of T).
(4) The same constraints as mentioned in 3 were applied to the LRM's.
(5) The period of oscillation is assumed to be between three and ten years.

This is in accordance with the observed length of business cycles in the period 1890–1920; see Historical statistics of the U.S. (1975).

In table 1 we show the prior means and standard deviations for eight different priors, which were obtained by combining the sets of extra constraints in several ways. The priors of β_3 and γ_3 hardly reacted on the constraints. The means and standard deviations of the remaining parameters decreased, with one exception (the mean of γ_1 in prior 2). The parameters most affected by the constraints were α_1, β_1 and γ_2.

Table 1

Prior means and standard deviations of structural parameters.

Prior	Sets of extra constraints	α_1	α_2	α_3	β_1	β_2	β_3	γ_1	γ_2	γ_3
1	No	0.51	0.51	0.51	0.50	0.50	0.50	0.50	0.50	0.50
		(0.28)	(0.28)	(0.28)	(0.29)	(0.29)	(0.29)	(0.28)	(0.29)	(0.29)
2	1	0.41	0.43	0.37	0.41	0.36	0.49	0.66	0.33	0.49
		(0.29)	(0.28)	(0.26)	(0.28)	(0.26)	(0.29)	(0.26)	(0.24)	(0.30)
3	2	0.34	0.33	0.50	0.34	0.34	0.50	0.34	0.33	0.50
		(0.23)	(0.23)	(0.28)	(0.24)	(0.24)	(0.29)	(0.23)	(0.24)	(0.29)
4	1, 2	0.27	0.33	0.45	0.29	0.29	0.50	0.42	0.26	0.50
		(0.21)	(0.23)	(0.27)	(0.21)	(0.21)	(0.30)	(0.24)	(0.20)	(0.29)
5	2, 3	0.27	0.36	0.46	0.27	0.37	0.50	0.35	0.33	0.50
		(0.20)	(0.24)	(0.27)	(0.20)	(0.24)	(0.29)	(0.23)	(0.24)	(0.29)
6	1, 2, 3	0.25	0.34	0.43	0.27	0.30	0.50	0.41	0.27	0.50
		(0.20)	(0.23)	(0.27)	(0.20)	(0.21)	(0.30)	(0.24)	(0.20)	(0.29)
7	1, 2, 3, 4	0.25	0.33	0.40	0.27	0.30	0.53	0.41	0.27	0.49
		(0.20)	(0.22)	(0.25)	(0.20)	(0.21)	(0.27)	(0.24)	(0.20)	(0.29)
8	1, 2, 3, 4, 5	0.26	0.33	0.39	0.27	0.29	0.55	0.42	0.28	0.49
		(0.20)	(0.22)	(0.24)	(0.20)	(0.21)	(0.27)	(0.24)	(0.20)	(0.29)

In table 2 we show the prior means and standard deviations for the multipliers of W_2, T and G in the reduced and final form equations for national income (Y), the dominant root and the period of oscillation, and the prior probabilities for the four states of the system. If only structural extra constraints are introduced (priors 1 and 3) the prior means of the SRM's have the correct signs but very large standard deviations. (Obviously, these standard deviations are very sensitive for the lower bound of $\|\Gamma\|$.) Accordingly the probabilities of explosive behavior are substantial. If the DRT constraint (1) is introduced (priors 2 and 4) the standard deviations are moderate, but only if the SRM constraints (3) are introduced the standard deviations are of acceptable size. In the same way the standard deviations of the LRM's are very large in priors 2, 4 and 6 where the LRM constraints (4)

Table 2

Prior means and standard deviations of selected multipliers; period of oscillation and dominant root; prior probabilities of states.

Prior	Sets of extra constraints	Short-run effects on Y: W_2	T	G	Damped: Oscillatory	Damped: Monotone	Explosive: Oscillatory	Explosive: Monotone
1	No	2.08 (9.20)	−2.36 (15.64)	1.98 (12.46)	0.22	0.01	0.46	0.31
2	1	1.98 (1.43)	−2.82 (1.96)	2.17 (1.55)	0.96	0.04	0	0
3	2	2.65 (6.77)	−2.97 (10.46)	2.85 (9.49)	0.52	0.02	0.32	0.14
4	1,2	2.24 (1.47)	−2.55 (1.69)	2.40 (1.53)	0.98	0.02	0	0
5	2,3	2.11 (0.88)	−2.34 (0.95)	2.24 (0.84)	0.66	0.02	0.29	0.03
6	1,2,3	1.97 (0.82)	−2.20 (0.88)	2.08 (0.76)	0.98	0.02	0	0
7	1,2,3,4	1.87 (0.73)	−2.16 (0.87)	2.01 (0.72)	0.98	0.02	0	0
8	1,2,3,4,5	1.86 (0.73)	−2.18 (0.86)	2.02 (0.71)	0.98	0.02	0	0

Prior	Sets of extra constraints	Long-run effects on Y: W_2	T	G	Period of oscillation	$\|DRT\|$
2	1	2.39 (20.33)	−2.51 (13.63)	2.35 (20.72)	6.82 (8.48)	0.78 (0.17)
4	1,2	2.87 (11.46)	−2.83 (5.01)	2.95 (11.44)	7.37 (8.30)	0.74 (0.18)
6	1,2,3	2.61 (11.34)	−2.66 (4.91)	2.69 (11.32)	7.32 (7.96)	0.74 (0.18)
7	1,2,3,4	1.96 (0.88)	−2.22 (0.76)	2.04 (0.80)	6.60 (6.71)	0.72 (0.18)
8	1,2,3,4,5	1.94 (0.88)	−2.22 (0.76)	2.02 (0.80)	5.71 (1.42)	0.72 (0.18)

were not applied, while priors 7 and 8 have acceptable standard deviations for the LRM's. In the same way, the standard deviations of the period of oscillation are large until it is constrained in prior 8.

We conclude this section by making five remarks:

(i) As a byproduct of our integration procedure we obtained marginal prior densities of the structural parameters. These were either uniform or skew. They are not shown for space considerations.

(ii) A disadvantage of the truncated priors is that one might fail to detect specification errors. This problem can easily be solved by carrying out sensitivity analysis on larger intervals than the unit interval, but this subject is outside the scope of the present paper. It should be added that the Monte Carlo approach can very well be applied in cases of smooth priors. Some simple examples can be found in Kloek and Van Dijk (1978). The only problem, in our opinion, is the difficulty of specifying such priors.

(iii) In monotone cases the period of oscillation was not computed. So the distributions of the period of oscillation as shown should be interpreted as conditional distributions, the condition being that the system is oscillating. Since the period of oscillation is positive by definition, it follows from the large standard deviations that the implied prior densities in the cases of priors 2, 4, 6 and 7 are very skew. This was the motivation for introducing set of constraints (5).

(iv) Our priors 1 and 2 admit the possibility that the long-run marginal propensity to consume from profit income is in the interval (1, 2). Few economists would accept such an outcome. Similar things can be said about the sums $\beta_1+\beta_2$ and $\gamma_1+\gamma_2$ in the other equations. This explains set of constraints (2). If one of these constraints was violated we made use of the following computational device. Start with an auxiliary vector θ^a and test whether the above mentioned sums are larger than one. If $\alpha_1^a+\alpha_2^a>1$ (say), then take $\alpha_1=1-\alpha_1^a$, $\alpha_2=1-\alpha_2^a$, otherwise take $\alpha_1=\alpha_1^a$, $\alpha_2=\alpha_2^a$. The reduction of computation time due to this device was substantial.

(v) The approach of presenting results for several priors was advocated, in a somewhat different context, by Leamer (1978). The above examples show that Monte Carlo may be a useful tool to find out the consequences of specifying a structural prior. In addition, the possibility is demonstrated to combine prior information of different types.

4. Construction of importance functions

Let $g(\theta)$ be a function of the structural parameters. Then in our integration by Monte Carlo approach the posterior mean of $g(\theta)$ is computed as

$$\frac{1}{N}\sum_{i=1}^{N} g(\theta_i)\frac{p'(\theta_i)}{I(\theta_i)}, \tag{3}$$

apart from a normalizing constant which is computed separately. Here $I(.)$ is the density of a distribution from which vectors $\theta_1, \theta_2, \ldots, \theta_N$ are drawn, $p'(.)$ is the posterior density and N is the number of drawings. The density I (the so-called importance function) is supposed to have convenient Monte

Carlo properties and to be an approximation of the posterior density. For an explanation, see Kloek and Van Dijk (1978, sect. 3).

If the posterior density is multivariate and unimodal the most obvious choice for the functional form of the importance function is the multivariate Student family with the multivariate normal density as a limiting case. A multivariate Student density may be written as

$$I(\theta)=c[\lambda+(\theta-\mu)'V^{-1}(\theta-\mu)]^{-\frac{1}{2}(s+\lambda)}, \tag{4}$$

where

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

and where μ is the center of the distribution, V a positive definite matrix, λ the degrees of freedom parameter and s the dimension of θ.

Random s-vectors $\theta_1,\theta_2,\ldots,\theta_N$ distributed according to (4) are generated as follows. One starts to generate an s-vector u_i of independent standard normally distributed random variables. Efficient techniques for this step can be found in Atkinson and Pearce (1976). Then u_i is premultiplied by a matrix A which satisfies $V=AA'$. We obtained A from the eigenvalues and eigenvectors of V but it is also possible to construct a triangular matrix A by a Cholesky technique; see e.g. Bard (1974). Finally, one draws a vector w_i of λ independent standard normal variables and obtains the s-vector θ_i from

$$\theta_i=\mu+Au_i(\lambda/w_i'w_i)^{\frac{1}{2}}. \tag{5}$$

The scalar in the last term can be deleted in case the importance function is multivariate normal.

If we adopt the multivariate Student density as a functional form for the importance function, we need a way to specify its parameters. If the sample is sufficiently large and certain conditions are satisfied, the likelihood function is approximately normally distributed. The same holds for the posterior density if the prior is locally uniform on a neighborhood of the maximum likelihood estimate. So it seems reasonable to take for μ the ML estimate of θ and for V its estimated asymptotic covariance matrix, possibly multiplied by a scalar ϕ. Alternatively, one may take for μ the posterior mode of θ and for V minus the inverse of the Hessian of the log posterior density, evaluated at the posterior mode, possibly multiplied by ϕ.

In the case of Klein's model I the FIML estimate is unacceptable in view of our prior information. Neither does the κ-function possess a maximum in the interior of the unit region. The consequence is that our log posterior density for prior 1 is not differentiable at the posterior mode. The same holds for the other priors, but for simplicity we focus our attention on prior 1. So we considered several other ways to find a μ and V for our importance function.

(i) *Sequential Monte Carlo.* We started with rough estimates of the posterior mean and covariance matrix for μ and V, and applied the Monte Carlo approach in an iterative fashion. The posterior results of the first stage were used to construct the importance function of the second stage, and so on. This is the procedure we advocated in Kloek and Van Dijk (1978). Now we have more experience with it and we have found that the convergence is often slow, sometimes very slow. The explanation for this phenomenon is as follows. If the importance function is a poor approximation of the posterior density, the variation in the ratios $p'(\theta_i)/I(\theta_i)$ for $i=1,\ldots,N$ may be very large. As a result, formula (3) is dominated by a small number of large values of $p'(\theta_i)/I(\theta_i)$ and hence the approximation error may be large, unless the sample size N is excessively large. This is particularly the case for the second-order moments. And if the posterior covariance matrix is not well approximated, the same phenomenon appears again in the next stage. Given this conclusion we looked for alternative approaches, which will be described now.

(ii) *Exact posterior mode.* Given our uniform prior and the fact that the κ-function has no interior maximum in the a priori acceptable region, the exact posterior mode is on the boundary of that region. It turns out that it can be obtained by maximizing κ under the equality constraint $\beta_1=0$. The Hessian of $-\log\kappa$, evaluated in that point, is positive definite and hence its inverse can be used for V. But this may not be the rule in similar cases.

(iii) *Normal approximation of prior.* We approximated our prior by an $N(\theta_0, V_0)$ density with means and standard deviations as specified in the third line of table 1. The covariances of the pairs (α_1,α_2), (β_1,β_2), (γ_1,γ_2) were set at $-1/36$ (this follows from the constraints of the type $\alpha_1+\alpha_2<1$) and the remaining covariances were taken to be zero. We then obtained an *information contract curve* by maximizing

$$\ln\kappa(\theta)-\tfrac{1}{2}k(\theta-\theta_0)'V_0^{-1}(\theta-\theta_0), \tag{6}$$

for a sequence of values for the scalar k. This procedure is similar to the approach suggested by Dickey (1975) and Leamer (1978). (The former uses the term 'curve décolletage'.) We chose a value for k such that the resulting approximate posterior mode was just inside the unit region. Additional details concerning this approach are given in the appendix.

(iv) *Uniform prior with polynomial transitions.* In this case we return to the idea of the uniform prior but we replace the discontinuities in the density by fifth-degree polynomials in order to obtain a log posterior with continuous derivatives of the first and second orders. An example of such a polynomial transition is

$$\begin{aligned} D(r) &= 0 && \text{if } \; r \leqq -1, \\ &= \frac{3}{16}r^5 - \frac{5}{8}r^3 + \frac{15}{16}r + \frac{1}{2} && \text{if } \; -1 \leqq r \leqq 1, \\ &= 1 && \text{if } \; 1 \leqq r, \end{aligned} \tag{7}$$

where $r = \alpha/\delta$, α denotes an arbitrary parameter and δ a small positive number. Formula (7) approximates the zero-one transiton at $\alpha = 0$. [Recently, the same polynomial was used by Tishler and Zang (1979) in the context of switching regressions.] The polynomial for the one-zero transition at $\alpha = 1$ can easily be derived. We experimented with several values of δ and obtained the modes of the corresponding posterior densities by numerical optimization.

We computed posterior moments of θ (based on prior 1) by means of (3) using Student importance functions (4) with $\lambda = 1$, $\phi^2 = 1$, while we used the approaches (i)–(iv) described above to find μ and V. For each case we computed the coefficients of variation as described in Kloek and Van Dijk (1978, sect. 6). The results for the cases (ii)–(iv) are shown in the first three lines of table 3. The lower half gives comparable results for $\lambda = 5$, $\phi^2 = 2$. For the exact posterior mode (ii) the results were not very good. A tentative

Table 3

Variation coefficients ($\times\sqrt{N}$) of the posterior means of θ and of the reciprocal of the normalizing constant.[a]

	Approach	α_1	α_2	α_3	β_1	β_2	β_3	γ_1	γ_2	γ_3	1
		Case 1: $\lambda = 1$, $\phi^2 = 1$									
(ii)	Exact posterior mode	1.64	1.13	0.15	1.72	0.53	0.65	0.68	1.01	0.84	2.46
(iii)	Normal approximation	2.22	2.13	0.17	2.34	0.63	0.62	0.50	0.75	0.70	3.28
(iv)	Polynomial transitions	1.18	0.89	0.11	1.83	0.31	0.42	0.40	0.53	0.51	1.68
(v)	Normal approximation[b]	1.25	0.91	0.11	1.58	0.32	0.48	0.52	0.65	0.53	1.71
		Case 2: $\lambda = 5$, $\phi^2 = 2$									
(ii)	Exact posterior mode	1.80	0.85	0.13	2.04	0.38	0.56	0.65	0.66	0.70	2.38
(iii)	Normal approximation	1.78	1.23	0.14	1.91	0.41	0.50	0.49	0.77	0.59	2.59
(iv)	Polynomial transitions	1.21	0.78	0.10	1.42	0.31	0.53	0.55	0.76	0.48	1.53
(v)	Normal approximation[b]	0.97	0.66	0.07	1.19	0.26	0.36	0.35	0.41	0.50	1.08

[a]The computations reported here were based on prior 1 and $N = 10{,}000$. The reciprocal of the normalizing constant is obtained by substituting $g = 1$ in (3).
[b]Second step.

explanation is that this approach disregards our prior information in the stage of importance function construction. For the normal approximation (iii) the results are sometimes better, but more often even worse. The problem with normal prior densities is that they are very informative in an undesired way. (We already discussed this point in section 3 for a one-dimensional example. The multivariate case creates more complications.) As a consequence, the resulting approximate posterior contained larger approximation errors than desired. In this respect the uniform prior with polynomial transitions (iv) with $\delta=0.1$ gave better results.

(v) *Two-step normal approximation of prior.* Finally we combined in a sense the approaches described under (i) and (iii), as follows. First we computed posterior moments using the importance function described under (iii). Then we used the resulting posterior mean vector and covariance matrix as the moments of a new normal approximate prior, and used this as the starting point for again applying approach (iii). In this way the sample information is used twice, so that the first-step prior implicitly becomes less informative. Note that we only used this approach to construct an importance function. Our original prior was maintained in the posterior density $p'(\theta_i)$. It is seen from table 3 that the results improved considerably compared with approach (iii). In case 1($\lambda=1$, $\phi^2=1$) it appears that (iv) is equally good or slightly better than (v); in case 2 ($\lambda=5$, $\phi^2=2$) (v) seems to be preferable.

This leads us to the problem of the optimal values of ϕ and λ. We computed a large number of cases but the results are ambiguous. The surfaces of the variation coefficients, interpreted as functions of ϕ and λ, appear to be rather flat. It is clear that $\lambda=1$, $\phi^2=1$ is not an optimal combination, but for values of λ between 2 and 15 there is no clear optimum value; neither is there for ϕ^2 values between 1.8 and 2.4. It should be added that we did not compute the variation coefficients of the variation coefficients. It might be the case that the latter are much less reliably computed than the posterior means.

Since (iv) is simpler to compute than (v), it is, as far as our presently available evidence goes, to be preferred. More experience is needed, however, before more definitive conclusions are possible.

This concludes our discussion of the importance functions we constructed for Klein's model. We add some remarks which may be applicable in more general situations. If the posterior density is bimodal or multimodal, the most appropriate specification seems to be a mixture of multivariate Student densities. (One might also consider the use of mixtures in case of a unimodal but very skew posterior.) A complication of this approach is that the normalizing constants of (4) can no longer be disregarded. In order to identify regions where the ratio $p'(\theta_i)/I(\theta_i)$ is large, one might try to apply cluster analysis. More research on this topic is needed.

5. Posterior results

In this section we present the posterior means and standard deviations of the structural parameters θ (table 4), the short-run multipliers (table 5), the long-run multipliers (table 6), the period of oscillation and the dominant root (table 7). Furthermore we present the posterior probabilities of the states of the system (table 8) and the marginal posterior densities of the structural parameters (fig. 1) and of the multipliers in the reduced form and final form equations for national income (fig. 2). The dotted vertical lines in the figures

Table 4

Posterior means and standard deviations of structural parameters.

	α_1	α_2	α_3	β_1	β_2	β_3	γ_1	γ_2	γ_3
FIML (no prior)	−0.23	0.39	0.80	−0.80	1.05	0.15	0.23	0.28	0.23
	(0.58)	(0.30)	(0.04)	(0.84)	(0.42)	(0.05)	(0.09)	(0.06)	(0.06)
Prior 2	0.12	0.19	0.79	0.06	0.64	0.15	0.34	0.23	0.19
	(0.08)	(0.08)	(0.04)	(0.06)	(0.10)	(0.03)	(0.05)	(0.05)	(0.04)
Prior 8	0.24	0.06	0.72	0.14	0.56	0.20	0.37	0.24	0.19
	(0.08)	(0.04)	(0.05)	(0.09)	(0.12)	(0.05)	(0.04)	(0.04)	(0.05)

Table 5

Posterior means and standard deviations of short-run multipliers (SRM).

Prior information	C	I	W_1	P	Y	K
	SRM values of government wage expenditure (W_2)					
FIML (no prior)	0.81	−0.31	0.12	0.38	1.50	−0.31
Prior 2	1.24	0.06	0.44	0.86	2.30	0.06
	(0.12)	(0.06)	(0.10)	(0.11)	(0.16)	(0.06)
Prior 8	1.35	0.14	0.55	0.95	2.49	0.14
	(0.17)	(0.11)	(0.12)	(0.17)	(0.26)	(0.11)
	SRM values of business taxes (T)					
FIML (no prior)	0.24	0.41	0.15	−0.51	−0.36	0.41
Prior 2	−0.23	−0.08	−0.11	−1.21	−1.32	−0.08
	(0.15)	(0.08)	(0.07)	(0.13)	(0.20)	(0.08)
Prior 8	−0.58	−0.22	−0.29	−1.51	−1.81	−0.22
	(0.22)	(0.16)	(0.13)	(0.22)	(0.34)	(0.16)
	SRM values of government non-wage expenditure (G)					
FIML (no prior)	0.01	−0.38	0.15	0.48	0.62	−0.38
Prior 2	0.58	0.07	0.57	1.09	1.65	0.07
	(0.14)	(0.07)	(0.13)	(0.12)	(0.19)	(0.07)
Prior 8	0.87	0.19	0.76	1.31	2.06	0.19
	(0.19)	(0.14)	(0.14)	(0.21)	(0.30)	(0.14)

Table 6

Posterior means and standard deviations of long-run multipliers (LRM).

Prior information	C	I	W_1	P	Y	K
	LRM values of government wage expenditure (W_2)					
FIML (no prior)	1.57	0	0.82	0.76	2.57	1.28
Prior 2	1.87	0	1.06	0.81	2.87	3.86
	(0.16)		(0.10)	(0.09)	(0.16)	(2.37)
Prior 8	1.63	0	0.99	0.64	2.63	2.39
	0.16)		(0.10)	(0.09)	(0.16)	(0.67)
	LRM values of business taxes (T)					
FIML (no prior)	−0.30	0	−0.16	−1.14	−1.30	−1.94
Prior 2	−0.73	0	−0.41	−1.32	−1.73	−6.27
	(0.23)		(0.12)	(0.11)	(0.23)	(3.43)
Prior 8	−0.67	0	−0.41	−1.27	−1.67	−4.74
	(0.20)		(0.12)	(0.09)	(0.20)	(1.33)
	LRM values of government non-wage expenditure (G)					
FIML (no prior)	0.96	0	1.02	0.94	1.96	1.59
Prior 2	1.38	0	1.35	1.03	2.38	4.90
	(0.14)		(0.11)	(0.10)	(0.14)	(2.79)
Prior 8	1.25	0	1.37	0.89	2.25	3.30
	(0.13)		(0.10)	(0.09)	(0.13)	(0.88)

Table 7

Posterior moments of the period of oscillation and the modulus of the dominant root.

	Period of oscillation	$\|DRT\|$
FIML (no prior)	34.83	0.76
Prior 2	15.06 (2.90)	0.84 (0.08)
Prior 8	9.61 (0.37)	0.77 (0.08)

Table 8

Posterior probabilities of states.

Prior	Oscillating	Monotone
2	0.9999	0.0001
8	0.9927	0.0073

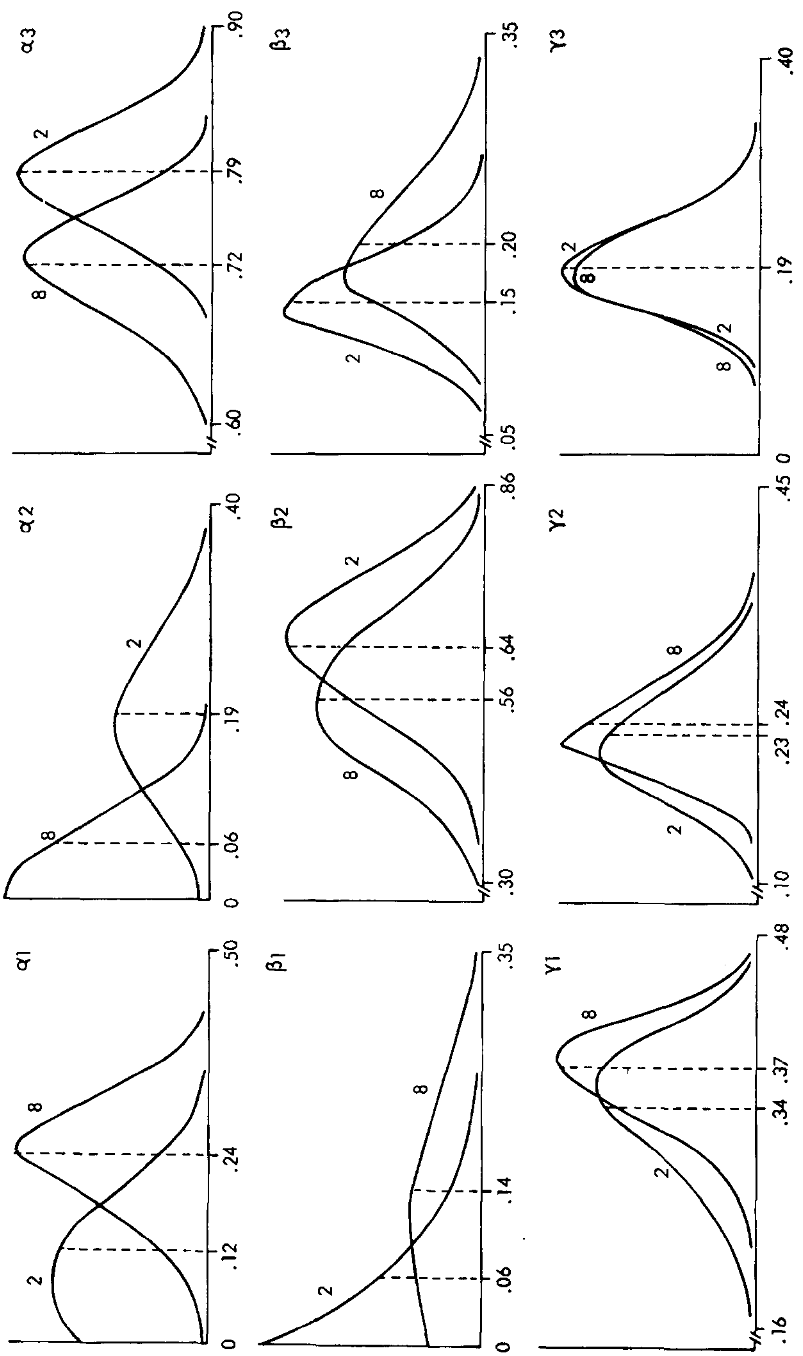

Fig. 1. Posterior densities of structural parameters.

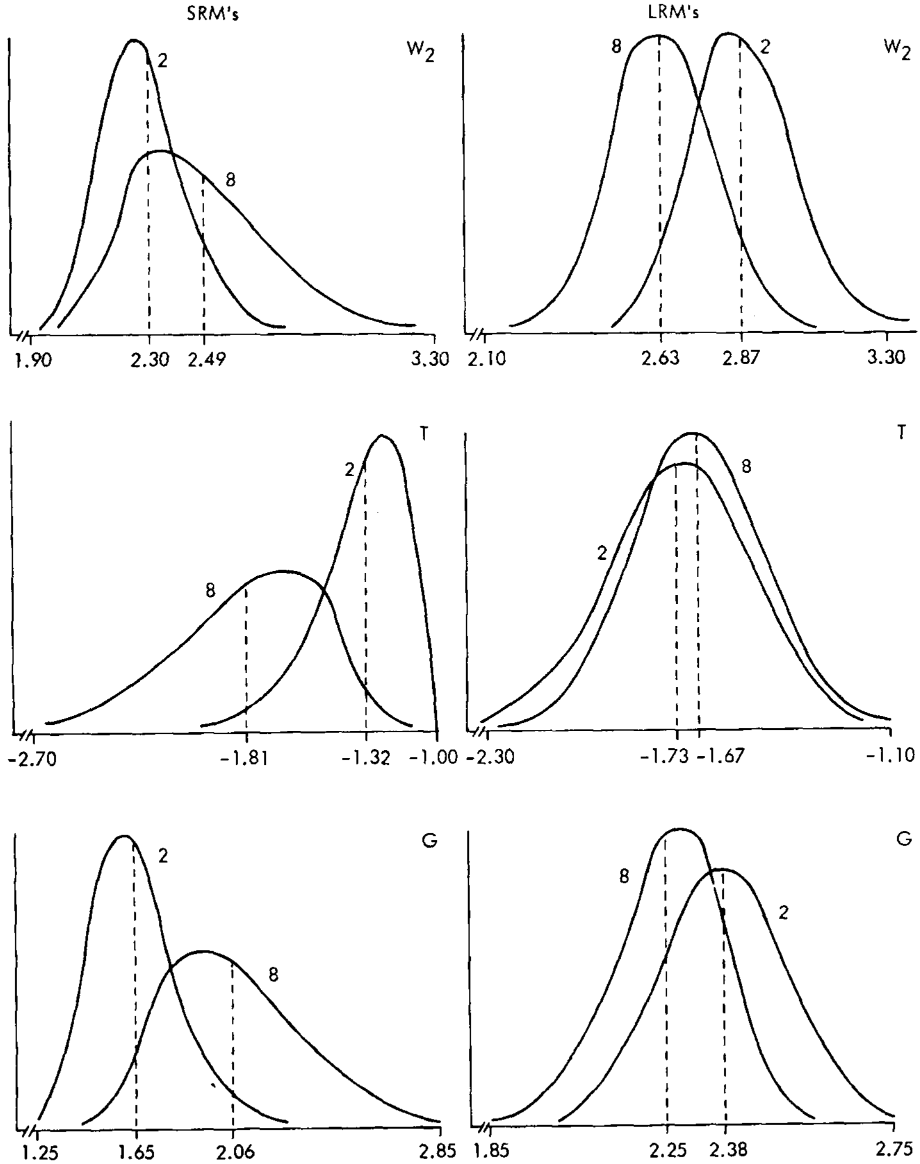

Fig. 2. Posterior densities of multipliers in reduced and final form equations of national income.

indicate the means. In all tables we give the FIML results (with asymptotic standard errors in table 4) for comparison. In all cases we confine ourselves to presenting the results based on priors 2 and 8. The reason is that the differences between the results for priors 1 through 7 for the structural parameters and SRM's and for priors 2, 4, 6 and 7 for the LRM's were very small. We shall discuss this point in more detail below. All results presented are based on $N=30{,}000$. For the construction of the importance functions we made use of an approximate prior, which was uniform with polynomial transitions; compare the preceding section, approach (iv). The results for prior 2 were based on $\lambda=5$, $\phi^2=2$ and those for prior 8 on $\lambda=3$, $\phi^2=2$. In a typical run for a given prior and a given importance function we compute 29 posterior means, 137 posterior second moments and 302 posterior probabilities. The computer time for such a run on a DEC 2050 computer varies from approximately 45 minutes CPU time for results based on prior 2 to approximately 75 minutes for results based on prior 8. This demonstrates the feasibility of the approach.

We start to observe that the FIML estimates of α_1 and β_1 have wrong signs. When analyzing this phenomenon it is found that three factors play a role. First, the data reveal collinearity of P and P_{-1}, which implies that the fit of the investment equation, for example, does not deteriorate much if β_1 decreases while β_2 increases at the same time. Second, there is a positive correlation between the residuals of the consumption and investment functions. If the covariance matrix Σ is postulated to be a diagonal matrix the wrong signs are not observed [Klein (1950)]. This hypothesis is, however, strongly rejected in a likelihood ratio test [$\chi^2(3)=28.46$]. Third, the Jacobian,

$$\|\Gamma\|=|1-(\alpha_1+\beta_1)(1-\gamma_1)-\alpha_3\gamma_1|, \tag{8}$$

is less than or equal to unity in the unit region but equals 1.60 in the FIML point. Recall that a factor $\|\Gamma\|^n$ occurs in the likelihood function and in the κ function. Note that in 3SLS, where the Jacobian factor is absent but the nondiagonal elements of Σ are present, β_1 has a wrong sign but only marginally so ($\hat{\beta}_1=-0.013$) [Theil (1971)]. If β_1 is restricted to be zero, α_1 gets the correct sign and this hypothesis is not rejected in a likelihood ratio test [$\chi^2(1)=3.20$]. So, there is no conflict between the sample information and our prior information which states that α_1 and β_1 should be non-negative. If we compare the FIML asymptotic standard errors and the posterior standard deviations of the structural parameters (table 4) we see that this prior information plays a large role.

Once we have accepted the prior information that all elements of θ are in the unit region, the extra constraints (1, 2, 3 and 4), introduced in section 3, turn out not to be restrictive. Given prior 1 the posterior probability that the system is explosive is 0.021. Given prior 2 the long-run effects in the structural equations are all in the unit interval (table 4). In this respect it should be noted that the relevant covariances are all negative. All SRM's (table 5) and LRM's (table 6) amply satisfy the upper bound constraints. They also satisfy the sign constraints, though some are close to zero. In these cases the posterior densities (not shown here in order to save space) turn out to be skew so that the probability of wrong signs is extremely small, even under prior 1. This explains why the differences between the posteriors are very small.

The only set of extra constraints which adds substantial information to the sample is set (5), which says that the period of oscillation should be between 3 and 10 years. It is seen in tables 5 through 8 and in figs. 1 and 2 that this set, introduced in prior 8, influences almost every parameter. In particular, if α_2 and β_2 are relatively large (which corresponds to negative or small positive values of α_1 and β_1) the lags become large and this, in turn, implies long periods of oscillation (compare tables 4 and 7) and relatively small absolute values of most of the SRM's (table 5).

So we have observed that the prior constraints on the period of oscillation have rather large effects. The question arises whether this information is acceptable. The posterior mean and standard deviation of the period of oscillation under prior 2 (table 7) suggest that the hypothesis of a ten-year period is acceptable, but closer inspection of the marginal posterior density (not shown) reveals that the distribution is positively skew and the posterior probability that the period is ten years or less equals 0.024. This suggests rejection of the prior constraint.

When considering these results we were tempted to look for specification errors and accordingly to respecify the model. But we deliberately refrained from doing so for three reasons. First, the main purpose of this paper is to demonstrate that the integrations can be done and how they are done. Second, so far Bayesian statistics lacks a well developed standard battery of diagnostic checks as has been developed for instance in the context of time-series analysis. Third, the available data set corresponding to Klein's model I (21 annual observations per variable) is rather limited if one would consider to use the traditional diagnostic tests.

6. Conclusions

(1) In this paper we have demonstrated the *feasibility* of nine-dimensional numerical integration by means of Monte Carlo in the context of Bayesian posterior analysis. A condition is that an importance function can be found which is a reasonable approximation of the posterior density. This was the case in the present examples.

(2) We have also given some guidelines for the *construction of importance*

functions. In particular, we have emphasized the importance of optimization in the initial stage, since integration based on an importance function which is a poor approximation tends to give very unreliable results. If the latter approach is repeated in an iterative fashion, one may encounter very slow convergence. We confined ourselves to using multivariate Student densities. The search for more flexible alternatives is an area for future research.

(3) Finally, we have shown how Monte Carlo may be used as a tool for *elicitation of prior information.* In our particular example we investigated how our initial prior information on structural parameters was modified by specifying prior information on multipliers and the period of oscillation. In particular, we confined ourselves to prior information which is uniform on a region in a space of nine structural parameters. The bounds of this region consist of inequality constraints on (functions of) these structural parameters. Many more examples can be thought of in cases where other types of prior information are available. A practical restriction is that for some types of prior information generating random variables by means of Monte Carlo may require complicated computer programs. There is ample room for experimentation in this area.

Appendix

In this appendix we present some numerical results on the information contract curves discussed in section 4. The first approach is to combine the normal approximation of the uniform prior, denoted by $N(\theta_0, (1/k)V_0)$, with a normal approximation of the likelihood based on the FIML point estimate and the corresponding asymptotic covariance matrix. In this case one can make use of formula (5.27) in Leamer (1978). Computationally this expression is simple.

The second approach is to combine the normal approximation to the prior with the exact κ-function. Then (6) is maximized numerically with respect to θ for several values of k by means of a variable metric method [see Broyden (1972)], and the references cited there]. Given the availability of good software these optimizations take only a few seconds CPU-time on modern computers. Therefore the second approach is computationally easy as well (though somewhat more time consuming than the first). It may have the advantage of being a more accurate approximation to the posterior.

Six information contract curves have been drawn in fig. 3 for three-parameter combinations for the case of prior 1. The dotted line indicates that the first approach has been used. The drawn line indicates the use of the exact approach. We have chosen the moments of prior 3 of section 3 as moments of the normal approximation to the uniform prior on the unit region. The more obvious choices of 0.5 as prior mean for each element of θ_0 were rejected because this turned out to be informative in an undesired way.

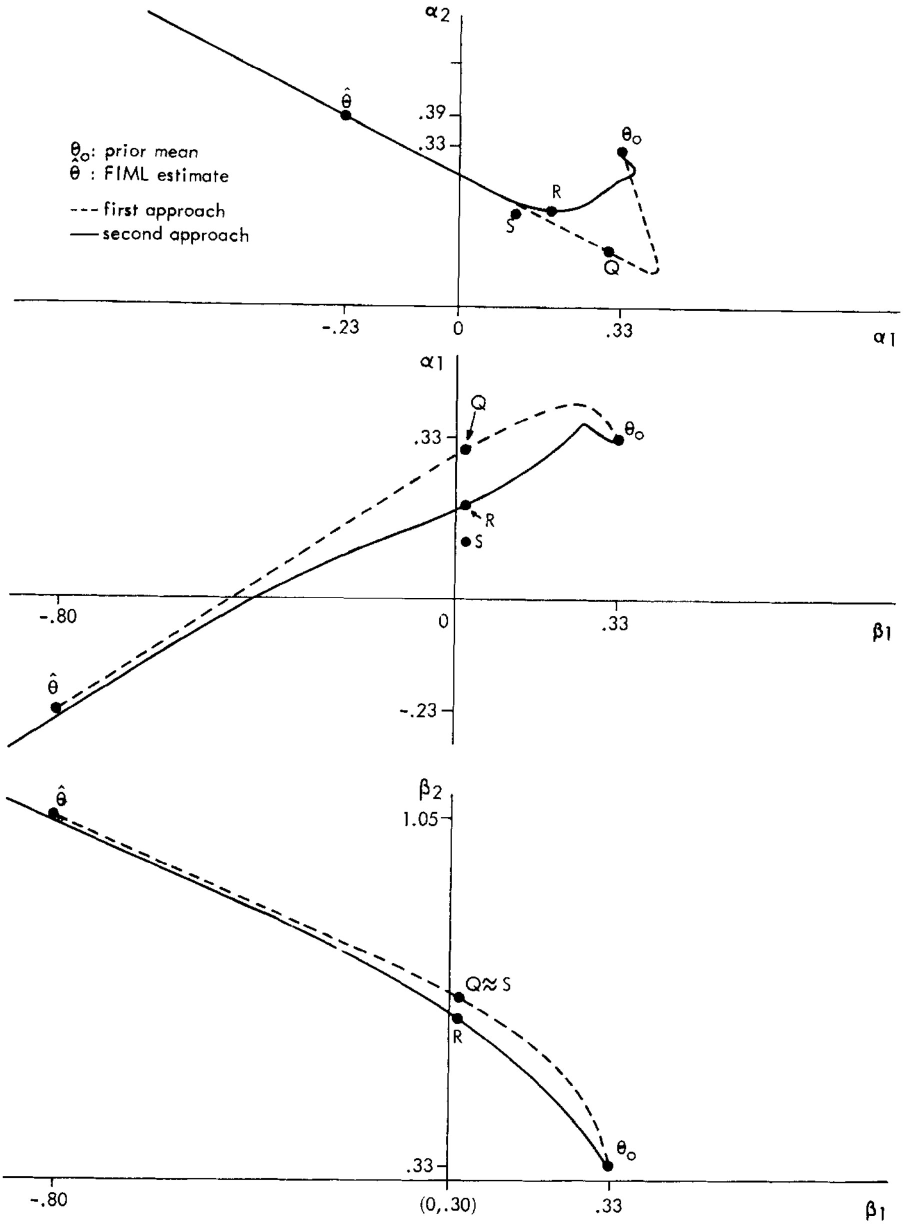

Fig. 3. Information contract curves for (α_1, α_2), (α_1, β_1) and (β_1, β_2).

Table 9
Three points on three information contract curves.[a]

	α_1	α_2	α_3	β_1	β_2	β_3	γ_1	γ_2	γ_3
Q	0.30	0.12	0.78	0.01	0.68	0.17	0.29	0.25	0.21
R	0.18	0.19	0.76	0.01	0.65	0.16	0.33	0.23	0.22
S	0.11	0.19	0.79	0.01	0.69	0.16	0.34	0.22	0.19
Posterior means (prior 1)	0.12	0.19	0.79	0.06	0.64	0.15	0.34	0.23	0.19

[a]The third information contract curve on which S is situated is not drawn in fig. 3 since it is very close to the second approach. The posterior means are given for comparison.

The FIML estimate $\hat{\theta}$ is given for comparison. In table 9 the cases Q, R and S represent points (on three information contract curves) which are just inside the unit region. The case of the asymptotic approximation to the κ function is represented by Q, the exact approach by R and the two-step approach with exact κ function, discussed in section 4, is represented by S. These points have been drawn in fig. 3. Summarising the results, the available evidence indicates that the exact approach is more accurate than the asymptotic approach [compare $\alpha_1 = 0.30$ and $\alpha_1 = 0.18$ with $E(\alpha_1) = 0.12$] and secondly that a two-step approximation procedure is better than a one-step, since the normal approximation to a uniform prior is still informative in an undesired way (the point R is pulled too far towards θ_0 for the case of α_1). Finally, note that the constraint $\beta_1 > 0$ is binding and that the mode of the marginal posterior density is zero in this case.

References

Atkinson, A.C. and M.C. Pearce, 1976, The computer generation of Beta, Gamma and Normal random variables, Journal of the Royal Statistical Society A 139, 431–448.

Bard, Y., 1974, Nonlinear parameter estimation (Academic Press, New York).

Broyden, C.G., 1972, Quasi-Newton methods, in: W. Murray, ed., Numerical methods for unconstrained optimization (Academic Press, London).

Dickey, J., 1975, Bayesian alternatives to the F-test and least squares estimate in the normal linear model, in: S.E. Fienberg and A. Zellner, eds., Studies in Bayesian econometrics and statistics (North-Holland, Amsterdam).

Historical statistics of the United States, Colonial times to 1970, Bi-centennial edition, 1975 (U.S. Bureau of the Census, Washington, DC).

Kiefer, N.M., 1979, Limited information analysis of two small underidentified macroeconomic models, Report 7929 (Center for Mathematical Studies in Business and Economics, University of Chicago, Chicago, IL).

Klein, L.R., 1950, Economic fluctuations in the United States, 1921–1941 (Wiley, New York).

Kloek, T. and H.K. van Dijk, 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 honor of Harold Jeffreys (North-Holland, Amsterdam).

Leamer, E.E., 1978, Specification searches (Wiley, New York).

Rothenberg, T.J., 1973, Efficient estimation with a priori information (Yale University Press, New Haven, CT).

Theil, H., 1971, Principles of econometrics (Wiley, New York).

Theil, H. and J.C.G. Boot, 1962, The final form of econometric equation systems, Review of the International Statistical Institute 30, 136–152. Reprinted in: A. Zellner, ed., 1968, Readings in economic statistics and econometrics (Little, Brown and Co., Boston, MA).

Theil, H., J.C.G. Boot and T. Kloek, 1965, Operations research and quantitative economics (McGraw-Hill, New York).

Tishler, A. and I. Zang, 1979, A switching regression method using inequality conditions, Journal of Econometrics 11, 259–274.

Uspensky, J.V., 1948, Theory of equations (McGraw-Hill, New York).

Van Dijk, H.K. and T. Kloek, 1977, Predictive moments of simultaneous econometric models, A Bayesian approach, in: A. Aykaç and C. Brumat, eds., New developments in the applications of Bayesian methods (North-Holland, Amsterdam).

Van Dijk, H.K. and T. Kloek, 1978, Posterior analysis of Klein's Model I, Report 7824/E (Econometric Institute, Erasmus University Rotterdam, Rotterdam).

Zellner, A., 1971, An introduction to Bayesian inference in econometrics (Wiley, New York).