Econometrica, Vol. 40, No. 6 (November, 1972)

THE STRUCTURAL ESTIMATION OF A STOCHASTIC DIFFERENTIAL EQUATION SYSTEM

BY P. C. B. PHILLIPS[1]

It is now popular to construct economic models in differential equation form. Perhaps the most serious econometric problem faced when dealing with a differential equation system is the practical difficulty of finding consistent estimates of the important structural parameters. In this paper a simple three-equation Phillips model is considered and consistent estimates of the structural parameters are provided by the minimum-distance procedure. The small-sample distributions of these estimates are investigated by the Monte Carlo method; and the results are then compared with those of the three-stage least-squares estimates found by making a discrete approximation to the system of differential equations.

1. INTRODUCTION

1.1. *Introductory Remarks*

MANY MACROECONOMIC MODELS are now constructed in a form in which time is considered a continuous variable.[2] This form is intuitively appealing because the actual movements of economic variables, as distinct from the observations on them, are dependent on the continuous passage of time. If the variables of a model are assumed to be continuous and differentiable functions of time, it seems natural to represent the direction and strength of movements in the variables by derivatives. The dynamics of a system involving several variables of this type can then be described by a number of interdependent differential equations. Differential equation models are currently used to explain various economic phenomena; for example, trade cycles in macroeconomic theory. They also have the advantage of prescribing a continuous time path for each variable in the model. This latter property is useful for purposes of prediction when we may be interested in estimating the value of a variable at any point in time.

The more precise specification of such a model for statistical purposes is known as a system of stochastic differential equations; and a typical statistical problem is that of estimating certain structural parameters by means of a series of observations on the variables of the model. It is important that econometric techniques be available to handle these models in the form suggested by economic theory. For, we can hope to obtain consistent estimators of the structural parameters that interest us only if the model as estimated is specified correctly. Some work has already been done on the estimation of parameters in stochastic differential equation systems.[3] The approach usually adopted is to make a discrete approximation to the differential equation system and apply well-tried procedures such as

[1] This paper was written from the research carried out for my M.A. thesis at Auckland University. I wish to acknowledge my great debt to Professor A. R. Bergstrom who suggested the topic and helped me as supervisor. I must also thank both referees for their penetrating comments and useful suggestions.

[2] Arguments for the use of such models have been advanced by several writers. See, for example, Koopmans [5] and Bergstrom [4].

[3] See Phillips [8] and Bergstrom [3].

three-stage least squares (3SLS)[4] on the resulting model. The undesirable feature of this method is the specification error, implicit in the approximation of the basic structural form, which causes such estimators as 3SLS to be asymptotically biased.

Moreover, the parameters of a stochastic differential equation system are unlikely to be completely free of a priori restrictions. For instance, theory sometimes suggests that certain variables be excluded from an equation and the corresponding parameters are then restricted to be zero. Full use must be made of such information if parameter estimates are to be efficient; and by estimating the model in a form consistent with its specification we have more opportunity for taking into account a priori restrictions than if we are dealing with an approximate model.

In this study a procedure is considered that provides consistent and asymptotically efficient estimators of the parameters in the structural form of a stochastic differential equation system. For a simple three-equation Phillips[5] model the small-sample properties of the estimators are investigated using the Monte Carlo simulation technique. In particular, we are interested in whether the asymptotic distributions of the estimators give a reasonably reliable guide to their small-sample distributions. Finally, these estimators are compared with the 3SLS estimators obtained from the discrete approximation to the specified differential equation system.

1.2. *The General First-Order Model*

The general first-order stochastic differential equation system can be written

$$Dy(t) = A(\delta)y(t) + b(\delta) + \zeta(t), \tag{1}$$

where D is the differential operator d/dt, $y(t)$ is an $n \times 1$ vector of random functions observable at discrete points in time (t), δ is a $p \times 1$ parameter vector whose elements are the key economic parameters of interest in the model, the matrix A and the vector b have elements that are functions of δ, and $\zeta(t)$ is a vector of disturbances. In the general stochastic hypothesis of the model (1), certain properties are attributed to the disturbance vector $\zeta(t)$. We shall assume the elements of $\zeta(t)$ satisfy:

$$E\left[\int_{t_1}^{t_2} f(t)\zeta_i(t)\,dt\right] = 0 \qquad (i = 1, \ldots, n), \tag{2}$$

$$E\left[\int_{t_1}^{t_2} f(t_2 - r)\zeta_i(r)\,dr \int_{t_1}^{t_2} g(t_2 - r)\zeta_j(r)\,dr\right] = \sigma_{ij}\int_{t_1}^{t_2} f(t_2 - r)g(t_2 - r)\,dr \qquad (i,j = 1, \ldots, n), \tag{3}$$

[4] See Zellner and Theil [**12**].

[5] The Phillips model (Phillips [7]) modified by a lag in the consumption equation.

and

(4) $$E\left[\int_{t_1}^{t_2} f(t_2 - r)\zeta_i(r)\,dr \int_{t_3}^{t_4} g(t_4 - r)\zeta_j(r)\,dr\right] = 0 \qquad (i,j = 1,\ldots,n),$$

where E represents the expected value, $t_1 < t_2 < t_3 < t_4$, f and g are weight functions, σ_{ij} are parameters, and $\sigma_{ii} > 0$ for all i.

Suppose A has distinct characteristic roots $\beta_1, \beta_2, \ldots, \beta_n$ all with negative real parts. Then there exists a nonsingular matrix P (the matrix with the characteristic vectors of A as its columns) such that

(5) $$P^{-1}AP = \text{diag}(\beta_1, \beta_2, \ldots, \beta_n) = \Delta, \quad \text{say.}$$

It can be shown[6] that equispaced observations generated by (1) satisfy the autoregressive scheme:

(6) $$y(t) = B(\delta)y(t-1) + A^{-1}(\delta)[B(\delta) - I]b(\delta) + \xi(t)$$

where

$$B(\delta) = e^{A(\delta)}, \qquad \xi(t) = \int_{t-1}^{t} P\,e^{\Delta(t-r)}P^{-1}\zeta(r)\,dr,$$

and

(7) $$E[\xi(t)\xi(t)'] = \int_0^1 P\,e^{\Delta r}P^{-1}\Sigma(P')^{-1}\,e^{\Delta r}P'\,dr = \Omega, \quad \text{say,}$$

where Σ is the $n \times n$ matrix with σ_{ij} its elements.

1.3. *A Simple Trade-Cycle Model*

Consider the modified Phillips model

(8) $$DC(t) = \alpha[(1-s)Y(t) + F - C(t)],$$

(9) $$DY(t) = \lambda[C(t) + DK(t) - Y(t)],$$

(10) $$DK(t) = \gamma[vY(t) - K(t)],$$

where $Y(t)$ is real net national income at time t, $C(t)$ is real consumption, $K(t)$ is the amount of fixed capital, F is the autonomous consumption component, and α, s, λ, γ, v are parameters of the model. In this study, F was regarded as a predetermined variable whose constant value ($F = 5$) was known. This assumption was made to simplify regression programmes. Of course, in practical econometric work the value of F would need to be estimated along with the other parameters.

By substituting equation (10) into (9) and introducing disturbances into the resulting equations we obtain the stochastic model

(11) $$Dy(t) = Ay(t) + b + \zeta(t)$$

[6] C.f. Bergstrom [3].

where

$$y(t) = \begin{bmatrix} C(t) \\ Y(t) \\ K(t) \end{bmatrix}, \qquad A = \begin{bmatrix} -\alpha & \alpha(1-s) & 0 \\ \lambda & \lambda(\gamma v - 1) & -\lambda\gamma \\ 0 & \gamma v & -\gamma \end{bmatrix}, \qquad b = \begin{bmatrix} \alpha F \\ 0 \\ 0 \end{bmatrix},$$

and we assume $\zeta(t)$ satisfies (2), (3), and (4). Let $\delta' = [\alpha, \lambda, \gamma, v, s]$ be the vector of parameters in this system.

It follows from (6) that equispaced observations generated by (11) also satisfy

$$y(t) = By(t-1) + A^{-1}(B - I)b + \xi(t). \tag{12}$$

1.4. *Generation of Data*

To estimate the parameters of the structural form (11), a sequence of observations $\{y(t); t = 1, \ldots, T\}$ must be made on the aggregate variables of the model. In applied work, time series are usually available for this purpose. However, in this study a simulation technique was used to generate observations through the system (12). The observations then satisfy the structural form (11).

To generate the data we need to specify true parameter values

$$\delta^{0'} = [\alpha^0, \lambda^0, \gamma^0, v^0, s^0] = [0.6, 4.0, 0.4, 2.0, 0.25] \tag{13}$$

and the autonomous consumption $F = 5$. Furthermore, we assume that the integrals of $\zeta(t)$ are normally distributed with a covariance matrix equal to the identity matrix. It follows that $\xi(t)$ is $N(0, \Omega)$ where Ω is given by (7) and

$$E[\xi(t)\xi(s)'] = 0$$

for $t \neq s$ (because of (4)).

With the parameter values (13)

$$A = \begin{bmatrix} -0.6 & 0.45 & 0 \\ 4.0 & -0.8 & -1.6 \\ 0 & 0.8 & -0.4 \end{bmatrix}, \qquad b = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix}, \tag{14}$$

and the characteristic values of A are

$$\begin{aligned} &\beta_1 = -1.56579, \qquad \beta_2 = -0.1171 + 0.37358i, \\ &\beta_3 = -0.1171 - 0.37358i. \end{aligned} \tag{15}$$

We may now calculate B, $A^{-1}(B - I)b$, and, using (7), Ω.

But before (12) can be used to generate observations, a starting point $y(0)$ in the series must be specified. The equilibrium value, $\lim E[y(t)]$, was used for this purpose. Artificial samples could now be generated by using variates sampled from $N(0, \Omega)$. However, in order to use $N(0, I)$ variates, Ω was diagonalised by the orthogonal matrix R:

$$R'\Omega R = \operatorname{diag}(\sigma_1^2, \sigma_2^2, \sigma_3^2)$$

and the new system

$$(16)\qquad \begin{aligned} z(t) &= R'By(t-1) + R'A^{-1}(B-I)b + \varepsilon(t), \\ y(t) &= Rz(t), \qquad \text{where} \quad \varepsilon(t) = R'\xi(t), \end{aligned}$$

was constructed. Clearly, the standardised variates $\varepsilon_i(t)/\sigma_i$ $(i = 1, 2, 3)$ are $N(0, 1)$.

The system (16) was used to generate 100 samples each of 25 observations on $y(t)$ and 7,500 random $N(0, 1)$ variates (taken from Wold's tables[7]) were needed to do this.[8] Calculations were done on the University of Auckland's IBM 1130 Version 2A computer and the machine was programmed to punch the sample observations on cards.

2. PARAMETER ESTIMATION

2.1. *The Minimum Distance Procedure*

Let $d(\delta) = A(\delta)^{-1}[B(\delta) - I]b(\delta)$. Then, the reduced form obtained in the last part is

$$(17)\qquad y(t) = N(\delta)z(t) + \xi(t)$$

where $N(\delta) = [B(\delta), d(\delta)]$ and $z(t)' = [y(t-1)', 1]$. It is convenient to rewrite (17) as

$$(18)\qquad y(t) = g_t(\delta) + \xi(t)$$

where $g_t(\delta)$ is a vector whose components are nonlinear functions of the parameters δ, the lagged endogenous variables $y(t-1)$, and exogenous consumption F. It is possible, therefore, to use a nonlinear procedure[9] to estimate the parameter vector δ directly from (18), given a series of observations $\{y(t); t = 1, \ldots, T\}$ on the variables.

We denote by $\delta_T(S_T)$ the vector which minimises the quadratic form

$$(19)\qquad \frac{1}{T}\sum_{t=1}^{T} [y(t) - g_t(\delta)]'S_T[y(t) - g_t(\delta)]$$

where S_T is some positive definite matrix. It is clear that the functions $g_{it}(\delta)$ are differentiable with respect to δ, at least to the third order, so that we may construct the matrix of derivatives

$$Q_t(\delta) = \left[\left(\frac{\partial}{\partial\delta_k} g_{it}(\delta)\right)_{ik}\right].$$

Then $\delta_T(S_T)$ satisfies the necessary condition

$$(20)\qquad H_T(S_T, \delta) = \frac{1}{T}\sum_{t=1}^{T} Q_t(\delta)'S_T[y(t) - g_t(\delta)] = 0.$$

[7] See H. Wold [**11**].
[8] The first sample generated is given in Table A-1 (Appendix).
[9] Such a procedure is described in Malinvaud [**6**, Ch. 9, Sections and 5].

This system involves five nonlinear simultaneous equations in the elements of δ. We shall now examine an iterative procedure[10] to obtain the solution vector of the system.

The general model (18) with which we are working may be reconsidered as the linear model

$$x(t) = Q_t\delta + \xi(t), \tag{21}$$

where $x(t) = y(t) - g_t(\delta^0) + Q_t\delta^0$ and $Q_t = Q_t(\delta^0)$, because, for the true value of the parameter vector δ^0, (21) reduces to the basic model (18). The fact that (21) is not a realistic model[11] is unimportant, for it is used here only to help solve (20) and later to develop an asymptotic theory for a practical estimator.

The vector $\bar{\delta}_T(S_T)$ which minimises the quadratic form

$$T^{-1}\sum_{t=1}^{T}[x(t) - Q_t\delta]'S_T[x(t) - Q_t\delta]$$

satisfies

$$H_T(S_T, \delta^0) - M_T(S_T, \delta^0)(\delta - \delta^0) = 0 \tag{22}$$

where

$$M_T(S_T, \delta^0) = T^{-1}\sum_{t=1}^{T} Q_t'S_TQ_t.$$

It is equation (22) that suggests an iterative procedure leading to the solution of (20). However, the elements of H_T and M_T involve derivatives of $g_t(\delta)$ which need to be calculated for values of δ at each iteration. Hence, we define

$$H_T(S_T, \delta^{(n)}) = \frac{1}{T}\sum_{t=1}^{T} Q_t(\delta^{(n)})'S_T[y(t) - g_t(\delta^{(n)})]$$

and

$$M_T(S_T, \delta^{(n)}) = \frac{1}{T}\sum_{t=1}^{T} Q_t(\delta^{(n)})'S_TQ_t(\delta^{(n)}).$$

Suppose $\delta^{(0)}$ is our initial value (later, a method of finding a suitable $\delta^{(0)}$ will be discussed). Then the first approximation is obtained by solving

$$H_T(S_T, \delta^{(0)}) - M_T(S_T, \delta^{(0)})(\delta^{(1)} - \delta^{(0)}) = 0.$$

[10] A Newton procedure of successive approximations was not used as it involved the calculation of second order partial derivatives which for this model proved too costly in terms of computer core storage space. The advantage this method probably has over the one we are about to derive is a more rapid convergence to the solution.

[11] The model is not realistic simply because the vector $x(t)$ and matrix Q_t cannot be calculated from the observations without knowledge of the true parameter vector δ^0. Consequently, no estimator that depends on sample values of $x(t)$ and Q_t can be obtained in practice. Malinvaud [**6**, p. 293], calls this model a "linear pseudo-model."

Similarly, the nth approximation satisfies

$$H_T(S_T, \delta^{(n-1)}) - M_T(S_T, \delta^{(n-1)})(\delta^{(n)} - \delta^{(n-1)}) = 0. \tag{23}$$

If the procedure converges, it follows that the limit of the sequence $\{\delta^{(n)}\}$ satisfies $H_T(S_T, \delta) = 0$ and is the required solution of (20). Therefore, (23) provides a practical iterative method for finding the estimator $\delta_T(S_T)$ in the general nonlinear model (18).

If the covariance matrix Ω of the disturbances were known, then an obvious choice of estimator would be $\delta_T(\Omega^{-1})$, the maximum likelihood estimator (MLE).[12] Because Ω remains unknown in most practical econometric work, it seems reasonable to select as S_T a matrix which will converge to Ω^{-1} in probability. Consequently, the objective of the following sequence of operations,[13] used in this study to estimate δ^0, is to obtain an estimator that is asymptotically the MLE.

1. Calculate $\delta_T(S)$ for $S = I$, the identity matrix.
2. Calculate the residuals $\tilde{\xi}(t) = y(t) - g_t(\delta_T(S))$ and the moment matrix

$$\tilde{M}_{\xi\xi} = \frac{1}{T}\sum_{t=1}^{T} \tilde{\xi}(t)\tilde{\xi}(t)'.$$

3. Calculate $\delta^* = \delta_T(\tilde{M}_{\xi\xi}^{-1})$.
4. Calculate the residuals $\xi(t)^* = y(t) - g_t(\delta^*)$ and the moment matrix

$$M_{\xi\xi}^* = \frac{1}{T}\sum_{t=1}^{T} \xi(t)^*\xi(t)^{*\prime}.$$

5. Calculate the final estimator $\delta^{**} = \delta_T(M_{\xi\xi}^{*-1})$.

This final estimator will subsequently be referred to as the minimum-distance estimator (MDE) and the procedure involved in its calculation as the minimum-distance (MD) procedure.

2.2. *Properties of the MDE*

Before the asymptotic properties of δ^{**} are investigated, it must be established that the true parameter vector δ^0 is identifiable in the reduced form structure (B_0, d_0), where $B_0 = B(\delta^0)$ and $d_0 = d(\delta^0)$. This problem can be broken up into two parts. In the first place, it is clear that if $A(\delta) = A_0$, where $A_0 = A(\delta^0)$, then $\delta = \delta^0$. On the other hand, it is by no means obvious that A_0 can be identified in the matrix B_0. In fact, the result is not generally true.[14] However, in this particular problem the à priori restrictions on A (we know that $a_{13} = a_{31} = 0$) are sufficient to ensure that there is only one feasible root A_0 of the equation $\exp(A) = B_0$.[15]

[12] The disturbances $\xi(t)$ are $N(0, \Omega)$. This assumption was made when the data were generated.
[13] This sequence of operations is suggested by Malinvaud [**6**, p. 297].
[14] See, for instance, Telser [**10**, p. 495]. This reference was given to me by a referee.
[15] This result is proved in my thesis, Phillips [**9**].

The consistency of δ^{**} can now be established by making some modifications to the proof of Malinvaud's Theorem 6 [**6**, p. 301].[16] From this result further properties can be derived. In particular, the moment matrix $M^*_{\xi\xi}$ tends in probability to Ω and hence δ^{**} is asymptotically equivalent to the MLE $\delta_T(\Omega^{-1})$. Also, δ^{**} is asymptotically equivalent to the MLE from the linear model (21), that is

$$\hat{\delta} = \left[\frac{1}{T}\sum_{t=1}^{T} Q_t'\Omega^{-1}Q_t\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T} Q_t'\Omega^{-1}x_t\right].$$

This estimator $\hat{\delta}$ is subsequently called the optimal estimator to avoid confusion with $\delta_T(\Omega^{-1})$. From this latter equivalence the asymptotic normality of $\sqrt{T}(\delta^{**} - \delta^0)$ is secured. Moreover, if φ is the joint frequent function of $x(1), \ldots, x(T)$, then it can be readily shown that

$$E\left(\frac{\partial \log \varphi}{\partial \delta}\frac{\partial \log \varphi}{\partial \delta'}\right) = E\left[\sum_{t=1}^{T} Q_t'\Omega^{-1}Q_t\right]$$

and that $\sqrt{T}\hat{\delta}$ has an asymptotic covariance matrix which is the limit in probability of

$$\left[\frac{1}{T}\sum_{t=1}^{T} Q_t'\Omega^{-1}Q_t\right]^{-1}.$$

Hence, δ^{**} can be regarded as an asymptotically efficient estimator in the sense that its asymptotic covariance matrix is the inverse of the information matrix.

3. RESULTS OF THE MONTE CARLO STUDY

3.1. *Small-Sample Estimates*

The final estimator δ^{**} has desirable asymptotic properties. But knowledge of an estimator's asymptotic behaviour is not necessarily a reliable guide to the performance of the estimator in small samples. For this reason, a simulation experiment was carried out to provide some picture of the small-sample distribution of the estimator and to determine whether this distribution could be satisfactorily approximated by the known asymptotic distribution.

[16] The proof differs from Malinvaud's in two respects. First, the presence of lagged endogenous variables means that the sequence of variables $\{z(t)\}$ (see (17)) is stochastic and not uniformly bounded. But the existence of high order moments of the disturbance $\xi(t)$ overcomes this problem. Second, the lower boundedness (by a positive number) of the Euclidean distance $\|N(\delta) - N(\delta^0)\|$ in any closed set not containing δ^0 must be proved and is not an assumption of the model. However, once the identification of δ^0 is established, this result follows fairly easily.

To obtain the final estimator δ^{**} I used the sequence of operations given at the end of Section 2.1. It was found that the distance $\|\delta^{**} - \delta^0\|$ was generally smaller than $\|\delta^* - \delta^0\|$. Consequently, two more steps were included in this sequence than in that prescribed by Malinvaud.[17] The iterative scheme (23) was used to solve the nonlinear system (20) and to find the vector $\delta_T(S_T)$ for $S_T = I$, $S_T = \tilde{M}_{\xi\xi}^{-1}$, and $S_T = M_{\xi\xi}^{*-1}$ at steps 1, 3, and 5 respectively.

For $\delta^{(0)}$, the initial value, I selected the true value of the parameter vector δ^0. Such a choice would naturally not be possible in practical work; in this case, the 3SLS estimate (obtained by making a discrete approximation to the basic differential equation system[18]) could be used as a starting value. In fact, when the approximate system had been estimated by 3SLS for the 100 samples, the first twelve samples were rerun using

$$\delta^{(0)} = \hat{\hat{\delta}},$$

the 3SLS estimate of δ^0, and the final estimate δ^{**} was the same as was first obtained when

$$\delta^{(0)} = \delta^0,$$

the true value. In other words, some variation of the starting value $\delta^{(0)}$ (as would be hoped) did not seem to affect the final MDE δ^{**}. This result suggests that the study was not prejudiced by the choice of $\delta^{(0)}$.

Iteration through (23) was continued until $|\delta_i^{(n)} - \delta_i^{(n-1)}| < 0.001$, for all i, at the final step 5 in the sequence of operations, and δ^{**} was taken as the vector $\delta^{(n)}$.[19] In this way, the estimates δ^{**} were obtained for each of the 100 samples.

There is no reason why, in this simulation experiment, the practical estimator δ^{**} should not be compared with the optimal estimator $\hat{\delta}$. The purpose of this comparison is to measure the efficiency, in small samples, of the estimator δ^{**} against that of $\hat{\delta}$, which is known to make use of information (specifically, the covariance matrix Ω of the disturbances) that is not taken into account by the MDE δ^{**}. Although $\hat{\delta}$ is not a practical estimator, we are in a position to calculate it because δ^0 and Ω are known. Accordingly, the optimal estimates $\hat{\delta}$ were found for each of the 100 samples.

The arithmetic means and standard deviations of the two observed sampling distributions (that of δ^{**} and $\hat{\delta}$) were calculated; and using the two matrices

$$\begin{aligned} \frac{1}{T}[M_T(M_{\xi\xi}^{*-1})]^{-1} &= \frac{1}{T}[M_T(M_{\xi\xi}^{*-1}, \delta^{**})]^{-1}, \\ \frac{1}{T}[M_T(\Omega^{-1})]^{-1} &= \frac{1}{T}[M_T(\Omega^{-1}, \delta^0)]^{-1}, \end{aligned} \tag{24}$$

[17] Malinvaud [**6**, p. 297].
[18] See Bergstrom [**3**]. This method is discussed in Section 4.
[19] The length of the vector $H_T(M_{\xi\xi}^{*-1}, \delta)$ at this point was generally of the order 0.01.

together with the corresponding estimates, confidence intervals[20] of the form

$$\text{(25)}\qquad \begin{aligned} &\delta_i^{**} \pm 2\sqrt{\left(\frac{1}{T}[M_T(M_{\xi\xi}^{*-1})]^{-1}\right)_{ii}}, \\ &\hat{\delta}_i \pm 2\sqrt{\left(\frac{1}{T}[M_T(\Omega^{-1})]^{-1}\right)_{ii}}, \end{aligned}$$

for $i = 1, \ldots 5$, were constructed (and the number of intervals that did not contain the true parameter δ_i^0 were recorded). The results have been tabulated and appear in Table I.

TABLE I

Parameter True Value	α 0.6	λ 4.0	γ 0.4	v 2.0	s 0.25
Minimum Distance					
Mean of the estimates	0.5734	4.0709	0.4016	2.0021	0.2537
Standard deviation of the estimates	0.1410	0.7077	0.0153	0.0149	0.0259
Number of wrong intervals[a]	7	7	7	9	7
Optimal Estimator					
Mean of the estimates	0.5756	3.9918	0.4012	2.0014	0.2510
Standard deviation of the estimates	0.1386	0.5968	0.0144	0.0140	0.0259
Number of wrong intervals	5	5	5	4	3

[a] Intervals not containing the true parameter value.

For each parameter the optimal estimates have a mean over the 100 samples which is only marginally better (i.e., closer to the true parameter value) than that of the MD estimates. If we accept the standard deviation as a measure of the observed sampling dispersion[21] in the 100 samples, then the optimal estimates are slightly more concentrated than the MD estimates, although for the last three parameters the standard deviations differ only after the third or fourth decimal place.

The confidence intervals (see (25)) constructed about the estimates for each parameter and sample are only approximate since the exact sampling distributions of the estimators are unknown and the standard deviations had to be estimated.

[20] If the elements of Q_t were not stochastic, $\hat{\delta}$ would be a linear estimate with the distribution $N(\delta^0, (1/T)[M_T(\Omega^{-1})]^{-1})$. However, the matrix Q_t involves lagged endogenous variables $y(t-1)$, and the exact sampling distribution of $\hat{\delta}$ is unknown. This also applies to δ^{**}. The moment matrices (24) are not the exact covariance matrices of the estimators but are acceptable practical measures because of the analogy with the linear model above and the fact that the asymptotic covariance matrix of both $\sqrt{T}\delta^{**}$ and $\sqrt{T}\hat{\delta}$ is plim $[M_T(M_{\xi\xi}^{*-1})]^{-1} = \text{plim}\,[M_T(\Omega^{-1})]^{-1}$.

[21] We should note carefully that there is no guarantee, a priori, that the sampling distributions of the estimators possess finite moments of any order. This point was considered by Basmann [2]. However, this is no reason why the standard deviation cannot be used as a measure of dispersion in the empirical distributions with which we are dealing here.

Consequently, the intervals were based on a rough 2 standard deviations which suggests a 95 per cent level of confidence; and over the 100 samples only 5 intervals not containing the true value would be expected for each parameter. The results observed in the case of the optimal estimator seem compatible with this. On the other hand, the number of wrong intervals constructed about the MD estimates was not seriously greater than for the optimal estimates.

In conclusion, the MD procedure provides reasonably precise estimates of all parameters. Moreover, these estimates do not appear to be subject to much more variation in the samples of size 25 considered here than the optimal estimates. By this standard, therefore, the MDE retains a fair degree of efficiency even for small samples.

3.2. *The Asymptotic and Small-Sample Distributions*

The asymptotic distribution of $\sqrt{T}(\delta^{**} - \delta^0)$ is $N(0, [M(\Omega^{-1})]^{-1})$ where $M(\Omega^{-1}) = \text{plim } T^{-1} \Sigma_{t=1}^{T} Q_t'\Omega^{-1}Q_t$. But before we can compare the asymptotic and small-sample distributions of $\sqrt{T}\delta^{**}$, we must calculate the asymptotic covariance matrix $[M(\Omega^{-1})]^{-1}$.

If M_{zz} is the sample moment matrix of the variables $\{z(t); t = 1, \ldots, T\}$, then M_{zz} tends in probability to the matrix[22]

$$M_0 = \begin{bmatrix} V_0 & v_2 \\ v_2' & 1 \end{bmatrix}$$

where $V_0 = \lim_{t\to\infty} E[y(t)y(t)']$ and $v_2 = \lim_{t\to\infty} E[y(t)]$. We know that $v_2 = [I - B]^{-1}d$ and V_0 satisfies $V_0 = BV_0B' + X$ where $X = Bv_2d' + dv_2'B' + dd' + \Omega$, so that the limit matrix M_0 can be easily found.

We may write $Q_t = [(n_{hk}^{0\prime}z(t))_{hk}]$ where $n_{hk}^{0\prime}$ is the derivative of the hth row of $N(\delta)$ with respect to δ_k evaluated at $\delta = \delta^0$. Then the (i, j)th element of the matrix $M(\Omega^{-1})$ is

$$\begin{aligned} \text{plim} \frac{1}{T} \sum_{t=1}^{T} \sum_{h,k=1}^{3} [n_{hi}^{0\prime}z(t)](\Omega^{-1})_{hk}[n_{kj}^{0\prime}z(t)] &= \text{plim}\,[\text{trace}\,\{N_i^{0\prime}\Omega^{-1}N_j^0 M_{zz}\}] \\ &= \text{trace}\,\{N_i^{0\prime}\Omega^{-1}N_j^0 M_0\} \end{aligned}$$

where $N_i^{0\prime} = [n_{1i}^0, n_{2i}^0, n_{3i}^0]$. Using this result, we find that

$$[M(\Omega^{-1})]^{-1} = \begin{bmatrix} 0.227195 & 0.056412 & -0.011191 & 0.002541 & 0.006694 \\ 0.056412 & 6.084874 & -0.089639 & 0.026894 & -0.021981 \\ -0.011191 & -0.089639 & 0.003517 & -0.001116 & 0.000129 \\ 0.002541 & 0.026894 & -0.001116 & 0.004882 & 0.004246 \\ 0.006694 & -0.021981 & 0.000129 & 0.004246 & 0.013601 \end{bmatrix}.$$

[22] A well known result for autoregressive systems. The weak conditions assumed by Anderson [1] are certainly satisfied by the model under consideration. This reference was given to me by a referee.

If we denote the diagonal elements of this matrix by σ_α^2, σ_λ^2, σ_γ^2, σ_v^2, and σ_s^2, then $\sqrt{T}(\alpha^{**} - \alpha^0)/\sigma_\alpha$, $\sqrt{T}(\lambda^{**} - \lambda^0)/\sigma_\lambda$, $\sqrt{T}(\gamma^{**} - \gamma^0)/\sigma_\gamma$, $\sqrt{T}(v^{**} - v^0)/\sigma_v$, and $\sqrt{T}(s^{**} - s^0)/\sigma_s$ all have an asymptotic $N(0, 1)$ distribution.

On the other hand, using the parameter estimates obtained over the 100 samples, the true parameter value, and the known asymptotic variance, we can construct, by setting $T = 25$, an empirical sampling distribution for each of the above statistics. From these observed distributions, frequency polygons can be sketched and smoothed into continuous curves. This has been done for each parameter and the resulting curves[23] are drawn against the asymptotic $N(0, 1)$ distribution to facilitate the comparison of observed sampling behaviour and known asymptotic properties.

The reader can draw his own conclusion from these figures. My own view is that the asymptotic distribution gives, under the limitations of the study,[24] a fair guide to the behaviour of the estimator in samples of the order 25; and, more specifically, there appears to be no great disparity between the sampling and asymptotic dispersions, especially in the case of the parameters γ, v, and s for which the asymptotic property of efficiency seems to carry over reasonably well.

4. ESTIMATION OF THE DISCRETE APPROXIMATION TO THE STRUCTURAL SYSTEM

4.1. *The Approximate Model*[25]

The trade-cycle model we have been considering (see (8), (9), and (10)) comprises three differential equations. We may integrate each of these equations from $t - 1$ to t and using the approximation $\int_{t-1}^{t} X(t)\,dt \sim 0.5\{X(t) + X(t-1)\}$ for the three variables $C(t)$, $Y(t)$, and $K(t)$, the following system is obtained:

$$C(t) - C(t-1) = \alpha\left[(1 - s)\frac{Y(t) + Y(t-1)}{2} - \frac{C(t) + C(t-1)}{2} + F\right],$$

$$Y(t) - Y(t-1) = \lambda\left[\frac{C(t) + C(t-1)}{2} + \{K(t) - K(t-1)\} - \frac{Y(t) + Y(t-1)}{2}\right],$$

$$K(t) - K(t-1) = \gamma\left[v\frac{Y(t) + Y(t-1)}{2} - \frac{K(t) + K(t-1)}{2}\right].$$

[23] See Figures 1, 2, 3, 4, and 5 in the Appendix.

[24] It must be remembered that only 100 samples have been used. Before generalising about sampling behaviour on this basis alone, we must be very careful. Undoubtedly 1,000 artificial samples would provide a more complete picture of the sampling distributions. But, even so, the results would still be empirical and the Monte Carlo method is, it must be stressed, only a poor substitute for the analytical determination of the exact sampling distributions.

[25] Discrete approximations of the type used in this section are discussed in Bergstrom [3].

We now define the new variables $u_t = C(t) - C(t-1)$, $v_t = Y(t) - Y(t-1)$, $w_t = K(t) - K(t-1)$, $x_t = [C(t) + C(t-1)]/2$, $y_t = [Y(t) + Y(t-1)]/2$, and $z_t = [K(t) + K(t-1)]/2$. Then the system above can be written as three structural equations:

$$u_t = \alpha F - \alpha x_t + \alpha(1-s)y_t,$$
$$v_t = \lambda p_t,$$
$$w_t = \gamma v y_t - \gamma z_t,$$

and four identities: $p_t = x_t + w_t - y_t$, $u_t = 2x_t - C(t-1)$, $v_t = 2y_t - Y(t-1)$, and $w_t = 2z_t - K(t-1)$. The jointly dependent variables are $u_t, v_t, w_t, x_t, y_t, z_t, p_t$, and the predetermined variables are $C(t-1)$, $Y(t-1)$, $K(t-1)$, and F. We may now use as an approximation to (11) the stochastic model:

(26) $$u_t = \beta_0 F + \beta_1 x_t + \beta_2 y_t + \eta_{1t},$$

(27) $$v_t = \beta_3 p_t + \eta_{2t},$$

(28) $$w_t = \beta_4 y_t + \beta_5 z_t + \eta_{3t},$$

where

(29) $$\beta_0 = \alpha, \quad \beta_1 = -\alpha, \quad \beta_2 = \alpha(1-s), \quad \beta_3 = \lambda, \quad \beta_4 = \gamma v, \quad \beta_5 = -\gamma,$$

and η_{it} $(i = 1, 2, 3)$ are random disturbances which, we assume, have zero means, finite variances, and satisfy $E[\eta_{it}\eta_{jt-r}] = 0$ for $r \neq 0$ and for all i, j.

The equations (26), (27), and (28) form a simultaneous equation system whose parameters can be estimated by a number of well-known procedures. By using (29), estimates can then be obtained of the structural parameters α, λ, γ, v, and s in which we are interested. The problem inherent in this approach is the specification error that results from using the approximate model to estimate the parameters, when the observations on the variables were generated by a different system. The most damaging effect of incorrect specification is that estimates of the structural parameters will be inconsistent. Therefore, as the number of sample observations grows larger we cannot necessarily expect the estimates obtained to move closer to the true parameter values. It is the significance of this bias, especially in small samples, that will be examined in this part. In particular, the question emerges whether the consistent estimators developed in Section 2 give better[26] small-sample results than standard estimators, such as 3SLS from the approximate model.

[26] Many different standards of comparison for estimators in small samples can be used. But, in general, by "better" we mean more concentrated about the true value of the parameter being estimated.

4.2. *Parameter Estimation by Three-Stage Least Squares*

Let β be the vector of parameters in the system defined by (26), (27), and (28); and suppose $\hat{\hat{\beta}}$ is the estimate of β obtained by the 3SLS method. The corresponding estimates of the structural parameters are

$$\hat{\hat{\delta}} = [-\hat{\hat{\beta}}_1, \hat{\hat{\beta}}_3, -\hat{\hat{\beta}}_5, -\hat{\hat{\beta}}_4/\hat{\hat{\beta}}_5, (\hat{\hat{\beta}}_1 + \hat{\hat{\beta}}_2)/\hat{\hat{\beta}}_1]. \tag{30}$$

Interval estimates of the parameters can also be constructed. We denote by $[(\psi_{ij})]$ the moment matrix generally used to estimate the covariance matrix of $\hat{\hat{\beta}}$.[27] It follows that we can estimate var $[\hat{\hat{\alpha}}]$ by ψ_{22}, var $[\hat{\hat{\lambda}}]$ by ψ_{44}, and var $[\hat{\hat{\gamma}}]$ by ψ_{66}. The remaining two parameter estimates $\hat{\hat{v}}$ and $\hat{\hat{s}}$ are simple functions of the elements of $\hat{\hat{\beta}}$:

$$\hat{\hat{v}} = -\hat{\hat{\beta}}_4/\hat{\hat{\beta}}_5 = f(\hat{\hat{\beta}}_4, \hat{\hat{\beta}}_5), \quad \text{say, and}$$

$$\hat{\hat{s}} = (\hat{\hat{\beta}}_1 + \hat{\hat{\beta}}_2)/\hat{\hat{\beta}}_1 = g(\hat{\hat{\beta}}_1, \hat{\hat{\beta}}_2), \quad \text{say.}$$

We may then estimate var $[\hat{\hat{v}}]$ by

$$\varphi_1 = \left(\frac{\partial f}{\partial \beta_4}\bigg|_{\hat{\hat{\beta}}_4, \hat{\hat{\beta}}_5}\right)^2 \psi_{55} + \left(\frac{\partial f}{\partial \beta_5}\bigg|_{\hat{\hat{\beta}}_4, \hat{\hat{\beta}}_5}\right)^2 \psi_{66} + 2\left(\frac{\partial f}{\partial \beta_4}\bigg|_{\hat{\hat{\beta}}_4, \hat{\hat{\beta}}_5}\right)\left(\frac{\partial f}{\partial \beta_5}\bigg|_{\hat{\hat{\beta}}_4, \hat{\hat{\beta}}_5}\right)\psi_{56},$$

and var $[\hat{\hat{s}}]$ by

$$\varphi_2 = \left(\frac{\partial g}{\partial \beta_1}\bigg|_{\hat{\hat{\beta}}_1, \hat{\hat{\beta}}_2}\right)^2 \psi_{22} + \left(\frac{\partial g}{\partial \beta_2}\bigg|_{\hat{\hat{\beta}}_1, \hat{\hat{\beta}}_2}\right)^2 \psi_{33} + 2\left(\frac{\partial g}{\partial \beta_1}\bigg|_{\hat{\hat{\beta}}_1, \hat{\hat{\beta}}_2}\right)\left(\frac{\partial g}{\partial \beta_2}\bigg|_{\hat{\hat{\beta}}_1, \hat{\hat{\beta}}_2}\right)\psi_{23}.$$

4.3. *The Specification Error*

The significance of the specification error mentioned in 4.1 can be brought out by examining the asymptotic bias of the 3SLS estimator: that is, the difference between the true parameter vector δ^0 and the probability limit of $\hat{\hat{\delta}}$. The elements of $\hat{\hat{\beta}}$ are rational functions of sample moments which converge in probability to known (or readily calculated) values.[28] Using these values, we can find the probability limit of $\hat{\hat{\beta}}$, and from this result and (30) we obtain

$$\operatorname{plim} \hat{\hat{\delta}} = \operatorname{plim} \begin{bmatrix} \hat{\hat{\alpha}} \\ \hat{\hat{\lambda}} \\ \hat{\hat{\gamma}} \\ \hat{\hat{v}} \\ \hat{\hat{s}} \end{bmatrix} = \begin{bmatrix} 0.61626 \\ 3.429465 \\ 0.401659 \\ 1.995571 \\ 0.242031 \end{bmatrix}.$$

The four parameters α, γ, v, and s do not seem to have been unduly affected by the error of specification implicit in the discrete approximation. On the other

[27] See, for instance, Zellner and Theil [**12**, p. 58]—the first member on the right side of (2.17). This matrix is obtained in the calculation of $\hat{\hat{\beta}}$.

[28] See Section 3.2.

hand, the response parameter λ in the income equation has taken the brunt of the error. The underestimation in value (the true value being 4.0) suggested by the probability limit of 3.429465 is borne out in the sample estimates (to be discussed in the next section) where the bias appears to be accentuated.[29]

4.4. *Practical Results*

It is clear that the MDE, which is consistent and asymptotically efficient, will be preferred, where large samples are available, to estimators such as 3SLS which are obtained from the discrete approximate model. In practical work, however, the number of reliable observations on the variables of a model is often less than 30. In Section 3.2 it was demonstrated that, for the model under consideration, the observed sampling distributions of the MD estimates seem to accord reasonably well with what we have established about the estimator's asymptotic behaviour. Conceivably, it would still be possible for the 3SLS procedure to give better small-sample results. Therefore, the next step is to compare the performance of the MDE in small samples with that of the 3SLS estimator.

From the 100 artificial samples of 25 observations on the variables $C(t)$, $Y(t)$, and $K(t)$ sample data could be obtained for the variables in the discrete model (26), (27), and (28). The 3SLS estimates $\hat{\hat{\delta}}$ were then calculated for each sample. The corresponding MD estimates δ^{**} that were used in Section 3 were already available. The interval estimates $\hat{\hat{\alpha}} \pm 2\psi_{22}$, $\hat{\hat{\lambda}} \pm 2\psi_{44}$, $\hat{\hat{\gamma}} \pm 2\psi_{66}$, $\hat{\hat{v}} \pm 2\varphi_1$, $\hat{\hat{s}} \pm 2\varphi_2$ were constructed, for each sample, from the 3SLS estimates themselves and the estimated variances given in 4.2. Finally, the number of intervals not containing the true parameter was noted. The results are presented in Table II.[30]

TABLE II

Parameter True Value	α 0.6	λ 4.0	γ 0.4	v 2.0	s 0.25
Minimum Distance					
Mean of the estimates	0.5734	4.0709	0.4016	2.0021	0.2537
Standard deviation of the estimates	0.1410	0.7077	0.0153	0.0149	0.0259
Root mean square error	0.1435	0.7112	0.0154	0.0150	0.0262
Number of wrong intervals	7	7	7	9	7
Three-Stage Least Squares					
Mean of the estimates	0.6652	2.7444	0.4182	1.9995	0.2767
Standard deviation of the estimates	0.1800	0.8015	0.0241	0.0311	0.0937
Root mean square error	0.1914	1.4896	0.0302	0.0311	0.0974
Number of wrong intervals	10	62	3	3	17

[29] It may be of interest to point out here that the probability limit of the two-stage least-squares (2SLS) estimate of this parameter was even lower, 2.92664 (while the asymptotic bias of the 2SLS estimates of the other parameters was of the same order as that of the 3SLS estimates given here).

[30] Recall the study by Basmann [**2**]. But even if the MD and 3SLS estimators do not possess finite first and second order moments, we may still compare the observed mean and root mean square error of the estimates to give us some guide as to the relative precision of the two methods of estimation. Cf. Footnote 21.

The MD estimates, in all cases, are subject to much less variation from sample to sample than 3SLS. What is more, they are concentrated about a mean which, for all parameters except the capital-output ratio v, is closer to the true value than the corresponding mean of the 3SLS estimates; and, of course, $\hat{\hat{v}}$ appears to have a sampling dispersion which is far greater than that of v^{**}. The most significant difference between the two methods occurs in the estimation of λ where the 3SLS estimates are serious biased. We have already mentioned the asymptotic bias of the estimator $\hat{\hat{\lambda}}$, which stems from the incorrect specification implicit in the approximate form of the model. It now seems, from the results above, that this bias is accentuated when the sample size is small.

Bias not only affects the point estimates of the parameters; it also disrupts the confidence level that should be associated with interval estimates. We would normally expect the intervals constructed about both the MD and 3SLS estimates to contain the true parameter value in approximately[31] 95 samples out of 100. This means that the number of wrong intervals recorded in the estimation of one parameter by one of the methods should ideally be of the order 5. 3SLS gives disappointing results on this criterion. Confidence intervals about the estimates $\hat{\hat{\lambda}}$ and $\hat{\hat{s}}$ provide very little guide to the true value of the parameter. In the first case, the bias of $\hat{\hat{\lambda}}$ seriously disrupted the interval estimate; and secondly, the estimated standard error of $\hat{\hat{s}}$ often led to an interval which was too wide to be of much practical use in estimating s. This latter point held also for the 3SLS estimates of the other parameters. Although the figures are not listed above, the estimated standard errors were generally much smaller for the MD than the 3SLS estimates; this led to correspondingly more precise confidence intervals about the MD estimates.

Finally, if we accept the root mean square error as a means of classifying the two estimation procedures, the above figures suggest that, for the sample size used in this study, the MD method provides better estimates of all the structural parameters than 3SLS.

5. CONCLUSION

In this paper, the MD method has been used to obtain consistent estimates of the structural parameters in a stochastic differential equation system. We have been concerned particularly with the behaviour of the MDE in small samples. The Monte Carlo technique was used to investigate the sampling distribution of the estimator, and it was found that the asymptotic theory seemed to carry over reasonably well to samples of size 25. The MD estimates were then compared with the 3SLS estimates (for the same 100 samples of 25 observations) that were obtained by making a discrete approximation to the differential equation system. The results indicated that the MD method gives estimates which are superior by most standard criteria. This is consistent with the order in which the methods would be classified according to their asymptotic properties. For a practical

[31] The exact sampling distributions of the estimators are unknown, and the intervals are only approximate.

estimation procedure, it is suggested that the 3SLS estimates of the structural parameters be obtained and that these estimates be used as a starting point in the sequence of operations required to find the final MDE.

It is important to remember that the results obtained and a fortiori the conclusions drawn in Sections 3 and 4 refer to the simple three-equation model selected for the Monte Carlo study. However, I am inclined to think that the results may remain valid for more involved models. In the first place, the identification of parameters in a linear differential equation system from the reduced form (the corresponding system of difference equations) is assured under conditions (simple a priori restrictions on the structural matrix) that will usually be satisfied.[32] The MDE is then almost certainly consistent, even for much larger models than the one we have considered. Furthermore, the presence of additional complexities (more equations, more variables) may well worsen the effect of the specification error that is caused by taking a discrete approximation of the model in order to find standard estimators such as 3SLS. These considerations would suggest that the MDE will retain its preferable asymptotic properties when the model is more complicated; and if the experience of this study is any guide, it will perform satisfactorily for finite samples as well.

I must point out that the MDE is obtained from a recursive model whereas the 3SLS estimator is obtained from a simultaneous equations model. However, no attempt has been made to discuss the relative merits of the two types of models; and it would certainly be wrong to deduce from the apparent superiority of the MDE over the 3SLS estimator (in the model considered) that the recursive form is the better specification. On many occasions, of course, a simultaneous equations model, which is viable in its own right, could be the appropriate specification. But, in this study, I have concerned myself with the situation in which a differential equation system is the given specification and have examined the consequences of estimating the parameters of such a system from an approximate simultaneous equations model.

All computing work was done on the Auckland University IBM 1130 Version 2A machine. The MD estimates were found for each sample, after an average of 8 iterations in about 25 minutes. The 3SLS estimates could be calculated much more quickly, and the programme took approximately 20 seconds per sample to execute. However, the time difference should not be considered significant in the evaluation of the two methods. For, the final MD estimate could be obtained on a larger machine in as little as 2 or 3 minutes.

University of Auckland

Manuscript received December, 1970; revision received April, 1971.

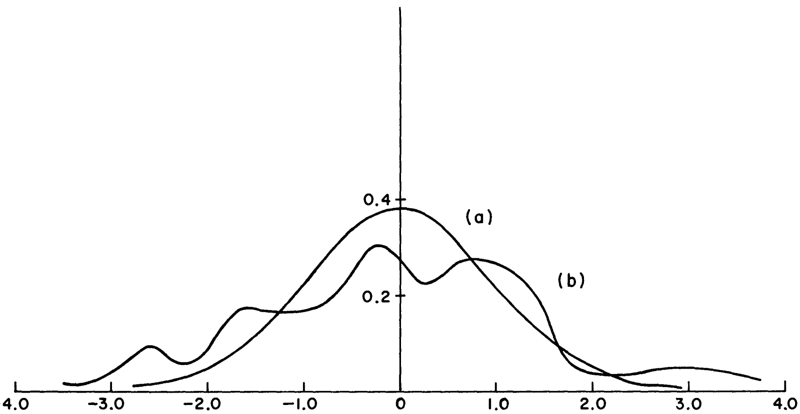
[32] See Section 2.2.

APPENDIX

TABLE A–I

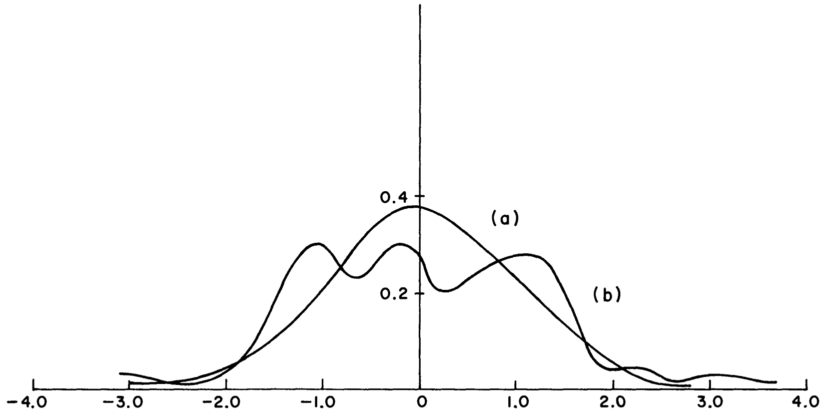
Observations $y(t)$ Generated in Sample 1

t	$C(t)$	$Y(t)$	$K(t)$
0	20.001465	20.001294	40.002760
1	20.724659	21.873653	40.319084
2	19.500591	20.517932	41.772445
3	17.740573	16.459774	40.902389
4	16.797718	12.794065	36.626464
5	16.304851	14.501026	32.963211
6	13.996612	13.712356	32.426635
7	15.690959	11.987287	32.672866
8	13.092237	9.756517	28.897232
9	14.220060	12.906684	27.340717
10	15.704090	16.940200	30.263877
11	17.718055	17.428257	30.470222
12	19.487106	23.287891	34.520500
13	21.442127	25.636001	37.044532
14	24.951919	30.484500	45.467407
15	26.961414	32.732826	48.174156
16	27.625804	34.344078	56.707367
17	29.802745	33.231643	62.022628
18	30.370506	29.752571	65.508667
19	27.389129	21.525676	60.045471
20	24.136940	18.777145	52.820404
21	21.597133	17.251140	47.556259
22	20.041355	14.616481	42.308754
23	18.967193	16.165267	35.948036
24	18.595157	19.830234	32.583084
25	20.269153	23.141529	37.360000



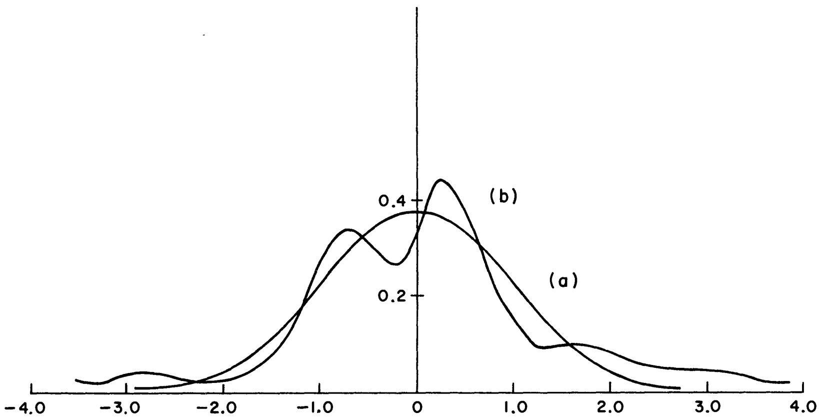
(a) The $N(0, 1)$ distribution which is the limiting distribution of $\sqrt{T}(\alpha^{**} - \alpha^0)/\sigma_\alpha$ where σ_α^2 is the asymptotic variance of $\sqrt{T}\alpha^{**}$.
(b) The observed sampling distribution of $\sqrt{T}(\alpha^{**} - \alpha^0)/\sigma_\alpha$ for the 100 samples of 25 observations.

FIGURE 1



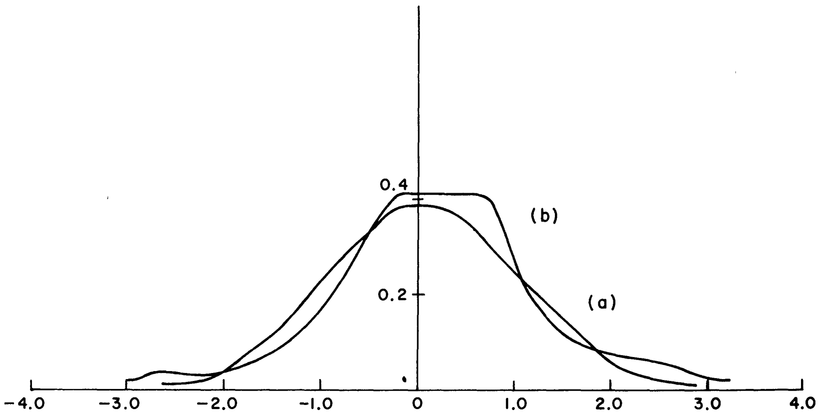
(a) The $N(0, 1)$ distribution which is the limiting distribution of $\sqrt{T}(\lambda^{**} - \lambda^0)/\sigma_\lambda$ where σ_λ^2 is the asymptotic variance of $\sqrt{T}\lambda^{**}$.
(b) The observed sampling distribution of $\sqrt{T}(\lambda^{**} - \lambda^0)/\sigma_\lambda$ for the 100 samples of 25 observations.

FIGURE 2



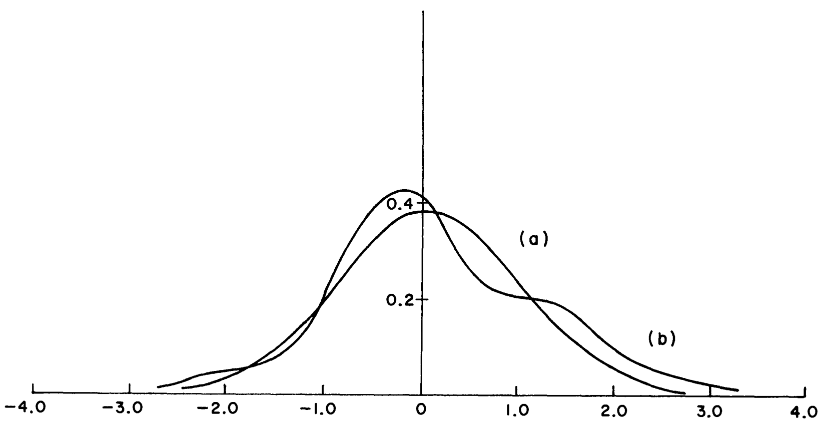
(a) The $N(0, 1)$ distribution which is the limiting distribution of $\sqrt{T}(\gamma^{**} - \gamma^0)/\sigma_\gamma$ where σ_γ^2 is the asymptotic variance of $\sqrt{T}\gamma^{**}$.
(b) The observed sampling distribution of $\sqrt{T}(\gamma^{**} - \gamma)/\sigma_\gamma$ for the 100 samples of 25 observations.

FIGURE 3



(a) The $N(0, 1)$ distribution which is the limiting distribution of $\sqrt{T}(v^{**} - v^0)/\sigma_v$ where σ_v^2 is the asymptotic variance of $\sqrt{T}v^{**}$.
(b) The observed sampling distribution of $\sqrt{T}(v^{**} - v^0)/\sigma_v$ for the 100 samples of 25 observations.

FIGURE 4



(a) The $N(0, 1)$ distribution which is the limiting distribution of $\sqrt{T}(s^{**} - s^0)/\sigma_s$ where σ_s^2 is the asymptotic variance of $\sqrt{T}s^{**}$.
(b) The observed sampling distribution of $\sqrt{T}(s^{**} - s^0)/\sigma_s$ for the 100 samples of 25 observations.

FIGURE 5

REFERENCES

[1] ANDERSON, T. W.: "On Asymptotic Distributions of Estimates of Parameters of Stochastic Difference Equations," *Annals of Mathematical Statistics*, 30 (1959), 676–87.

[2] BASMANN, R. L.: "A Note on the Exact Finite Sample Frequency Functions of Generalized Classical Linear Estimators in Two Leading Over-Identified Cases," *Journal of the American Statistical Association*, 56 (1961), 619–636.

[3] BERGSTROM, A. R.: "Nonrecursive Models as Discrete Approximations to Systems of Stochastic Differential Equations," *Econometrica*, 34 (1966), 173–182.

[4] ———. *The Construction and Use of Economic Models*. London: The English Universities Press, 1967.

[5] KOOPMANS, T. C.: "Models Involving a Continuous Time Variable" in *Statistical Inference in Dynamic Economic Models*, ed. by T. C. Koopmans. New York: Wiley, 1950.

[6] MALINVAUD, E.: *Statistical Methods of Econometrics*. Amsterdam: North-Holland Publishing Company, 1966, translation of a book first published by Dunod in French in 1964.

[7] PHILLIPS, A. W.: "Stabilization Policy in a Closed Economy," *Economic Journal*, 64 (1954), 283–99.

[8] ———: "The Estimation of Parameters in Systems of Stochastic Differential Equations," *Biometrika*, 46 (1959), 67–76.

[9] PHILLIPS, P. C. B.: "The Structural Estimation of Stochastic Differential Equation Systems," unpublished thesis, University of Auckland, 1970.

[10] TELSER, L. G.: "Discrete Samples and Moving Sums in Stationary Stochastic Processes," *Journal of the American Statistical Association*, 62 (1967), 484–99.

[11] WOLD, H.: *Random Normal Deviates*, from the series *Tracts for Computers*. Cambridge: Cambridge University Press, 1948.

[12] ZELLNER, A., AND H. THEIL: "Three-Stage Least Squares: Simultaneous Estimation of Simultaneous Equations," *Econometrica*, 30 (1962), 54–78.