THE ESTIMATION OF SOME CONTINUOUS TIME MODELS[1]

By P. C. B. Phillips

When a continuous time model is estimated from its non-recursive discrete approximation, the presence of identities and exogenous variables in the system does not preclude the use of standard procedures. However, if we wish to use the exact discrete model for estimation purposes, the treatment of identities and exogenous variables is not so straightforward. It is found that the procedure based on the exact discrete model is unlikely to be affected by the presence of identities, but when exogenous variables occur in the system some sort of approximation is usually necessary before the model can be estimated with discrete data. An approximate model is constructed to deal with the latter case and the asymptotic properties of estimators derived from this model are investigated.

1. INTRODUCTION

UNDER CERTAIN CONDITIONS, a stochastic model represented by a system of continuously distributed lags can be regarded as the solution of a system of linear stochastic differential equations. Two general approaches are available if we wish to estimate the parameters of such a system by conventional methods and with discrete data.[2] The first approach (see [1 and 2]) is to take a discrete approximation to the model and estimate the approximate model by standard methods. The second approach makes use of the discrete model which is known to correspond to the continuous time model in the sense that observations at equidistant points in time that are generated by the latter system also satisfy the former. The main advantage of the second approach is that no specification error is involved, so that it is possible in some cases to obtain consistent and asymptotically efficient estimators of the parameters in the model. In addition to the arguments of asymptotic theory, the results of a previous study [8] have given some recommendation to the second approach on the basis of small sampling performance. However, the model used in the sampling experiment of this study was relatively simple and it is the aim of the present paper to discuss the use of the second approach in more complicated models. The complications with which we will be concerned are the presence of identities and exogenous variables; both these complications may be expected to occur in more realistic economic models.

Before the procedure is viable when there are identities in the model, we must ascertain whether the disturbance in the exact discrete model has a non-singular

---

[1] In writing this paper I have benefited from discussions with Professor J. D. Sargan. I am also very grateful to both referees for their comments and suggestions on earlier drafts of this paper.

[2] The following comments are not intended to suggest that these two approaches exhaust all possibilities. Indeed, Phillips [7] has indicated a different procedure, the essence of which is to estimate the spectral density matrix of the system variates from a sufficient number of sample autocovariance matrices and then use this estimate to obtain estimates of the parameters of the model. Unfortunately, this procedure makes no allowance for possible a priori restrictions on the parameters, so that estimates obtained in this way will not be asymptotically efficient. The procedure also neglects the aliasing or identification problem. On this last point the reader is referred to [9].

distribution. A simple test that should be widely applicable is developed for this purpose in Section 2. If the distribution is non-singular, then we can proceed to estimate the parameters of the model in the usual way.[3]

When exogenous variables occur in a system of stochastic differential equations, the corresponding discrete time model no longer has a simple autoregressive form. In fact, the exogenous variable component in the exact discrete model depends on a continuous time record of the exogenous variables. In most cases such a record is not available, so that some sort of approximation is usually necessary before the structural parameters of the system can be estimated with discrete data. An approximate model is constructed to deal with this case and the asymptotic properties of quasi-maximum likelihood (QML) estimators derived from this model are investigated.

The methodological approach we adopt in this section of the paper is to develop an asymptotic theory on the hypothesis that the variables in the model are observed at equispaced intervals in time measured by some positive real number h. If N is the number of unit time intervals over which we have observations and T is the total number of observations available, then $N = hT$. As N increases we can expect our sampling interval h to decrease so that we have the functional dependence $h = h(N)$; but it is very difficult to say anything more precise about this function since the function may well differ according to the time series under consideration. For many economic series h remains fixed for a reasonable number of unit time periods, so that T can be fairly large before h changes and the series is observed more frequently. For instance, if the time unit is taken to be a year and $h = 1/12$, then we may have several hundred post war observations of a series that is available monthly but, as yet, not more frequently. Thus, the line of argument taken in Section 3 of the paper is to consider first an asymptotic theory as T becomes infinitely large for fixed h. We then turn our attention to what happens to the asymptotic bias and limiting distribution of the estimators we are considering as h tends to zero. This theory can tell us whether our estimators are likely to be satisfactory when the number of observations is large and the sampling interval is small. Since Sargan [11] has developed a similar theory for estimators based on the discrete approximation, the present theory can also help us to discriminate between the different procedures on the basis of asymptotic properties.[4]

---

[3] Since the problem is one of non-linear regression, an iterative technique will be required. See Phillips [8].

[4] It would be possible, no doubt, to construct an asymptotic theory in a different way. We could, for instance, consider first the effect of letting h and T approach their limits while N remains fixed, so that our data becomes closer to a continuous time record in a given time period; and then we could allow N to tend to infinity. However, this approach tends to contradict the manner in which economic variables are observed. It would be much more complicated to construct a theory which allowed $h = h(N)$ and $T = T(N)$ to converge to their respective limits as N increased indefinitely. Although this latter situation comes closer to the facts, it is, as we have indicated above, very difficult to be specific about the functional dependence $h(N)$ and at the same time expect this function to be realistic enough to apply to all variables in the model.

2. LINEAR MODELS WITH IDENTITIES

The type of system with which we are frequently concerned involves identities as well as stochastic equations. To accommodate this situation, the general model may be formally written

$$(1) \qquad Dy(t) = Ay(t) + H\zeta(t)$$

where the matrix A is $n \times n$ with distinct characteristic roots, all of which have negative real parts, the matrix H is $n \times r$ with $r < n$, and the elements of both A and H belong to the real number field. D is the differential operator d/dt, taken in the mean square sense, the disturbance $\zeta(t)$ is assumed to be a pure noise process,[5] and the covariance matrix of the integral of $\zeta(t)$ over a unit time interval is assumed to be the positive definite matrix Σ. Thus, if the last $n - r$ equations of (1) are identities, we may write $H' = [I_r : 0]$. Unlike H some elements of A are unknown and are to be estimated; others are restricted a priori to be zero, in accordance with the specification of the model suggested by theory.

The proper interpretation of a system of stochastic differential equations such as (1) has been extensively discussed in the literature on stochastic processes and was the subject of a number of early papers. One of the mathematical difficulties that arises in treating (1) as it stands is due to the fact that $\zeta(t)$ is not really a second order process at all[6] and hence the derivative $Dy(t)$ does not exist in mean square. Some authors[7] might prefer to write the system as $dy(t) = Ay(t) dt + Hdu(t)$ where $u(t)$ is a homogenous random process with independent increments (i.e., $u(t)$ is a process of independent stationary increments for which the elements of $E\{u(t) - u(s)\}\{u(t) - u(s)\}'$ are proportional to the length $|t - s|$ of the time interval), which is a symbolic way of writing the stochastic integral equation

$$y(t) = y(a) + A \int_a^t y(\tau) \, d\tau + H \int_a^t du(\tau)$$

for some fixed a. The reason why (1) is regarded as the underlying specification in this paper is that it seems the most convenient way of formulating an econometric model to correspond to a system of differential equations that have been developed in economic theory.

The discrete model corresponding to (1) is

$$(2) \qquad y(t) = \exp(hA)y(t - h) + \xi(t)$$

where $\xi(t) = \int_0^h \exp(sA)H\zeta(t - s) \, ds$ and $h > 0$; and the covariance matrix of $\xi(t)$ is given by

$$(3) \qquad \Omega = \int_0^h \exp(sA)H\Sigma H' \exp(sA') \, ds$$

[5] A pure noise process is usually described as a process with constant spectral density. Just as a purely random process in discrete time $\zeta(t)$ is independent of $\zeta(s)$ whenever $t \neq s$.

[6] Since the spectral density matrix of $\zeta(t)$ is constant, the covariance matrix of $\zeta(t)$ does not exist.

[7] See, for instance, Doob [3] or, in the econometric literature, Wymer [13].

which is not necessarily positive definite.[8] However, it turns out that Ω is positive definite in most of the important cases. This fact is of some importance when we come to estimate A using the discrete model (2).[9]

2.1. *General Results*

We first prove the following:[10]

THEOREM 1: *The covariance matrix* (3) *of* $\xi(t)$ *in the discrete model* (2) *is positive definite if and only if the subspace spanned by the vectors* $\{A^k He_i | i = 1, 2, \ldots, r;$ $k = 0, 1, 2, \ldots\}$, *where* e_i *denotes the ith column of the* $r \times r$ *identity matrix, is the entire n-dimensional Euclidean space.*

Clearly, Ω is positive semi-definite. Let R^n denote the n-dimensional Euclidean space and M the subspace spanned by the vectors defined in the statement of the theorem. To prove the necessary condition we need to show that the orthogonal complement $M^\perp$ of M is empty but for the zero vector. If v lies in $M^\perp$, then it is orthogonal to the range space of the matrix $\exp(sA)H$ for all s. It follows that $v'\Omega v = 0$, which by assumption implies that $v = 0$. Hence, $M = R^n$.

Conversely, suppose $M = R^n$ and let v be any vector of R^n such that $v'\Omega v = 0$. The integrand of (3) is a continuous function so that $v' \exp(sA)H\Sigma H' \exp(sA')v = 0$ for all s in the interval $[0, h]$. Therefore, for this range of values of s, v is orthogonal to the range space of $\exp(sA)H$. If we now take a first order Taylor series expansion of the matrix $\exp(sA)H$ about the value at $s = 0$ we obtain

$$\exp(sA)H = H + sA \exp(s\theta A)H$$

where $0 < \theta < 1$. Since v is orthogonal to the columns of H we have, for any $r \times 1$ vector β, $sv'A \exp(s\theta A)H\beta = 0$. That is, $v'\gamma(s) = 0$ where $\gamma(s) = A \exp(s\theta A)H\beta$; and from the continuity of $\gamma(s)$, it follows that v is orthogonal to the columns of AH. By taking successive higher order Taylor series expansions, it can be seen that v is orthogonal to the columns of $A^k H$ for all positive integral values of k. Consequently, v belongs to $M^\perp$. By hypothesis, therefore, $v = 0$ and Ω is positive definite. This proves the theorem.

It may be of interest to point out that the statement of Theorem 1 can be somewhat modified by using an alternative representation of $\exp(A)$. For, if q is any polynomial such that $q(\lambda_i) = e^{\lambda_i}$ where $\lambda_i (i = 1, 2, \ldots, n)$ are the characteristic roots of A, then $q(A) = \exp(A)$. We may, for instance, choose the Lagrange interpolation formula so that

$$(4) \qquad q(A) = \sum_{k=1}^{n} e^{\lambda_k} \prod_{\substack{i=1 \\ i \neq k}}^{n} \frac{A - \lambda_i I}{\lambda_k - \lambda_i} = \sum_{k=0}^{n-1} \alpha_k A^k.$$

---

[8] For instance, in the trivial case where $H' = [I_r \,\vdots\, 0]$ and A is the direct sum of two matrices, the first of which is $r \times r$, it can be readily verified that Ω is singular.

[9] Notice that the consistent estimation of A depends on whether A is identifiable in the reduced form (2). This problem is discussed in [9].

[10] I am grateful to Professor A. R. Bergstrom for pointing out an error in an earlier version of the proof of Theorem 1.

More generally,

$$(5) \qquad \exp(sA) = \sum_{k=0}^{n-1} \alpha_k(s)A^k$$

for all finite s and, when s is real, the coefficients $\alpha_k(s)$ $(k = 0, 1, \ldots, n - 1)$ are real. Returning to Theorem 1 we need only make use of (5) in the proof of that Theorem and we have the following:[11]

THEOREM 2: *The covariance matrix* (3) *of* $\xi(t)$ *in the discrete model* (2) *is positive definite if and only if the subspace spanned by the vectors* $\{A^k He_i | i = 1, 2, \ldots, r; k = 0, 1, \ldots, n - 1\}$ *is the entire n-dimensional Euclidean space.*

A condition equivalent to that in this theorem has been used for some time in the theory of control engineering to establish the controllability of a system (see, for instance, [14]); and a recent article by Erickson [4] is relevant to the problem with which we are concerned. Erickson mentions the possible singularity of the covariance matrix (although, in his paper, the system variates rather than the disturbances in the discrete model are being considered) and states a condition for singularity which compares with that of the above theorem.

We now examine more closely a system in which the number of stochastic equations is at least as great as the number of identities. This situation can be expected when, for example, the original model is largely of the first order. The reduced system may be regarded as a particular case of (1) by setting

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad \text{and} \quad H = \begin{bmatrix} I_r \\ 0 \end{bmatrix},$$

where the square matrices A_{11} and A_{22} are of order r and $n - r$ respectively, and $r \geqslant n - r$ in accordance with the stated hypothesis. A simple sufficient condition that will enable us to test in this case whether the distribution of the disturbances in the discrete model is non-singular is contained in our next result.

THEOREM 3: *The matrix* (3) *is positive definite if the submatrix* A_{21} *of the structural matrix* A *has rank* $n - r$.

The proof is straightforward. For if v is such that $v'\Omega v = 0$, then v is orthogonal to the leading r-dimensional subspace of R^n as well as to the columns of the matrix

$$A = \begin{bmatrix} I_r \\ 0 \end{bmatrix}.$$

It follows from the condition of the theorem that $v = 0$ and Ω is positive definite.

[11] Theorem 2 is a restatement of Theorem 1 in which we take account of the fact that the subspace M in the proof of Theorem 1 is spanned by a finite number of the vectors $\{A^k He_i | i = 1, \ldots, r; k = 0, 1, \ldots,\}$. We could derive the fact that M is spanned by the vectors $\{A^k He_i | i = 1, \ldots, r; k = 0, 1, \ldots, n - 1\}$ directly from the Cayley Hamilton Theorem. For, the characteristic polynomial of A is an annihilating polynomial, and we can represent A^n as a linear combination of powers of A of lower degree than n.

2.2. Higher Order Models

We consider the model

(6) $\qquad D^k y(t) = \sum_{i=1}^{k} A_i D^{i-1} y(t) + \zeta(t)$

which is a special case of (1) in which

$$
A = \begin{bmatrix} A_k & A_{k-1} & \cdots & A_2 & A_1 \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & \cdots & 0 & 0 \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & \cdots & I & 0 \end{bmatrix} \quad \text{and} \quad H = \begin{bmatrix} I \\ 0 \\ 0 \\ \cdot \\ 0 \end{bmatrix}.
$$

The fact that the disturbances in the corresponding discrete model have a non-singular distribution is immediate. For, if v is orthogonal to the columns of $A^m H(m = 0, 1, \ldots, n-1)$, then every element of v must be zero. Although, in the model (6), all equations have the same order, we can easily verify that the result remains valid for mixed order models.

There is one serious practical limitation, however, to the use of the discrete model corresponding to the reduced first order system of a higher order model. This is simply that some of the variables of the discrete model are derivatives, and we do not usually have observations of derivatives in econometrics. Therefore, if we wish to use the discrete model for estimation, derivatives must be approximated, say by first differences. An alternative is to use the higher order difference equation system that is satisfied by equispaced observations of the basic variables (see [1]); but this system is known to involve moving average disturbances, and although consistent estimation is possible when the sample size is large enough it will be expensive computationally.

3. LINEAR MODELS WITH EXOGENOUS VARIABLES

A simple extension of the closed model used in the previous section follows from the introduction of exogenous variables. But parameter estimation of the new system from its exact discrete model is no longer straightforward. To fix ideas, we consider the model

(7) $\qquad Dy(t) = Ay(t) + Bz(t) + \zeta(t)$

where the matrix A and the process $\zeta(t)$ are as in (1),[12] the matrix B whose elements belong to the real number field is $n \times m$, and $z(t)$ is a vector of exogenous variables, observable at discrete points in time, which we assume to be non-random and to have continuous derivatives up to the third order. Some conditions on the

---

[12] It is implicitly assumed that no identities occur in (7). If identities do occur, then the results of the last section are relevant, and the procedure we are about to consider remains viable so long as the disturbances in the discrete time model have a non-singular distribution.

exogenous variables are essential if two later assumptions[13] we will make in order to develop an asymptotic theory are to be satisfied. The assumption that the elements of $z(t)$ have continuous derivatives to the third order is quite strong, especially in view of the fact that the mean square derivative of $y(t)$ does not exist. But this assumption, is not necessary and is made here only to help us develop an approximate model for estimation purposes and obtain some intuitive idea of the specification error involved in this model. Various weaker assumptions that would be sufficient for our purposes are outlined in the appendix.

If the initial values are in the infinite past, we can write the solution of (7) as a system of convolution integrals, which corresponds to a model first suggested by Koopmans [6]. More recently, the model has been the subject of a long paper by Sargan [11], who examined the asymptotic bias of various estimators obtained from the discrete approximation to (7) and, more particularly, the order of this bias as the interval between successive observations approaches zero. The alternative procedure is based on the discrete time model corresponding to (7). But, in most cases and unlike the discrete form of a closed model, this model must be approximated before the parameters can be estimated. The effect of this approximation is to make the estimators asymptotically biased. The question of how this bias compares with the bias involved in estimating the discrete approximation to (7) will be considered later.

From the solution of (7) we derive the system

$$(8) \qquad y(t) = \exp(hA)y(t - h) + \int_0^h \exp(sA)Bz(t - s)\,ds + \xi(t)$$

where $\xi(t) = \int_0^h \exp(sA)\zeta(t - s)\,ds$ and, as in (2), h is some positive real number which represents the time interval between successive observations of the variables y and z. By defining $y_r = y(rh)$ and $\xi_r = \xi(rh)$ for integral r, we may write (8) as

$$(9) \qquad y_t = \exp(hA)y_{t-1} + \int_0^h \exp(sA)Bz(th - s)\,ds + \xi_t.$$

To estimate the parameters of A and B from this model when only discrete observations of the variables are available we must, in general, approximate the integral involving the exogenous variables. The special case occurs when $z(t)$ is a simple integrable function of time such as a polynomial, trigonometric, or exponential function; we can then integrate out in (9) to obtain a model that can be estimated directly. We now proceed to consider the general case in which an approximation is necessary.

3.1. An Approximate Model

Expanding $z(th - s)$ in a Taylor series about the value $s = 0$ and using the notation $z_r = z(rh)$ for integral r, we obtain

$$z(th - s) = z_t - sz_t^{(1)} + s^2 z_t^{(2)}/2! - s^3 z_t^{(3)}(\tau)/3!$$

---

[13] Assumptions 1 and 2 in Section 3.2.

where $th - s < \tau < th$. One way of approximating z in the interval $(th - h, th)$ would be to truncate this expansion at the third term and use the approximations $z_t^{(1)} \sim (z_t - z_{t-1})/h$ and $z_t^{(2)} \sim (z_t - 2z_{t-1} + z_{t-2})/h^2$. But, this approximation is fairly crude and a better approximation is obtained if we replace $z(th - s)$ by a quadratic in s and express the coefficients of this quadratic in terms of the three consecutive observations z_{t-2}, z_{t-1}, and z_t.[14] This method is equivalent to using a form of numerical differentiation more refined than that just mentioned. We approximate $z(th - s)$ by a three-point Lagrange interpolation formula and then differentiate once to obtain $z_t^{(1)} \sim (z_{t-2} - 4z_{t-1} + 3z_t)/2h$ and twice to obtain $z_t^{(2)} \sim (z_t - 2z_{t-1} + z_{t-2})/h^2$. Thus, we can write the approximation as

$$(10) \qquad \hat{z}(th - s) = z_t - s(z_{t-2} - 4z_{t-1} + 3z_t)/2h + s^2(z_t - 2z_{t-1} + z_{t-2})/2h^2.$$

The error involved in using this approximation is well known and we have

$$(11) \qquad \psi(th - s) = z(th - s) - \hat{z}(th - s)$$

$$= -s(-s + h)(-s + 2h)z^{(3)}(\theta)/3!$$

where θ, which is an unknown function of s, lies in the interval $(th - 2h, th)$.

As an approximation to (9) we may now construct the model

$$(12) \qquad y_t = \exp(hA)y_{t-1} + \int_0^h \exp(sA)B\hat{z}(th - s)\,ds + \eta_t$$

where, for estimation purposes, it may be assumed that η_t is a vector of serially independent random variables with zero means and non-singular covariance matrix. From (10) and (12) we obtain

$$(13) \qquad y_t = E_1 y_{t-1} + E_2 z_t + E_3 z_{t-1} + E_4 z_{t-2} + \eta_t$$

where the coefficient matrices are

$$E_1 = \exp(hA),$$

$$E_2 = h[\{\tfrac{1}{2}(hA)^{-2} + (hA)^{-3}\}\exp(hA) - (hA)^{-1} - 3(hA)^{-2}/2 - (hA)^{-3}]B,$$

$$E_3 = h[\{(hA)^{-1} - 2(hA)^{-3}\}\exp(hA) + 2(hA)^{-2} + 2(hA)^{-3}]B, \quad \text{and}$$

$$E_4 = h[\{-\tfrac{1}{2}(hA)^{-2} + (hA)^{-3}\}\exp(hA) - \tfrac{1}{2}(hA)^{-2} - (hA)^{-3}]B.$$

Clearly, (13) can be used to estimate the parameters of A and B from the discrete sample data $\{y_t, z_t; t = 1, 2, \ldots, T\}$. On the other hand, the estimators obtained in this way will not be consistent because the model is not exact. But we can expect the misspecification bias to be small if $\hat{z}(th - s)$ is a good approximation to $z(th - s)$ in the interval $(0, h)$; and the smaller the time interval h the better the approximation is likely to be. More precisely, for s in the interval $(0, h)$ it follows from (11) that $\psi(th - s)$ is of $O(h^3)$ as h tends to zero.[15] Moreover, the condition

---

[14] Professor J. D. Sargan suggested this approximation.
[15] The order of magnitude symbol O is used in the usual sense.

under which the two models (9) and (13) are equivalent is contained in the relationship

$$(14) \qquad \eta_t = \xi_t + \int_0^h \exp(sA)B\psi(th - s)\, ds$$

which reduces to

$$\eta_t = \xi_t + O(h^4)$$

as long as the elements of $\exp(sA)$ and B remain bounded as h tends to zero. This requirement will be discussed later. For the moment, it is sufficient to remark that the bias involved in using the approximate model (13) for estimation purposes and, thus, treating η_t as a random disturbance with zero mean seems to be of $O(h^4)$. We might add that (13) has the advantage of being exact when the elements of $z(t)$ are polynomials in t of degree at most two. This follows from the fundamental property of the Lagrange interpolation formula.

We now turn to investigate the effect of the specification error that results from using the approximate model (13) on typical estimators obtained from this model. The argument that follows in Sections 3.2 and 3.3 is somewhat abbreviated to keep the paper to a reasonable length. For more detail the reader is referred to [**10**].

3.2. The Asymptotic Bias of the QML Estimators

Before proceeding we must be specific about the parameters to be estimated. In general, the elements of A and B in (7) are simple functions of a smaller set of parameters which we can represent by the p-vector δ. If we wish to emphasize this dependence we may write $A(\delta)$, $B(\delta)$, and, similarly, $E_i(\delta)(i = 1, \ldots, 4)$ so that (13) can be rewritten

$$(15) \qquad y_t = G(\delta)x_t + \eta_t$$

where $G = [E_1 \vdots E_2 \vdots E_3 \vdots E_4]$ and $x_t' = (y_{t-1}', z_t', z_{t-1}', z_{t-2}')$.

The QML estimator of δ is obtained by numerically minimizing

$$(16) \qquad \log \det (Y'Y - GX'Y - Y'XG' + GX'XG')$$

where $Y' = [y_1, y_2, \ldots, y_T]$ and $X' = [x_1, x_2, \ldots, x_T]$. Some elements of δ may be just non-zero elements of A and B, others may be more involved functions of these elements. Nevertheless, it is often convenient to minimize (16) with respect to the non-zero elements[16] of A and B first and then solve to find the corresponding value of δ (given that the transformation is one to one, as it frequently will be). This being the case, we may as well write $G(A, B)$ where it is understood that the functional dependence is on the non-zero elements of A and B. Similarly, by the estimators $\hat{A}$ and $\hat{B}$ we mean the matrices with the estimators of the non-zero elements of A and B in their appropriate positions and zeros elsewhere. This convention will be retained in the rest of the paper.

---

[16] Strictly, we mean those elements of A and B which are unknown a priori; for some elements of A and B may be known to have constant values other than zero.

We return to a remark made earlier about the requirement that $\exp(sA)$ and B remain bounded as h tends to zero. In general, the elements of A and B depend not only on the units in which the variables are measured but also on the unit of time. Since a typical element of A or B involves the product of a response parameter, which is proportional to the unit of time, and a simple function of other parameters, some of which may be invariant with respect to the unit of time, we may anticipate that many elements of A and B become smaller or at least remain constant as the time unit decreases. This is not to say that, for any given time unit, the speed of response parameters must be small. In fact, if the model is at all disaggregated we may very well expect some equations to have large response parameters, representing fast rates of reaction.

However, it is important to distinguish between the unit of time and the time interval between observations, which we have denoted by h. This distinction is often blurred by the convenient practice of using h as the unit of time when we construct a model. If we do take h to be the unit of time, then, as we have suggested, many elements of A and B will decrease with h. But the possibility of some elements of A and B becoming progressively larger as h decreases is not completely ruled out.[17] Since we have assumed that A is stable, it is reasonable to conclude that even in this case $\exp(sA)$ is bounded as h decreases; but the conclusion does not follow for B. Consequently, (14) does not lead to the simple specification error of $O(h^4)$ for the model (13). Another important consequence of identifying h with the time unit is that those elements of y and z which are proportional to the unit of time (such as flows) tend to zero as the time unit decreases; this would prevent a later assumption[18] being satisfied. Finally, it is worth mentioning that the convergence of h to zero is not rigorously defined if h itself is taken to be the unit of time.

We now assume, therefore, that the time unit remains fixed as h, the interval between observations, decreases to zero. It follows that A, B and $\exp(sA)$ where s lies in the interval $(0, h)$ are bounded as h tends to zero, so that the specification error of the model (13) is of $O(h^4)$. This result compares favorably with the specification error of the discrete approximation to (7) which is known (see [11]) to be of $O(h^2)$.

We define $Y'_{-1} = [y_0, \ldots, y_{T-1}]$, $Z' = [z_1, \ldots, z_T]$, $Z'_{-1} = [z_0, \ldots, z_{T-1}]$, $Z'_{-2} = [z_{-1}, \ldots, z_{T-2}]$, and

$$J' = \begin{bmatrix} I_n & 0 & 0 & 0 \\ 0 & I_m/3 & I_m/3 & I_m/3 \end{bmatrix}$$

so that

$$J'X'XJ = \begin{bmatrix} Y'_{-1}Y_{-1} & Y'_{-1}Z^* \\ Z^{*'}Y_{-1} & Z^{*'}Z^* \end{bmatrix}$$

[17] Notice that if all the variables of a linear model have the same time dimension, then the elements of A and B will certainly decrease with h. However, most macro-economic models anyway involve variables having time dimensions that are not all the same.

[18] Assumption 1, Section 3.2.

where $Z^* = (1/3)(Z + Z_{-1} + Z_{-2})$. The following additional assumptions are now made in order that we may find the asymptotic bias of the QML estimators of A and B. They are also sufficient to ensure that these estimators have a limiting non-singular distribution, which will be discussed later.

ASSUMPTION 1: *The matrices*

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} = \underset{T \to \infty}{plim} \frac{1}{T} \begin{bmatrix} Y'Y & Y'X \\ X'Y & X'X \end{bmatrix}$$

and $\overline{M} = lim_{h \to 0} M$ *both exist.* M_{22} *is assumed to be positive definite for* $h > 0$ *and the limit matrix* $lim_{h \to 0} J'M_{22}J = J'\overline{M}_{22}J$ *is also assumed to be positive definite.*

ASSUMPTION 2: (i) *The elements of* $z(t)$ *are bounded uniformly in* t. (ii) *The matrix* $plim_{T \to \infty} (1/T) \Sigma_{t=1}^{T} x_t \phi'_{t-r}$, *where* $\phi_t = \int_0^h exp(sA)B\psi(th - s)ds$, *exists for all integral* r *and when* $r = 0$ *it is of* $O(h^4)$ *as* h *tends to zero. The matrix* $lim_{T \to \infty} (1/T)$ $\Sigma_{t=1}^{T} \phi_t \phi'_{t-r}$ *also exists for all integral* r *and when* $r = 0$ *it is of* $O(h^3)$.

From the last part of Assumption 2 we obtain, by Cauchy's inequality,

$$\left| \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \phi_{it}\phi_{jt-r} \right| \leqslant \left\{ \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \phi_{it}^2 \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \phi_{jt}^2 \right\}^{\frac{1}{2}}.$$

Hence $lim_{T \to \infty} (1/T) \Sigma_{t=1}^{T} \phi_t \phi'_{t-r}$ is bounded uniformly in r and of $O(h^3)$ as h tends to zero.

ASSUMPTION 3: *The elements of the disturbance* ξ_t *in the model* (9) *have finite moments up to the fourth order.*

ASSUMPTION 4: *The matrix function* $G = G(A, B)$ *does not have a singularity at the true values* A^0 *and* B^0. *This assumption requires, in particular, that the matrix of derivatives of* G *with respect to the non-zero elements of* A *and* B *has full rank at the point defined by* (A^0, B^0).

The conditions on the model implied by Assumptions 1 and 2 are not obvious and it is worthwhile to consider how they can be derived from more fundamental hypotheses about the components of the model, particularly the exogenous variables. However, we leave this discussion to the appendix of this paper and now derive explicitly the asymptotic bias of the QML estimators of A and B.

We define $G^* = M_{12}M_{22}^{-1}$ and $G^0 = G(A^0, B^0)$. Since

$$(G^* - G^0)M_{22} = M_{12} - G^0 M_{22}$$

and

$$M_{12} = G^0 M_{22} + \underset{T \to \infty}{plim} \frac{1}{T} \sum_{t=1}^{T} \xi_t x'_t + \underset{T \to \infty}{plim} \frac{1}{T} \sum_{t=1}^{T} \phi_t x'_t,$$

it follows from Assumption 2 that

(17) $(G^* - G^0)M_{22} = O(h^4)$.

Postmultiplying (17) by J, we obtain

(18) $(G^* - G^0) \plim\limits_{T \to \infty} \dfrac{1}{T} \begin{bmatrix} Y'_{-1}Y_{-1} & Y'_{-1}Z^* \\ Z' \ Y_{-1} & Z' \ Z^* \\ Z'_{-1}Y_{-1} & Z'_{-1}Z^* \\ Z'_{-2}Y_{-1} & Z'_{-2}Z^* \end{bmatrix} = O(h^4)$,

and if we let $S_{12} = \plim_{T \to \infty} (1/T)Y'_{-1}Z$, $S_{21} = S'_{12}$, and $S_{22} = \lim_{T \to \infty} (1/T)Z'Z$, we can write (18) as

(19) $(G^* - G^0) \begin{bmatrix} M_{11} & S_{12} + O(h) \\ S_{21} & S_{22} + O(h) \\ S_{21} + O(h) & S_{22} + O(h) \\ S_{21} + O(h) & S_{22} + O(h) \end{bmatrix} = O(h^4)$.

Since M_{11}, S_{12}, S_{21}, and S_{22} are of $O(1)$ as h tends to zero, it follows from the normal properties of order relations that (19) implies the system

$$\left[E_1^* - E_1^0 : \sum_{i=2}^{4} (E_i^* - E_i^0) \right] J'M_{22}J = O(h^4).$$

By Assumption 1, $J'M_{22}J$ is positive definite and has a positive definite limit as h tends to zero so that

(20) $E_1^* - E_1^0 = O(h^4)$ and $\sum\limits_{i=2}^{4} (E_i^* - E_i^0) = O(h^4)$.

We let $\hat{A}$ and $\hat{B}$ denote the limits in probability of the QML estimators of A and B and we define $\hat{G} = G(\hat{A}, \hat{B})$, $H = [I : -G]$, $\hat{H} = [I : -\hat{G}]$, $H^0 = [I : -G^0]$, and $H^* = [I : -G^*]$. Then, from the definition of G^*, we obtain the relationship

$$H^0MH^{0\prime} = H^*MH^{*\prime} + (G^0 - G^*)M_{22}(G^0 - G^*)'$$

which, in view of (20), we can write as

(21) $H^0MH^{0\prime} = H^*MH^{*\prime} + O(h^8)$.

We know also that

(22) $\hat{H}M\hat{H}' = H^*MH^{*\prime} + (\hat{G} - G^*)M_{22}(\hat{G} - G^*)'$

and since $\hat{A}$ and $\hat{B}$ minimize (16) we have

(23) $\det(H^0MH^{0\prime}) \geqslant \det(\hat{H}M\hat{H}')$.

Using a lemma proved by Sargan [11] we can deduce from (21), (22), and (23) that

$$(\hat{G} - G^*)M_{22}(\hat{G} - G^*)' = O(h^8),$$

which implies the system

$$(24) \qquad [\hat{E}_1 - E_1^* \vdots \sum_{i=2}^{4} (\hat{E}_i - E_i^*)]J'M_{22}J \begin{bmatrix} (\hat{E}_1 - E_1^*)' \\ \sum_{i=2}^{4} (\hat{E}_i - E_i^*)' \end{bmatrix} = O(h^8)$$

where $\hat{G} = [\hat{E}_1 \vdots \hat{E}_2 \vdots \hat{E}_3 \vdots \hat{E}_4]$. Since $J'M_{22}J$ is positive definite and tends to a positive definite limit as h tends to zero, it follows from (20) and (24) that

$$(25) \qquad \hat{E}_1 - E_1^0 = O(h^4) \qquad \text{and} \qquad \sum_{i=2}^{4} (\hat{E}_i - E_i^0) = O(h^4).$$

From the definition of the matrix G it can be seen that the elements of G are differentiable up to the third order with respect to the non-zero elements of A and B. Clearly, $\partial E_1/\partial a_{ij}$ is of $O(h)$ and, expanding $\exp(hA)$ in a power series, we find from (13) that[19]

$$E_2 = h[\{I + \tfrac{1}{2}(hA)\}C + \tfrac{1}{4}I]B,$$

$$E_3 = h[\{(hA)^2 - 2I\}C + \tfrac{1}{2}(hA) + I]B, \qquad \text{and}$$

$$E_4 = h[\{I - \tfrac{1}{2}(hA)\}C - \tfrac{1}{4}I]B,$$

where $C = \Sigma_{r=0}^{\infty} (hA)^r/(r+3)!$ Thus, the derivatives of G with respect to the non-zero elements of A and B tend to zero as quickly as h.

The elements of the matrix functions $E_1(A, B)$ and $F(A, B) = \Sigma_{i=2}^{4} E_i(A, B) = A^{-1}\{\exp(hA) - I\}B$ may now be expanded in Taylor series to the first order about the true values A^0 and B^0. Using the fact that the first derivatives of E_1 and F are of $O(h)$ as h tends to zero, we find from (25) that

$$(26) \qquad \hat{A} - A^0 = O(h^3) \qquad \text{and} \qquad \hat{B} - B^0 = O(h^3).$$

We can expect, therefore, that the asymptotic bias of the QML estimators diminishes rapidly as the interval between observations decreases. The order of magnitude of this bias, $O(h^3)$, is better than that of similar estimators obtained from the discrete approximation to (7).[20] This suggests that the use of the approximate model (13) for estimation purposes may be worthwhile in spite of the computational difficulties. We should emphasize, however, that the final result (26) is conditional on the assumptions we have made earlier. When these assumptions (particularly Assumption 2(ii)) are not satisfied, the order of magnitude of the asymptotic bias of our estimators may be larger than that given by (26). We take up this problem in the Appendix. We show there, inter alia, that if the first derivatives

[19] It may be of interest to point out here that the following expressions for E_2, E_3, and E_4 should be used when we come to write a program to estimate (13). For the first derivatives of the expressions for E_2, E_3, and E_4 given earlier are unnecessarily complicated and can lead to serious rounding errors in computation.
[20] Sargan [11] has shown that the QML estimators of the coefficient matrices A and B obtained from the discrete approximation to (7) have an asymptotic bias of $O(h^2)$.

of the elements of $z(t)$ do not exist at a countable set of points on the real line, then the order of magnitude of the asymptotic bias is no longer given by (26) but is of $O(h)$. Thus, when our assumptions are not satisfied, asymptotic theory does not lead us to prefer the approximate model (13) rather than models derived from simpler approximations such as the discrete approximation to (7). We now turn to discuss the limiting distribution of the estimators.

3.3. *The Limiting Distribution of the QML Estimators*

Since the exogenous variables are non-random we know that the disturbance η_t in (13) has the same covariance matrix as ξ_t. This matrix is $\Omega = h\Sigma + O(h^2)$ and thus the distribution of η_t is degenerate in the limit as h tends to zero. If we assume that the diagonal elements of Σ are non-zero, then Ω^{-1} is of $O(1/h)$ as h tends to zero.[21]

Given the sample observations $\{y_t, x_t; t = 1, \ldots, T\}$, we know that the QML estimators of A and B satisfy the necessary conditions

$$(27) \qquad \text{trace}\,(V^{-1}\,dV) = 0$$

where $V = (1/T)\Sigma_{t=1}^T (y_t - Gx_t)(y_t - Gx_t)'$. If we denote by c the column vector formed by taking the direct sum of the non-zero elements of successive rows of the matrices A and B, the system (27) can be written

$$(28) \qquad H(c) = (1/T) \sum_{t=1}^{T} W_t' V^{-1}(y_t - Gx_t) = 0$$

where W_t' is the matrix whose (i,j)th element is $w_{ijt} = \Sigma_{k=1}^{n+3m} (\partial g_{jk}/\partial c_i)x_{kt}$ and g_{jk} is the (j,k)th element of G. In passing we note that the matrix W_t is of $O(h)$ as h tends to zero.[22]

The ith row of (28) has the following limited expansion[23] in the neighborhood of c^0, the true value of c:

$$H_i(c) = H_i(c^0) + H_i^1(c^0)'(c - c^0) + \tfrac{1}{2}(c - c^0)'H_i^2(c^1)(c - c^0)$$

where $H_i^1(c) = \partial H_i(c)/\partial c$, $H_i^2(c) = \partial^2 H_i(c)/\partial c\partial c'$, and the vector c^1 lies between c and c^0. If $\hat{c}$ denotes the QML estimator of c, it follows that $\hat{c}$ satisfies

$$(29) \qquad H(c^0) + Q_T(\hat{c} - c^0) = 0$$

where the (i, j)th element of the matrix Q_T is

$$(30) \qquad \partial H_i(c^0)/\partial c_j + \tfrac{1}{2}(\hat{c} - c^0)'\partial^2 H_i(c^1)/\partial c\partial c_j.$$

---

[21] Of course, some elements of Ω^{-1} may be $O(1)$ if Σ has zero elements and this case is not excluded. The assumption made ensures that there is at least one element in each row of Ω^{-1} that is $O(1/h)$ and not $O(1)$.

[22] A little reflection shows that there is at least one element in each row of W_t' which tends to zero no faster than $O(h)$.

[23] It may be worth mentioning that in the system (28) the elements of W_t', V, and G are all functions of c.

Under the assumptions we have made, (30) has a finite limit in probability. In fact, we can readily show that Q_T converges in probability to a matrix which is dominated, as h tends to zero, by

$$(31) \qquad -\operatorname*{plim}_{T \to \infty} (1/T) \sum_{t=1}^{T} W_t^{0\prime} N^{-1} W_t^0$$

where the affix 0 denotes evaluation at the true value c^0 and

$$N = \operatorname*{plim}_{T \to \infty} V^0 = \Omega + \lim_{T \to \infty} (1/T) \sum_{t=1}^{T} \phi_t \phi_t'.$$

From an earlier remark it follows that N^{-1} is of $O(1/h)$, so that (31) is of $O(h)$. Assumption 4 suffices to ensure that (31) is non-singular. Hence, if we write $\bar{Q} = \operatorname{plim}_{T \to \infty} Q_T$, $\bar{Q}^{-1}$ is of $O(1/h)$.

From (29) we obtain:

$$\hat{c} - \operatorname{plim} \hat{c} = -Q_T^{-1} H(c^0) + \bar{Q}^{-1} \operatorname{plim} H(c^0)$$

so that $\sqrt{T}(\hat{c} - \operatorname{plim} \hat{c})$ has the same asymptotic distribution as

$$(32) \qquad -\bar{Q}^{-1}(1/\sqrt{T}) \sum_{t=1}^{T} W_t^{0\prime} N^{-1} \xi_t - \bar{Q}^{-1} \sqrt{T} \left[(1/T) \sum_{t=1}^{T} W_t^{0\prime} N^{-1} \phi_t \right.$$
$$\left. - \operatorname*{plim} (1/T) \sum_{t=1}^{T} W_t^{0\prime} N^{-1} \phi_t \right].$$

The first term of (32) has a limiting normal distribution[24] with zero mean and covariance matrix

$$(33) \qquad \bar{Q}^{-1} \left\{ \operatorname*{plim}_{T \to \infty} (1/T) \sum_{t=1}^{T} W_t^{0\prime} N^{-1} \Omega N^{-1} W_t^0 \right\} \bar{Q}'^{-1}$$

which is of $O(1/h)$ as h tends to zero.

Since $y_t = \Sigma_{s=0}^{\infty} \exp(shA)\xi_{t-s} + u_t$, where $u_t \doteq \Sigma_{s=0}^{\infty} \exp(shA)\chi_{t-s}$ and $\chi_t = \int_0^h \exp(sA)Bz(th-s)\,ds$, the limiting distribution of the second term of (32) depends on that of[25]

$$\sum_{j,k=1}^{n} \sum_{r=1}^{n} \sum_{s=0}^{\infty} \sum_{p=1}^{n} l_{ijr}^0 n^{jk} [\exp(shA)]_{rp} (1/\sqrt{T}) \sum_{t=1}^{T} \phi_{kt} \xi_{pt-1-s}$$

for $(i = 1, 2, \ldots, n)$. But the elements of ϕ_t are bounded, so that $(1/\sqrt{T}) \Sigma_{t=1}^{T} \phi_{kt} \times \xi_{pt-1-s}$ has an asymptotic normal distribution with mean zero and variance $[\Omega]_{pp} \lim_{T \to \infty} (1/T) \Sigma_{t=1}^{T} \phi_{kt}^2$. We can deduce that the limiting distribution of the

---

[24] The ith element of the first term of (32) is a finite linear combination of elements such as $(1/\sqrt{T}) \Sigma_{t=1}^{T} x_{rt} \xi_{kt}$, and x_{rt} includes lagged endogenous variables as well as exogenous variables for various r. To show that such elements have a limiting normal distribution when x_{rt} is a lagged endogenous variable, we must assume the existence of moments of ξ_t of higher order than the second (hence our Assumption 3). We can then appeal to the recent results of Schönfeld [12].

[25] In the following expression, we denote the derivative $\partial g_{jk}/\partial c_i$ by l_{ijk}.

second term of (32) is normal with mean zero and covariance matrix of $O(h^2)$ as h tends to zero.

We can similarly show that the asymptotic covariance between the first and second terms of (32) is of $O(h^2)$. Hence, the dominant term as h tends to zero, of the asymptotic covariance matrix of $\sqrt{T}(\hat{c} - \text{plim } \hat{c})$ is the matrix (33). We have now proved the following result:

THEOREM 4: *The limiting distribution of $\sqrt{hT}(\hat{c} - \text{plim } \hat{c})$ as T tends to infinity is normal for each fixed h. The mean of this limiting distribution is zero for all h and the limit, as h tends to zero, of the covariance matrix of this limiting distribution is*

$$\lim_{h \to 0} h \left\{ \plim_{T \to \infty} \left[(1/T) \sum_{t=1}^{T} W_t^{0\prime} \Omega^{-1} W_t^0 \right]^{-1} \right\}$$

$$= [(\lim_{h \to 0} (1/h^2)\, trace\, \{(\partial G^{0\prime}/\partial c_i)\Sigma^{-1}(\partial G^0/\partial c_j)\overline{M}_{22}\})_{ij}]^{-1}$$

$$= [S(\Sigma^{-1} \otimes J'\overline{M}_{22}J)S']^{-1}$$

where $\otimes$ denotes the right-hand Kronecker product and S is a selection matrix used to delete those rows and columns of $\Sigma^{-1} \otimes J'\overline{M}_{22}J$ which correspond to the zero elements of A and B.

4. CONCLUDING REMARKS

From the results in Section 2, it seems unlikely that the presence of identities in a continuous time system will cause any estimation problem. A spot test on the structure of the model to see if the disturbances in the discrete model have a non-singular distribution is given by the rank criterion of Theorem 3. If the test is affirmative, then the usual procedure based on the discrete model is viable.

Exogenous variables are a more serious complication. In Section 3.1 we found that the discrete time model can still be used for estimation purposes after we have constructed an approximation to the component of the model involving the exogenous variables. The specification error implicit in this approximation is small, and so too is the asymptotic bias of estimators derived from the approximate model. The QML estimators considered in this paper have a limiting normal distribution, but with a biased mean and covariance matrix. This bias disappears as the interval between successive observations goes to zero. The conclusion of asymptotic theory leads us to favor the discrete time model rather than the discrete approximation for estimation purposes, but only in certain cases. This qualification depends on the results of the Appendix to this paper where it is shown that the crucial Assumptions 1 and 2(ii) in Section 3.2 will be satisfied when the exogenous variables are, essentially, either smooth non-random functions of time or stationary stochastic processes that are differentiable in mean square. As yet we have no guide to the small sampling performance of the procedure based on the discrete time model when exogenous variables occur in the model.

The approach we have adopted in this paper to the estimation of continuous time models rests on the hypothesis that simple finite-order finite-parameter models

can satisfactorily represent many processes that we wish to explain in the real economic world. Undoubtedly, our a priori knowledge about the underlying mechanism is incomplete and on many occasions, for instance, we may be unsure of the appropriate order of an equation in the model. But this is a symptom of the general problem an investigator faces in empirical economic research when he discovers that theory alone does not provide a statisfactory guide to the precise formulation of an econometric model. However, this should not deter us from developing parametric methods; and it seems to me that as long as we have reason to believe that models in which time is treated as a continuous variable are important in economics and as long as we continue to build such models in economic theory then we should continue to develop and test techniques of estimating such systems from observed data.

Several questions are left unanswered by this paper. One is the effect of having in econometrics variables whose observations are integrals over time. For, when we integrate the discrete time model the stochastic disturbance becomes a moving average and is serially correlated. Another is the treatment of models in which the assumption of a pure noise disturbance is relaxed. In this case, a more general assumption is that the disturbances are generated by a simple parametric model, such as a first order differential equation system driven by pure noise. By raising the order of the basic system we then end up with the same situation as that discussed in Section 2.2. But these questions are the subject of another paper.

University of Essex

APPENDIX[26]

The purpose of this Appendix is to indicate how Assumptions 1 and 2 in Section 3.2 of the paper can be derived from more basic assumptions about the exogenous variables. In treating the discrete approximate model, Sargan [11] has already done this and the argument here follows a similar line. We consider first the case where the elements of $z(t)$ are non-random functions of time.

ASSUMPTION A: *(i) The elements of $z(t)$ are non-random continuous functions of time that are bounded uniformly in t. (ii) The first three derivatives of $z(t)$ exist and are bounded uniformly in t. (iii) The matrix $\lim_{N \to \infty} (1/N) \int_0^N z(\tau)z(\tau - r)' \, d\tau$ exists for all real r and when $r = 0$ it is positive definite. (iv)[27] The matrix $\lim_{T \to \infty} (1/T) \Sigma_{t=1}^T z(th - s)z(th - r)'$ exists for all real s and r. (v) The submatrix*

$$\lim_{T \to \infty} \frac{1}{T} \begin{bmatrix} Z'Z & Z'Z_{-1} & Z'Z_{-2} \\ Z'_{-1}Z & Z'_{-1}Z_{-1} & Z'_{-1}Z_{-2} \\ Z'_{-2}Z & Z'_{-2}Z_{-1} & Z'_{-2}Z_{-2} \end{bmatrix}$$

of the matrix M_{22} given in Assumption 1 of the paper is positive definite for $h > 0$.

[26] Some of the arguments in this Appendix are rather condensed and for more detail we refer the reader again to [10].

[27] We must make A(iv) an explicit assumption because, although the existence of $\lim_{T \to \infty} (1/T) \times \Sigma_{t=1}^T z(th)z(th)'$ is sufficient to ensure that the upper and lower limits of $(1/T) \Sigma_{t=1}^T z(th - s)z(th - r)'$ are finite, it is not sufficient to ensure that these are equal for $h > 0$.

In view of A(ii) there exist finite positive quantities d_0 and d_1 for which $^{28}\|z(t)\| < d_0$ and $\|z^{(1)}(t)\| < d_1$ for all t. Then, for $s > 0$, $\|z(t) - z(t - s)\| < d_1 s$ and if we let $N = Th$ we have

$$(34) \qquad \left\| \frac{1}{N} \int_0^N z(\tau)z(\tau)' \, d\tau - \frac{1}{T} \sum_{t=1}^T z(th)z(th)' \right\| < d_0 d_1 h + d_1^2 h^2 / 3.$$

It follows from A(iv) and (34) that

$$\lim_{h \to \infty} \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^T z(th)z(th)' = \lim_{N \to \infty} \frac{1}{N} \int_0^N z(\tau)z(\tau)' \, d\tau,$$

which is, by A(iii), positive definite.

We know that

$$(35) \qquad \frac{1}{T} \sum_{t=1}^T y_t y_t' = \frac{1}{T} \sum_{t=1}^T \int_0^\infty \int_0^\infty \exp(sA)Bz(th - s)z(th - r)'B' \exp(rA') \, ds \, dr$$

$$+ \frac{1}{T} \sum_{t=1}^T \int_0^\infty \int_0^\infty \exp(sA)\zeta(th - s)\zeta(th - r)' \exp(rA') \, ds \, dr$$

and since the second term on the right side of (35) converges in probability to

$$(36) \qquad \Omega^* = \int_0^\infty \exp(sA) \sum \exp(sA') \, ds$$

which is positive definite, it follows that

$$\operatorname*{plim}_{T \to \infty} \frac{1}{T} Y_{-1}' Y_{-1} = \lim_{T \to \infty} \frac{1}{T} Y_{-1}^{*'} Y_{-1}^* + \Omega^*$$

where $Y_{-1}^* = [y_0^*, \ldots, y_{T-1}^*]$ and $y_t^* = \int_0^\infty \exp(sA)Bz(th - s) \, ds$. The matrix M_{22} given in Assumption 1 of the paper is then

$$M_{22} = \lim_{T \to \infty} \frac{1}{T} \begin{bmatrix} Y_{-1}^{*'} Y_{-1}^* & Y_{-1}^{*'} Z & Y_{-1}^{*'} Z_{-1} & Y_{-1}^{*'} Z_{-2} \\ Z' Y_{-1}^* & Z'Z & Z'Z_{-1} & Z'Z_{-2} \\ Z_{-1}' Y_{-1}^* & Z_{-1}'Z & Z_{-1}'Z_{-1} & Z_{-1}'Z_{-2} \\ Z_{-2}' Y_{-1}^* & Z_{-2}'Z & Z_{-2}'Z_{-1} & Z_{-2}'Z_{-2} \end{bmatrix} + \begin{bmatrix} \Omega^* & 0 \\ 0 & 0 \end{bmatrix}$$

which is positive definite in view of (36) and $A(v)$. The fact that the matrix $\lim_{h \to 0} J'M_{22}J = J'\overline{M}_{22}J$ exists and is positive definite then follows from A(iii). This establishes Assumption 1.

Under Assumptions A(i) and A(ii), ϕ_t is of $O(h^4)$ uniformly in t and Assumption 2 in the body of the paper is clearly satisfied. However, A(i) and A(ii) are quite restrictive, and we may consider how the order of ϕ_t is affected by relaxing these assumptions. If the elements of $z(t)$ are not necessarily smooth functions of time, then we may expect fairly simple approximations to perform as well in terms of asymptotic bias as the interpolation formula developed in the paper, which uses current and two lagged values. As an alternative to A(ii) suppose we make the following assumption.

ASSUMPTION A'(ii): *The first derivatives of the elements of $z(t)$ exist and are bounded except at a countable set of isolated points on the real line.*

By subdividing the interval $(th - 2h, th)$ into subintervals within which the first derivatives of each element of $z(t)$ exist, we find from the mean value theorem that $\psi(th - s)$ is of $O(h)$. Then ϕ_t is at most of $O(h^2)$ and the estimators considered in the paper have an asymptotic bias of $O(h)$ as h tends to zero. Since the simple approximation $z(th - s) \sim z_t$, for s in the interval $(0, h)$, is sufficient to yield this result, it seems that in the case of exogenous variables satisfying A'(ii) rather than A(ii) the approximate model (13) has no real advantage over simpler approximate models.

Assumption A(iii) might also be regarded as a rather restrictive assumption because it excludes important cases such as exogenous variables which are simple polynomial functions of time. However,

---

^{28} We take the norm $\|D\|$ of a matrix $D = (d_{ij})$ to be the sum $\Sigma_{i,j} |d_{ij}|$.

we can take account of these cases by appropriate normalization. Instead of A(iii) we can substitute the following assumption:

ASSUMPTION A'(iii): *The matrix*

$$\lim_{N \to \infty} D_N^{-1} \int_0^N z(\tau)z(\tau - r)' \, d\tau \, D_N^{-1},$$

where $D_N = diag\,(\sqrt{d_{11}^N}, \ldots, \sqrt{d_{mm}^N})$ *and* $d_{ii}^N = \int_0^N z_i(\tau)^2 \, d\tau$ $(i = 1, \ldots, m)$, *exists for all real r, and when* $r = 0$ *it is positive definite.*

The sums in Assumption 2(ii) of the paper would also have to be normalized. Although the exogenous variables are not uniformly bounded when they are polynomial functions of time, the interpolation formula (10) should give a good approximation to $z(th - s)$ for s in the interval $(0, h)$. The error $\psi(th - s)$ will be of $O(h^3)$ and Assumption 2(ii) of the paper will be satisfied when the sums are appropriately normalized.

We now consider the second case where $z(t)$ is stochastic and we assume the following:

ASSUMPTION B: (i) *$z(t)$ is a strictly stationary ergodic process that is stochastically independent of the process $\zeta(t)$. (ii) The autocovariance matrix $R_{zz}(\tau) = E\{z(t)z(t - \tau)'\}$ has continuous derivatives up to the third order and $R_{zz}(0)$ is positive definite. (iii) There exist no non-zero vectors a, b, and c such that*

$$a'z(t) + b'z(t - \tau) + c'z(t - 2\tau) = 0, \qquad \tau > 0,$$

with probability l.

Given that the initial conditions are in the infinite past we have

$$R_{yz}(\tau) = E\{y(t)z(t - \tau)'\} = \int_0^\infty \exp{(sA)}BR_{zz}(\tau - s) \, ds$$

and

$$(37) \qquad R_{yy}(\tau) = E\{y(t)y(t - \tau)'\}$$

$$= \int_0^\infty \int_0^\infty \exp{(sA)}BR_{zz}(\tau - s + r)B' \exp{(sA')} \, ds \, dr$$

$$+ \int_\tau^\infty \exp{(sA)} \sum \exp{\{(s - \tau)A'\}} \, ds.$$

We denote the first matrix on the right side of (37) by $R_{yy}^{**}(\tau)$. It follows from B(i) that

$$(38) \qquad M_{22} = \begin{bmatrix} R_{yy}^{**}(0) & R_{yz}(0) & R_{yz}(h) & R_{yz}(2h) \\ R_{yz}'(0) & R_{zz}(0) & R_{zz}(h) & R_{zz}(2h) \\ R_{yz}'(h) & R_{zz}'(h) & R_{zz}(0) & R_{zz}(h) \\ R_{yz}'(2h) & R_{zz}'(2h) & R_{zz}'(h) & R_{zz}(0) \end{bmatrix} + \begin{bmatrix} \Omega^* & 0 \\ 0 & 0 \end{bmatrix}$$

where Ω^* is given by (36). We can deduce that M_{22} is positive definite from B(iii) and the fact that both $R_{zz}(0)$ and Ω^* are positive definite. It is also clear from (38) that the limit matrix $J'\overline{M}_{22}J$ is positive definite. This establishes Assumption 1 of the paper.

Since ϕ_t is a linear filter of the stationary ergodic process $z(t)$, we know that

$$(39) \qquad \plim_{T \to \infty} \frac{1}{T} \sum_{t=1}^T x_t \phi_{t-r}' = E(x_t \phi_{t-r}')$$

and

$$(40) \qquad \plim_{T \to \infty} \frac{1}{T} \sum_{t=1}^T \phi_t \phi_{t-r}' = E(\phi_t \phi_{t-r}')$$

for all integral r. To find the order of magnitude of (39) when $r = 0$, we consider first the mean value

$$(41) \qquad E(z_t \phi_t') = \int_0^h E\{z_t \psi(th - s)'\} B' \exp(sA') \, ds.$$

From (10) and (11) we obtain

$$(42) \qquad E\{z_t \psi(th - s)'\} = R_{zz}(s) - [R_{zz}(0) - s\{R_{zz}(2h) - 4R_{zz}(h) + 3R_{zz}(0)\}/2h + s^2\{R_{zz}(0)$$
$$- 2R_{zz}(h) + R_{zz}(2h)\}/2h^2].$$

By expanding each of $R_{zz}(s)$, $R_{zz}(h)$, and $R_{zz}(2h)$ in a Taylor series to the third order about the origin, we find that the right side of (42) reduces to

$$\{s^3 R_{zz}^{(3)}(\theta_1) - sh(2h - s)R_{zz}^{(3)}(\theta_2) + 4sh(h - s)R_{zz}^{(3)}(\theta_3)\}/3!$$

where[29] $0 < \theta_1 < s$, $0 < \theta_2 < h$, and $0 < \theta_3 < 2h$. Hence, for s in the interval $(0, h)$, $E\{z_t \psi(th - s)'\}$ is of $O(h^3)$ as h tends to zero, and it follows from (41) that $E(z_t \phi_t')$ is of $O(h^4)$. We can similarly show that the other elements of (39) are of $O(h^4)$ when $r = 0$.

We now turn to (40). If we put $r = 0$, we have

$$(43) \qquad E(\phi_t \phi_t') = \int_0^h \int_0^h \exp(s_1 A) BE\{\psi(th - s_1)\psi(th - s_2)'\} B' \exp(s_2 A') \, ds_1 \, ds_2,$$

and we can verify in a way similar to the above that (43) is at most of $O(h^3)$ as h tends to zero. Thus, it is clear that Assumption 2(ii) of the paper is satisfied. Since Assumption 2(i) is needed only to help derive the limiting distribution of the QML estimators, Assumption B is sufficient to ensure that these estimators have an asymptotic bias of $O(h^3)$.

However, Assumption B(ii) is quite restrictive. For, if $F(\lambda)$ is the spectral distribution matrix of the process $z(t)$, the existence of the second derivative of $R_{zz}(\tau)$ at the origin implies that

$$\text{trace} \left\{ \int_{-\infty}^{\infty} \lambda^2 \, dF(\lambda) \right\} < \infty.$$

Thus, the process $z(t)$ is differentiable in mean square (see, for instance, [5]). It may appear unrealistic to assume that the exogenous process has this property when from the formulation of the model it is clear that the mean square derivative of the endogenous process $y(t)$ does not exist. We therefore consider the following weaker assumption as an alternative.

ASSUMPTION B'(ii): *The autocovariance matrix $R_{zz}(\tau)$ has right and left derivatives up to the third order at $\tau = 0$ and continuous derivatives to the third order elsewhere. $R_{zz}(0)$ is positive definite.*

Since we do not assume in B'(ii) that $R_{zz}(\tau)$ is differentiable at $\tau = 0$, we do not exclude processes that have no mean square derivative. Thus the elements of $z(t)$ may, for example, be generated by a first order stochastic differential equation system driven by pure noise.

This alternative assumption B'(ii) does not affect the derivation given earlier of Assumption 1 of the paper. On the other hand, Assumption 2(ii) is now not generally true. For, under B'(ii) we can use only one sided Taylor expansions of $R_{zz}(\tau)$ about the origin, and although it is still true that $E(z_t \phi_t')$ and $E(z_{t-2} \phi_t')$ are of $O(h^4)$, we now find that

$$(44) \qquad E\{z_{t-1} \psi(th - s)'\} = \frac{s}{2h}(h - s)\{R_{zz}^{(1)}(0+) - R_{zz}^{(1)}(0-)\} + O(h^2)$$

where $R_{zz}^{(1)}(0+)$ and $R_{zz}^{(1)}(0-)$ denote the right and left derivatives of $R_{zz}(\tau)$ at the origin. It follows from the first mean value theorem for integrals that

$$E(z_{t-1} \phi_t') = \frac{h^2}{12}\{R_{zz}^{(1)}(0+) - R_{zz}^{(1)}(0-)\} B' e^{\theta A'} + \theta(h^3)$$

where $0 \leq \theta \leq h$. In the same way $E(y_{t-1} \phi_t')$ is of $O(h^2)$ and we deduce that the matrix $\text{plim}_{T \to \infty} (1/T) \times \Sigma_{t=1}^T x_t \phi_t'$ in Assumption 2(ii) of the paper is generally of $O(h^2)$ as h tends to zero. On the other hand, we can verify that $E(\phi_t \phi_t')$ is of $O(h^3)$ so that the second part of Assumption 2(ii) is still true under B'(ii).

---

[29] Each element of $R_{zz}^{(3)}(\theta_1)$, for instance, is evaluated at some θ_1 satisfying $0 < \theta_1 < s$. The relation $R_{zz}^{(3)}(\theta_1)$ is not meant to imply that each element is evaluated at the same θ_1.

The effect of replacing B(ii) by the weaker assumption B'(ii), therefore, is to increase the asymptotic bias of the estimators of the paper from $O(h^3)$ to $O(h)$. This leads us to the conclusion based on asymptotic theory that the approximate model (13) offers no real advantage over simpler approximations (such as the discrete approximation) when the exogenous variables are stochastic processes which are not differentiable in mean square.

REFERENCES

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

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

[3] DOOB, J. L.: *Stochastic Processes*. New York: Wiley, 1953.

[4] ERICKSON, R. V.: "Linear Differential Equations Driven by White Noise," *Annals of Mathematical Statistics*, 42 (1971), 820–823.

[5] HANNAN, E. J.: *Multiple Time Series*. New York: Wiley, 1970.

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

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

[8] PHILLIPS, P. C. B.: "The Structural Estimation of a Stochastic Differential Equation System," *Econometrica*, 40 (1972), 1021–1041.

[9] ——: "The Problem of Identification in Finite Parameter Continuous Time Models," *Journal of Econometrics*, 1 (1973), 351–362.

[10] ——: "The Estimation of Some Continuous Time Models," University of Essex Discussion Paper No. 49, 1973.

[11] SARGAN, J. D.: "Some Discrete Approximations to Continuous Time Stochastic Models," *Journal of the Royal Statistical Society*, forthcoming.

[12] SCHÖNFELD, P.: "A Useful Central Limit Theorem for *m*-Dependent Variables," *Metrika*, 15 (1971), 116–128.

[13] WYMER, C. R.: "Econometric Estimation of Stochastic Differential Equation Systems," *Econometrica*, 40 (1972), 565–577.

[14] ZADEH, L. A., AND C. A. DESOER: *Linear System Theory*. New York: McGraw-Hill, 1963.