ECONOMETRIC INSTITUTE

BAYESIAN ESTIMATES OF EQUATION SYSTEM PARAMETERS AN APPLICATION OF INTEGRATION BY MONTE CARLO

T. KLOEK and H.K. van DIJK

REPORT 7622/E

ERASMUS UNIVERSITY ROTTERDAM, P.O. BOX 1738, ROTTERDAM, THE NETHERLANDS

Erasmus University Rotterdam
ECONOMETRIC INSTITUTE

Report 7622/E

BAYESIAN ESTIMATES OF EQUATION SYSTEM PARAMETERS
An Application of Integration by Monte Carlo

by T. Kloek and H.K. van Dijk

November, 1976

BAYESIAN ESTIMATES OF EQUATION SYSTEM PARAMETERS*

An Application of Integration by Monte Carlo

by T. Kloek and H.K. van Dijk

ABSTRACT

Monte Carlo (MC) is used to draw parameter values from a distribution defined on the structural parameter space of an equation system. Making use of the prior density, the likelihood, and Bayes Theorem it is possible to estimate posterior moments of both structural and reduced form parameters. The MC method allows a rather liberal choice of prior distributions. The number of elementary operations to be performed need not be an explosive function of the number of parameters involved. The method overcomes some existing difficulties of applying Bayesian methods to medium size models.

The method is applied to a small scale macro-model. The prior information used stems from considerations regarding short and long run behavior of the model and from extraneous observations on empirical long term ratios of economic variables. Likelihood contours for several parameter combinations are plotted and some marginal posterior densities are assessed by MC.

Contents

* Earlier versions of this paper (Report 7511) were presented, a.o., at the World Meeting of the Econometric Society, Toronto, 1975; at CORE, Louvain, and at the First European Congress on Foundations and Applications of Bayesian Methods, Fontainebleau, 1976. The authors are indebted to J. Drèze, R. Harkema, J.F. Richard, and A. Zellner for valuable comments and to an anonymous referee for several useful suggestions.

1. INTRODUCTION

In recent years several Bayesian methods of estimating parameters of simultaneous equation systems have been introduced; see, e.g., Drèze [5], Zellner [21], Harkema [12], Rothenberg [18, 19], and Richard [17], and the references cited there. An important motive for research in this area is the analysis of economic policy problems from a decision theoretic point of view. It appears that in this context Bayesian estimates are more satisfactory than classical ones. The analysis of these problems requires the use of numerical methods, for, in order to obtain analytically tractable results, restrictions have to be imposed which are less attractive from an economic point of view; see Rothenberg [18, pp.139-144] and Harkema [12].

The application of numerical methods appears to be hampered by the amount of computational work involved; see Rothenberg [18, p.140]. However, the numerical work for several econometric problems is restricted to the computation of first and second order moments, e.g., analysis of economic policy problems based on a quadratic loss function, see Zellner [21, Chapter 11], or MELO estimators of ratios of parameters; see Zellner [22].

Usually, standard numerical integration methods, like Cartesian product rules based on Gaussian or Newton-Cotes quadrature formulas, are used. We propose a Monte Carlo method, which enables one to compute the moments mentioned above in the following way. One starts with specifying a so-called importance function. This is a density function defined on the space of structural parameters, or on the space of a subset of these parameters in case part of the integration is carried out analytically. It should have convenient Monte Carlo properties, in the sense that it is not difficult to generate drawings from such a distribution. In addition, the importance function should be an approximation of the posterior density. In some cases the prior density can be used. Making use of the prior density, the likelihood of a given sample, and Bayes' theorem it is possible to obtain estimates of the posterior moments of both structural and reduced form parameters. In case the numerical accuracy of these estimates (which can also be estimated) is not satisfactory, one performs a second round of Monte Carlo using the preliminary estimates of the first round as moments of a new importance function. We name this technique: Posterior Moments computed by means of a Monte Carlo method (PMMC). We also use a Monte Carlo technique to derive marginal posterior distributions for some particularly interesting parameters. In fact, their approximations

can be interpreted as moments of certain functions of the parameters.

The important advantage of Monte Carlo is that a large number of posterior moments can be estimated at a reasonable computational effort and that estimates of the numerical accuracy of these results are obtained in a simple way. There are several indications that Monte Carlo is computationally efficient in problems with many dimensions, say, more than five or six.

Since Monte Carlo is a sampling method, the error goes to zero as $N^{-\frac{1}{2}}$, where N is the number of points where the integrand f is evaluated. Haber [10, p.515] comments on this as follows:[1]

> "This convergence does not seem to be very rapid until we note that neither the dimensionality s of the integration region, nor any specification of the degree of smoothness of f, entered into the determination of the error estimate. All that is required of f is that the integrals entering into the quantity $\sigma(f)$ exist and are finite - f need not even be continuous. In this situation no deterministic error bound is available at all. By Bahvalov's theorem, even if we assumed that $f \in C_1^s$, the best that we could say about the error of any non-probabilistic quadrature formula would be that it is $(N^{-1/s})$ - and if s is, say, 5, this is much slower convergence than is given by the Monte Carlo method."

This comparison is a simplification in two respects. On the one hand more sophisticated methods of integration require amendments to the statements made above. On the other hand, in most cases σ will be an increasing function of d. To what extent σ will increase with d is, in general, unknown and will depend on the properties of the integrand. The same holds for the improvements that can be obtained by employing more sophisticated integration methods. This explains why it is so difficult to give general conclusions. Some of our own experiments with a nine-dimensional integrand suggest that the advantage of Monte Carlo is a real one. For expository reasons, however, we prefer to present a three-dimensional example.

By using a numerical method, we get rid of the restrictions on the prior distribution imposed by the use of analytical techniques; therefore, a liberal choice of prior distributions is possible. To demonstrate this we used prior information of two types in our example. Firstly, we experimented with prior distributions on structural parameters; in most cases the larger part of prior information will pertain to structural parameters. Secondly, we used prior information on short term and long

[1] Haber uses the symbol $\sigma^2(f)$ for the variance of the integrand and C_1^s for the class of functions of s real variables of whose first order partial derivatives exist.

term multipliers. Also, it was possible to handle a case of nonlinear dependence between two structural parameters.

In section 2 we describe the model assumptions and the class of prior distributions considered. The method we used is formally described in section 3. Sections 4 and 5 deal with an illustration (prior specification and posterior moments, respectively). In section 6 problems of numerical precision are discussed and illustrated. In section 7, we investigated the information conveyed by the likelihood contours of structural parameters and by the marginal posterior densities of some interesting parameters. Section 8 contains concluding remarks.

2. MODEL AND CLASS OF PRIOR DISTRIBUTIONS

Our starting point is the standard version of a linear simultaneous equation system

$$Y\Gamma + ZB = U \tag{2.1}$$

where the matrix Y consists of n observations on G current endogenous variables and the matrix Z of n observations on K predetermined variables; Γ is a G×G matrix and B a K×G matrix of constants, some of which are known a priori; U is a matrix of disturbances. The system (2.1) is supposed to satisfy the following assumptions: (1) $|\Gamma| \neq 0$; (2) the n rows of U are independently and identically distributed as $N(0, \Sigma)$, where Σ is a positive-definite[2] symmetric matrix; (3) Z has full column rank; (4) the row vectors z_s', u_t', u_{t+1}', ..., u_n' are independently distributed for any s, t = 1, ..., n with $s \leq t$.

These assumptions enable us to specify the likelihood function of (2.1)

$$\begin{aligned} \ell(\Gamma, B, \Sigma \mid Y, Z) \propto\ & |\Sigma|^{-\frac{1}{2}n} ||\Gamma||^{n} \\ & \times \exp\{-\tfrac{1}{2}\mathrm{tr}[(n\Gamma'\hat{\Omega}\Gamma + (B - \hat{B})'Z'Z(B - \hat{B}))\Sigma^{-1}]\} \end{aligned} \tag{2.2}$$

where the exponent has been rewritten and $\hat{\Pi} = (Z'Z)^{-1}Z'Y$, $\hat{B} = -\hat{\Pi}\Gamma$, and $n\hat{\Omega} = (Y - Z\hat{\Pi})'(Y - Z\hat{\Pi})$.

[2] This means that (2.1) is not supposed to contain identities. Possible identities may be removed by a preliminary substitution procedure; see Rothenberg [18, Chapter 4, Appendix B]. A simple solution is to ignore the identities, except in the Jacobian $||\Gamma||^{n}$; see Rothenberg and Leenders [20, Section 7].

We next define a rather wide class of prior distributions on the elements of Γ, B, and Σ. We distinguish four types of structural parameters.

1. Some parameters are supposed to be known exactly a priori. These may include the unit diagonal elements of Γ following from normalization and a number of zero elements implied by identifying restrictions[3]. (We follow the classical approach to the identification problem.) No sample information can change our prior information on these parameters, so that no posterior computations are needed. Hence, we substitute these known parameter values in our likelihood function. This will be done throughout the paper without further discussion.

2. The so-called constant terms. We assume that each of the equations in (2.1) contains a nonzero constant term. We arrange these in a vector denoted by ϕ. Since usually little is known a priori, we assume a (locally) uniform prior. As a consequence we can handle these parameters by analytical integration.

3. The elements of the variance-covariance matrix, Σ. We shall assume that little is known a priori and that our prior information is adequately described by the expression $|\Sigma|^{-\frac{1}{2}(G+1)}$; compare Zellner [21, p.225 and 226].[4]

4. All remaining elements of Γ and B; that is, all elements which are a priori unknown, apart from the constant terms. We arrange these in a vector denoted by the symbol θ. Prior distributions of the elements of θ will be discussed later.

Assuming independence, we can summarize our prior density by

$$D(\theta, \phi, \Sigma) \propto p(\theta)|\Sigma|^{-\frac{1}{2}(G+1)} \tag{2.3}$$

where $p(\theta)$ is the prior density of θ, to be specified later. Combining the prior density (2.3) and the likelihood (2.2) one obtains, according to Bayes' theorem, the joint posterior density.

The computational burden can be considerably reduced by handling ϕ and Σ by analytical procedures. So we eliminate Σ and ϕ from the posterior distribution. This is performed in two steps. We firstly integrate the

---

[3] No problem arises when one wants to introduce more complicated restrictions, such as general linear restrictions, nonlinear restrictions, or restrictions across equations.

[4] Alternatively, one may specify an Inverted-Wishart function. In this case one also has to specify a matrix H of prior parameters; see Zellner [21, p.395]. This leads to slight modification in the κ function defined in (2.5) below.

posterior density with respect to the elements of Σ, which yields the joint marginal distribution of (θ, ϕ). This distribution is, except for $p(\theta)$ and $||\Gamma||^n$ of the generalized Student t form.[5] As a second step we eliminate the elements of ϕ by integrating this distribution with respect to ϕ. According to a theorem by Dickey [3] we then obtain for the marginal posterior of θ

$$p'(\theta|Y,Z) \propto \kappa(\theta|Y,Z)p(\theta) \tag{2.4}$$

where

$$\kappa(\theta|Y,Z) = ||\Gamma||^n \cdot |n\Gamma'\hat{\Omega}\Gamma + [B_1 - \hat{B}_1]'Z_1'NZ_1[B_1 - \hat{B}_1]|^{-\frac{1}{2}(n-1)} \tag{2.5}$$

Here B and Z have been partitioned according to $B' = [\phi \vdots B_1']$, $Z = [\iota \vdots Z_1]$, where ι is a column vector of unit elements and $N = I - \iota\iota'/n$. Note that Γ and B_1 depend on θ but not on ϕ.

3. POSTERIOR MOMENTS COMPUTED BY A MONTE CARLO METHOD

Starting from the marginal posterior of θ, given in (2.4), we want to find posterior moments of the structural parameters. For the time being we disregard the constant terms. Hence, all moments to be computed are expectations of functions of θ, $g(\theta)$, say. Notice, that $g(\theta)$ may be a scalar, a vector, or a matrix. The existence of these moments depends on both the likelihood and the prior distribution of θ. If the existence conditions are satisfied, such expectations are given by

$$E[g(\theta)|Y,Z] = \frac{\int g(\theta)\kappa(\theta|Y,Z)p(\theta)d\theta}{\int \kappa(\theta|Y,Z)p(\theta)d\theta} \tag{3.1}$$

where the region of integration is a subspace of the parameter space.

A simple sufficient condition for the existence of the moments considered is that all integrals required are defined on bounded regions and have integrands of bounded variation. Recall the assumption that $|\Gamma| \neq 0$. Then it is easily seen that κ is bounded on any bounded region

[5] See Dickey [3] or Zellner [21, p.273 and Appendix B5].

if $\hat{\Omega}$ is positive definite.[6] The bounded region condition can always be satisfied by choosing truncated prior distributions. For the structural parameters this is obvious. For the reduced form parameters it can be met by choosing the truncation in such a way that the prior density is zero on an open set containing all values of θ, where $|\Gamma| = 0$. This condition implies that $||\Gamma|| > 0$ in the region of integration. The extension to other functions of θ such as long term multipliers is obvious. Examples of these conditions are discussed in section 4.

Next, we consider the computation of the moments just defined by means of a Monte Carlo procedure. Let $I(\theta)$ be a density function defined on the parameter space, to be called importance function. The choice of $I(\theta)$ will be discussed below. Let $M(\theta)$ be defined by

$$M(\theta) = \frac{g(\theta)\kappa(\theta|Y,Z)p(\theta)}{I(\theta)} \tag{3.2}$$

This function is defined on the region where $I(\theta) > 0$. Then the numerator of (3.1) can be written as

$$\int M(\theta)I(\theta)d\theta = E[M(\theta)] \tag{3.3}$$

where the expectation is taken with respect to $I(\theta)$. The denominator is obtained by taking $g(\theta) = 1$.

Now, by means of standard Monte Carlo procedures[7], parameter values θ are drawn at random from the distribution with density $I(\theta)$. For each drawn value of θ, the function $M(\theta)$ is evaluated.[8] Let $\theta_1, \theta_2, \ldots, \theta_n$ be our sequence of random drawings. Then we have approximately

$$E[M(\theta)] = \frac{1}{n}\sum_{i=1}^{n} M(\theta) \tag{3.4}$$

---

[6] The latter condition may cause problems in large models if the number of observations is small. This problem is well known from classical estimation; see, e.g., Fisher [8]. The most natural Bayesian way to solve the problem is by specification of a proper prior distribution on Σ; compare footnote 4. Finally, the possibility that gk is bounded but not of bounded variation does not seem to be realistic in the type of problems we consider.

[7] See, for instance, Hammersley and Handscomb [11].

[8] One may have to scale $M(\theta)$ in order to avoid overflow. This can be done by scaling the data or by means of a preliminary optimization applied to $M(\theta)$.

for sufficiently large n. This is the basic formula of our method.

Next, we discuss the choice of the importance function $I(\theta)$. Obviously, the first requirement is that it should have convenient Monte Carlo properties; that is, generating random drawings θ_i should be relatively simple. Many standard distributions satisfy this requirement. Among the univariate families of distributions we mention the uniform, exponential, Gamma, Beta, Normal, and Student t families. For details see Naylor [15] and the references cited there. If multivariate densities can be factored as products of independent marginal densities or marginal and conditional densities, the above families provide a large number of possibilities. The most obvious choices are the multivariate Normal and Student t families.

The second requirement for the choice of $I(\theta)$ is that the variance σ^2 of $M(\theta)$ [with respect to $I(\theta)$] should be kept small. If one takes a uniform $I(\theta)$, one obtains a few drawings for which $M(\theta)$ is important and a large number of drawings for which M is very close to zero. So, even if one has drawn a sample of several thousands of drawings, the right hand side of (3.4) may mainly depend on a very few values $M(\theta_i)$ which are not close to zero. This explains both why in this case one obtains an unreliable estimate and also why $I(\theta)$ is called importance function. It serves to select as many important drawings as possible. The same conclusion may also be reached in a more formal way. The sample size n required to obtain a given amount of accuracy is proportional to σ^2, as is easily seen from (3.4).

Now, for every estimation problem we have one function κ, one function p, but as many functions g as the number of moments we want to compute plus the zero order moment required for the denominator of (3.1). So two different strategies may be chosen: (i) find a density I for each g to reduce var M as much as possible; (ii) find a density I which is proportional to a good approximation of the posterior kernel κp.

The second approach has several advantages. First, κp is a kernel of a density while $g\kappa p$ is not. (In some cases g may change sign in the region of integration.) So it is probably simpler to find a density which is a good approximation. Second, finding a density I which is a good approximation to a function which is not very well known may be difficult. For that reason it seems preferable if one can confine oneself to solving such a problem only once. Third, if the posterior density is not too far from normal we may start with a rough approximation to the posterior distribution and use the so obtained posterior moments as moments of a

multivariate Normal (or Student) importance function in the second stage. If the prior density is informative and not conflicting with the likelihood, it may be used as importance function in the first stage. Finally, it should be noted that if the sample is not extremely small and if the prior density is not very informative the variation of κ will be much greater than that of either g or p. For that reason one also may start to maximize κ and to evaluate the Hessian of log κ at the maximum as a basis for constructing a normal approximation to κ which may serve as importance function in the Monte Carlo procedure.[9]

This concludes the introduction of the PMMC method. More details will be commented upon in the discussion of the illustration.

4. ILLUSTRATION: MODEL AND PRIOR INFORMATION

For illustration purposes we shall make use of a small scale demand-oriented macroeconomic model described by Johnston[10] [13, p.269]. The structural equations read as follows

$$C_t = \alpha_1 + \beta_1 Y_t + u_{1t} \tag{4.1}$$

$$I_t = \alpha_2 + \beta_2 Y_t + \gamma_2 I_{t-1} + u_{2t} \tag{4.2}$$

$$Y_t = C_t + I_t + Z_t \tag{4.3}$$

C_t represents consumer expenditure, Y_t total expenditure, I_t investment, Z_t exogenous expenditure. The interpretation of the parameters will play an important role when specifying the prior distributions. It is seen that β_1 is the marginal propensity to consume with respect to total expenditure, β_2 the short run marginal propensity to spend on investment goods (not the fixed accelerator), and γ_2 an adjustment parameter of investment. It seems reasonable to assume that these three parameters satisfy

$$0 < \beta_1 < 1, \quad 0 < \beta_2 < 1, \quad 0 < \gamma_2 < 1 \tag{4.4}$$

The determinant of Γ, which is not allowed to vanish, equals $1 - \beta_1 - \beta_2$. The short term income multiplier (STM) of Z_t is given by $\pi_{33} = \partial Y_t / \partial Z_t = 1/(1 - \beta_1 - \beta_2)$. Evidence from many investigations suggests

---

[9] We are indebted to A. Zellner for this suggestion.

[10] Lyttkens [14] used the same model to compare estimates based on several estimation techniques. Some of his results will be commented upon below.

that π_{33} is positive and not extremely high, say less than a fixed number η_1 to be specified in Section 5. So we obtain the prior restriction

$$(4.5) \qquad 1 < \text{STM} = \frac{1}{1 - \beta_1 - \beta_2} < \eta_1$$

The lower bound both in (4.5) and in (4.6) below is implied by the conditions (4.4) and the positivity restrictions on the multipliers.

We also investigated the long term multiplier (LTM) of autonomous expenditure with respect to total expenditure. We shall assume that it is positive and less than η_2 (to be specified in Section 5). So our restriction is

$$(4.6) \qquad 1 < \text{LTM} = \frac{1}{1 - \beta_1 - \beta_2/(1 - \gamma_2)} < \eta_2$$

These restrictions imply that the stability condition of the final form is satisfied. Though the present model is not very realistic and the conclusions must be handled with caution, it is interesting to investigate the implied dynamic characteristics and verify their a priori acceptability; see also Dhrymes [2, p.542].

For the importance function we started to choose the prior densities. This worked well in two out of three cases.[11] In the third case we applied a two-stage procedure: we took the prior in the first stage and the resulting rough approximation of the posterior in the second stage.

The generating process of a sequence of a priori acceptable values of $(\beta_1, \beta_2, \gamma_2)$ runs as follows. Monte Carlo is used to draw a sequence of values of structural parameters and these are tested for some or all of the restrictions (4.4)-(4.6).[12] If a value of $(\beta_1, \beta_2, \gamma_2)$ does not pass such a test a new value is drawn and tested. So the restrictions may be used to truncate the prior densities (and the importance functions).

The prior distributions we use differ widely with respect to the amount of prior information incorporated. We start with a rather weak prior, viz., a uniform distribution on the unit region specified by (4.4).

---

[11] As was pointed out to us by J.F. Richard, this result may not be representative. The accordance between two of our priors and our likelihood seems to be greater than usual.

[12] The results of Lyttkens [14, p.362] satisfy (4.4) and (4.5) in all cases; the FIML estimates satisfy (4.6), the Fix-point estimates imply a negative LTM, while the limited information techniques lead to relatively large LTM values (even 28.33 for 2SLS).

We also experimented with prior distributions reflecting somewhat stronger information. On the basis of various results stated in the empirical literature, we specify a prior 95 per cent interval[13] for the marginal propensity to consume as $.2 < \beta_1 < .8$. With respect to γ_2 we do not know very much, except that it will be positive, less than 1 and probably not close to 1; so we specify $0 < \gamma_2 < .8$ as a prior 95 per cent interval. The short term marginal propensity to spend on investment goods is not very well known, but its long term analogue appears to show a fairly good empirical stability. So we specify $.05 < \beta_2/(1 - \gamma_2) < .25$ as a prior 95 per cent interval.

We make use of both the Normal and the Beta family of prior distributions. The generating process of a sequence of normal drawings runs as follows. Draw at random a value of β_1 from $N(.5, .0225)$, a value of γ_2 from $N(.4, .04)$ and a value of the auxiliary variable u from $N(.15, .0025)$. Then compute β_2 from $\beta_2 = (1 - \gamma_2)u$. The Beta family has the property that the range of the parameter is restricted to the interval $[0, 1]$. To specify parameters for the Beta distributions corresponding with the prior 95 per cent intervals specified above, we made use of the tables of Pearson [16]. This resulted in the choice of $B(5,5)$ for β_1, $B(2,3)$ for γ_2, and $B(7,40)$ for $\beta_2/(1 - \gamma_2)$.

In the case of the uniform prior, the prior density differred too much from the posterior to be acceptable as importance function. The function κ, defined in (2.5) for the Johnston model, has a sharp peak on a small subregion of the unit interval. Only two per cent of the number of drawings had a κ-value exceeding 13.6 per cent (e^{-2}) of the maximum value of κ. The remaining drawings got a negligible weight and a large number of drawings would be necessary to obtain a reasonable degree of accuracy. So we used the two stage procedure mentioned above. In the second stage we experimented with Normal and Cauchy densities as importance function. The rather thick tails of the latter prevent explosive behavior of the ratio $\kappa p/I$. We took the posterior moments of the first stage with a uniform distribution to specify the parameters[14] of $I(\theta)$. The second order moments from the first stage were multiplied by a factor $k > 1$ in

[13] More precisely: a 95 per cent "highest prior density" interval; compare Zellner [21, p.27]. Note that the intervals were slightly modified for the case of the Beta distribution.

[14] Of course, the Cauchy density has no moments of order ≥ 1. In fact, we drew from a multivariate Normal distribution with the moments mentioned in the text and divided each vector of drawings by an independent drawing from an $N(0, 1)$ distribution.

order to enlarge the region from which most of the drawings were made. The optimal value of k was experimentally determined as one which yields the smallest coefficient of variation of the weights $\kappa p/I$. This was reached by the Cauchy type with a k value of 1.5.

Another variance reduction technique, the antithetic variable method, was tried in combination with importance sampling. This did not yield satisfactory results. The probable reason was lack of symmetry of both the κ function for the Johnston model and the (truncated) importance functions (prior densities).

5. NUMERICAL POSTERIOR MOMENTS[15]

Prior and posterior means and standard deviations were computed for the structural parameters $(\beta_1, \beta_2, \gamma_2)$, for the reduced form parameters (apart from the constant terms) and for the long term multiplier. The three prior distributions discussed in section 4 were used. Furthermore, we investigated the sensitivity of the results with respect to changes in η_1 and η_2 of the restrictions (4.5) and (4.6).

Tables 1 and 2 show some results for the posterior moments of the structural and reduced form parameters and the LTM. The restriction (4.5) with $\eta_1 = 100$ was applied. The FIML point estimates have been presented for comparison. The results based on the Normal and Beta priors are virtually the same. In a number of cases the results based on the two informative priors are somewhere between the FIML point estimates and the posterior results based on the uniform prior. This is due to the skewness of the κ function which stretches out into the negative quadrant of (β_1, β_2), due to the factor $||\Gamma||^n$. The standard deviations are much smaller in the cases were an informative prior has been used, as was to be expected. In addition, Table 2 shows the importance of prior information for the posterior standard deviations, in particular that of the LTM. Note, however, that restriction (4.6) was not used in this case.

Next, we investigated the sensitivity of the posterior moments with respect to restrictions (4.5) and (4.6) for various values of η_1 and η_2. If one uses the informative Normal and Beta prior distributions, the results are not sensitive. In case of the uniform prior the structural

[15] The authors are indebted to Mr. A.S. Louter of the Econometric Institute for valuable advice and assistance in preparing the necessary computer programs.

TABLE 1. STRUCTURAL POSTERIOR MEANS AND STANDARD DEVIATIONS*

Prior Distribution	β_1	β_2	γ_2
Classical FIML	0.458	0.089	0.363
estimates **	(0.095)	(0.035)	(0.066)
Uniform (stage 2)	0.341	0.054	0.372
	(0.121)	(0.034)	(0.144)
Normal	0.432	0.079	0.401
	(0.073)	(0.025)	(0.113)
Beta	0.428	0.076	0.407
	(0.076)	(0.025)	(0.128)

* The numbers within brackets are standard deviations.

** Lyttkens results [14, p.361] are slightly different.

TABLE 2; REDUCED FORM* AND LTM POSTERIOR MEANS AND STANDARD DEVIATIONS

Prior Distribution	π_{21}	π_{31}	π_{22}	π_{32}	π_{23}	π_{33}=STM	LTM
FIML. No prior	0.37	1.01	0.43	0.20	0.80	2.21	2.49
Uniform (stage 2)	0.25	0.66	0.41	0.11	0.67	1.77	1.99
	(0.20)	(0.41)	(0.16)	(0.10)	(0.34)	(0.50)	(1.16)
Normal	0.38	0.94	0.47	0.18	0.84	2.12	2.44
	(0.17)	(0.33)	(0.12)	(0.08)	(0.27)	(0.41)	(0.64)
Beta	0.37	0.92	0.47	0.17	0.85	2.09	2.41
	(0.18)	(0.34)	(0.14)	(0.08)	(0.30)	(0.41)	(0.67)

* See the notes of Table 1. The row indices 2 and 3 of π refer to I_{-1} and Z, the column indices 1, 2, 3 to C, I, Y, respectively.

TABLE 3. SENSITIVITY ANALYSIS OF STM AND LTM FOR RESTRICTIONS*

LTM < η_2, with STM < 10	STM	LTM	n_1/n**
η_2 = 100	1.76 (0.49)	1.99 (1.07)	0.72
η_2 = 50	1.76 (0.49)	1.99 (1.03)	0.72
η_2 = 25	1.76 (0.49)	1.99 (1.00)	0.73
η_2 = 12.5	1.76 (0.50)	1.97 (0.81)	0.75
η_2 = 6.25	1.76 (0.48)	1.95 (0.74)	0.82

* A uniform prior on $(\beta_1, \beta_2, \gamma_2)$ is used, with importance sampling.

** n is the number of drawings, which passed the restrictions, n_1 is the number of rejected drawings.

and reduced form posteriors are not sensitive (provided that (4.5) is maintained for $n_1 = 100$) but the posterior standard deviation of the LTM is; see Table 3.

6. NUMERICAL ACCURACY

We are interested in the trade-off between the level of numerical precision of Monte Carlo (MC) estimators of integrals and the number (n) of drawings performed. For example, the results for the first order posterior moments, presented in tables 1 and 2, are based on 1500 drawings. We want to know the accuracy of these results and we want to obtain an idea about the number of drawings sufficient for a given level of accuracy.

MC estimators make use of mean values of a random sample, compare equation (3.4). For sufficiently large n, they possess the property of approximate normality, under the usual conditions of the Central Limit Theorem[16] and the existence conditions of the moments, discussed in section 3. Therefore an asymptotically valid 95 per cent confidence interval for a 1 per cent relative error in the MC estimate indicates a required number of drawings in the following way. Let H be an MC estimator for the value of an integral μ and let σ^2 denote the variance of each drawing, then

$$(6.1) \qquad P\left(\left|\frac{H - \mu}{\sigma/\sqrt{n}}\right| < 1.96\right) \geq 0.95$$

When one imposes

$$(6.2) \qquad \left|\frac{H}{\mu} - 1\right| < 0.005$$

with a 95 per cent confidence level, it follows that

$$(6.3) \qquad n \geq \left(\frac{1.96}{0.005}\right)^2 \left(\frac{\sigma}{\mu}\right)^2 \approx 160{,}000\left(\frac{\sigma}{\mu}\right)^2$$

is a sufficient number of drawings. This result illustrates the importance of variance reduction; compare section 3.

In our case, we deal with MC estimators of ratios of integrals, as shown in equation (3.1). The coefficient of variation of this ratio is derived as follows. Let t_i be an MC estimator of the i-th element of the numerator of (3.1) and let t_0 be an MC estimator of the denominator. So we have in our case $H = t_i/t_0$. Under certain regularity conditions, see

[16] Cramér [1, Chapters 16 and 17].

Cramér [1, pp.353-359], we have

$$(6.4) \quad \text{var } H \approx \left(\frac{\partial H}{\partial t_i}\right)^2 \text{var } t_i + 2\frac{\partial H}{\partial t_i} \cdot \frac{\partial H}{\partial t_0} \text{cov}(t_i, t_0) + \left(\frac{\partial H}{\partial t_0}\right)^2 \text{var } t_0$$

and for the coefficient of variation (squared) one obtains

$$(6.5) \quad \frac{\text{var } H}{H^2} \approx \frac{\text{var } t_i}{t_i^2} + \frac{\text{var } t_0}{t_0^2} - 2\rho(t_i, t_0) \frac{\sqrt{\text{var } t_i \text{ var } t_0}}{t_i \cdot t_0}$$

Some numerical results for the structural parameters (β_1, β_2, γ_2) and the short and long term multiplier are presented in tables 4 and 5.

The results show clearly that the coefficients of variation of the first order PMMC estimates are much smaller than the coefficients of variation of the numerator of such estimates, due to the substantial correlation between numerator and denominator. It is interesting to observe that a good importance function (as used in stage 2) decreases the correlation between numerator and denominator (which is plausible) but this is more than offset by the decrease in the variation coefficients of numerator and denominator. It is also seen from Table 4 that the coefficients of variation for the denominator (for which the importance function was constructed) is not considerably smaller than the corresponding coefficients of variation for the numerator. This justifies our decision to work with one importance function for all integrals.

Table 5 shows the number of required drawings at the prescribed level of accuracy. Notice that if one is satisfied with a two per cent relative error, the numbers have to be divided by four. Similarly if one is satisfied with a 68 per cent confidence level. In fact a round of 1500 drawings was performed which gave reasonably accurate results, except in the case of a uniform prior without importance sampling.

Special attention should be given to β_2, which has a posterior mean of 0.054 if the prior is uniform. In fact, most economists would be satisfied to have an interval estimate with an interval width of 0.01 of such a parameter. But this interval is eighteen times as large as the interval required in Table 5. So for this level of accuracy we may divide n by $18^2 = 324$ so that 426 drawings would suffice, instead of the 1500 we actually used in computing Table 1 or the 138,000 (of Table 5) required to reach the one per cent relative accuracy discussed in the beginning of this section.

TABLE 4. SQUARED VARIATION COEFFICIENTS (TIMES n) OF FIRST ORDER PMMC ESTIMATES AND CORRELATION COEFFICIENTS OF NUMERATOR AND DENOMINATOR OF SUCH ESTIMATES OF (β_1, β_2, γ_2, STM, LTM)*

Coefficient of Variation (squared) for:	β_1	β_2	γ_2	STM	LTM
Uniform Prior (Stage 1)	3.19	12.00	2.36	1.87	19.00
Uniform Prior (Stage 2)	0.25	0.86	0.21	0.15	3.28
Normal Prior	0.09	0.30	0.20	0.11	0.19
Beta Prior	0.09	0.30	0.24	0.11	0.19
Coefficient of Variation (squared) of Numerator for:	β_1	β_2	γ_2	STM	LTM
Uniform Prior (Stage 1)	40.36	46.42	35.39	36.93	53.42
Uniform Prior (Stage 2)	0.87	1.39	0.70	0.82	3.93
Normal Prior	3.37	3.56	3.20	3.31	3.21
Beta Prior	3.63	3.81	3.47	3.55	3.46
Correlation Coefficient of Numerator and Denominator	β_1	β_2	γ_2	STM	LTM
Uniform Prior (Stage 1)	0.96	0.86	0.97	0.97	0.80
Uniform Prior (Stage 2)	0.84	0.62	0.84	0.90	0.41
Normal Prior	0.99	0.96	0.97	0.98	0.97
Beta Prior	0.99	0.96	0.97	0.98	0.97

Coefficient of Variation (squared) of Denominator for:	all parameters
Uniform Prior (Stage 1)	35.62
Uniform Prior (Stage 2)	0.64
Normal Prior	3.40
Beta Prior	3.63

* Based on 80,000 drawings.

TABLE 5. REQUIRED NUMBER (× 1,000) OF DRAWINGS FOR A 1 % ACCURACY OF PMMC ESTIMATES*,**

Prior Distribution	β_1	β_2	γ_2	STM	LTM
Uniform Prior (Stage 1)	510	1,920	378	299	3,140
Uniform Prior (Stage 2)	40	138	34	24	525
Normal Prior	14	48	32	18	30
Beta Prior	14	48	38	18	30

* The numbers are rounded off in thousands

** The accuracy is based on a (asymptotically valid) 95 % confidence interval estimate.

Finally, we mention a much simpler way to get some idea about the accuracy of the results. One may always print the results at, say, $\frac{1}{4}n$, $\frac{1}{2}n$, and $\frac{3}{4}n$ and check the stability in the answers.

7. MESSAGES OBTAINED FROM LIKELIHOOD CONTOURS AND MARGINAL POSTERIOR DENSITIES

Up to now, we concentrated on the computation of first and second order posterior moments. These are useful distribution characteristics for normal and nearnormal distributions, but their value is doubtful for multimodal or very skew distributions. Furthermore, many researchers are reluctant to include prior information in the estimation of parameters of a model, because of the (supposed) subjectiveness of such information. For these reasons, the analyst, who has confined himself to computing low order moments, may feel the need to obtain some messages from the material which warn him in cases where the likelihood surface is multimodal or very skew, or exhibits nonlinear ridges.[17]

Some of these problems can be detected by a careful study of the shape of the likelihood function. In our opinion it is a good strategy to start with computing full-information maximum-likelihood estimates. Numerical optimization routines usually indicate such problems as flat segments of the likelihood function and secondary maxima. The flat likelihood function

[17] If the ridges are linear (or near-linear) and relatively flat they are reflected by large correlation coefficients, which can easily be found by the moment approach, but are not presented here to save space.

reflects near-identification, or multicollinearity or both.[18] Furthermore the Jacobian of the transformation from the structural disturbances to the dependent variables usually introduces skewness in the likelihood function.

In order to gain some insight into the likelihood function of the Johnston model, we plotted likelihood contours for several parameter combinations of the well known concentrated likelihood function

$$(7.1) \qquad \ell'(\theta|Y,Z) \propto ||\Gamma||^{n}|n\Gamma'\hat{\Omega}\Gamma + (B_1 - \hat{B}_1)'Z_1'MZ_1(B_1 - \hat{B}_1)|^{-\frac{1}{2}n}$$

obtained by algebraic maximization of equation (2.2) of section 2 with respect to ϕ and Σ. Note that

$$\ell'(\theta|Y,Z) \propto [\kappa(\theta|Y,Z)/||\Gamma||]^{n/(n-1)}$$

Three graphs of the likelihood surface of the Johnston model are presented for the following combinations: (β_1, β_2), (β_1, γ_2) and (β_2, γ_2); see figures 1 A, 1 B, and 1 C. These contours are conditional with respect to the third parameter, which is fixed at its FIML-value. The figures are sufficiently suggestive about the shape of the contours. They reveal that the only prior restriction which plays a serious role in the analysis is $\beta_2 > 0$.

One can assess the influence of the prior information from the shape of the posterior densities.[19] These posterior densities are obtained in the same run as the posterior moments. In fact, if appropriate functions $g(\theta)$ are defined, they <u>are</u> posterior moments. The posterior probability $P(a < \beta_2 < b)$, for example, is obtained by (3.1) with

$$(7.2) \qquad g(\theta) = 1 \quad \text{if} \quad a < \beta_2 < b$$
$$= 0 \quad \text{otherwise}$$

It may be estimated by the method described in section 3. And, provided that the intervals (a, b) are small enough, the marginal posterior density of β_2 evaluated at $\frac{1}{2}(a + b)$ is approximately given by

$$(7.3) \qquad \frac{P(a < \beta_2 < b)}{b - a}$$

---

[18] In many cases it is difficult to distinguish between these phenomena. For illustrations, see Goldfeld and Quandt [9] and Fair and Jaffee [7].

[19] Alternatively one may compute higher order moments, compare Van Dijk and Kloek [6].

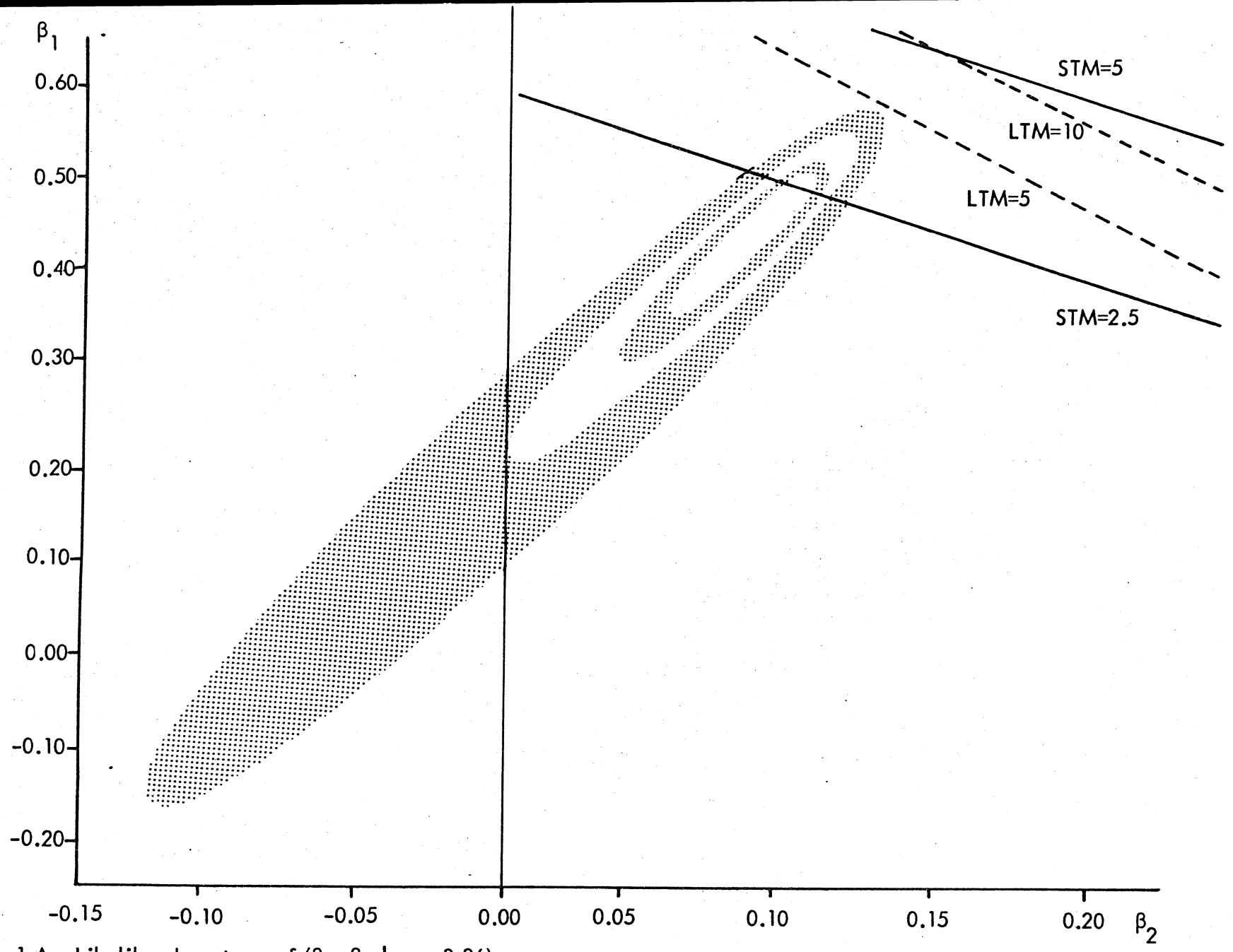

Fig. 1 A. Likelihood contours of $(\beta_1, \beta_2 \mid \gamma_2 = 0.36)$

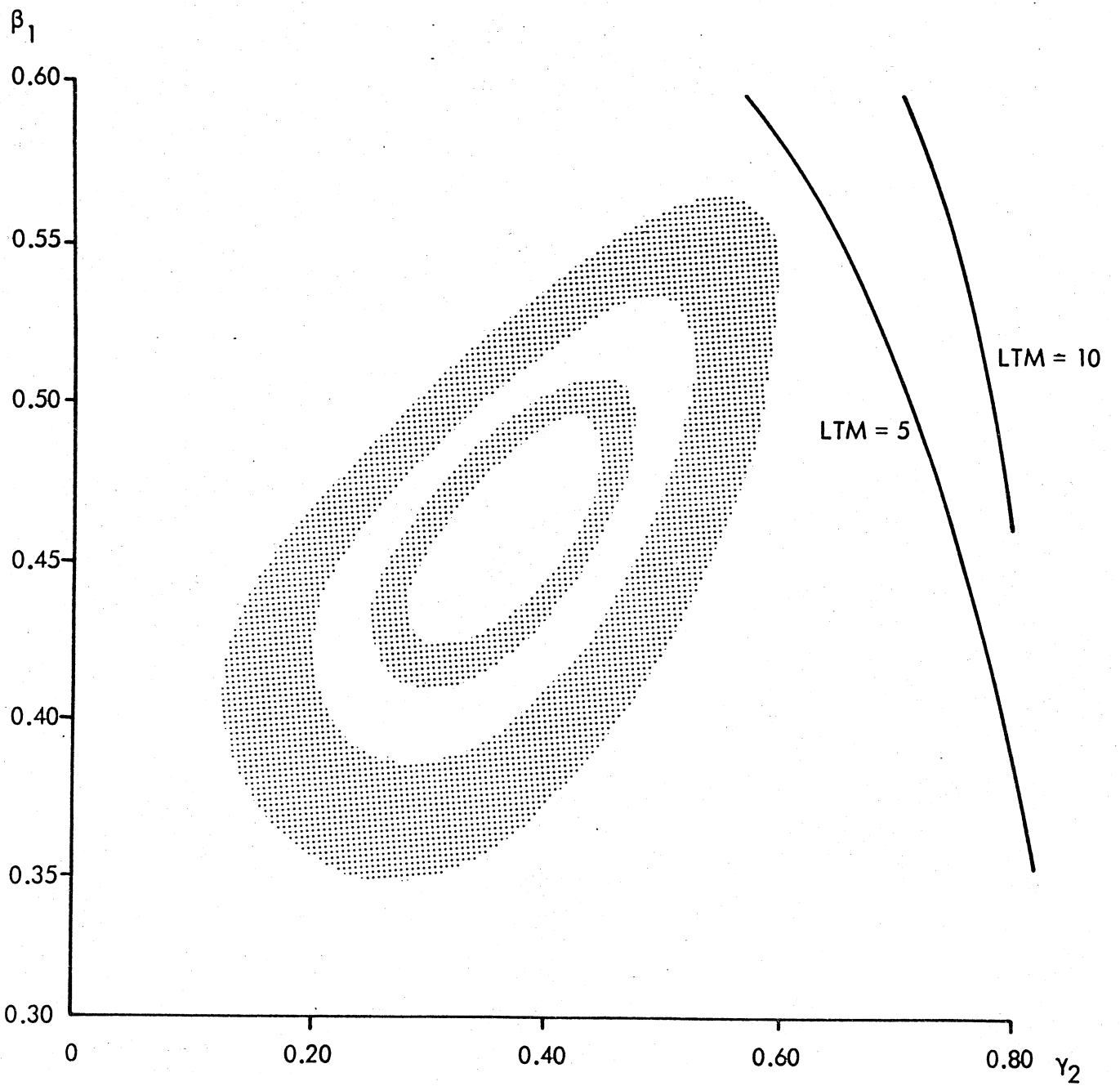

Fig. 1 B. Likelihood contours of $(\beta_1, \gamma_2 | \beta_2 = 0.09)$

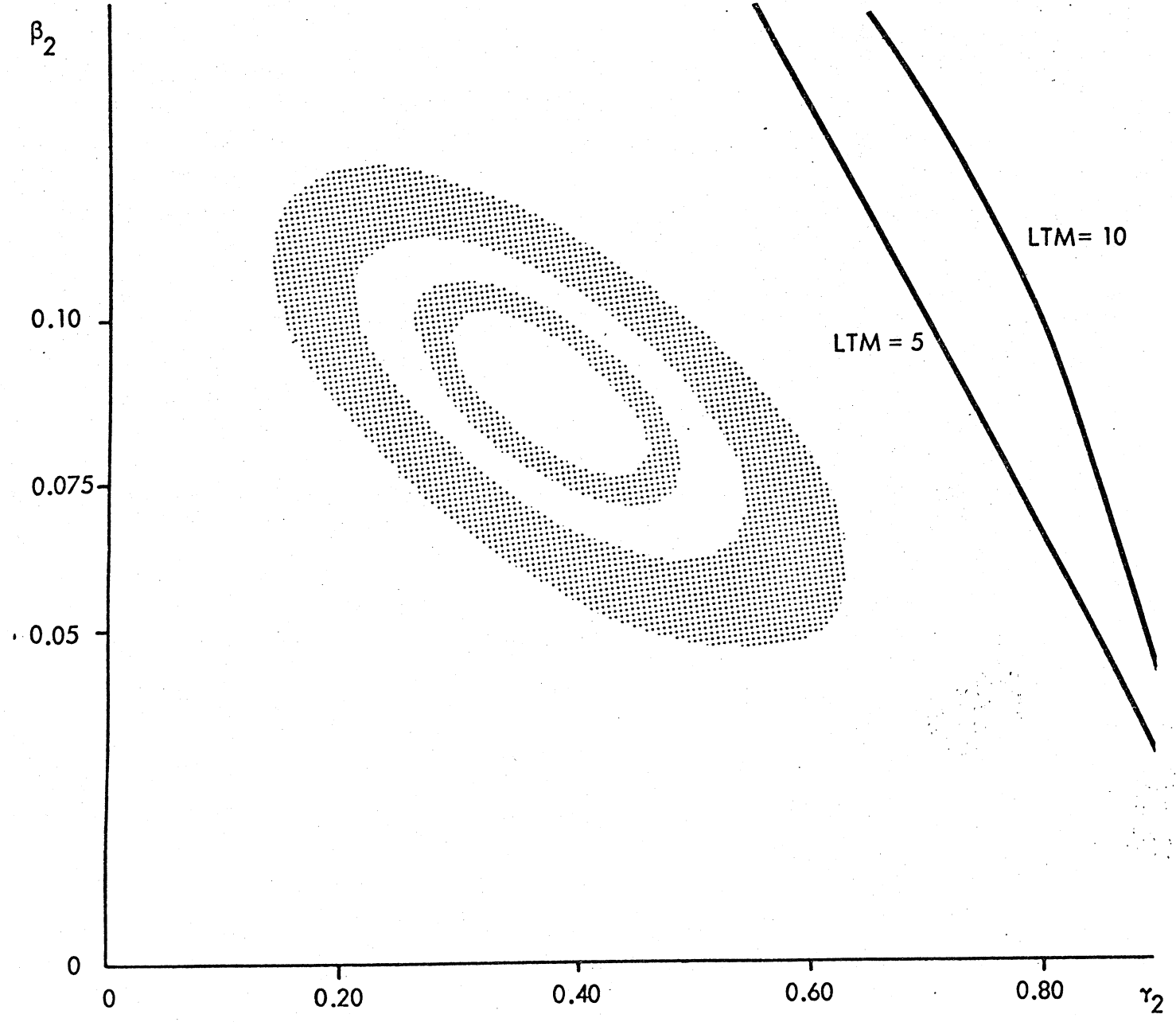

Fig. 1 C. Likelihood contours of $(\beta_2, \gamma_2 | \beta_1 = 0.46)$

The marginal posterior densities of the structural parameter β_2 and of the STM and LTM parameters are presented in figure 2. The prior information used has been described in section 4. Figure 2 reveals that the posteriors based on the uniform prior are skewed to the right as was to be expected because of the Jacobian. The significance of the restriction $\beta_2 > 0$ is clearly seen in the marginal posterior density of β_2 for the uniform prior.[20] The Normal prior shifts the mean of the posterior densities somewhat to the right. The STM and LTM posterior densities indicate very small probabilities for STM and LTM values greater than 4. This is in contrast to several results obtained by classical estimators.

This section gives some procedures to get an overall picture of the functions to be integrated. Not all aspects, however, could be covered and, obviously, more work is needed in this area.

8. CONCLUDING REMARKS

In this paper we applied Monte Carlo methods in order to obtain estimates of posterior moments of structural and reduced form parameters of simultaneous equation systems. The MC methods allow the analyst to make use of several types of exact and stochastic prior information. The MC methods carry a computational workload in high dimensional problems (say, more than five), which appears to be efficient, compared to other methods. In addition, estimates of numerical errors can directly be obtained. Our illustrative example was a small equation system, which served to indicate different ways of using prior information in a Bayesian analysis of a simultaneous equation system. The approach is general enough, however, to be used in the context of other types of models.

We also used Monte Carlo methods in order to compute marginal posterior distributions of some particularly interesting parameters. It is a problem in any numerical integration technique to obtain accurate results in cases of flat tails of the posterior densities. It seems a sensible strategy first to obtain posterior moments, especially in cases of high dimensionality. In such cases summarizing quantities are called for, as a rule; compare Dickey [4].

It may be possible to apply this numerical method to medium size

---

[20] The effect on the marginal posterior densities of β_1 and γ_2 was much less pronounced. Figures for marginal posterior densities of these parameters are therefore omitted.

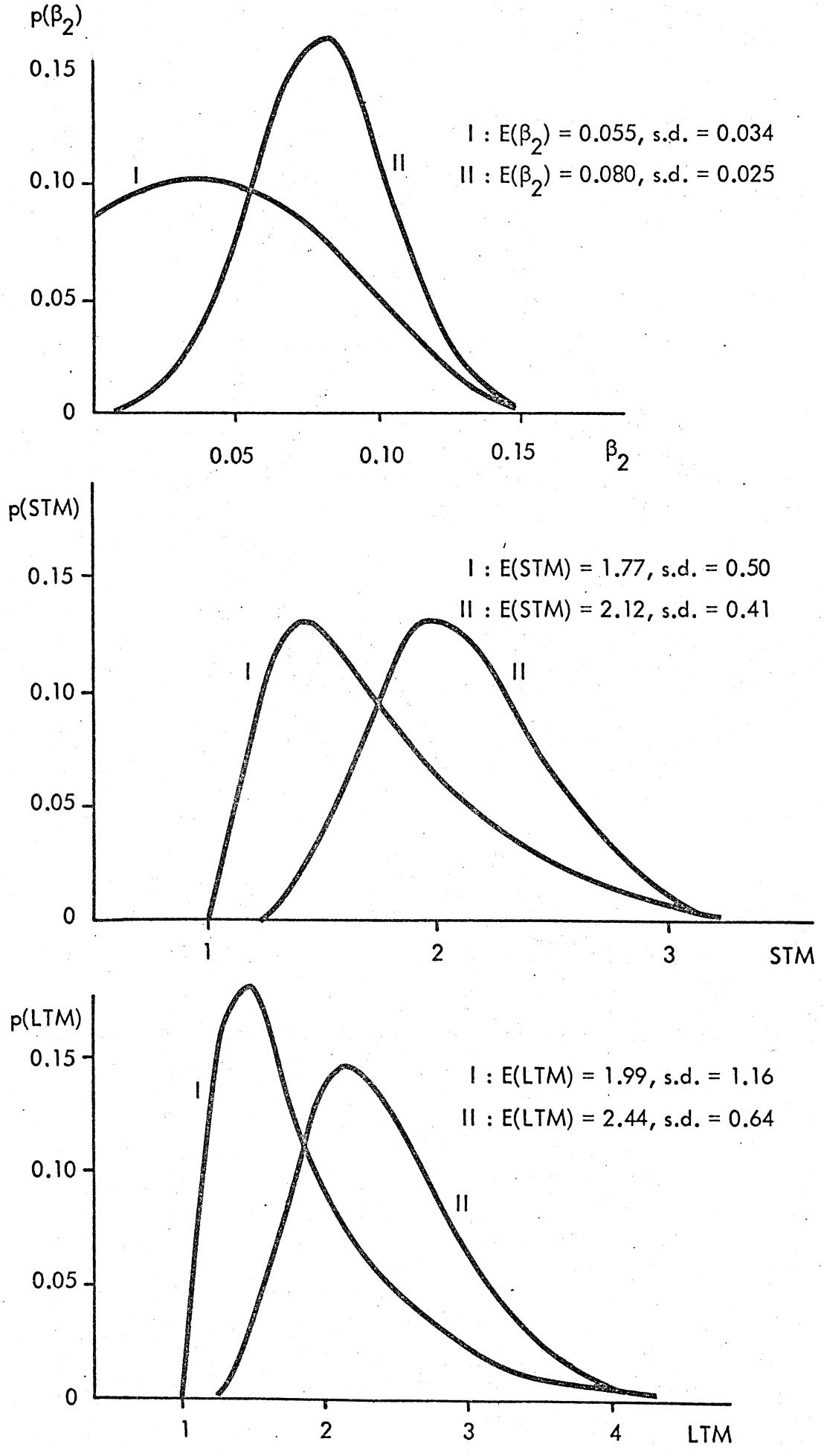

Fig. 2. Marginal posterior densities of β_2, STM, and LTM, with a uniform prior (I) and a normal prior (II).

models. Such a possibility has been a subject of doubt in the Bayesian literature; see Rothenberg [18, p.153] and Richard [17, p.10]. It seems then a good strategy to use partly analytical integration methods and partly the Monte Carlo methods. Such a principle of reducing the computational workload, by using analytical integration whenever possible, is often advocated; see, e.g., Hammersley and Handscomb [11, p.74]. Here we will mention two examples.[21]

(i) We eliminated the constant terms and the Σ matrix by integration from the posterior distributions of the structural and reduced form parameters. It is possible to evaluate the moments of these parameters by using analytical integration methods and numerical results of the Monte Carlo methods; see Van Dijk and Kloek [6]. Then one can use the posterior moments in some prediction and decision problems.

(ii) In case the researcher has prior information, which allows him to restrict his attention to the estimation of subsystems of equation systems, he may use Drèze's [5] limited information analysis to derive a posterior density for the parameters of the subsystem. Next, one uses Monte Carlo methods to evaluate posterior moments of the subsystem parameters using prior information of several types.

Finally, we want to emphasize that Monte Carlo methods are a branch of experimental mathematics; compare Hammersley and Handscomb [11, p.2]. Although we are satisfied with the results reported in this paper, much experimentation is still needed before a final answer about the usefulness of Monte Carlo methods in Bayesian simultaneous equation estimation is possible.

REFERENCES

[1] Cramér, H., Mathematical Methods of Statistics, Princeton, N.J.: Princeton University Press, 1946.

[2] Dhrymes, P.J., Econometrics: Statistical Foundations and Applications, New York: Harper and Row, 1970.

[3] Dickey, J.M., "Matricvariate Generalizations of the Multivariate t Distribution and the Inverted Multivariate t Distribution", Annals of Mathematical Statistics, 38 (1967), 511-519.

[4] Dickey, J.M., "Three Multidimensional-Integral Identities with Bayesian Applications", Annals of Mathematical Statistics, 39 (1968), 1615-1628.

[21] The second example was suggested to us by J. Drèze.

[5] Drèze, J.H., "Bayesian Limited Information Analysis of the Simultaneous Equation Model", Econometrica, 46 (1976), 1045-1075.

[6] Dijk, H.K. van, and T. Kloek, "Predictive Moments of Simultaneous Econometric Models, A Bayesian Approach", Report 7614/E of the Econometric Institute, Rotterdam.

[7] Fair, R.C. and D.M. Jaffee, "Methods of Estimation for Markets in Disequilibrium", Econometrica, 40 (1972), 497-514.

[8] Fisher, F.M., "Dynamic Structure and Estimation in Economy-wide Econometric Models", Ch. 15 of J.S. Duesenberry et al., The Brookings Quarterly Econometric Model of the United States, Chicago: Rand-McNally, 1965.

[9] Goldfeld, S.M., and R.E. Quandt, Nonlinear Methods in Econometrics, Amsterdam: North Holland, 1971.

[10] Haber, S., "Numerical Evaluation of Multiple Integrals", SIAM Review 12 (1970), 481-526.

[11] Hammersley, J.M., and D.C. Handscomb, Monte Carlo Methods, London: Methuen and New York: John Wiley, 1964.

[12] Harkema, R., Simultaneous Equations, A Bayesian Approach, Rotterdam: University Press, 1971.

[13] Johnston, J., Econometric Methods, 1st ed., New York: McGraw-Hill, 1963.

[14] Lyttkens, E., "The Fix-Point Method for Estimating Interdependent Systems with the Underlying Model Specification", Journal of the Royal Statistical Society, 136 (1973), 353-375.

[15] Naylor, T.H., Computer Simulation Experiments with Models of Economic Systems, New York: John Wiley, 1971.

[16] Pearson, K., Tables of the Incomplete Beta-Function, Cambridge: University Press, 1956 printing.

[17] Richard, J.F., Posterior and Predictive Densities for Simultaneous Equation Models, Berlin: Springer Verlag, 1973.

[18] Rothenberg, T.J., Efficient Estimation with A Priori Information, New Haven and London: Yale University Press, 1973.

[19] Rothenberg, T.J., "Bayesian Analysis of Simultaneous Equations Models", Chapter 9.3 of Fienberg S.E. and A. Zellner (eds.), Studies in Bayesian Econometrics and Statistics, Amsterdam: North Holland, 1975.

[20] Rothenberg, T.J., and C.T. Leenders, "Efficient Estimation of Simultaneous Equation Systems, Econometrica", 32 (1964), 57-76.

[21] Zellner, A., An Introduction to Bayesian Inference in Econometrics, New York: John Wiley, 1971.

[22] Zellner, A., "Estimation of Functions of Population Means and Regression Coefficients including Structural Coefficients: A Minimum Expected Loss (MELO) Approach", Manuscript, H.G.B. Alexander Research Foundation, Graduate School of Business, University of Chicago, 1975.

REPORTS 1976

7600 List of Reprints, nos. 172-178; List of Reports, 1975.

7601/ES "Profit Structure in Retailing", by B. Nooteboom.

7602/M "On a Diophantine Equation of E. Bombieri", by R.J. Stroeker.

7603/S "Limit Theory for Multivariate Sample Extremes", by L. de Haan and S.I. Resnick.

7604/S "On Typical Characteristics of Economic Time Series and the Relative Qualities of Five Autocorrelation Tests", by C. Dubbelman, A.P.J. Abrahamse and A.S. Louter.

7605/E "The Two-Level C.E.S. Production Function in Input-Output Analysis: Some Preliminary Results of an Emperical Study for the Netherlands, 1948-1958", by P.M.C. de Boer.

7606/E "Solution of Econometric Equation Systems by Means of a Modified Gauss-Seidel Procedure", by A. Kunstman and T. Kloek.

7607/E "Linearization and Estimation of the Leser-Somermeyer-Houthakker Model of Optimal Allocation of Consumer Expenditure", by P.M.C. de Boer.

7608/O "A Study on Maintenance Systems for Locomotives", by J. Telgen.

7609/M "Moduli and Canonical Forms for Linear Dynamical Systems. II: The Topological Case", by M. Hazewinkel.

7610/M "Moduli and Canonical Forms for Linear Dynamical Systems. III: The Algebraic-Geometric Case", by M. Hazewinkel.

7611/M "On n-Isoclinic Groups", by M. Hazewinkel.

7612/S "Sample Extremes: An Elementary Introduction", by L. de Haan.

7613/ES "Production Functions and Mark-Up Pricing: with a Methodological Enquiry and an Emperical Study of Grocers", by B. Nooteboom.

7614/E "Predictive Moments of Simultaneous Econometric Models, A Bayesian Approach", by H.K. van Dijk and T. Kloek.

7615/ES "Economics of Scale in Retailing", by B. Nooteboom.

7616/E "Efficient Estimation of Income Distribution Parameters", by T. Kloek and H.K. van Dijk.

7617/S "A Characterization of Multidimensional Extreme-Value Distribution", by L. de Haan.

7618/E "On Bayesian and Non-Bayesian Estimation of C.E.S. Production Functions within the Context of Two Alternative Models of Firm Bahaviour", by S. Schim van der Loeff and R. Harkema.

7619/O "A Reduction Method for Nonlinear Programming", by G. van der Hoek.

7620/O "Projective Methods to Find a Feasible Solution to Systems of Linear Constraints", by J. Telgen and G. van der Hoek.

7621/M "A Class of Diophantine Equations Connected with Certain Elliptic Curves over $Q(\sqrt{-13})$", by R. Stroeker.

7622/E "Bayesian Estimates of Equation System Parameters, An Application of Integration by Monte Carlo", by T. Kloek and H.K. van Dijk.