



AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
<http://ageconsearch.umn.edu>
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

ECONOMETRIC INSTITUTE

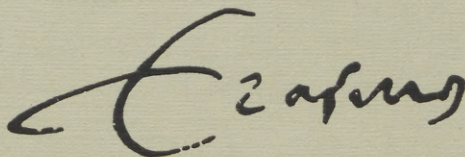
ERASMUS FOUNDATION OF
SOCIAL ECONOMICS
WITHDRAWN

FEB 25 1988

SOME ADVANCES IN BAYESIAN ESTIMATION
METHODS USING MONTE CARLO INTEGRATION

H.K. VAN DIJK

REPORT 8704/A



SOME ADVANCES IN BAYESIAN ESTIMATION METHODS
USING MONTE CARLO INTEGRATION*

by

Herman K. van Dijk**

Abstract

In this paper some Monte Carlo integration methods are discussed that can be used for the efficient computation of posterior moments and densities of parameters of econometric and, more generally, statistical models. The methods are based on the principle of importance sampling and are intended for the evaluation of multi-dimensional integrals where the integrand is unimodal and multivariate skew. That is, the integrand has different tail behavior in different directions. Illustrative results are presented on the dynamic behavior and the probability of explosion of a small scale macro-economic model. This application involves nine-dimensional numerical integration.

Contents

	Page
1. Introduction	1
2. Direct simulation and simple rejection	4
3. Basics of importance sampling	14
4. Prior and posterior analysis of Klein's Model I	17
5. Mixed integration	38
6. Remarks	45
References	46

* I am indebted to A.S. Louter of the Institute of Social Studies in The Hague and to J.P. Hop, J. Martens and J. van Dijk of Erasmus University for their assistance in preparing the necessary computer programs and the list of figures.

** Econometric Institute, Erasmus University Rotterdam and Center for Operations Research and Econometrics (CORE), Université Catholique de Louvain.

1. INTRODUCTION

Two related problems in the development of Bayesian statistical methods for econometric models are the specification of prior information and the efficient computation of the corresponding posterior moments and densities. A particularly difficult problem in this context is the conflict between two - apparently reasonable - requirements on prior information, i.e., analytical tractability and richness. An important advantage of the class of analytically tractable priors is that the integrals defined in the implied posterior moments can be evaluated analytically. In other words, the integrals are known in terms of elementary functions. As a consequence the computations involved are relatively simple. This is important in view of the problems that arise when numerical integration methods have to be applied for the computation of posterior moments in spaces with high dimensionality. An example of a class of analytically tractable prior densities is the natural conjugate family [see Raiffa and Schlaifer (1961)]. However, Rothenberg (1963, 1973, Section 6.4) has pointed out that the natural conjugate family is not rich enough for an important class of econometric models, namely, the class of simultaneous equation models (SEM). Further, the structural parameters of the SEM are not identified in case a natural conjugate prior is used and all structural parameters are unrestricted.

In this paper we do not discuss the formal specification of prior information but assume that a posterior density is available. Our objective is to describe some methods that are computationally efficient and that are flexible enough to allow for a rich set of possible prior densities for the parameters of interest to economists. The price for increased flexibility of the prior densities is that in several cases the integrals defined in the posterior moments and densities of the parameters of econometric models have to be computed by numerical integration methods. Examples include models where the prior is uniform and the likelihood function is skew or where the prior density is informative and skew. This occurs, for instance, in the structural form of a SEM, in a linear regression model with serially correlated errors, in disequilibrium models or, more generally, in nonlinear models where no linear approximation is available [see, e.g., Bauwens (1984), van Dijk and Kloek (1980, 1983, 1985, 1986), Boender and van Dijk (1987), and Kooiman, van Dijk and Thurik (1985)].

Well known numerical integration methods are Cartesian product rules that

are based on Gaussian or Newton-Cotes quadrature formulas [see, e.g., Abramowitz and Stegun (1964)]. The application of such methods appears to be hampered by the amount of computational work involved in dimensions greater than five, say. Consider the example where use is made of a ten-point Gaussian quadrature formula. In K dimensions one has to evaluate 10^K points. When K is greater than 5 or 6 the computational workload is heavy. Due to the advances in modern computer technology the problem of the computational workload may become less important. However, there will still exist a relative advantage in efficient computation for the Monte Carlo methods discussed below. One may argue that a two- or three-point Gaussian quadrature formula can be used instead of a ten-point formula. In such cases one makes quite often use of the Cartesian product rules in an iterative way in order to check the numerical accuracy. That is, after the first round of numerical integration, one makes use of 2^K times the number of points in a second round, and so forth. So, Cartesian product rules suffer from what is sometimes referred to as 'the curse of dimensionality'.

In this paper we make use of Monte Carlo numerical integration methods in order to compute the integrals defined in the posterior moments and the marginal posterior densities mentioned above. The Monte Carlo approach is concerned with experiments on pseudo-random numbers where use is made of a computer¹ [see, e.g., Hammersley and Handscomb (1964)]. Numerical integration is an important area of application of the Monte Carlo approach. Basically, the integrals to be computed are interpreted as expectations of certain random variables in the Monte Carlo approach and the numerical integration problem is changed into a statistical estimation problem.

In econometrics, the Monte Carlo integration approach has been used traditionally for the investigation of properties of the finite sample distributions of classical estimators in the context of a given model and a given point in the parameter space. As a consequence, the results of such Monte Carlo experiments are specific for one parameter point. A recent survey of this application of the Monte Carlo approach, which contains several suggestions for further research, has been given by Hendry (1984).

We make use of Monte Carlo integration in a different way than the approach mentioned above. First, there is a conceptual difference. Monte Carlo

1. We note that there exist also physical devices that generate pseudo-random numbers. However, modern use of Monte Carlo integration methods involves usually a computer procedure for the generation of pseudo-random numbers.

applied in classical estimation implies integration over the data space (drawing repeatedly samples of artificial economic data). Monte Carlo applied in Bayesian estimation implies integration over the parameter space. Second, in our case the values of the parameters of interest of an econometric model are not known. Third, the present analysis is empirical in the sense that we do not generate artificial economic data.

Our way of using Monte Carlo may be succinctly stated as follows. The starting point is the specification of a so-called importance function. This is a density function defined on the space of the structural parameters (or on the space of a subset of these parameters in case part of the integration is carried out analytically). There are two requirements on this importance function. It should have convenient Monte Carlo properties in the sense that it is relatively easy to generate pseudo-random drawings from a probability distribution with a density function that is equal (or proportional) to the importance function. In addition, the importance function should be a good approximation of the posterior density. One can describe the basic steps of Monte Carlo integration as follows. A sample of random drawings of parameters of interest θ is drawn from the distribution mentioned above.¹ Each random drawing and each function of the random drawing, where one is interested in, are multiplied by the ratio of the posterior density and the importance function. This ratio serves as a weight function. Then one computes posterior moments and densities by means of simple formulas that are based on standard sampling theory. For more details we refer to Section 3.

The important advantage of Monte Carlo is that a large number of posterior moments can be estimated at a reasonable computational effort. For instance, in a single Monte Carlo integration procedure one is able to compute the posterior first-order moments, second-order moments, univariate marginal densities and bivariate marginal densities of a vector of parameters of interest and of a vector (or matrix) of nonlinear functions of the parameters of interest. For some illustrative results using a nine-dimensional vector we refer to Section 4. There are several indications that Monte Carlo is computationally efficient in problems with many dimensions, say more than five or six. The basic reason is that Monte Carlo is a sampling method and hence the error goes to zero as $N^{-\frac{1}{2}}$, where N is the number of sample points. This rather informal statement is explained in Section 2.

1. We shall in most cases use the term random drawing instead of the more accurate but tedious expression of pseudo-random drawing.

The contents of this paper is organized as follows. In the next section we discuss the basic computational steps that are part of most Monte Carlo integration procedures by means of two simple examples. In Section 3 we summarize some basic concepts of importance sampling. Illustrative results using this method are presented in Section 4 for Klein's Model I, which involves a nine-dimensional numerical integration problem. Section 5 contains a description of an algorithm that is based on a combination of one-dimensional Gaussian quadrature and importance sampling, which we have named mixed integration. Some suggestions for further work are given in Section 6.

2. DIRECT SIMULATION AND SIMPLE REJECTION

In this section we discuss two elementary Monte Carlo (MC) integration methods. Our purpose is to illustrate the sequence of computational steps of most MC integration methods through simple examples. Further, we give a listing of a computer program that is intended as an introduction to the more complex computer programs for importance sampling and mixed integration.

The multivariate integrals that we consider may be described briefly as follows. Let θ be an l -vector of parameters of interest and let $g(\theta)$ be an integrable function of θ . The posterior mean of $g(\theta)$ is defined as

$$Eg(\theta) = \frac{\int g(\theta)p(\theta)d\theta}{\int p(\theta)d\theta} \quad (2.1)$$

where $p(\theta)$ is a kernel of a posterior density function. That is, $p(\theta)$ is proportional and not equal to a density function and the denominator of (2.1) plays the role of integrating constant, similar to the role of $\sqrt{2\pi}$ in the case of the normal distribution. Simple examples of $g(\theta)$ are $g(\theta) = \theta$ and $g(\theta) = \theta\theta'$. Note that g may be a vector or a matrix. We emphasize that $g(\theta)$ may also be a complicated nonlinear function of θ such as the implied multipliers of the structural parameters of a simultaneous equation model [see, e.g., van Dijk and Kloek (1980) and Section 4]. There exist several other examples of nontrivial nonlinear functions of θ . For an example in the statistical literature we refer to Kass (1985), and for some examples in the econometric literature we refer to van Dijk (1985), Zellner (1985), and Geweke (1986).

Consider the problem of the computation of the integrals in equation (2.1) for the case where a computer procedure is available that enables one to generate a sample of pseudo-random drawings from a distribution function $F(\theta)$

with a density function equal (or proportional) to $p(\theta)$. Let $\theta^{(1)}, \dots, \theta^{(N)}$ denote the generated random sample. Given that $g(\theta)$ has a certain regularity property (i.e., g is measurable), it follows that $g(\theta^{(1)}), \dots, g(\theta^{(N)})$ is also a random sample. Then we may approximate the posterior mean (1.1) by the sample mean, which is defined as

$$\bar{g} = \frac{1}{N} \sum_{i=1}^N g(\theta^{(i)}) \quad (2.2)$$

The computation of $Eg(\theta)$ by means of this procedure is referred to in the literature as direct simulation, since one is able to generate a random sample directly from the distribution studied by making use of a computer procedure that simulates a sequence of random numbers (otherwise stated, the computer procedure generates a sequence of pseudo-random numbers). For more examples on direct simulation and for references on computer procedures that generate sequences of pseudo-random numbers for many families of distributions we refer to, Hammersley and Handscomb (1964, Chapter 3), Newman and Odell (1971), Atkinson and Pearce (1976), Kinderman and Ramage (1976), Kinderman and Monahan (1980), Rubinstein (1981, Chapter 3), Marsaglia (1984), Ripley (1983), Bauwens (1984), and the references cited there.

A flow diagram for direct simulation is given in Figure 1. Note that we make use of an arrow sign instead of an equality sign in Figure 1. For instance, one interprets $S^{(0)} \leftarrow 0$ as: 'the value zero is assigned to the variable $S^{(0)}$ '. Apart from the computer procedure that generates $\theta^{(i)}$, the basic computational steps are given as

$$S^{(0)} \leftarrow 0 \quad (2.3)$$

$$S^{(i)} \leftarrow S^{(i-1)} + g(\theta^{(i)}) \quad (i = 1, \dots, N) \quad (2.4)$$

$$\bar{S}_N \leftarrow \frac{S^{(N)}}{N} \quad (2.5)$$

The symbol $S^{(0)}$ stands for the initial zero-value of the sum of the sequence of random variables $g(\theta^{(1)}), \dots, g(\theta^{(N)})$; $S^{(i)}$ denotes the i -th partial sum of this sequence, given as $S^{(i)} \leftarrow g(\theta^{(1)}) + g(\theta^{(2)}) + \dots + g(\theta^{(i)})$. Equation (2.4) and Figure 1 illustrate that one does not have to store the large set of (pseudo-) random numbers $g(\theta^{(1)}), g(\theta^{(2)}), \dots, g(\theta^{(i)})$ in a computer as is suggested by the formula for the i -th partial sum, but one can make repeatedly

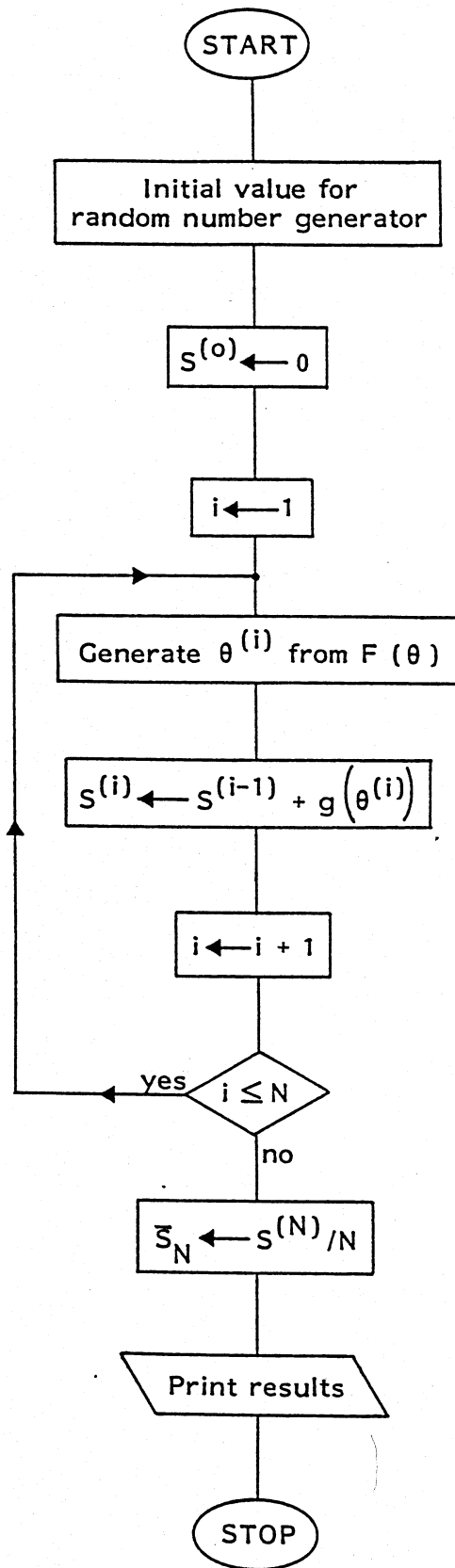


Figure 1. Flow diagram for direct simulation

use of a computer procedure that generates a pseudo-random number.

The accuracy of the approximation (2.2) may be studied by increasing the size of the sample from N to $2N$, $3N$, ..., MN . The results can be printed at each value of jN , with $j = 1, \dots, M$. Given certain regularity conditions, the Monte Carlo estimator \bar{g} converges with probability one to $Eg(\theta)$. Some examples of this convergence process are presented below.

An other measure of numerical accuracy can also be derived from large sample theory. Under certain regularity conditions it follows from central limit theory that the estimator \bar{g} , equation (2.2), is approximately normally distributed with mean $Eg(\theta)$ and variance σ_g^2/N , where σ_g^2 is given as

$$\sigma_g^2 = Eg^2(\theta) - [Eg(\theta)]^2 \quad (2.6)$$

Under the assumption that the integrals in (2.6) exist, one may estimate σ_g^2 by the sample variance

$$\hat{\sigma}_g^2 = \frac{1}{N} \sum_{i=1}^N g^2(\theta^{(i)}) - \bar{g}^2 \quad (2.7)$$

Given an estimator $\hat{\sigma}_g^2$ for σ_g^2 , one can define a 95 percent confidence interval for $Eg(\theta)$ in the usual way as $[\bar{g} - 1.96\hat{\sigma}_g/\sqrt{N}, \bar{g} + 1.96\hat{\sigma}_g/\sqrt{N}]$. For an introduction to the sampling theory results that we use, we refer to Mood, Graybill and Boes (1974, Chapters 2 and 6) and for a more advanced treatment we refer to Cramér (1946, Chapters 25 and 27). An example of a sequence of confidence intervals is presented below.

In order to illustrate the computational steps of an MC integration method we consider the computation of a truncated five-dimensional standard normal integral by means of MC.¹ So, in this case we have $l = 5$, $p(\theta)$ is equal to a multivariate standard normal density function, and

$$\begin{aligned} g(\theta) &= 1 && \text{if } \theta < a \\ &= 0 && \text{elsewhere} \end{aligned} \quad (2.8)$$

where a is a five-dimensional vector of known constants. Direct simulation is not a suitable integration method for this problem. An other Monte Carlo integration method for this problem may be formulated as follows:

1. We note that in this case the value of the integral can be determined by making use of the table of the standard normal integral. More comments are given at the end of this section.

Figure 2.

```

c -----
c -----
c
c A Fortran-77 program for the evaluation of a truncated
c multivariate normal integral.
c
c ndim      : the dimension of the multivariate normal integral,
c mround    : the number of times that an intermediate result is
c             printed,
c n         : the number of random drawings of the multivariate normal
c             distribution for each round,
c nacc      : the number of accepted random drawings of the
c             multivariate normal distribution for each round,
c ncum      : cumulative number of random drawings,
c nacum     : cumulative number of accepted random drawings.
c value     : computed value of integral in each round
c cumvalue  : computed value of integral after jrounds
c
c
c THETA(ndim) : a random drawing of the multivariate normal
c               distribution,
c BOUND(ndim) : a vector with upper bounds for the elements of
c               THETA.
c
c GENERATE(THETA,ndim) : a procedure that returns in THETA a random
c                       drawing taken from a ndim-dimensional
c                       multivariate normal distribution,
c TEST(THETA,BOUND,ndim,tacc): a procedure that assigns the value
c                               .true. to tacc if the values of vector THETA are
c                               smaller then the upperbounds in vector BOUND,
c                               else tacc becomes .false.
c
c -----
c
c integer ndim,mround,n
c parameter (ndim=5, mround=100, n=10000)
c
c logical tacc
c integer nacc,jround,i,ncum,nacum
c real*8 THETA(ndim),BOUND(ndim),value,cumval,p,q,ub,lb
c
c data BOUND/1.0d0,0.0d0,2.0d0,-1.0d0,-2.0d0/
c
c p = 0.0014838428435809d0
c q = 1.0d0 - p
c
c 1001 format('0','The evaluated integral for round',i3,' is: ',f8.6,
c &         ' the cumulative value of the integral is: ',f8.6,', (' ,
c &         f8.6,',',f8.6,')')
c open(unit=6,file='[e.ect.hkvdijk]example1.res',status='new')
c
c initial value of the pseudo-random number generator
c
c call GO5CBF(100000*ndim)

```

```

c
c   initial zero value of number of accepted random drawings at round zero
c
c   nacum = 0
c
c   do jround = 1,mround
c
c   initial zero value of number of accepted random drawings for each round
c
c       nacc = 0
c
c       do i = 1,n
c           call GENERATE(THETA,ndim)
c           call TEST(THETA,BOUND,ndim,tacc)
c           if (tacc) nacc = nacc + 1
c       end do
c       value = dfloat(nacc) / dfloat(n)
c
c       nacum = nacum + nacc
c       ncum = jround * n
c       cumval = dfloat(nacum) / dfloat(ncum)
c
c       lb = p - 1.9599639177322388d0 * dsqrt(p * q / dfloat(ncum))
c       ub = p + 1.9599639177322388d0 * dsqrt(p * q / dfloat(ncum))
c       write(6,1001) jround,value,cumval,lb,ub
c   end do
c
c   stop
c   end
c
c -----
c
c   subroutine GENERATE(THETA,ndim)
c   integer ndim,i
c   real*8 THETA(ndim)
c   do i = 1,ndim
c       THETA(i) = G05DDF(0.0d0,1.0d0)
c   end do
c
c   return
c   end
c
c -----
c
c   subroutine TEST(THETA,BOUND,ndim,tacc)
c   logical tacc
c   integer ndim,i
c   real*8 THETA(ndim),BOUND(ndim)
c   tacc = .true.
c   do i = 1,ndim
c       if (THETA(i) .gt. BOUND(i)) then
c           tacc = .false.
c       return
c   end if
c   end do
c
c   return
c   end

```

"Generate a vector of unrestricted normal random variables and test whether such a vector satisfies the restrictions of (2.8). Suppose that N_1 random drawings (out of a total of N drawings) satisfy the restrictions. Then we estimate the value of the integral P by means of $\hat{P} = N_1/N$."

This method of computing a multivariate integral is an example of a Monte Carlo integration method known as simple rejection. [A standard example of simple rejection is given in, e.g., Rubinstein (1981, pp. 115-116).] The rejection step can be inserted in a simple way in the flow diagram of Figure 1 after the step where a random vector θ is generated from $F(\theta)$. A computer program, written in Fortran-77 is listed in Figure 2 and the results of some experiments are presented in Figures 3 and 4. The computer program illustrates the basic computational steps of most Monte Carlo integration procedures. That is, it starts with some statements that refer to initial values, in particular, the initial value of a random number generator and initial zero-values. [We make use of the normal random number generator from Brent (1974), given as NAG-subroutine G05DDF.] The central part of the program refers to two so-called do loops. The inner loop, with the index i , refers to the computational steps given in Figure 1 and equations (2.2)-(2.4). [Note that in our program an estimate of the probability P in each round is denoted by the term value and the number of successes is denoted by the term nacc.] The outer loop refers to the number of times that intermediate results are printed. Let the index j refer to the number of times that a sample of size N is generated, with $j = 1, \dots, M$ [j is labeled as jround in the program in order to avoid possible confusion between the integers i and j]. Let \hat{P}_j denote the estimated value of the integral P after j samples of size N . The values \hat{P}_j , $j = 1, \dots, M$, are related in a recursive way, i.e., we can write

$$\hat{P}_0 = 0 \quad (2.9)$$

$$j\hat{P}_j = (j-1)\hat{P}_{j-1} + \bar{S}_{N,j} \quad (j = 1, \dots, M) \quad (2.10)$$

where

$$\bar{S}_{N,j} = \frac{s(jN) - s(j-1)N}{N} \quad (2.11)$$

and $s(jN)$ is defined in equations (2.3)-(2.4). Note that for $j=1$, it follows that (2.11) is equal to (2.5). Using (2.9) and (2.10), it is seen that \hat{P}_j can be computed in a recursive way as

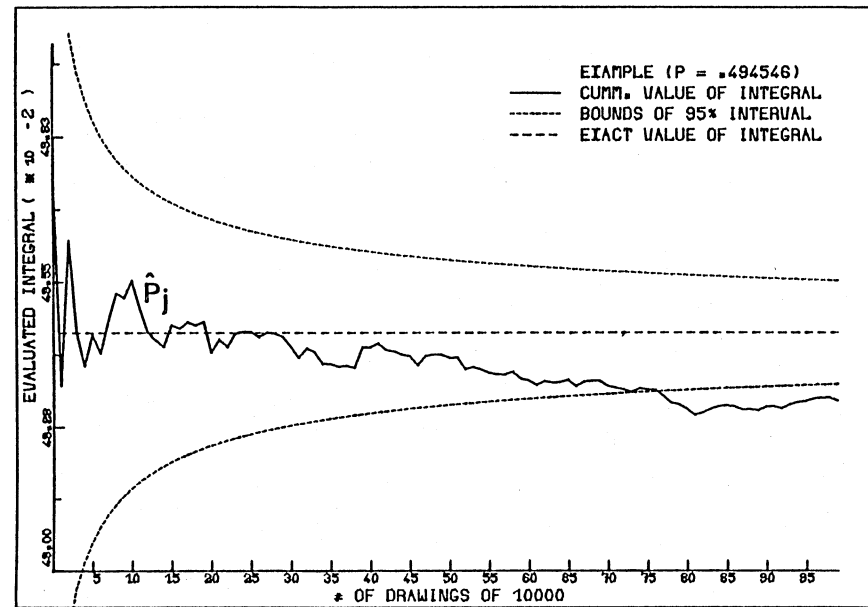
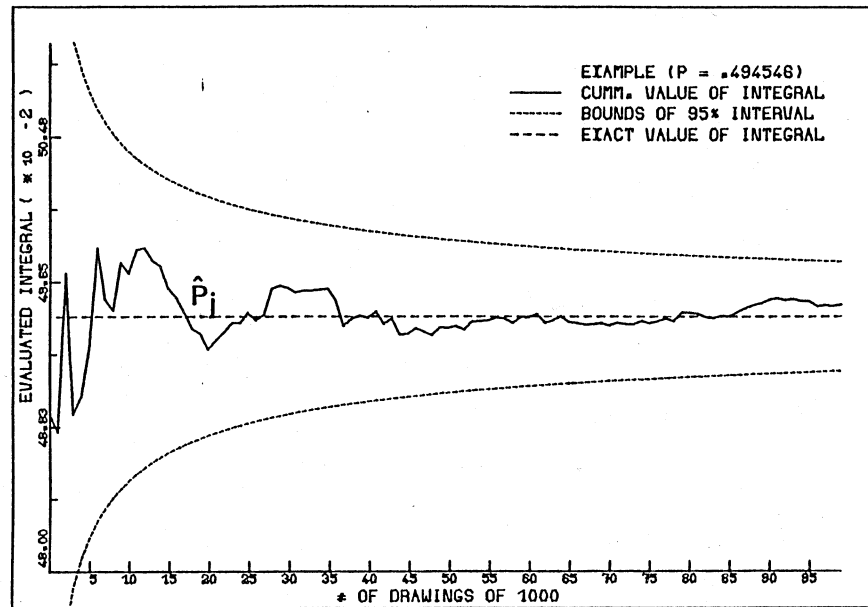
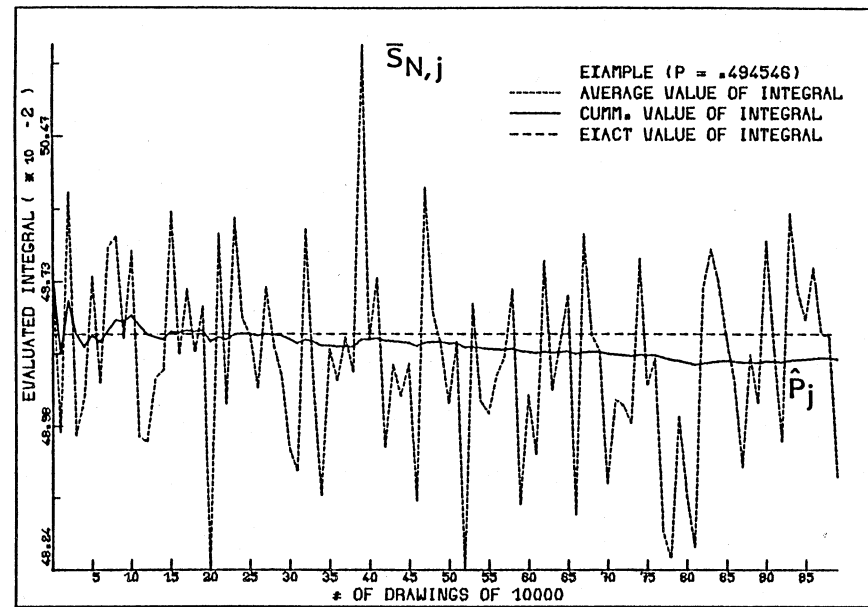
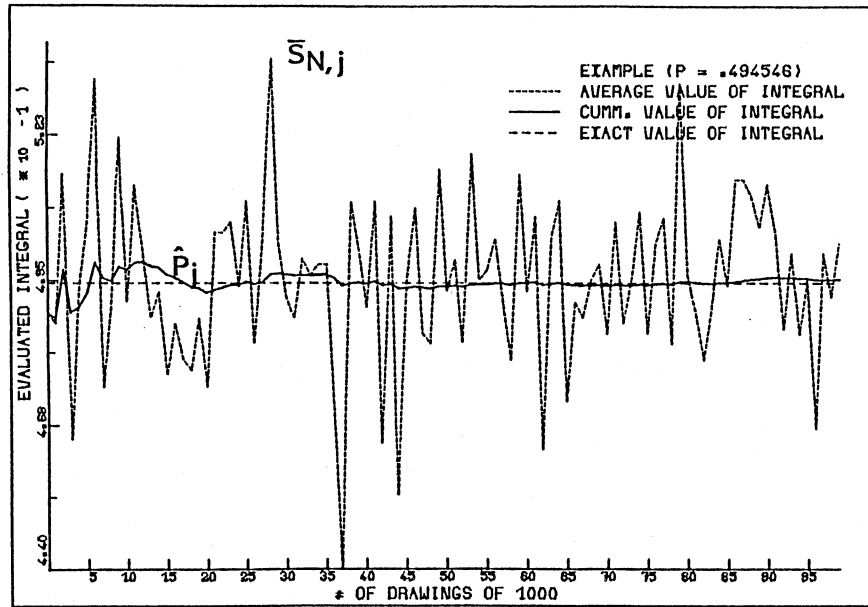


Figure 3. Sequences of estimates of the truncated multivariate normal integral $P = .49546$

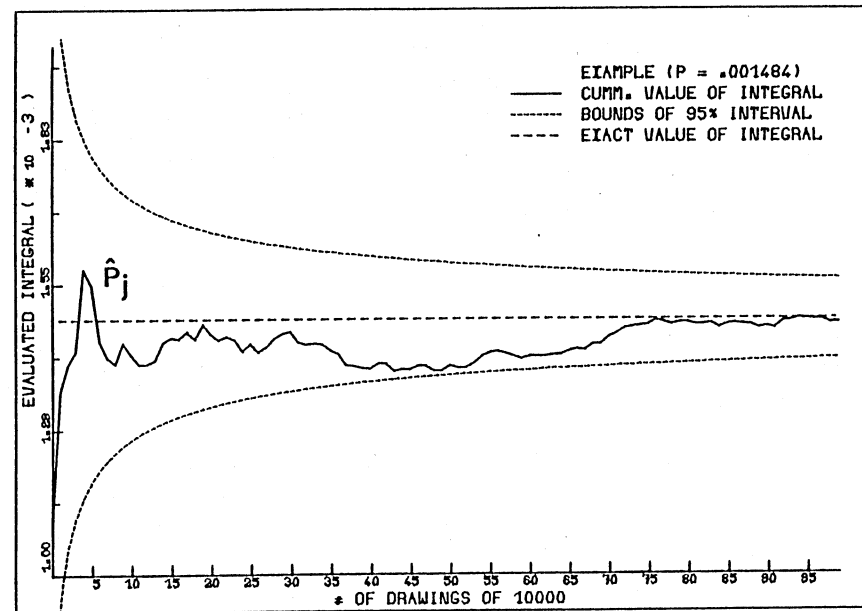
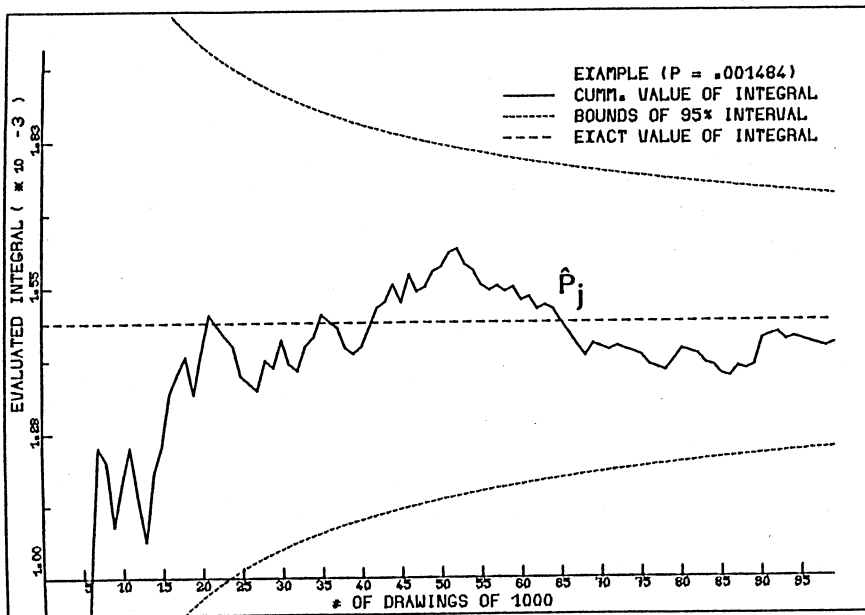
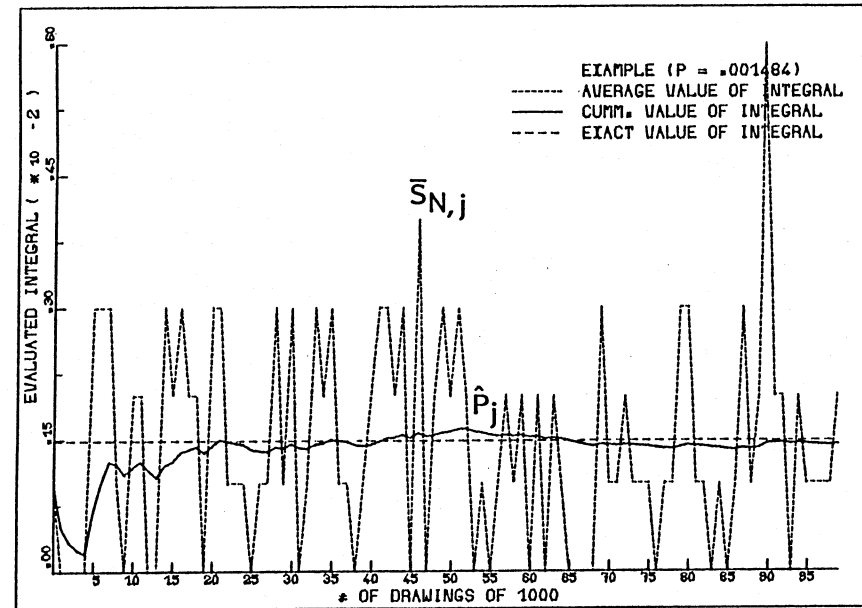
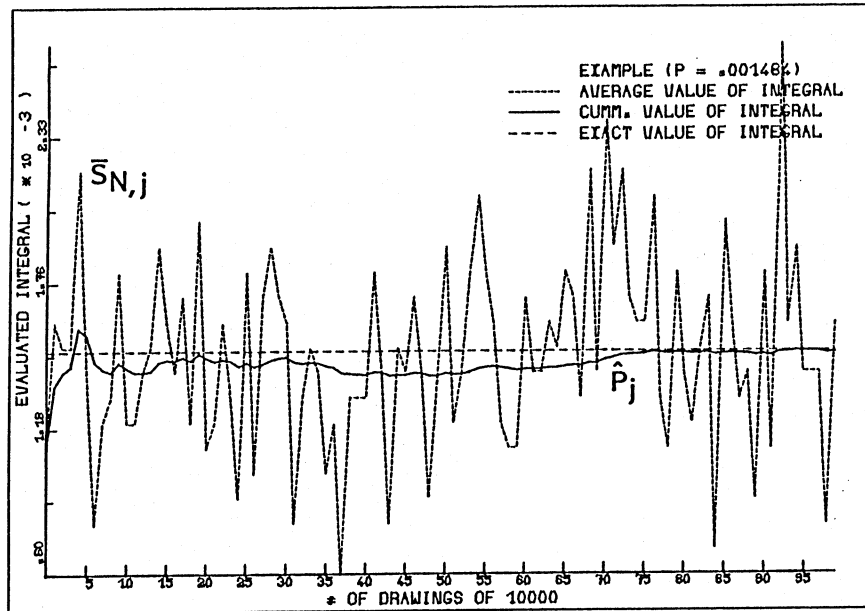


Figure 4. Sequences of estimates of the truncated multivariate normal integral $P = .001484$

$$\hat{P}_j = \left(\frac{j-1}{j}\right)\hat{P}_{j-1} + \frac{\bar{S}_{N,j}}{j} \quad (j = 1, \dots, M) \quad (2.12)$$

Equation (2.12) shows that convergence of \hat{P}_j is guaranteed when j tends to infinity since $((j-1)/j) < 1$. Note, however, that $\bar{S}_{N,j}/j$ is not a constant but the realized value of a random variable. Even at a large value of j the additional term $\bar{S}_{N,j}/j$ may throw the sequence temporarily off-track, but this event should have less and less effect as j increases.

Some examples of sequences \hat{P}_j and $\bar{S}_{N,j}$ are shown in Figures 3 and 4. In Figure 3 the true value of the integral P is given as $P = .49456$, which stems from taking the vector of constants a as $a = 1.12$; in Figure 4 the true value of P is $.001484$ and the value of the vector a is given in the computer program, listed in Figure 2. The sample size N is taken as $N = 1,000$ and as $N = 10,000$. The results in Figure 3 indicate that the sequence \hat{P}_j fluctuates around P due to the random character of the sequence $\bar{S}_{N,j}$. Further \hat{P}_j stays well within the 95 per cent confidence interval when $N = 1,000$. When N is increased to 10,000, it is seen that \hat{P}_j is outside the 95 per cent confidence interval for a number of consecutive times when $j \geq 80$. This reflects the slow convergence process when j is large. We note that the scales on the vertical axis in Figures 3 and 4 have been determined in each figure by the difference between the largest and smallest numbers in order to show the variation in the sequences. The sequences \hat{P}_j in Figure 4 fluctuate much less around their true value than the sequences of Figure 3. This is due to the fact that adding a very small number to a sequence of numbers has only a minor effect, in particular, when j is large. One may argue that the sequence \hat{P}_j in Figure 4 mimics in certain intervals on the horizontal axis, to a certain extent, the behavior of a linear difference equation with constant coefficients.

As an exercise one can determine the required size of the sample for a preassigned level of numerical accuracy. Suppose one is satisfied with a two-digit accuracy at a 95 per cent confidence level. That is, the required confidence interval bounds are given as $.001484 - .00005$ and $.001484 + .00005$ and we impose $1.96[P(1-P)/N]^{1/2} \leq .00005$ [compare the comment after equation (2.6)]. This implies that, roughly stated, N must be greater or equal to 20,000.

We end this section with a remark. Monte Carlo is not the most efficient method for the evaluation of a truncated multivariate normal integral [see, e.g., Quandt (1983, Section 8.3), and the references cited there]. We repeat

that our main purpose in this section is to indicate the basic computational steps of Monte Carlo integration and to illustrate the structure of a computer program using Monte Carlo.

3. BASICS OF IMPORTANCE SAMPLING

The posterior kernels of the parameters of interest of several econometric models have, to the best of our knowledge, no known Monte Carlo properties. That is, it is not known how one can generate random drawings from a distribution function with a density function equal (or proportional) to the kernels mentioned above. A well known example of such an econometric model is the linear simultaneous equation model where some of the structural parameters are known exactly and where the prior information on the unrestricted parameters is taken from a noninformative approach. Further examples are dynamic regression models with serially correlated errors and disequilibrium models [see the references cited in Section 1]. As a consequence, direct simulation is not a suitable numerical integration method.

In such a case one can make use of the following simple solution. Interpret the integral given in the numerator of (2.1) as the expectation of the function $g(\theta)p(\theta)$ with respect to the uniform distribution $U(\theta)$, defined on the region of integration S . So, the numerator of (2.1) can be rewritten as $E_U[g(\theta)p(\theta)]$. Similarly, the integral in the denominator of (2.1) can be rewritten as $E_U[p(\theta)]$. Next, a sample of uniform random drawings of the vector θ is generated on the region of integration S . Let $\theta^{(1)}, \dots, \theta^{(N)}$ denote the sequence of generated random drawings. Then, one can approximate (2.1) by

$$\bar{g} = \frac{\frac{1}{N} \sum_{i=1}^n g(\theta^{(i)})p(\theta^{(i)})}{\frac{1}{N} \sum_{i=1}^N p(\theta^{(i)})} \quad (3.1)$$

which can be written as

$$\bar{g} = \sum_{i=1}^N g(\theta^{(i)})p^*(\theta^{(i)}) \quad (3.2)$$

where

$$p^*(\theta^{(i)}) = \frac{p(\theta^{(i)})}{\sum_{i=1}^N p(\theta^{(i)})} \quad (3.3)$$

Note that $p^*(\theta^{(i)})$ is an estimator of the value of the normalized density function in the point $\theta^{(i)}$. The estimator (3.1) is a ratio of random variables. Given certain regularity conditions, it follows that (3.1) [and (3.2)] converges with probability one to (2.1) and that it is approximately normally distributed. The accuracy of the approximation (3.1) depends on the size N of the sample and on the variances of $g(\theta)p(\theta)$ and $p(\theta)$. These variances were in the cases that we studied very large due to the large variation of $p(\theta)$. As a consequence, one needs a very large sample of random drawings in order to achieve an acceptable level of numerical accuracy [compare Section 2].

In order to improve on the approximation (3.1) [and (3.2)] one may proceed as follows. Define

$$w(\theta) = \frac{p(\theta)}{i(\theta)} \quad (3.4)$$

where $i(\theta)$, labelled importance function, is a density function on the region of integration S . [So, $p(\theta) = w(\theta)i(\theta)$, with $i(\theta) > 0$ on S .] Then one can replace $p(\theta)$ in (2.1) by $w(\theta)i(\theta)$. There are two requirements on $i(\theta)$. First, it should be a good approximation to $p(\theta)$ and, second, one should be able to generate a sample of random drawings from a distribution with density equal or proportional to $i(\theta)$. Let $\theta^{(1)}, \dots, \theta^{(N)}$ be the generated sample. Then one can approximate (2.1) by

$$\hat{g} = \frac{\frac{1}{N} \sum_{i=1}^N g(\theta^{(i)}) w(\theta^{(i)})}{\frac{1}{N} \sum_{i=1}^N w(\theta^{(i)})} \quad (3.5)$$

which can be rewritten as

$$\hat{g} = \frac{\sum_{i=1}^N g(\theta^{(i)}) w^*(\theta^{(i)})}{\sum_{i=1}^N w^*(\theta^{(i)})} \quad (3.6)$$

where

$$w^*(\theta^{(i)}) = \frac{w(\theta^{(i)})}{\sum_{i=1}^N w(\theta^{(i)})} \quad (3.7)$$

So, $w^*(\theta^{(1)})$ is the relative weight given to each random drawing $g(\theta^{(1)})$. Obviously, if $w^*(\theta^{(1)})$ is approximately equal to unity everywhere on the region S , it follows that one is (almost) back in the situation of direct simulation [compare the estimators (2.2) and (3.7)]. More practically stated, if the variance of $w(\theta)$ is much smaller than the variance of $p(\theta)$, one has achieved a large increase in numerical accuracy at a given sample size [compare (3.2) and (3.7)]. Details on the exact formulas that are needed to compute posterior first- and second-order moments, univariate and bivariate marginal posterior densities, and numerical error estimates are given in van Dijk, Hop, and Louter (1986).

Next, we discuss briefly some proposals for the importance function $i(\theta)$.

1. The (truncated) multivariate Student t density. If the surface of the posterior kernel is reasonably well-behaved but there are some heavy tails, one can make use of the multivariate Student t density with the multivariate normal as a limiting case if the sample is large [see, e.g., Zellner (1971, Appendix B2)]. As location and scale parameters of the multivariate Student t density, one can take the posterior mode and minus the inverse of the Hessian of the log posterior density, evaluated at the posterior mode. The mode can be determined by numerical optimization methods and the Hessian by means of numerical differentiation [see, e.g., van Dijk (1984)]. For details on computer procedures that can be used to generate Student t random drawings we refer to Bauwens (1984), Kinderman and Monahan (1980), van Dijk, Hop, and Louter (1986), and the references cited there. In case the posterior mode is on the boundary of the region of integration S , one has to make use of constrained numerical optimization methods. Some proposals for this are discussed in van Dijk and Kloek (1980).
2. Poly-t density functions. In case the posterior density is not well-behaved, e.g., bimodal, one can make use of the poly-t class of density functions. In particular, when the posterior density is also a member of this class [see Bauwens (1984)]. More research is needed in order to determine the usefulness of this family of density functions.
3. Mixtures of normals or Student t densities. Mixtures are very flexible and, therefore, suitable as importance function for irregularly shaped posterior densities. A difficult problem is, however, the specification of the values of the parameters of mixtures. A simple solution is given in van Dijk and Kloek (1985). More research is needed on this topic.

4. PRIOR AND POSTERIOR ANALYSIS OF KLEIN'S MODEL I ¹4.1. Model and prior information

In this section we apply the Monte Carlo approach to Klein's Model I, which is a small simultaneous equation model based on annual data for the United States economy for the period 1921 to 1941 [see Klein (1950)]. Given our prior information, this model has nine structural parameters of interest to economists, which implies that we compute nine-dimensional integrals numerically.

The statistical model can be summarized as follows. Our starting point is the well known linear simultaneous equation model (SEM)

$$YB + Z\Gamma = U^* \quad (4.1)$$

where Y is an $T \times (G+G')$ matrix of observations on $G+G'$ current endogenous variables and Z an $T \times K$ matrix of observations on the K predetermined variables. The rows of $U^* = (U \ 0)$ are assumed to be independently normally distributed with mean zero and covariance matrix

$$\begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \quad (4.2)$$

where Σ (a nonsingular $G \times G$ matrix) corresponds to the stochastic equations. Further, current values of the disturbances are assumed to be independently distributed from current and lagged values of the predetermined variables and the data matrix (Y,Z) has full column rank.

The prior information on B , Γ , and Σ is specified as follows. The elements of the parameter matrices B , Γ , and Σ are divided into three groups: (i) Nuisance parameters (constant terms, denoted by the vector α , and the covariance matrix Σ); (ii) Exactly known elements of B and Γ ; (iii) Unrestricted elements of B and Γ , denoted by the vector θ . So we have

$$B = B(\theta), \quad \Gamma = \Gamma(\alpha, \theta) \quad (4.3)$$

The prior specification with respect to the nuisance parameters is taken

1. This section is an extension of van Dijk and Kloek (1980).

from a noninformative approach. The prior density of the constant terms α is locally uniform and the prior density of Σ is proportional to $|\Sigma|^{-\frac{1}{2}h}$, where $h = -\frac{1}{2}(G+1)$ and G is the number of stochastic equations. We have opted for this relatively low value of h so that the information contained in the likelihood function determines the posterior. As a consequence, analytical integration with respect to α and Σ is possible [see, e.g., van Dijk (1984, Chapter 2)].

With respect to the prior specification of the exactly known parameters we proceed as follows. Identification is treated in the traditional way [for an alternative approach, see Kiefer (1981)]. 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 exactly known parameter values of B and Γ are implied by the specification of Klein's Model I. The structural equations of Klein I read

$$C = \theta_1 P + \theta_2 P_{-1} + \theta_3 W + \alpha_1 + u_1 \quad (4.4)$$

$$I = \theta_4 P + \theta_5 P_{-1} - \theta_6 K_{-1} + \alpha_2 + u_2 \quad (4.5)$$

$$W_1 = \theta_7 X + \theta_8 X_{-1} + \theta_9 t + \alpha_3 + u_3 \quad (4.6)$$

$$X = C + I + G \quad (4.7)$$

$$P = X - W_1 - T \quad (4.8)$$

$$K = K_{-1} + I \quad (4.9)$$

$$W = W_1 + W_2 \quad (4.10)$$

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 θ_6 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 (l , P_{-1} , X_{-1} , K_{-1} , G, T, W_2 , t). All variables (except l 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. 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.

Finally, we discuss in this subsection the prior information on the parameters of interest θ . 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 first and simplest prior for the vector θ is uniform on the nine-dimensional unit region¹ minus the region where $\|B\| < .01$. The latter region has been subtracted in order to guarantee that the implied prior moments of the multipliers exist. The likelihood determines the posterior in the truncated uniform region. 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 $\|B\| > .01$. In case this condition was not satisfied, the vector was rejected and replaced by a new vector. Each experiment was stopped when 20,000 θ vectors satisfying the constraint were obtained.² 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 are used for the computation of the implied prior means and

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

² We emphasize that this number is large so that the prior densities are reasonably accurate. The Monte Carlo estimates for the posterior moments are already reasonably accurate at 2000 random drawings.

second-order moments and the implied univariate prior densities.

The reduced form equations form a system of linear difference equations. The three roots of the characteristic polynomial of this system summarize the dynamic properties of the system. If all roots are real, the system is monotone. If there is one real root and a pair of conjugate complex roots, the system is oscillatory. Further, if all roots are less than one in absolute value, the system is damped. In case there is, at least, one root greater than or equal to one, the system is explosive. [For more details, see Theil and Boot (1962)]. There exist four possible states of the system: damped oscillatory, damped monotone, explosive oscillatory and explosive 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 set of constraints 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. The value of the dominant root of this third degree polynomial may be computed by making use of analytical formulas [see, e.g., Abramowitz and Stegun (1964, p.17)] or by making use of numerical methods. We made use of a numerical method given by the NAG-Library routine F02AFF.

2. The long-run effects in the structural equations: $\theta_1 + \theta_2$, $\theta_4 + \theta_5$, $\theta_7 + \theta_8$ are all assumed to be in the unit interval.

3. The SRM's are assumed to be less than five in absolute value and to have the correct sign (positive for effects of W_2 and G , negative for effects of T).

4. The same set of constraints as mentioned in 3 was applied to the LRM's (with an exception for the final form equation of K , where an upper bound of ten was adopted).

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)]. Eight different priors were obtained by combining the sets of extra constraints, 1

to 5, in several ways. They are given as prior 1 (no sets of extra constraints), prior 2 (1), prior 3 (2), prior 4 (1,2), prior 5 (2,3), prior 6 (1,2,3), prior 7 (1,2,3,4), prior 8 (1,2,3,4,5). [The numbers between () refer to the sets of extra constraints.] We note that, due to space limitations, we present only results based on priors 2 and 8. More details are given in van Dijk (1984).

The posterior density of (θ, α, Σ) is obtained by combining the likelihood function of the SEM and the prior density by means of Bayes rule. The marginal posterior density of the parameters of interest θ reads in this case,

$$p(\theta | Y, Z) \propto p(\theta) \|B\|^T |S|^{-\frac{1}{2}(T-1)} \quad (4.11)$$

where S is defined as $S = U'NU$, with U given below (4.1) and N given as $N = I - (1/T)11'$. For details on the derivation of (4.11) we refer to van Dijk (1984) or van Dijk and Kloek (1977).

4.2. Prior and posterior results

In this subsection we present the prior and posterior means and standard deviations of the nine structural parameters θ (Table 1), the multipliers in the reduced and final form equations for national income (Table 2), the period of oscillation and the dominant root (Table 3), and we present the prior and posterior probabilities of the four states of the system (Table 3). Further, the marginal prior and posterior densities of the structural parameters are shown in Figure 5, univariate and bivariate marginal prior and posterior densities of the multipliers mentioned above are shown in Figures 6 and 7, and the prior and posterior densities of the period of oscillation and the dominant root are shown in Figure 8. We make use of a prime to denote a posterior density, e.g., $2'$ denotes the marginal posterior density based on prior 2. In all tables we give the Full Information Maximum Likelihood (FIML) results (with asymptotic standard errors in Table 1) for comparison. In all cases we confine ourselves to presenting the results based on priors 2 and 8 (except for the bivariate densities, where we confine ourselves to prior 2 due to space limitations). 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

TABLE 1
MEANS AND STANDARD DEVIATIONS OF STRUCTURAL PARAMETERS

	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9
FIML	-.23	.39	.80	-.80	1.05	.15	.23	.28	.23
(no prior)	(.58)	(.30)	(.04)	(.84)	(.42)	(.05)	(.09)	(.06)	(.06)
Prior 2	.40	.42	.37	.41	.36	.50	.65	.33	.50
	(.28)	(.28)	(.26)	(.28)	(.27)	(.29)	(.26)	(.24)	(.29)
Posterior 2'	.12	.19	.79	.06	.64	.15	.34	.23	.19
	(.08)	(.08)	(.04)	(.06)	(.10)	(.03)	(.05)	(.05)	(.04)
Prior 8	.25	.33	.39	.27	.30	.55	.43	.27	.50
	(.20)	(.23)	(.25)	(.20)	(.22)	(.26)	(.24)	(.20)	(.29)
Posterior 8'	.24	.06	.72	.14	.56	.20	.37	.24	.19
	(.08)	(.04)	(.05)	(.09)	(.12)	(.05)	(.04)	(.04)	(.05)

more detail below. All results presented are based on $N = 20,000$. In contrast, application of a ten-point Gaussian product rule of numerical integration requires 10^9 function evaluations. As importance function we make use of a truncated multivariate Student t density. For details we refer to van Dijk (1984, Chapters 3 and 4) and to van Dijk and Kloek (1980).

We start to observe that the FIML estimates of θ_1 and θ_4 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 θ_4 decreases while θ_5 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. FIML results based on a diagonal matrix Σ are presented by Klein (1950). The hypothesis of a diagonal covariance matrix Σ is, however, strongly rejected in a likelihood ratio test [$\chi^2(3) = 28.46$]. Third, the Jacobian

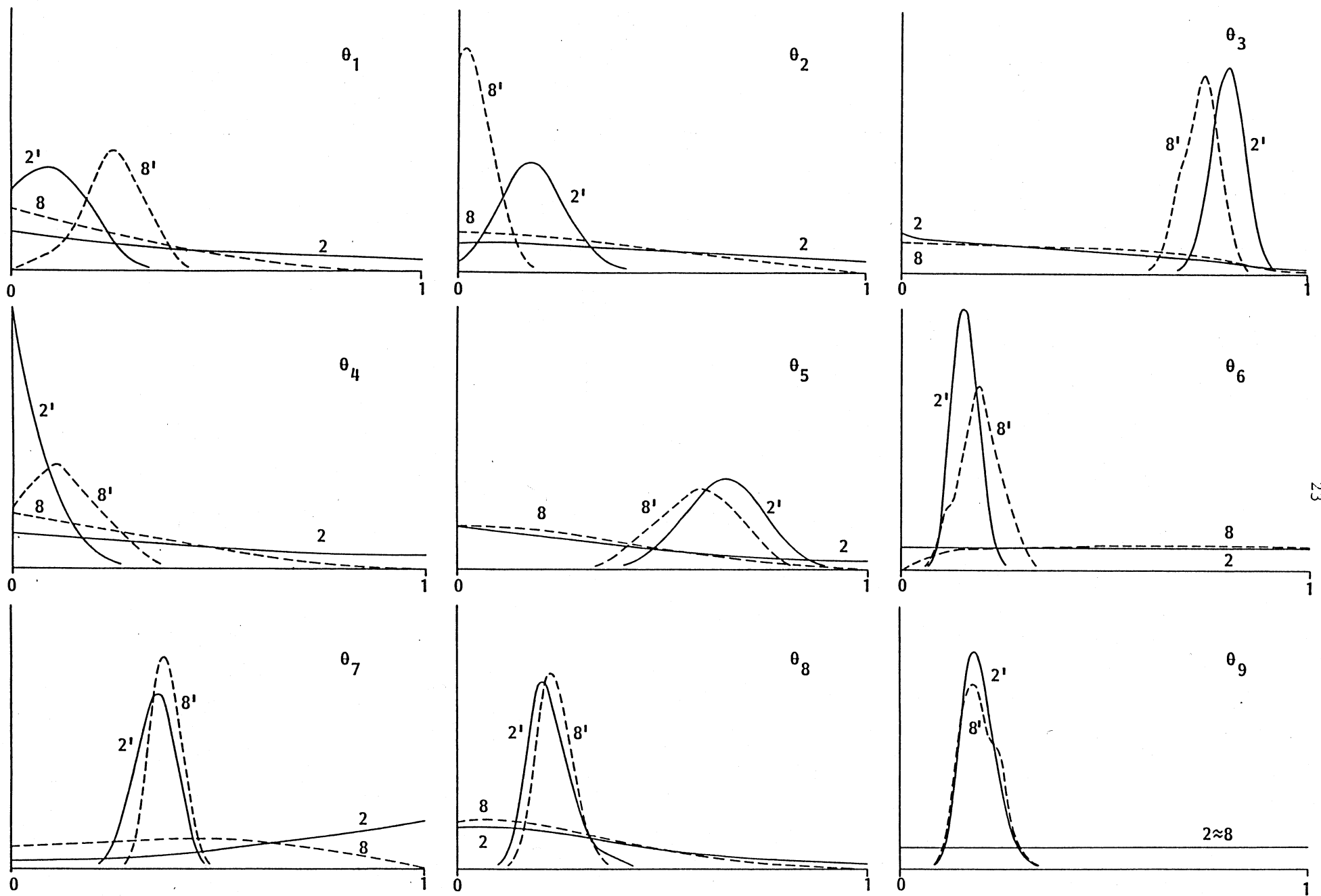


Figure 5. Marginal prior and posterior densities of structural parameters.

$$\|B\| = |1 - (\theta_1 + \theta_4)(1 - \theta_7) - \theta_3\theta_7| \quad (4.12)$$

is less than or equal to unity in the unit region but equals 1.60 in the FIML point. Recall that a factor $\|B\|^T$ occurs in the likelihood function and in the marginal posterior density $p(\theta | Y, Z)$, equation (4.11). We note that in 3SLS, where the Jacobian factor is absent in the function to be minimized but the nondiagonal elements of Σ are present, θ_4 has a wrong sign but only marginally so ($\hat{\theta}_4 = -.013$) [Theil (1971, p.517)]. If θ_4 is restricted to be zero, θ_1 gets the correct sign and this hypothesis is not rejected in a likelihood ratio test [$\chi^2(1) = 3.20$] According to this diagnostic result there is no conflict between the sample information and our prior information, which states that θ_1 and θ_4 should be nonnegative. If we compare the FIML asymptotic standard errors and the posterior standard deviations of the structural parameters (Table 1) we see that this prior information plays a large role. This conclusion is confirmed in Figure 5.

Once we have accepted the prior information that all elements of θ are in the unit region, the extra sets of constraints 1, 2, 3 and 4, introduced in Subsection 4.1, turn out not to be restrictive. Given prior 1, the posterior probability that the system is explosive is .021. Given prior 2, the long run posterior effects $\theta_1 + \theta_2$, $\theta_4 + \theta_5$, and $\theta_7 + \theta_8$ in the structural equations are all in the unit interval (Table 1). In this respect we note that the relevant covariances (not shown in order to save space) are all negative. All SRM's and LRM's (Table 2) amply satisfy the upper bound constraints. They also satisfy the sign constraints, though some are close to zero. In these cases the prior and posterior densities (Figures 6 and 7) turn out to be skew so that the probability of wrong signs is extremely small. This explains why the differences between the posteriors 2 to 7 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 1 through 3 and in Figures 5, 6 and 8 that this set, introduced in prior 8, influences almost every parameter. In particular, if θ_2 and θ_5 are relatively large (which corresponds to negative or small positive values of θ_1 and θ_4) the lags become large and this, in turn, implies long periods of oscillation (compare Tables 1 and 3) and relatively small absolute values of most of the SRM's (Table 2).

So we have observed that the prior constraints on the period of oscillation have rather large effects. The question arises whether this

TABLE 2
 MEANS AND STANDARD DEVIATIONS OF MULTIPLIERS IN THE REDUCED AND FINAL FORM
 EQUATIONS FOR NATIONAL INCOME

	Short run effects on Y (SRM's)			Long run effects on Y (LRM's)		
	G	T	W ₂	G	T	W ₂
	FIML (no prior)	0.62	-0.36	1.50	1.96	-1.30
Prior 2	2.20 (1.57)	-2.83 (1.94)	1.99 (1.45)	3.12 (151.)	-2.55 (58.0)	3.06 (134.)
Posterior 2'	1.65 (0.19)	-1.32 (0.20)	2.30 (0.16)	2.38 (0.14)	-1.73 (0.23)	2.87 (0.16)
Prior 8	2.00 (0.71)	-2.14 (0.83)	1.85 (0.73)	2.01 (0.78)	-2.20 (0.74)	1.93 (0.86)
Posterior 8'	2.06 (0.30)	-1.81 (0.34)	2.49 (0.26)	2.25 (0.13)	-1.67 (0.20)	2.63 (0.16)

TABLE 3

MEANS AND STANDARD DEVIATIONS OF PERIOD OF OSCILLATION AND DOMINANT ROOT;
PROBABILITIES OF STATES

	Period of oscillation (years)	DRT	Damped		Explosive	
			Oscillatory	Monotone	Oscillatory	Monotone
FIML (no prior)	34.83	.76	(not available)			
Prior 2	5.22 (4.74)	.78 (.17)	.96	.04	0	0
Posterior 2'	15.06 (2.90)	.84 (.08)	.9999	.0001	0	0
Prior 8	5.42 (1.57)	.72 (.18)	.98	.02	0	0
Posterior 8'	9.61 (0.37)	.77 (.08)	.9927	.0073	0	0

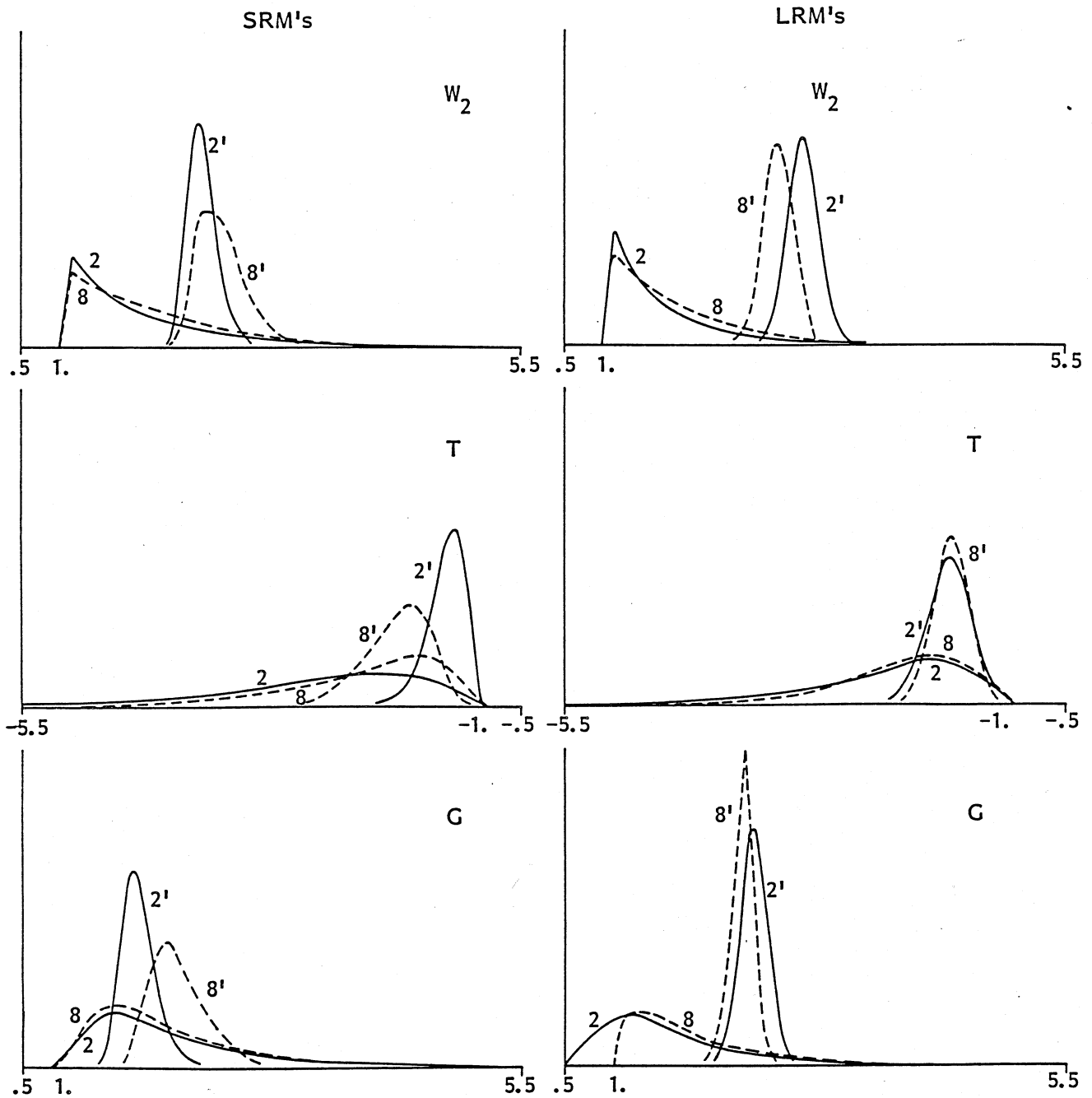
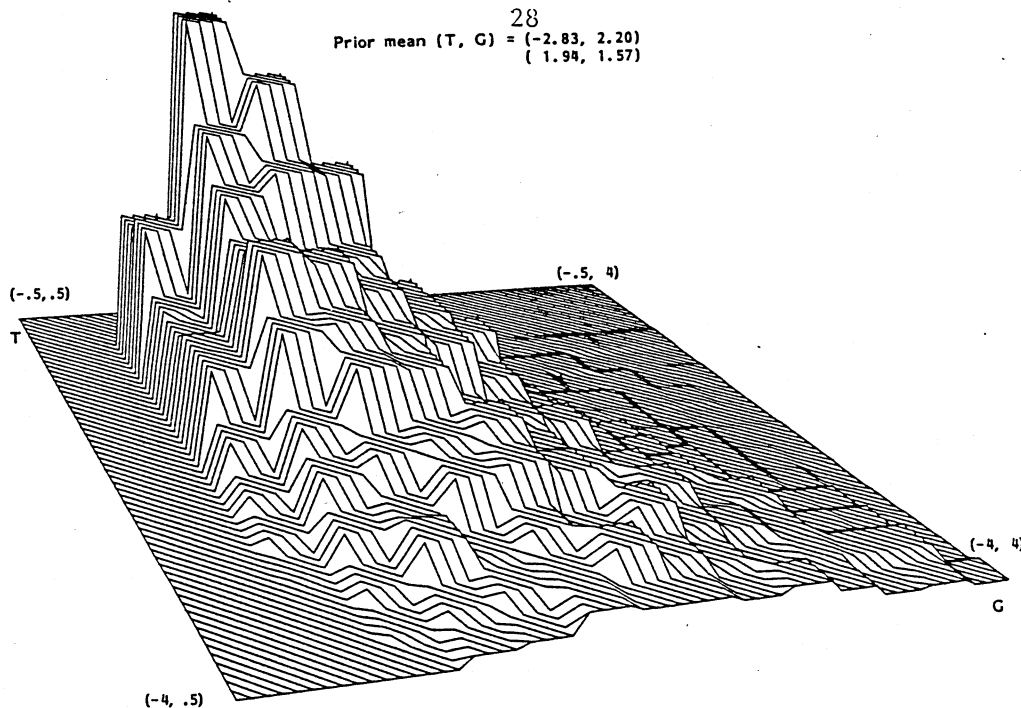


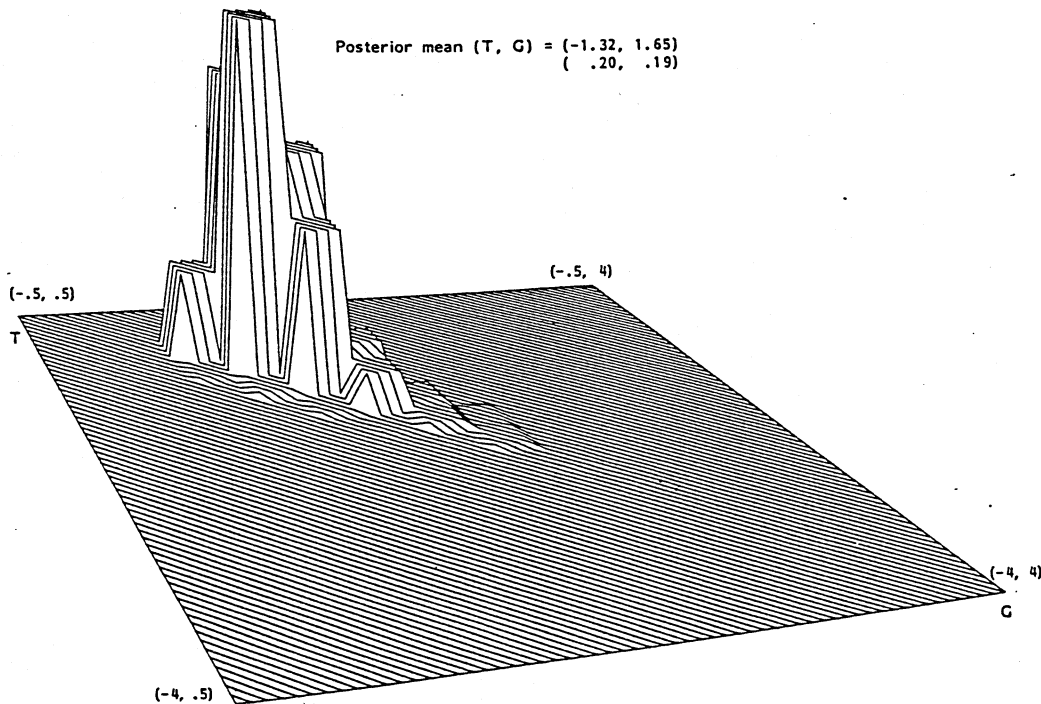
Figure 6. Prior and posterior densities in reduced and final form equations of national income.

28
Prior mean (T, G) = (-2.83, 2.20)
(1.98, 1.57)

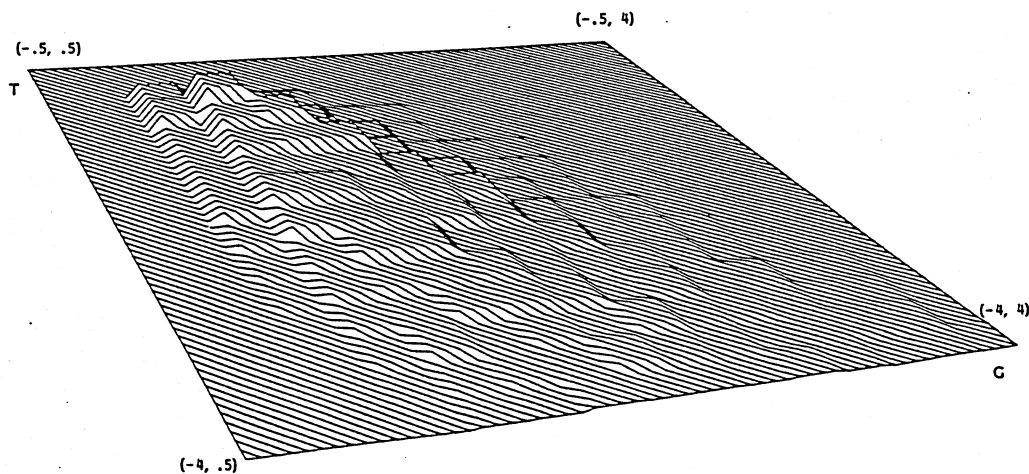


Prior density of short-run government-expenditures (G) and tax (T) multipliers

Posterior mean (T, G) = (-1.32, 1.65)
(.20, .19)

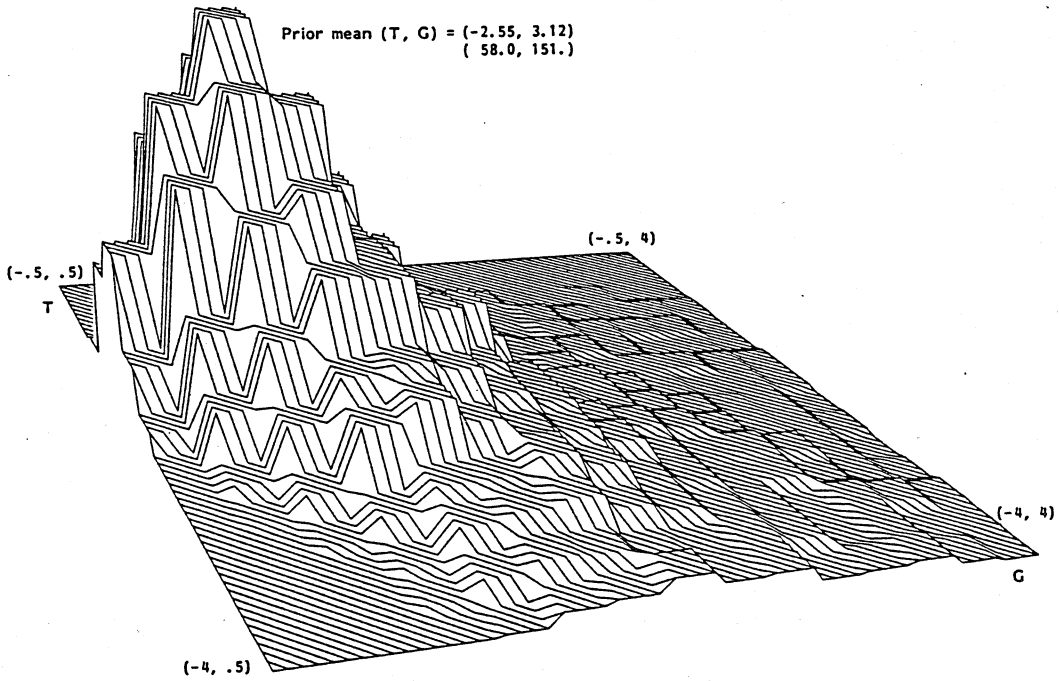


Posterior density of short-run government-expenditures (G) and tax (T) multipliers

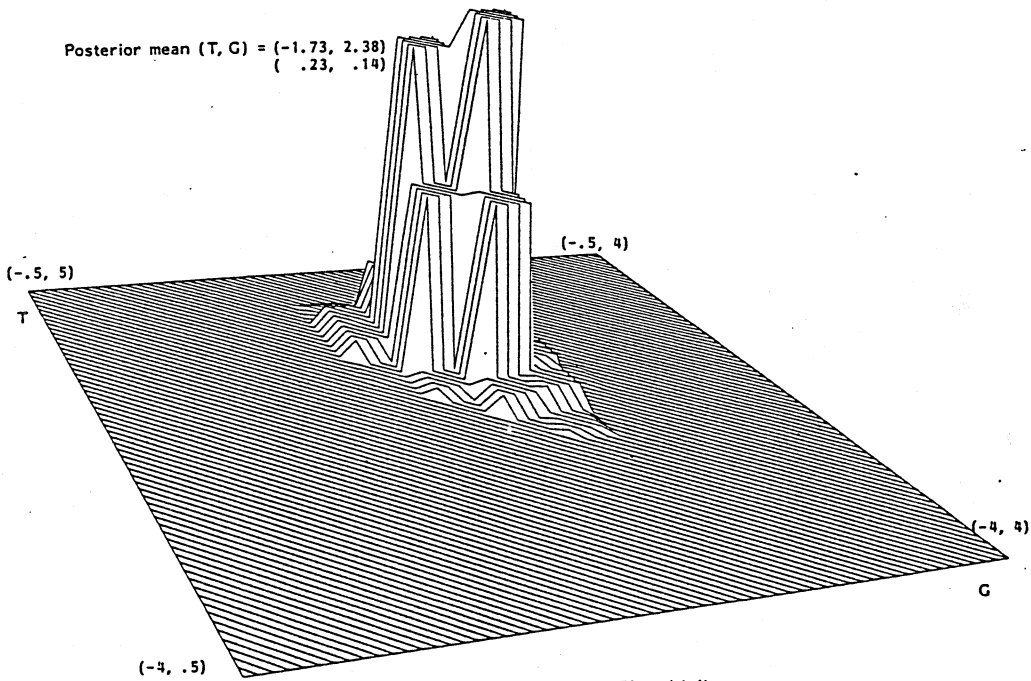


Scaled prior density

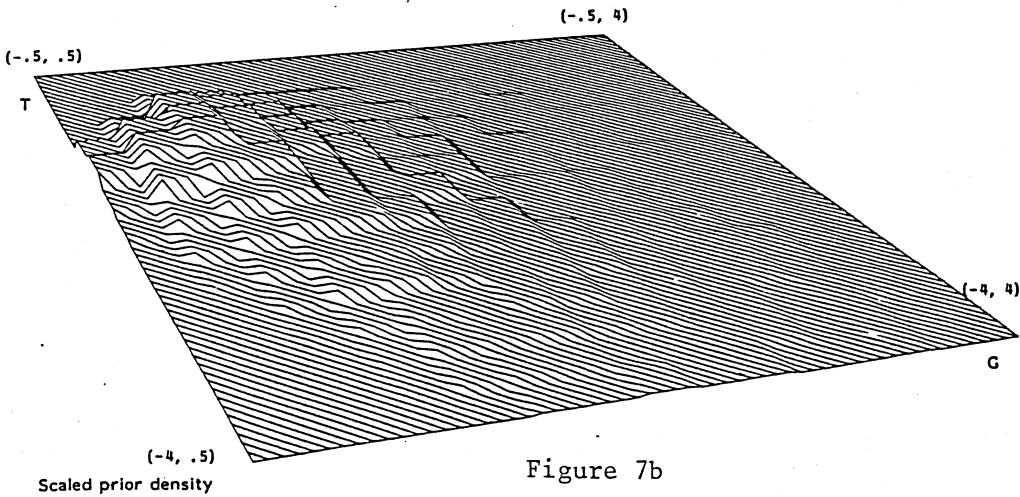
Figure 7a



Prior density of long-run government-expenditures (G) and tax (T) -multipliers

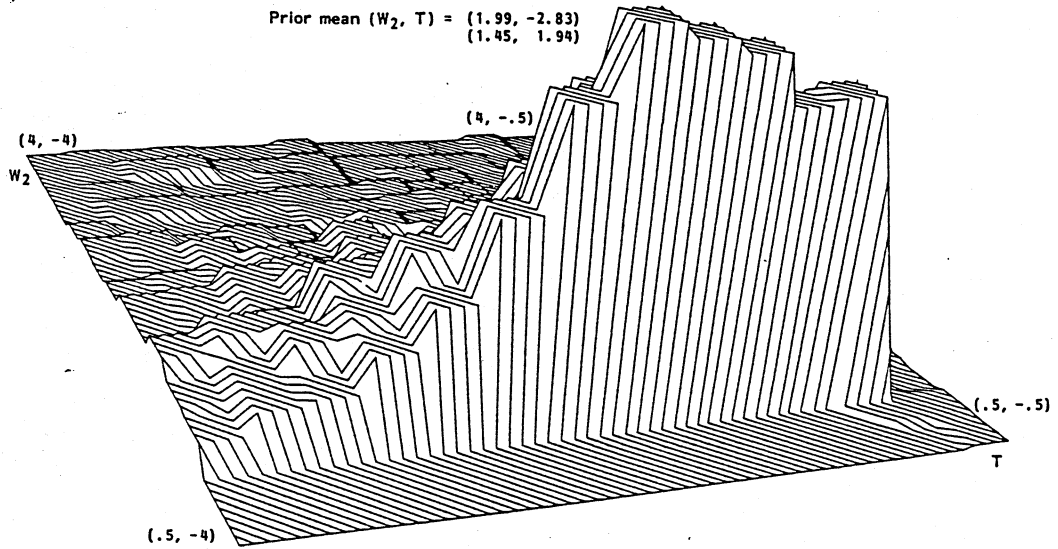


Posterior density of long-run government-expenditures (G) and tax (T) multipliers

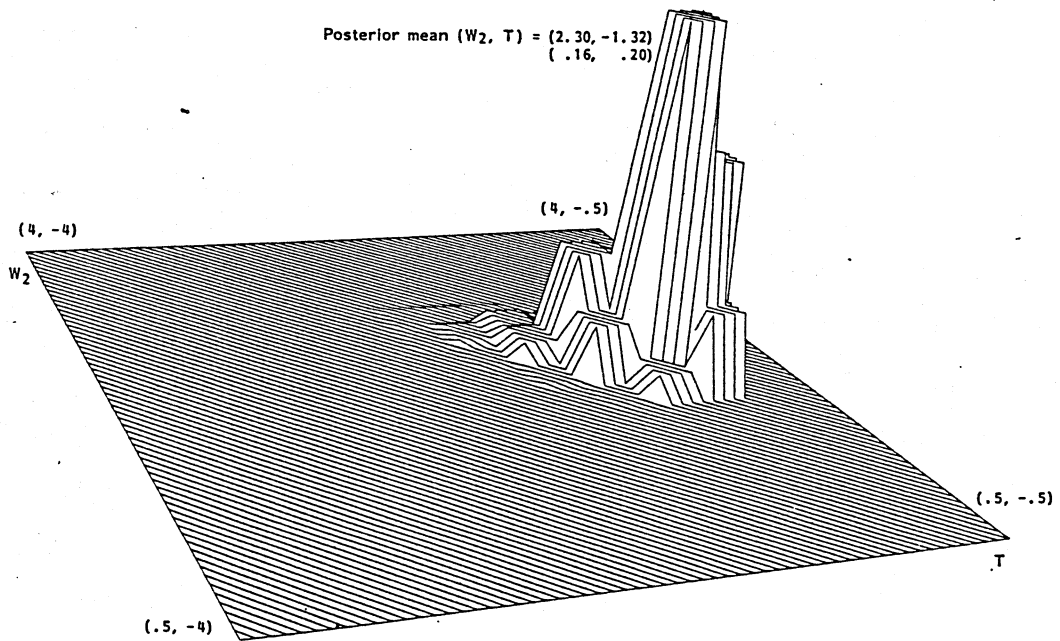


Scaled prior density

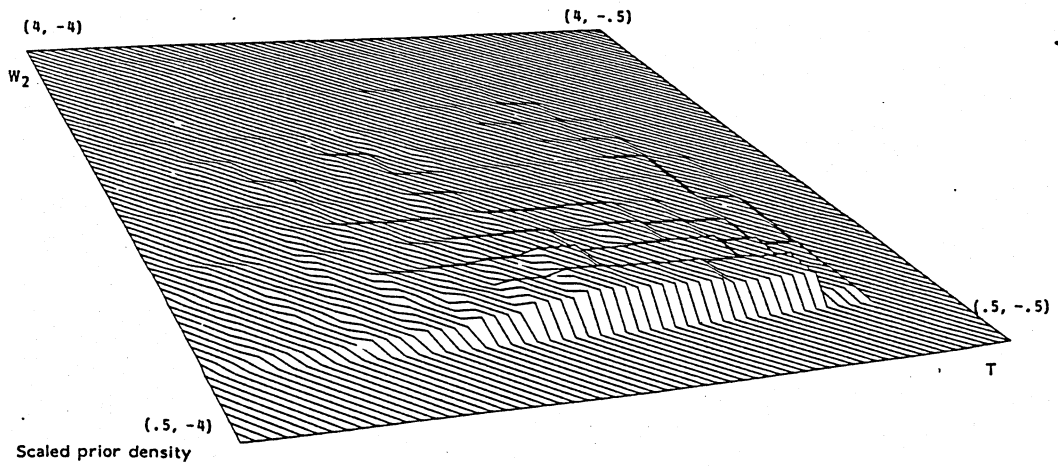
Figure 7b



Prior density of short-run tax (T) and government-wages (W_2) multipliers

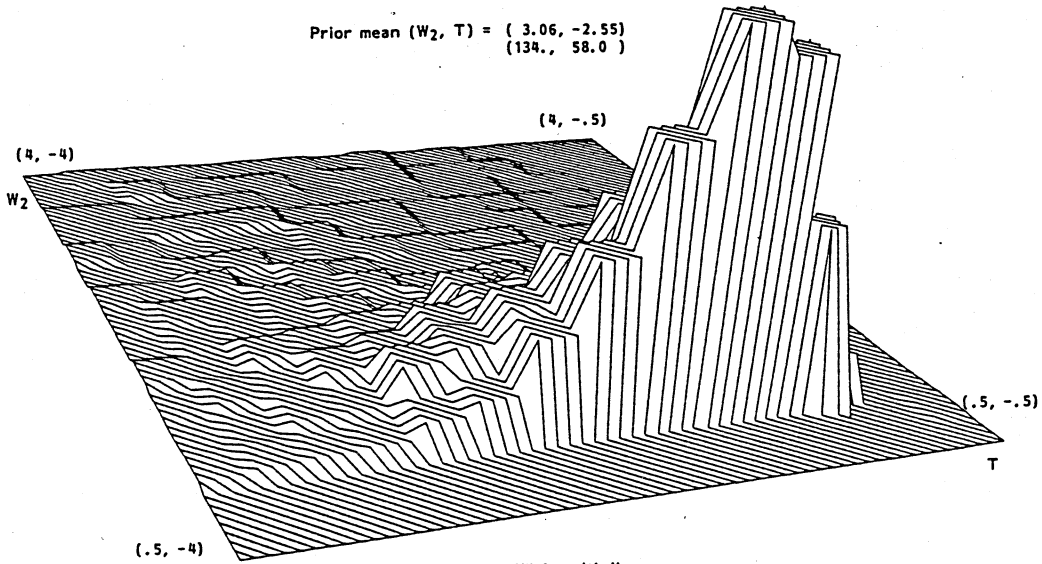


Posterior density of short-run tax (T) and government-wages (W_2) multipliers

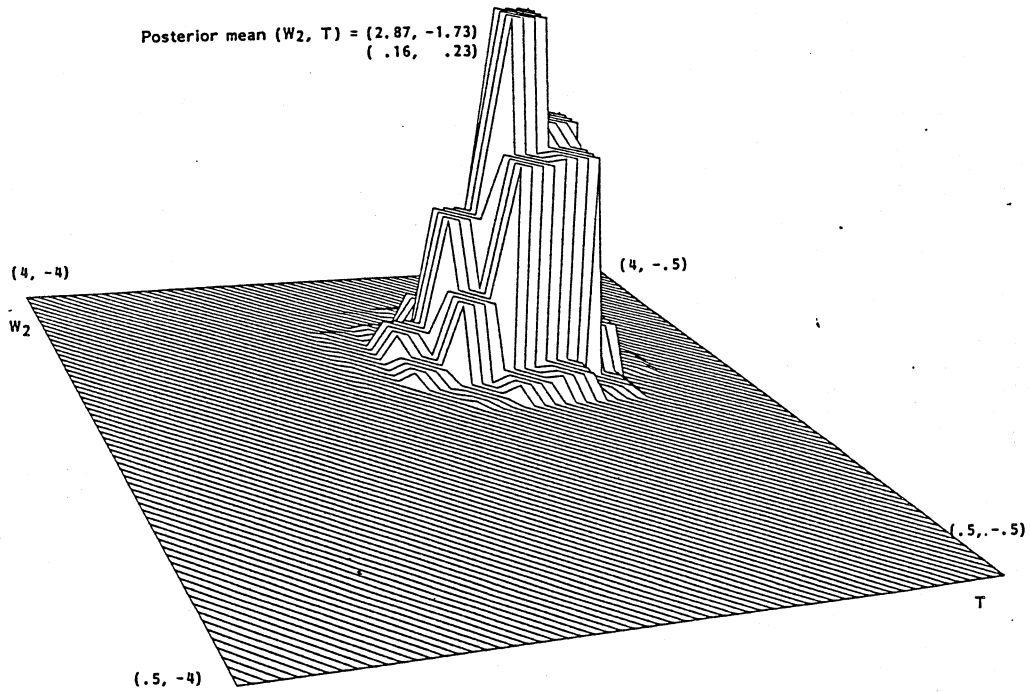


Scaled prior density

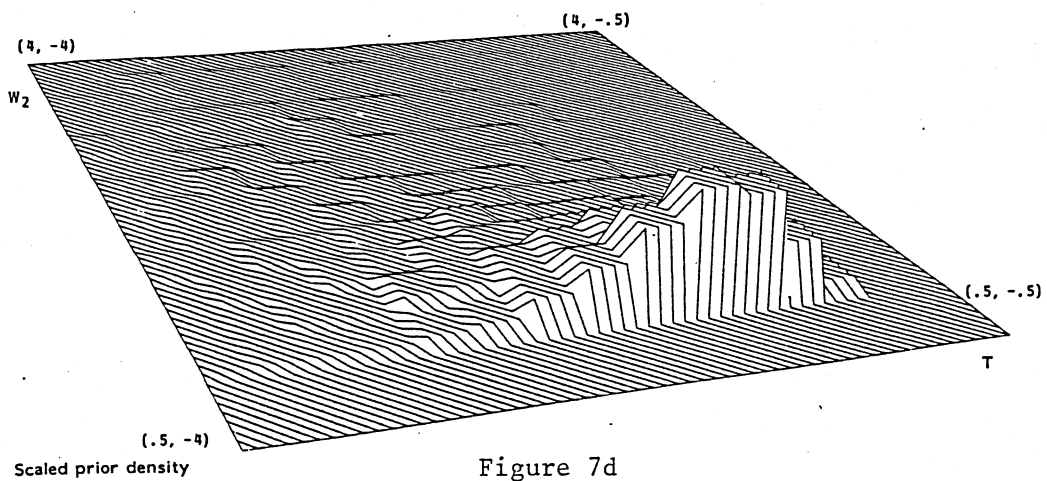
Figure 7c



Prior density of long-run tax (T) and government-wages (W_2) multipliers



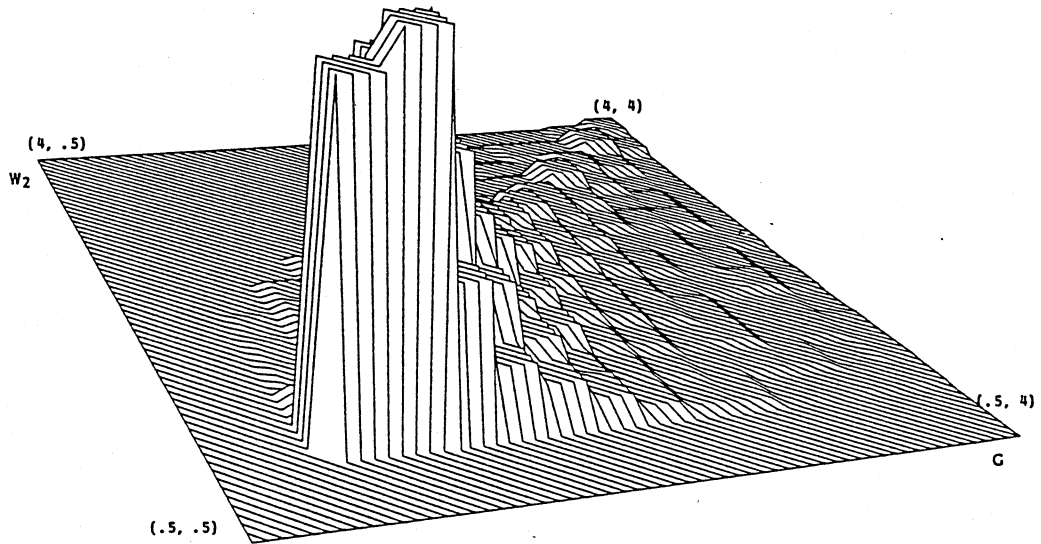
Posterior density of long-run tax (T) and government-wages (W_2) multipliers



Scaled prior density

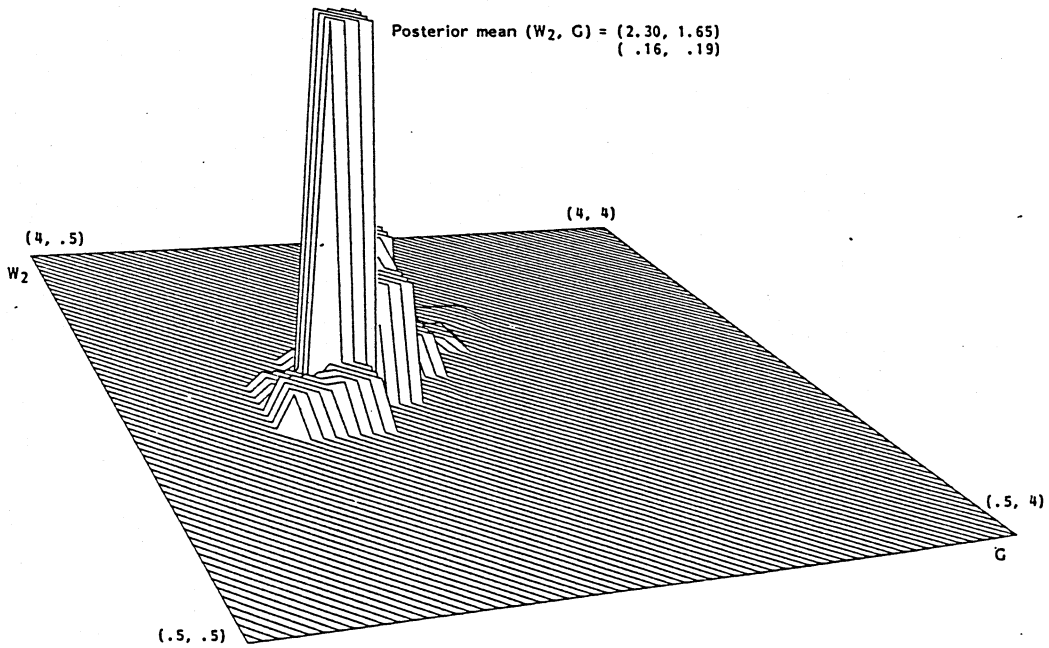
Figure 7d

Prior mean $(W_2, G) = (1.99, 2.20)$
 $(1.45, 1.57)$

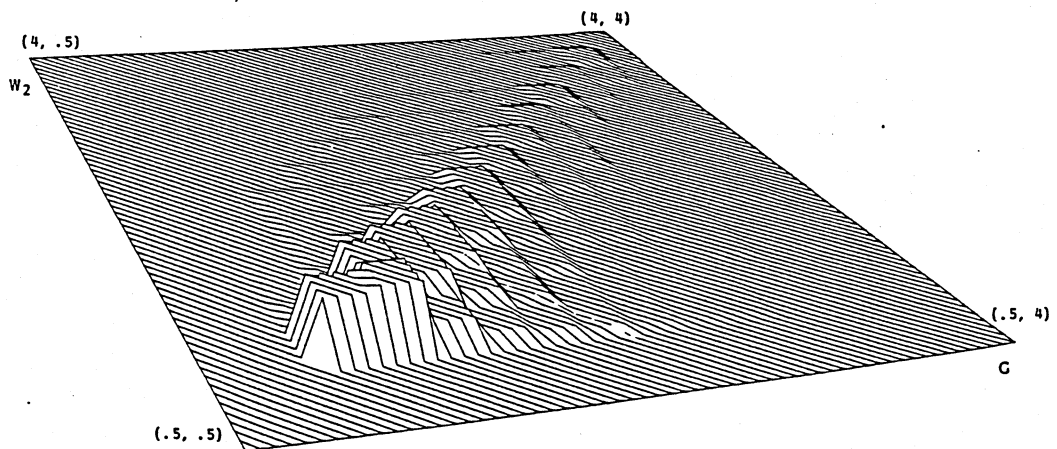


Prior density of short-run government-expenditures (G) and government-wages (W_2) multipliers

Posterior mean $(W_2, G) = (2.30, 1.65)$
 $(.16, .19)$

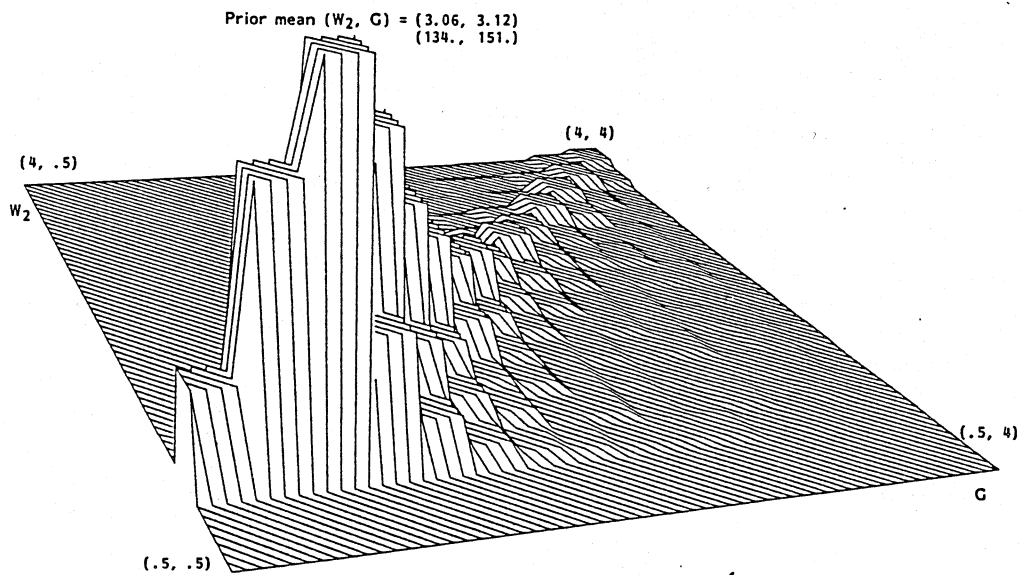


Posterior density of short-run government-expenditures (G) and government-wages (W_2) multipliers

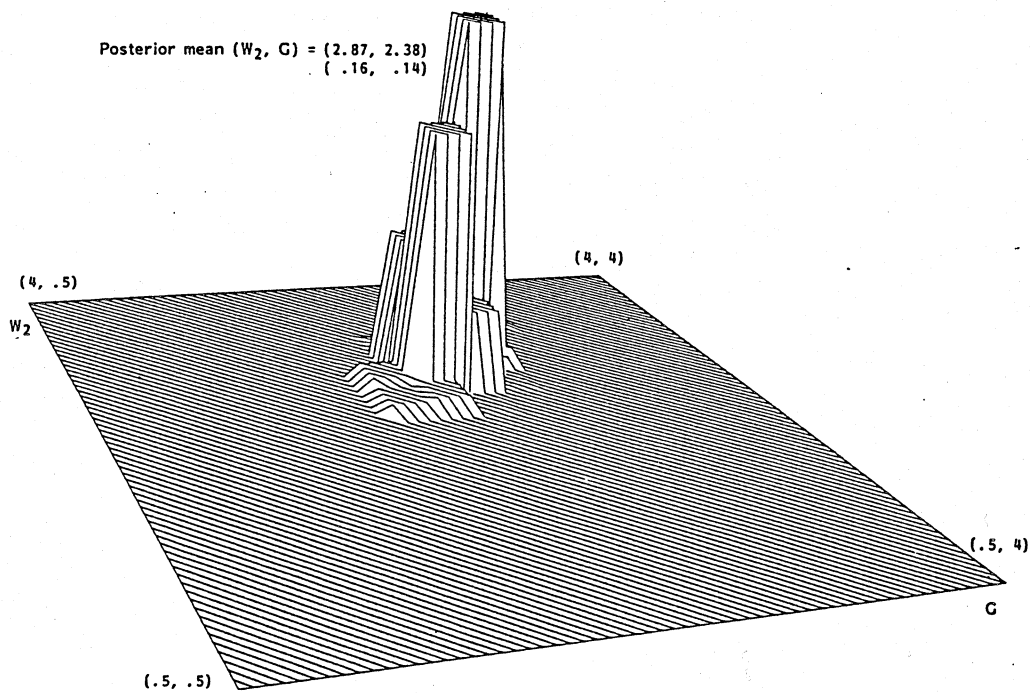


Scaled prior density

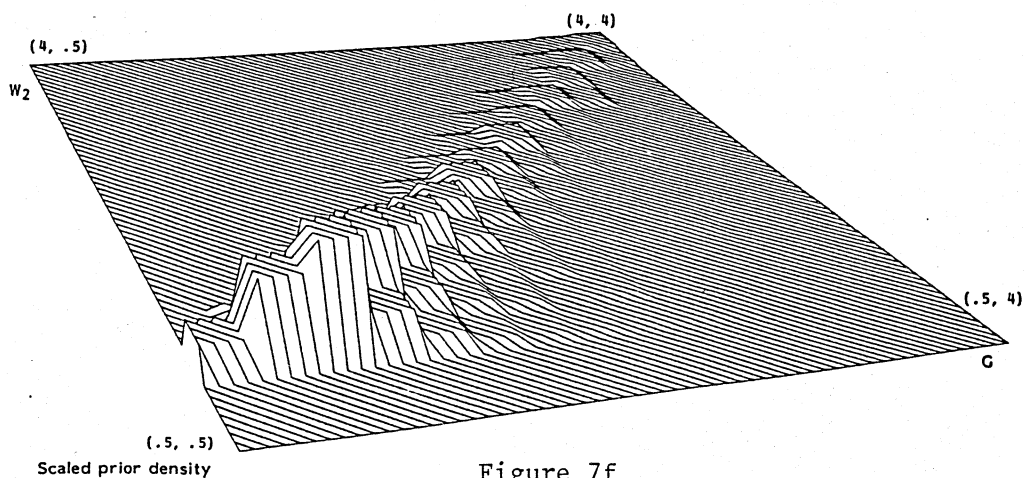
Figure 7e



Prior density of long-run government-expenditures (G) and government-wages (W_2) multipliers



Posterior density of long-run government-expenditures (G) and government-wages (W_2) multipliers



Scaled prior density

Figure 7f

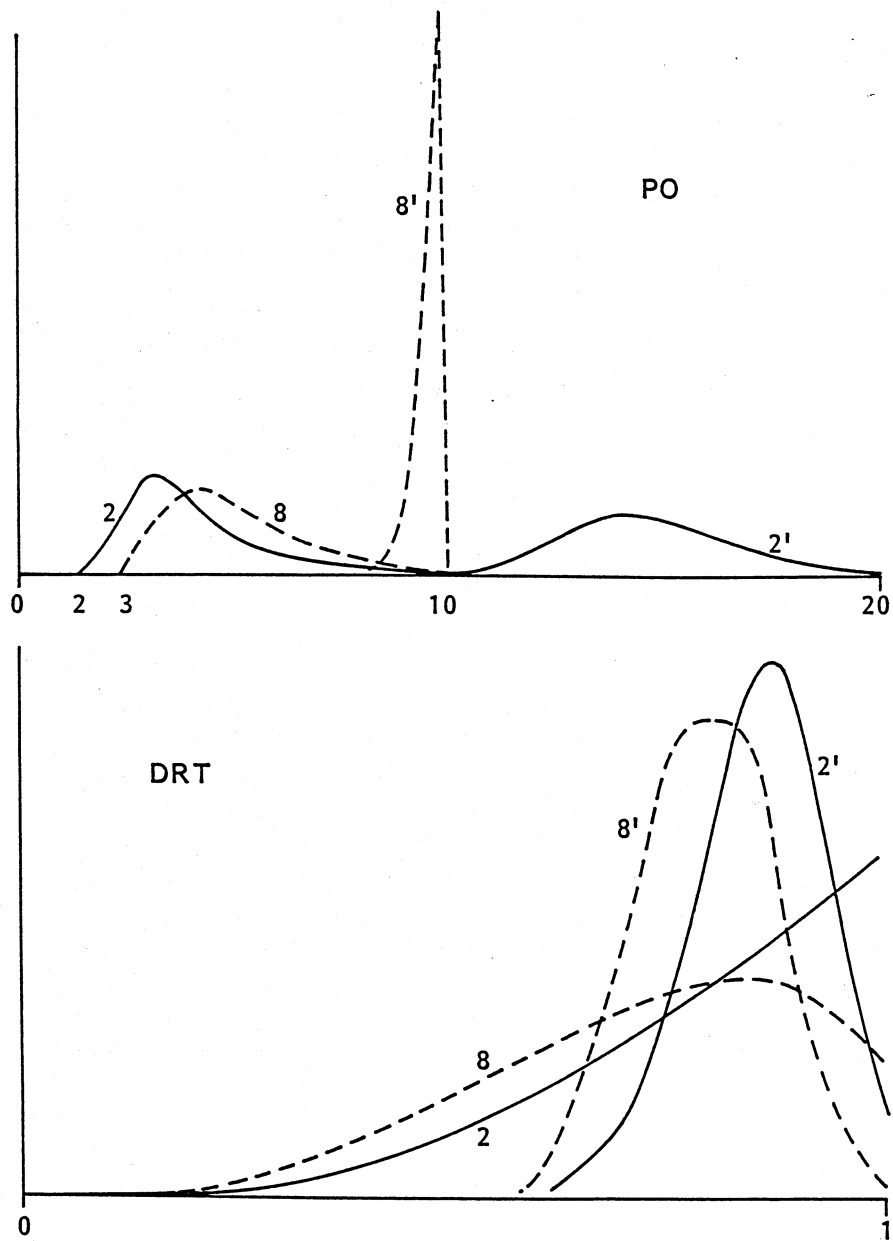


Figure 8.
Prior and posterior densities of period of oscillation and dominant root.

information is acceptable. The posterior mean and standard deviation of the period of oscillation under prior 2 (Table 3) suggest that the hypothesis of a ten year period is acceptable. Inspection of the prior and posterior densities of the period of oscillation in Figure 8 reveals that for the case of prior 2 the information from the likelihood function has modified the prior information substantially. The posterior probability that the period of oscillation is less than or equal to ten years is less than .02. Further, the effect of constraint 5 is clearly reflected in the posterior density 8'. These results suggest rejection of the constraint 5.

We summarize the posterior results as follows. The prior restrictions that θ_1 and θ_4 should be nonnegative have a large effect on the posterior results. Given these prior restrictions, we have shown that only constraint 5 has a substantial, but undesired, effect on the posterior results, which suggests rejection of this prior constraint.

When considering these results we were tempted to look for specification errors. 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. A first set of diagnostic results is based on an analysis of the expected values of the posterior residuals that are defined as $EU = Y_s EB_s(\theta) + Z_s E\Gamma_s(\alpha, \theta)$ where the index s refers to the set of variables and parameters that occur in the stochastic equations of Klein's Model I. Estimates of the posterior expected values of u_1 , u_2 and u_3 [equations (4.4)-(4.6)] are presented in Table 4, together with the residuals based on the FIML estimates of the vector θ . Further, the posterior densities of two functions of the posterior residuals are shown in Figure 9. That is, the serial correlation coefficients

$$\rho_j = \frac{\sum_{t=2}^T u_{jt} u_{jt-1}}{\sum_{t=2}^T u_{jt-1}^2} \quad j = 1, 2, 3 \quad (4.13)$$

and the well-known statistic of Durbin and Watson have been chosen as two diagnostic tools for the detection of correlation in the posterior residuals. The results indicate that there are errors in the dynamic specification of the consumption function. This finding is in accordance with results presented by Kiefer (1971). So, instead of reducing the parameter space by making use of

the sets of prior constraints 1-5, we have to enlarge the parameter space by including, e.g., lagged consumption in equation (4.2). Preliminary results obtained with an enlarged version of Klein's Model I confirm this. This will be reported on in future work. We emphasize that one has to interpret these diagnostic results with care. They indicate, however, that there exists a need for diagnostic analysis on the correct specification of the information contained in the prior density and in the likelihood function.

TABLE 4
FIML RESIDUALS AND MEAN VALUES OF POSTERIOR 2' RESIDUALS

	u_1		u_2		u_3	
	FIML	Post.2'	FIML	Post.2'	FIML	Post.2'
1921	-1.07	-0.37	-3.85	-1.50	-1.57	-1.04
1922	-1.02	-0.91	2.13	0.49	0.77	0.88
1923	-1.07	-1.54	2.19	1.11	2.44	2.04
1924	-0.02	-0.54	-0.00	-1.33	0.05	0.05
1925	0.47	-0.08	2.06	0.54	0.45	0.06
1926	0.90	0.69	2.19	1.38	0.44	-0.05
1927	1.60	1.23	2.31	1.13	-0.23	-0.56
1928	1.80	1.03	2.57	0.36	0.71	0.42
1929	0.01	-0.69	4.24	2.04	1.98	1.46
1930	-1.86	-0.34	-4.61	-1.32	-1.02	-0.73
1931	-2.31	-0.57	-5.81	-1.41	-1.23	-0.34
1932	-2.99	-0.59	-8.23	-1.77	-2.68	-1.14
1933	0.76	0.80	-0.08	0.97	-1.05	-0.01
1934	-0.43	0.10	-2.28	-0.45	-0.46	0.15
1935	-0.03	0.15	-0.82	-0.01	-0.46	-0.09
1936	2.62	1.92	3.48	1.89	-0.32	-0.59
1937	-0.65	-0.49	-0.33	-0.18	0.80	0.73
1938	-0.76	0.03	-5.21	-3.47	-1.93	-1.39
1939	2.17	1.30	2.76	0.51	0.41	-0.16
1940	1.73	0.90	2.75	0.21	-0.25	-1.04
1941	-0.87	-2.04	4.56	0.81	3.16	1.33

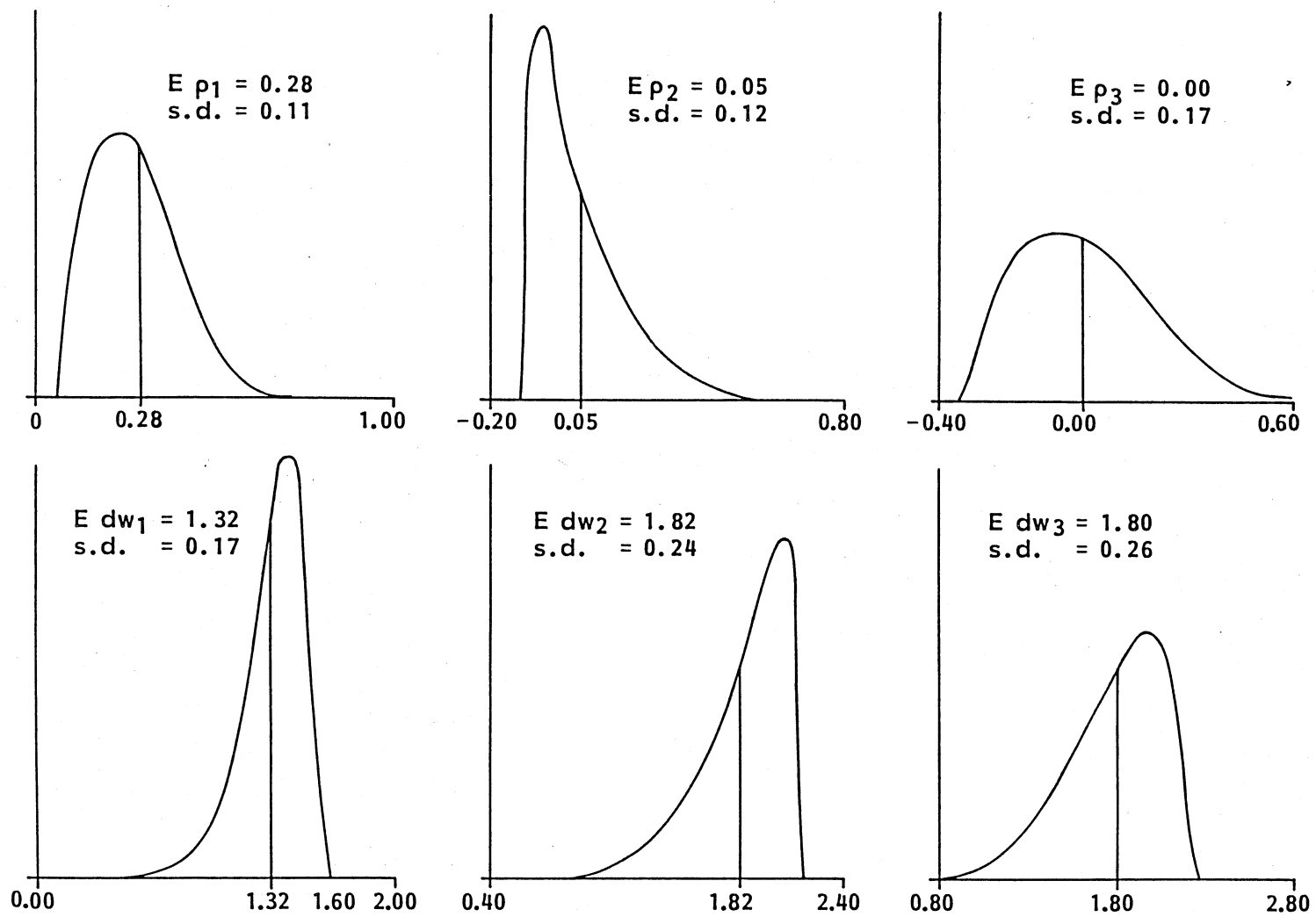


Figure 9. Marginal posterior densities of serial correlation coefficients (ρ) and Durbin-Watson coefficients (dw).

5. MIXED INTEGRATION

Mixed integration, henceforth referred to as MIN, is a numerical integration method for the evaluation of multivariate integrals where the integrand is multivariate skew, that is, the integrand has different tail behavior in different directions. Such integrands contrast with integrands that have symmetric tail behaviour, e.g., the multivariate normal density function and the multivariate Student-t density function. Some examples of contours of multivariate skew functions are presented in Figure 10. The distinctive feature of MIN is that it employs a mixture of one-dimensional classical numerical quadrature and importance sampling. The method consists of two main steps. First, one generates a point $\theta^{(1)}$ from a multivariate normal distribution that has the posterior mode θ^0 as its center and minus the inverse of the Hessian of the logposterior, evaluated at the posterior mode, as its covariance matrix V . (We assume that θ^0 has been estimated by a preliminary optimization procedure.) A generated point $\theta^{(1)}$ defines a line through $\theta^{(1)}$ and θ^0 . As a second step of the MIN-method one performs one-dimensional numerical integration along the line mentioned above, where the integrand is the posterior kernel multiplied by a particular factor that is specified below. One may argue that the MIN-technique conditions on skewness. That is, generating lines, or, more precisely stated, generating directions, by means of a multivariate normal sampling procedure occurs in a symmetric way. Conditional upon a generated direction, one performs a one-dimensional numerical integration step, which takes account of the possible skewness in the integrand.

We summarize the main idea of mixed integration as follows. MIN is based on a transformation of the random vector θ . That is, partition $\theta - \theta^0$ and V^{-1} as

$$\theta - \theta^0 = \begin{pmatrix} u \\ v \end{pmatrix}, \quad V^{-1} = \begin{pmatrix} p & q \\ q' & r \end{pmatrix} \quad (5.1)$$

where v and r are scalars. Define

$$d := (\theta - \theta^0)' V^{-1} (\theta - \theta^0)^{\frac{1}{2}} \quad (5.2)$$

$$\bar{v} := -u'q/r \quad (5.3)$$

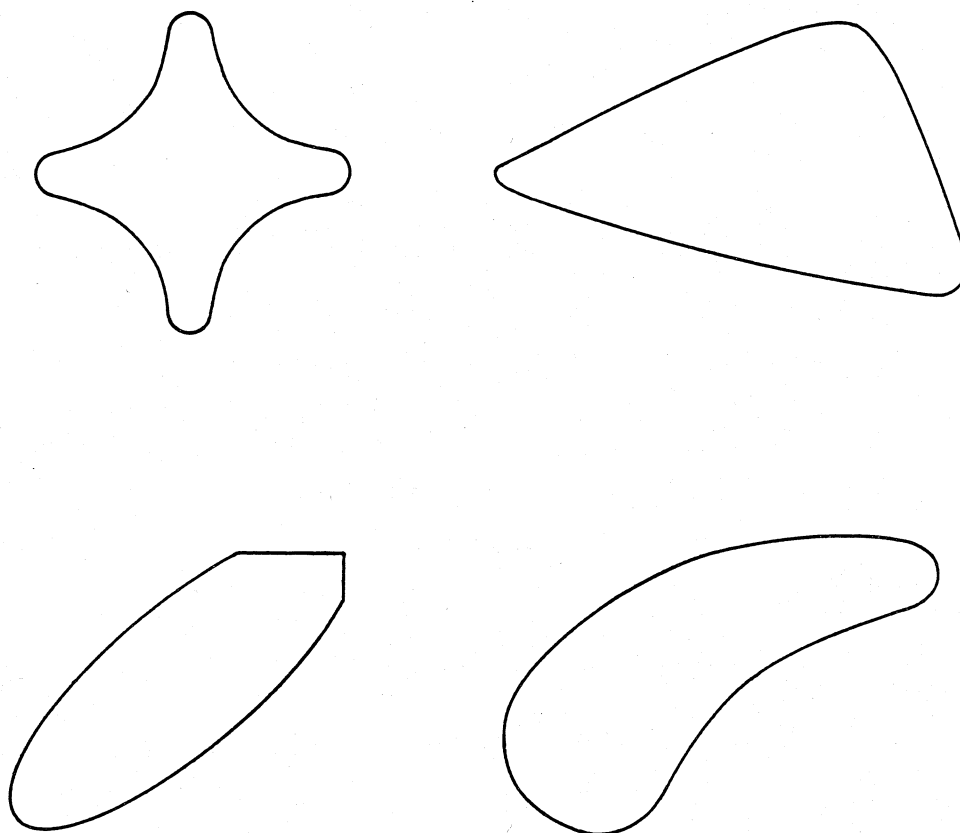


Figure 10. Examples of contours of multivariate skew functions

Then θ is transformed into (η, ρ) where

$$\begin{aligned} \eta &:= u/d, & \rho &:= d & \text{if } v \geq \bar{v} \\ \eta &:= -u/d, & \rho &:= -d & \text{if } v < \bar{v} \end{aligned} \quad (5.4)$$

It is seen from (5.4) that the ℓ -vector of parameters of interest θ is changed into a pair (η, ρ) , where the $(\ell-1)$ -vector η describes the direction of the vector $\theta - \theta^0$ and the scalar ρ describes the distance between θ and the posterior mode θ^0 in a metric that makes use of a covariance matrix V . That is, the scalar ρ satisfies $\rho^2 = (\theta - \theta^0)' V^{-1} (\theta - \theta^0)$ and a sign convention for ρ is added in order to guarantee that the transformation of θ into (η, ρ) is one-to-one. Let T denote the transformation formulae (5.4) that carry θ into (η, ρ) . Then we can write $\theta = T^{-1}(\eta, \rho)$. The actual transformation employed involves a Jacobian determinant

$$|J| = |\rho^{\ell-1}| |J(\eta)| \quad (5.5)$$

where $|J(\eta)|$ is a determinant that depends only on η [compare van Dijk (1984) and van Dijk, Kloek, and Boender (1985)]. If one applies this transformation of variables to the case of the zero-th order moment of θ , one obtains the following result.

$$\int p(\theta) d\theta = \phi_{01} + \phi_{02} \quad (5.6)$$

where

$$\begin{aligned} \phi_{01} &= \int_{\Omega} \left[\int_{R^+} p(T^{-1}(\eta, \rho)) |\rho^{\ell-1}| d\rho \right] |J(\eta)| d\eta \\ \phi_{02} &= \int_{\Omega} \left[\int_{R^-} p(T^{-1}(\eta, \rho)) |\rho^{\ell-1}| d\rho \right] |J(\eta)| d\eta \end{aligned} \quad (5.7)$$

The set Ω is the region of integration of η and R^+ and R^- are, respectively, the positive and negative real line, which are the regions of integration of ρ .

As a next step we make use of an other feature of mixed integration. That is, one can apply the transformation of the random vector θ into the pair (η, ρ) also to the case where θ is normally distributed with mean θ^0 and

covariance matrix V . The transformed density is proportional to $f(\rho)|J(\eta)|$ where $|J(\eta)|$ is the same determinant as given in (5.5). [For details on the specific forms of $f(\rho)$ and $J(\eta)$, see van Dijk, Kloek and Boender (op.cit.).] It follows that random drawings $\eta^{(i)}$ can be generated from a distribution with a density function proportional to $|J(\eta)|$ by simply generating random drawings $\theta^{(i)}$ from a multivariate normal distribution with mean θ^0 and covariance matrix V and then applying the transformation involved. So, given a sample of random drawings $(\theta^{(1)}, \dots, \theta^{(N)})$ one obtains two subsamples, say $(\eta_1^{(1)}, \dots, \eta_1^{(N_1)})$ and $(\eta_2^{(1)}, \dots, \eta_2^{(N_2)})$ with $N_1 + N_2 = N$. [Whether a random drawing of η belongs to the first or second set depends on the inequality conditions (5.4).]

In the actual computations we make use of the property that the generation of directions occurs in a symmetric way. That is, if $\theta^{(i)} - \theta^0$ is a generated point, one can take $\theta^0 - (\theta^{(i)} - \theta^0)$ as a next generated point since these two points are symmetric around θ^0 and they define opposite directions. The effect of such a sampling scheme, which is called antithetic sampling, is that each generated direction is used twice and that the one-dimensional integrals in ϕ_{01} and ϕ_{02} are computed on the entire real line. Therefore, we need not distinguish in the actual computations between ϕ_{01} and ϕ_{02} , except for the estimation of the numerical errors [see van Dijk and Hop (1987)]. So, given a random sample $\theta^{(1)}, \dots, \theta^{(N)}$, we have one set of generated directions $\eta^{(1)}, \dots, \eta^{(N)}$, and the left hand side of (5.6) is estimated by

$$\hat{\phi}_0 = \frac{1}{2N} \sum_{i=1}^N w_0(\eta^{(i)}) \quad (5.8)$$

where

$$w_0(\eta^{(i)}) = \int_{-\infty}^{+\infty} p(T^{-1}(\eta^{(i)}, \rho)) |\rho|^{\ell-1} d\rho \quad (5.9)$$

We divide in (5.8) by $2N$ since each generated direction $\eta^{(i)}$ is used twice. Note that we make use of a proportionality constant since the random sample $\eta^{(1)}, \dots, \eta^{(N)}$ stems from a distribution with a density proportional (and not equal) to $|J(\eta)|$ and we have not written the proportionality constant in an explicit way. Since we deal always with ratios of integrals these proportionality constants cancel.

Next, we discuss the computation of the posterior first-order moments by means of MIN, which illustrates the statement that MIN conditions on skewness. We start with rewriting θ in the following way. Let y be an auxiliary random

variable, defined as

$$y := \frac{\theta - \theta^0}{\rho} \quad (5.10)$$

so that

$$\theta = \theta^0 + \rho y \quad (5.11)$$

Since θ^0 may be interpreted as a vector of known constants, it follows from (5.11) that

$$E\theta = \theta^0 + E\rho y \quad (5.12)$$

Equation (5.12) illustrates that the term $E\rho y$ represents skewness. In the case of symmetry, when $E\theta = \theta^0$, one has the property $E\rho y = 0$. The computation of the posterior mean in case of skewness proceeds as follows. We write

$$E\theta = \theta^0 + \frac{\int \rho y p(\theta) d\theta}{\int p(\theta) d\theta} \quad (5.13)$$

A mixed integration estimator for the integral in the numerator of equation (5.13) can be derived by going through the same transformation of variables as for the case of the denominator. By making use of results that are similar to the ones given in equations (5.6)-(5.9) and (5.11) it follows that the vector of integrals in the numerator of (5.13) can be approximated by

$$\hat{\phi}_j \approx \frac{1}{2N} \sum_{i=1}^N y_j^{(i)} w_1(\eta^{(i)}) \quad (j = 1, \dots, \ell) \quad (5.14)$$

where

$$w_1(\eta^{(i)}) = \int_{-\infty}^{+\infty} \rho p(\theta^0 + \rho y^{(i)}) |\rho|^{\ell-1} |d\rho| \quad (5.15)$$

Note that the integrand in (5.15) is equal to the integrand of (5.9) premultiplied by a factor ρ . A MIN-estimator for $E\theta_j$, $j = 1, \dots, \ell$, can now be written as

$$\hat{\theta}_j = \theta_j^0 + \frac{\hat{\phi}_j}{\hat{\phi}_0} \quad (j = 1, \dots, \ell) \quad (5.16)$$

[compare (5.6)-(5.9) and (5.13)-(5.15)]. Note that we made use of a proportionality sign in (5.8) and (5.14), but that we use an equality sign in (5.16). Since the numerator and denominator in (5.16) have been estimated using the same random sample, it follows that the numerical constants, which were omitted in (5.8) and (5.14), cancel and one can make use of an equality sign in (5.16). Second-order moments can be computed in a similar way as first-order moments. However, for the case of second-order moments one makes use of an integral $w_2(\eta^{(i)})$ [compare (5.15)] that has as integrand the function given in (5.9) premultiplied by a factor ρ^2 .

The structure of the computer program for mixed integration is shown in the flow diagram in Figure 11. It is seen that random drawings $\theta^{(i)}$ are generated from a normal distribution function. After the transformation of variables, which carries θ into (η, ρ) , three one-dimensional integrals are computed. Particular examples of the function $h(\eta^{(i)})$ are given in (5.9) and (5.15).

We emphasize that in our applications the region of integration is bounded. As a consequence, the line integrals with respect to ρ are computed on a bounded interval. The upper and lower bound of this interval may be determined as follows. Given that $a_j < \theta_j < b_j$, $j = 1, \dots, \ell$, one can make use of (5.11) and write

$$a_j - \theta_j^0 < \rho y_j < b_j - \theta_j^0 \quad (j = 1, \dots, \ell) \quad (5.17)$$

For each of the ℓ -dimensions one can compute two values of ρ such that the inequalities are binding constraints. That is, if $y_j > 0$, then we define

$$\rho_j^* = \frac{a_j - \theta_j^0}{y_j}, \quad \rho_j^{**} = \frac{b_j - \theta_j^0}{y_j} \quad (j = 1, \dots, \ell) \quad (5.18)$$

If $y_j < 0$, then ρ_j^* and ρ_j^{**} are interchanged in (5.18). As a next step one determines the minimum value of $(\rho_1^{**}, \dots, \rho_\ell^{**})$ and the maximum value of $(\rho_1^*, \dots, \rho_\ell^*)$. These extreme values are the limits of integration for the line integrals.

Finally, we mention that mixed integration can also be used for the computation of moments of nonlinear functions of θ , such as implied multipliers of simultaneous equations models. The restriction of mixed integration is that for each nonlinear function one has to compute a one-dimensional integral with respect to ρ , given a generated direction $\eta^{(i)}$.

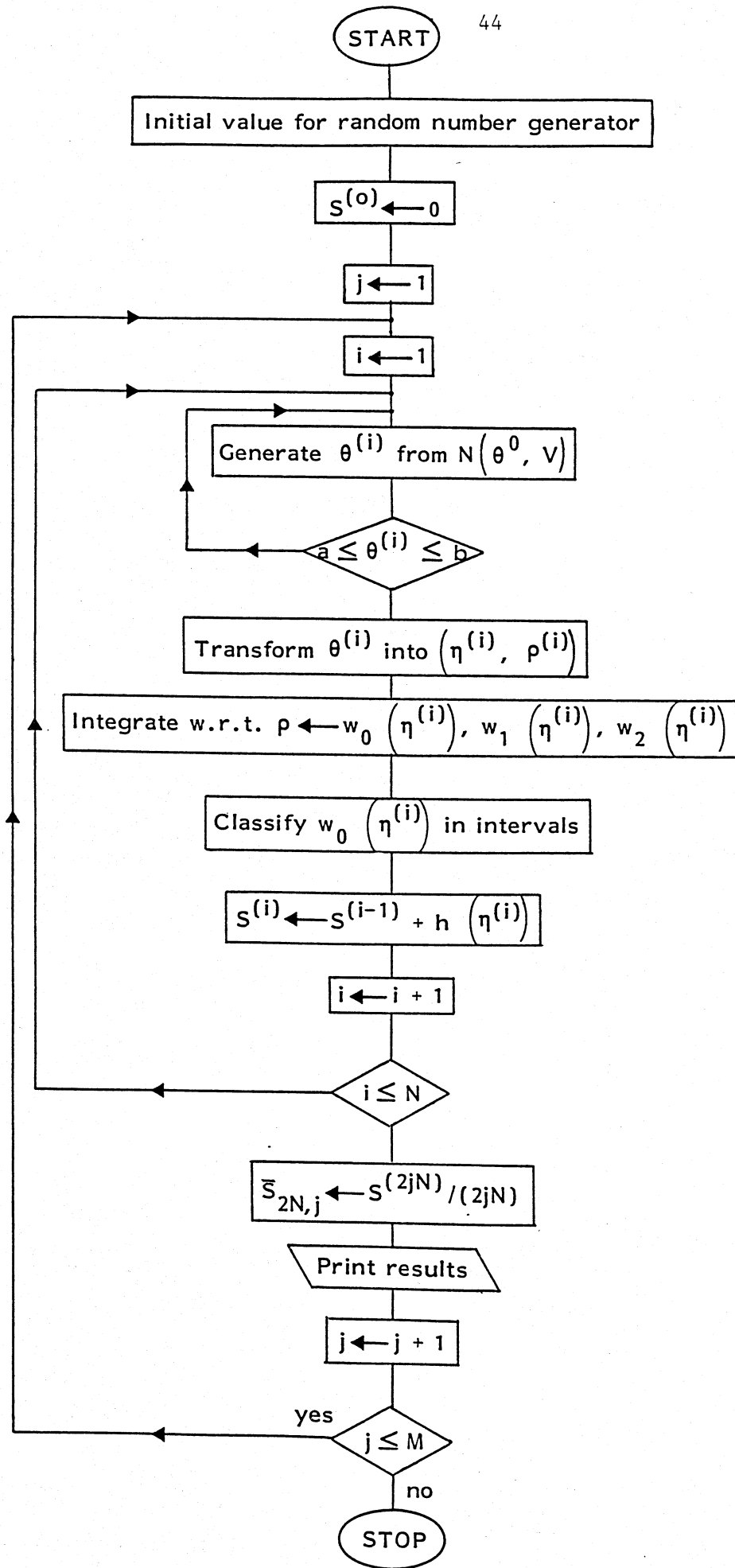


Figure 11. Flow diagram for mixed integration

For an application of mixed integration which involves a thirty-dimensional numerical integration problem we refer to van Dijk and Kloek (1986).

6. REMARKS

In this paper we have discussed some algorithms that can be used for the computation of posterior moments and marginal posterior densities of a vector of parameters θ for the case where the posterior kernel of θ is not a member of known family of density functions. If the surface of the posterior kernel is reasonably well-behaved one may make use of the method of importance sampling with the (truncated) multivariate Student t density as importance function. If the posterior surface is not well-behaved (but unimodal) one may use the method of mixed integration. Given our experience this method is flexible, robust, and parsimonious. That is, it is flexible since one follows the posterior kernel exactly with the computed line integrals; it is robust in the sense that it has handled several cases where the multivariate Student t density failed as importance function; and it is parsimonious in the sense that mixed integration uses the same number of parameters as a normal importance function (i.e., location and scale parameters). Mixed integration may not always be efficient in terms of CPU-time. A topic of further research is the search for flexible functional forms that can serve as importance function. Finite mixtures appear to be very flexible but they have also a large number of parameters that have to be specified a priori. A simple proposal in this area will be reported in a forthcoming paper. In order to experiment with the algorithms mentioned in this paper, we have prepared a set of standard programs [van Dijk and Hop (1987)]. The development of Bayesian software is an active area of research.¹

An other part of this paper was the application of Bayesian estimation methods to a particular econometric model. The results cast some doubt on the specification of the prior densities and on the specification of the likelihood function. So, apart from the development of Bayesian estimation procedures there is a need for procedures that evaluate the information in the prior and the likelihood function.

1. Recently, Tierney and Kadane (1986) proposed a method for the computation of posterior moments and densities that avoids numerical integration.

References

- Abramowitz, M. and I.A. Stegun, 1964, Handbook of mathematical functions (National Bureau of Standards, Washington, D.C.).
- 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.
- Bauwens, L., 1984, Bayesian full information analysis of simultaneous equation models using integration by Monte Carlo (Springer-Verlag, Berlin).
- Boender, C.G. and H.K. van Dijk, 1987, Bayesian analysis of a multi-criteria decision model, Report of the Econometric Institute (Erasmus University Rotterdam) to appear.
- Brent, R.P., 1974, Algorithm 488, A Gaussian pseudo-random number generator [G5], Communications of the ACM 17, 704-706.
- Cramér, H., 1946, Mathematical methods of statistics (Princeton University Press, Princeton, N.J.).
- Geweke, J., 1986, Exact inference in the inequality constrained normal linear regression model, Journal of Applied Econometrics 1, pp. 127-141.
- Hammersley, J.M. and D.C. Handscomb, 1964, Monte Carlo methods (Methuen, London).
- Hendry, D.F., 1984, Monte Carlo experimentation in economics, in: Z. Griliches and M.D. Intriligator, eds., Handbook of Econometrics, Vol. 2 (North-Holland, Amsterdam).
- Historical statistics of the United States; Colonial times to 1970, Bicentennial Edition, 1975 (U.S. Bureau of the Census, Washington, D.C.).
- Kass, R.E., 1985, Inferences about principal components and related quantities using a numerical delta method and posteriors calculated by simulation, Technical report 346 (Department of Statistics, Carnegie-Mellon University).
- Kiefer, N.M., 1981, Limited information analysis of a small underidentified macroeconomic model, International Economic Review 22, 429-442.
- Kinderman, A.J. and J.F. Monahan, 1980, New methods for generating Students' t and Gamma variables, Computing 25, 369-377.
- Kinderman, A.J. and J.G. Ramage, 1976, Computer generation of normal random variables, Journal of the American Statistical Association 71, 893-896.
- 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* (North-Holland, Amsterdam).
- Kooiman, P., H.K. van Dijk, and A.R. Thurik, 1985, Likelihood diagnostics and Bayesian analysis of a micro-economic disequilibrium model for retail services, *Journal of Econometrics* 29, 121-148.
- Marsaglia, G., 1984, The exact-approximation method for generating random variables in a computer, *Journal of the American Statistical Association* 79, 218-221.
- Mood, A.M., F.A. Graybill and D.C. Boes, 1974, *Introduction to the theory of statistics*, third edition (McGraw-Hill, New York).
- Numerical Algorithms Group (NAG), 1983, Mark 10 (Oxford University Computing Laboratory, Oxford).
- Newman, T.G. and P.L. Odell, 1971, *The generation of random variates* (Griffin, London).
- Quandt, R., 1983, Computational problems and methods, Chapter 12 in: Z. Griliches and M. Intriligator, eds., *Handbook of econometrics*, Vol. 1 (North-Holland, Amsterdam).
- Raiffa, H. and R. Schlaifer, 1961, *Applied statistical decision theory* (Graduate School of Business Administration, Harvard University, Boston).
- Ripley, B.D., 1983, Computer generation of random variables: A tutorial, *International Statistical Review* 51, 301-319.
- Rothenberg, T.J., 1963, A Bayesian analysis of simultaneous equation systems, Report 6315 (Econometric Institute, Rotterdam).
- Rothenberg, T.J., 1973, *Efficient estimation with a priori information* (Yale University Press, New Haven).
- Rubinstein, R.Y., 1981, *Simulation and the Monte Carlo method* (Wiley, New York).
- Tierney, L., and J.B. Kadane, 1986, Accurate approximations for posterior moments and densities, *Journal of the American Statistical Association* 81, 82-86.
- 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).
- van Dijk, H.K., 1984, Posterior analysis of econometric models using Monte

- Carlo integration, doctoral dissertation (Erasmus University Rotterdam).
- van Dijk, H.K., 1985, A note on simple algebraic relations between structural and reduced form parameters, Report 8549 (Econometric Institute, Erasmus University Rotterdam).
- van Dijk, H.K. and J.P. Hop, 1987, PMMC: A set of computer programs for the computation of posterior moments and densities using Monte Carlo integration, Report of the Econometric Institute, to appear.
- van Dijk, H.K., J.P. Hop, and A.S. Louter, 1986, An algorithm for the computation of posterior moments and densities using simple importance sampling, Report 8625/A (Econometric Institute, Erasmus University Rotterdam), to appear in abridged form in *The Statistician*.
- 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 application of Bayesian methods* (North-Holland, Amsterdam).
- van Dijk, H.K. and T. Kloek, 1980, Further experience in Bayesian analysis using Monte Carlo integration, *Journal of Econometrics* 14, 307-328.
- van Dijk, H.K. and T. Kloek, 1983, Monte Carlo analysis of skew posterior distributions: An illustrative econometric example, *The Statistician* 32, 216-223.
- van Dijk, H.K. and T. Kloek, 1985, Experiments with some alternatives for simple importance sampling in Monte Carlo integration (with discussion), in: J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M. Smith, eds., *Bayesian Statistics 2* (North-Holland, Amsterdam), 511-530.
- van Dijk, H.K. and T. Kloek, 1986, Posterior moments of the Klein-Goldberger model, in: P.K. Goel and A. Zellner, eds., *Bayesian inference and decision techniques with applications* (North-Holland, Amsterdam), 95-108.
- van Dijk, H.K., T. Kloek, and C.G.E. Boender, 1985, Posterior moments computed by mixed integration, *Journal of Econometrics* 29, 3-18.
- Zellner, A., 1971, *An introduction to Bayesian inference in econometrics* (Wiley, New York).
- Zellner, A., 1985, Further results on Bayesian minimum expected loss (MELO) estimates and posterior distributions for structural coefficients, Manuscript (H.G.B. Alexander Research Foundation, Graduate School of Business, University of Chicago).

