CHAPTER 7

*The Estimation of Linear Stochastic Differential Equations with Exogenous Variables**

P. C. B. PHILLIPS

1. Introduction

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. Such a record is not usually available, so that some sort of approximation is necessary in most cases before the structural parameters of the system can be estimated with discrete data. In the present paper an approximate model is constructed and the asymptotic properties of quasi-maximum likelihood (QML) estimators derived from this model are investigated. An alternative procedure which uses instrumental variables and should be useful in particular cases is also discussed.

The methodological approach we adopt in this 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.

* This paper is based on chapter 4 of my thesis [5] and represents an extended treatment of a problem that was considered in section 3 of [4]. I am very grateful to my thesis adviser Professor J. D. Sargan for many comments and suggestions on this work and to the referees of Econometrica for their helpful reports on an earlier version of [4].

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, the time unit is taken to be a year and $h = \frac{1}{12}$, 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 we take 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 [6] 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.[1]

2. *The structural system and the corresponding discrete time model*

We shall consider the model

$$Dy(t) = Ay(t) + Bz(t) + \zeta(t), \tag{1}$$

where the coefficient matrices A and B have dimensions $n \times n$ and $n \times m$, respectively, and both have elements which belong to the real number field. The characteristic roots of A are taken to be distinct and to lie in the left half plane. The vectors of endogenous variables $y(t)$ and exogenous variables $z(t)$ are observable at discrete points in time t, and we assume that the elements of $z(t)$ are non-random and have continuous derivatives to the third order. D is the mean square differential operator

[1] 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. But 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.

d/dt, and $\zeta(t)$ is a vector of pure noise disturbances whose spectral density matrix is the positive definite matrix $\Sigma/2\pi$.[2]

Some conditions on the exogenous variables are essential if two later assumptions[3] 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 appendix A.

From the solution of (1) we derive the system

$$y(t) = \exp(hA)y(t-h) + \int_0^h \exp(sA)Bz(t-s)\,ds + \xi(t), \tag{2}$$

where $\xi(t) = \int_0^h \exp(sA)\zeta(t-s)\,ds$, and h is a 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 (2) as

$$y_t = \exp(hA)y_{t-1} + \int_0^h \exp(sA)Bz(th-s)\,ds + \xi_t. \tag{3}$$

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 (3) to obtain a model that can be estimated directly. We now proceed to consider the general case in which an approximation is necessary.

[2] The mathematical difficulties involved in treating a system such as (1) are discussed elsewhere ([1], [4], [8]). It is implicitly assumed that no identities occur in (1). If identities do occur then the results of section 2 of [4] 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.

[3] Assumptions 1 and 2 in section 5.

3. 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

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

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

$$z_t^{(1)} \sim (z_t - z_{t-1})/h \quad \text{and} \quad z_t^{(2)} \sim (z_t - 2z_{t-1} + z_{t-2})/h^2.$$

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.[4] 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

$$\begin{aligned} \hat{z}(th - s) = z_t &- s(z_{t-2} - 4z_{t-1} + 3z_t)/2h \\ &+ s^2(z_t - 2z_{t-1} + z_{t-2})/2h^2. \end{aligned} \tag{4}$$

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

$$\begin{aligned} \psi(th - s) &= z(th - s) - \hat{z}(th - s) \\ &= -s(-s + h)(-s + 2h)z^{(3)}(\theta)/3!, \end{aligned} \tag{5}$$

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

As an approximation to (3) we may now construct the model

[4] Professor J. D. Sargan suggested this approximation.

$$y_t = \exp(hA)y_{t-1} + \int_0^h \exp(sA)B\hat{z}(th - s)\,\mathrm{d}s + \eta_t, \tag{6}$$

where, for extimation 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 (4) and (6) we obtain

$$y_t = E_1 y_{t-1} + E_2 z_t + E_3 z_{t-1} + E_4 z_{t-2} + \eta_t, \tag{7}$$

where the coefficient matrices are

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

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

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

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

Clearly, (7) 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 (5) that $\psi(th - s)$ is of $O(h^3)$ as h tends to zero. Moreover, the condition under which the two models (3) and (7) are equivalent is contained in the relationship

$$\eta_t = \xi_t + \int_0^h \exp(sA)B\psi(th - s)\,\mathrm{d}s, \tag{8}$$

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 (7) 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 (7) 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.

4. *The use of intermediate observations*

Interpolation formulae of higher degree in s than (4) could be used to approximate $z(th - s)$, but this would involve a further reduction of the effective sample size[5] and has the undesirable feature of greatly increasing the computational complexity of the approximate model used in the estimation of parameters. However, in some case, intermediate observations of the exogenous variables may be available; and the extra data can then be used without losing any degrees of freedom.

Suppose, for instance, that k additional equispaced observations of $z(t)$ are available in each interval $(th - h, th)$ for $t = 2, \ldots, T$. A suitable polynomial that can now be used to approximate $z(th - s)$ is

$$\hat{z}(th - s) = \sum_{j=0}^{k+2} z(a_j) \prod_{\substack{i=0 \\ i \neq j}}^{k+2} \frac{th - s - a_i}{a_j - a_i}, \tag{9}$$

where $a_i = th - h(k + 2 - i)/(k + 1)$. The error involved in using (9) is

$$\begin{aligned} \psi(th - s) &= z(th - s) - \hat{z}(th - s) \\ &= \frac{1}{(k+3)!} \prod_{i=0}^{k+2} \left(-s + h\frac{k + 2 - i}{k + 1} \right) z^{(k+3)}(\theta), \end{aligned}$$

where $th - (k + 2)h/(k + 1) < \theta < th$ and assuming, of course, that the derivative of z of order $k + 3$ exists everywhere in the appropriate interval. Since $0 < s < h$ it follows that $\psi(th - s)$ is of $O(h^{k+3})$. When $k = 0$, (9) reduces to the case of the quadratic approximation considered above.

We could now proceed to derive the approximate model based on (6) and (9) but this is merely a routine exercise. The specification error in this model, under the same conditions as before, will be of $O(h^{k+4})$.

[5] One degree of freedom has already been lost by the use of a quadratic approximation.

It should be clear, however, that it will be much more expensive computationally to estimate the approximate model when $k > 1$, and this must be taken into account before it is decided to make use of additional data in this way. An obvious alternative is to use a simple Newton–Cotes formula like Simpson's rule, for which we would require k to be at least 3. The error involved in this approximation will be of $O(h^5)$ as h tends to zero, which is not as good as that of the polynomial above. But we can expect the approximate model to be considerably simpler and this is a great computational advantage. Moreover, when $k = 3$ and Simpson's rule is used, the approximate model will be exact if $z(t)$ is a polynomial of degree at most 2. In practice, therefore, this may provide a good procedure when extra observations of the exogenous variables are available.

We now return to the approximate model (7) developed in section 3 and investigate the effect of the specification error implicit in this model in terms of the asymptotic bias of typical econometric estimators.

5. *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 (1) 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 (7) can be rewritten

$$y_t = G(\delta)x_t + \eta_t, \tag{10}$$

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 minimising

$$\log \det (Y'Y - GX'Y - Y'XG' + GX'XG'), \tag{11}$$

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 minimise (11) with respect to the non-zero elements[6] of A and B

[6] 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.

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 A and 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.

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.[7] 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, (8) does not lead to the simple specification error of $O(h^4)$ for the model (7). 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[8] being satisfied. Finally, it is worth mentioning that the convergence of h

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

[8] Assumption 1.

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 (7) is of $O(h^4)$. This result compares favourably with the specification error of the discrete approximation to (1) which is known[9] to be of $O(h^2)$.

We define

$$Y'_{-1} = [y_0, \ldots, y_{T-1}],$$

$$Z' = [z_1, \ldots, z_T],$$

$$Z'_{-1} = [z_0, \ldots, z_{T-1}],$$

$$Z'_{-2} = [z_{-1}, \ldots, z_{T-2}],$$

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

so that

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

where $Z^* = \frac{1}{3}(Z + Z_{-1} + Z_{-2})$. 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

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} = \operatorname*{plim}_{T\to\infty} T^{-1} \begin{bmatrix} Y'Y & Y'X \\ X'Y & X'X \end{bmatrix},$$

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.

[9] See Sargan [6].

Assumption 2

(i) The elements of $z(t)$ are bounded uniformly in t.

(ii) The matrix $\text{plim}_{T\to\infty} T^{-1}\sum_{t=1}^{T} x_t\phi_{t-r}'$, 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.

(iii) The matrix $\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\phi_t\phi_{t-r}'$ also exists for all integral r and when $r=0$ it is at most of $O(h^8)$.

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

$$
\begin{aligned}
\left|\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\phi_{it}\phi_{jt-r}\right|
&= \lim_{T\to\infty} T^{-1}\left|\sum_{t=1}^{T}\phi_{it}\phi_{jt-r}\right| \\
&\leqq \limsup_{T\to\infty} T^{-1}\left\{\sum_{t=1}^{T}\phi_{it}^2\sum_{t=1}^{T}\phi_{jt-r}^2\right\}^{\frac{1}{2}} \\
&\leqq \left\{\limsup_{T\to\infty}\left[T^{-1}\sum_{t=1}^{T}\phi_{it}^2\right]\right. \\
&\qquad\left.\times\limsup_{T\to\infty}\left[\frac{T+|r|}{T}\frac{1}{T+|r|}\sum_{t=1-r}^{T+|r|}\phi_{jt}^2\right]\right\}^{\frac{1}{2}} \\
&= \left\{\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\phi_{it}^2 \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\phi_{jt}^2\right\}^{\frac{1}{2}}.
\end{aligned}
$$

Hence, $\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\phi_t\phi_{t-r}'$ is bounded uniformly in r and of $O(h^8)$ as h tends to zero.

Assumption 3

The elements of the disturbance ξ_t in the model (3) 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).

Assumption 5

The pair (A^0, B^0) lies in a compact set ψ in $n(n + m)$-dimensional Euclidean space. Moreover, the a priori restrictions on the model (4.1) confine (A, B) to ψ and are sufficient to ensure that there is in ψ a unique solution, A^0, to the matrix equation $\exp(hA) = \exp(hA^0)$ for all $h > 0$.

The conditions on the model implies 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 appendix A, and now derive explicitly the asymptotic bias of estimators derived from (10). We look first at the unrestricted least squares estimator of G and then consider the QML estimators of A and B.

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

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

and

$$M_{12} = G^0M_{22} + \operatorname*{plim}_{T\to\infty} T^{-1} \sum_{t=1}^{T} \xi_t x_t' + \operatorname*{plim}_{T\to\infty} T^{-1} \sum_{t=1}^{T} \phi_t x_t', \tag{13}$$

it follows from assumption 2 that

$$(G^* - G^0)M_{22} = O(h^4). \tag{14}$$

Hence, from the normal properties of order relations,[10] we have

$$(G^* - G^0)M_{22}^2(G^* - G^0)' = O(h^8). \tag{15}$$

We will need the following lemma.

Lemma

If h is a non-negative scalar and the square matrix $A = A(h)$ of order n is positive definite for all $h > 0$ and $\lim_{h\to 0} A(h) = \bar{A}$ exists and is positive semi-definite, then given a small positive quantity ε satisfying $0 < \varepsilon < 1$, there exists a positive real number h_0 such that the matrix

$$A(h) - (1 - \varepsilon)\bar{A}$$

is positive semi-definite for all $h < h_0$.

[10] In particular, if $f(h) = O(h^r)$ and $g(h) = O(h^s)$ then $f(h)\,g(h) = O(h^{r+s})$ and, if $r = s$, $f(h) + g(h) = O(h^r)$.

Proof

We denote by P and $\bar{P}$ the orthogonal matrices for which

$$P'AP = \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n),$$

and

$$\bar{P}'\bar{A}\bar{P} = \bar{\Lambda} = \mathrm{diag}(\bar{\lambda}_1, \ldots, \bar{\lambda}_n).$$

We can now write $A(h) - (1 - \varepsilon)\bar{A}$ as

$$P\Lambda^{\frac{1}{2}}\Lambda^{\frac{1}{2}}P' - (1-\varepsilon)\bar{P}\bar{\Lambda}\bar{P}' = P\Lambda^{\frac{1}{2}}\{I - (1-\varepsilon)\bar{\Lambda}^{-\frac{1}{2}}P'\bar{P}\bar{\Lambda}\bar{P}'P\Lambda^{-\frac{1}{2}}\}\Lambda^{\frac{1}{2}}P',$$

and since

$$P = \bar{P} + R,$$

where $\lim_{h\to 0} R = 0$, we have

$$A(h) - (1 - \varepsilon)\bar{A} = L\{I - (1 - \varepsilon)\Lambda^{-\frac{1}{2}}\bar{\Lambda}\Lambda^{\frac{1}{2}}\}\, L' + S,$$

where $L = P\Lambda^{\frac{1}{2}}$, and

$$S = -(1 - \varepsilon)P\{R'\bar{P}\bar{\Lambda} + \bar{\Lambda}\bar{P}'R + R'\bar{P}\bar{\Lambda}\bar{P}'\bar{R}\}\,P'.$$

Clearly, the elements of S tend to zero with h so the eigenvalues of $A(h) - (1 - \varepsilon)\bar{A}$ can be written as

$$\mu_i = 1 - (1 - \varepsilon)\frac{\bar{\lambda}_i}{\lambda_i} + s_i(h), \qquad i = 1, \ldots, n,$$

where $s_i(h) = o(h)$ for all i.[11]

If $\bar{\lambda}_i > 0$ we select h_i such that

$$|\lambda_i - \bar{\lambda}_i| < \tfrac{1}{2}\varepsilon\bar{\lambda}_i.$$

for all $h < h_i$. It follows readily that

$$\mu_i > \frac{\varepsilon}{2 + \varepsilon} + s_i(h),$$

whenever $h < h_i$. We now select h_i^* such that for $h < h_i^*$,

$$|s_i(h)| < \frac{\varepsilon}{2 + \varepsilon}.$$

Thus, $\mu_i > 0$ for $h < \min_i\{h_i, h_i^*\}$.

[11] We use the order symbol o in the usual sense.

If $\bar{\lambda}_j = 0$ we have

$$\mu_j = 1 + s_j(h),$$

and selecting h_j^{**} such that

$$|s_j(h)| < 1,$$

for $h < h_j^{**}$ we have $\mu_j > 0$ when h is small enough. The lemma is proved with $h_0 = \min_{i,j}\{h_i, h_i^*, h_j^{**}\}$.

Theorem 1

Under the additional assumptions 1, 2(i) and 2(ii), the probability limit $G^* = [E_1^* \vdots E_2^* \vdots E_3^* \vdots E_4^*]$ of the unrestricted least squares estimator of the coefficient matrix G in the model (10) satisfies

$$E_1^* - E_1^0 = O(h^4) \quad \text{and} \quad \sum_{i=2}^{4} (E_i^* - E_i^0) = O(h^4).$$

Proof

Given $\varepsilon > 0$, it follows from the above lemma that the matrix

$$M_{22}^2 - (1 - \varepsilon)\overline{M}_{22}^2$$

is positive semi-definite for small enough h. We consider the matrix

$$(G^* - G^0)\overline{M}_{22}^2(G^* - G^0)',$$

and denote the i'th row of $G^* - G^0$ by e_i'. We have

$$e_i'\overline{M}_{22}^2 e_i \leq \frac{1}{1 - \varepsilon} e_i' M_{22}^2 e_i = O(h^8)$$

by (15). By the Cauchy inequality

$$|e_i'\overline{M}_{22}^2 e_j| \leq \{(e_i'\overline{M}_{22}^2 e_i)(e_j'\overline{M}_{22}^2 e_j)\}^{\frac{1}{2}},$$

and thus

$$(G^* - G^0)\overline{M}_{22}^2(G^* - G^0)' = O(h^8). \tag{16}$$

We now set $D = (G^* - G^0)\overline{M}_{22}$ so that from (16) we have

$$d_{ij}^2 \leq D_i'D_i = O(h^8),$$

where D_i' is the i'th row of D and d_{ij} the (i, j)th element of D. It follows that

$$(G^* - G^0)\overline{M}_{22} = O(h^4). \tag{17}$$

In fact, we can write $\overline{M}_{22}$ as

$$\overline{M}_{22} = \begin{bmatrix} S_{11} & S_{12} & S_{12} & S_{12} \\ S_{21} & S_{22} & S_{22} & S_{22} \\ S_{21} & S_{22} & S_{22} & S_{22} \\ S_{21} & S_{22} & S_{22} & S_{22} \end{bmatrix} = \begin{bmatrix} S_{11} & i' \otimes S_{12} \\ i \otimes S_{21} & ii' \otimes S_{22} \end{bmatrix},$$

where $S_{11} = \lim_{h\to 0} M_{11}$, $S_{12} = \lim_{h\to 0}\{\operatorname{plim}_{T\to\infty} T^{-1} Y'_{-1}Z\}$, $S_{21} = S'_{12}$, $S_{22} = \lim_{h\to 0}\{\lim_{T\to\infty} T^{-1}Z'Z\}$ and i' is the sum vector (1, 1, 1). To verify this expression for $\overline{M}_{22}$ we show that, for instance,

$$\lim_{T\to\infty} \frac{Z'Z}{T} - \lim_{T\to\infty} \frac{Z'Z_{-1}}{T} = O(h), \tag{18}$$

and

$$\operatorname*{plim}_{T\to\infty} \frac{Y'_{-1}Z}{T} - \operatorname*{plim}_{T\to\infty} \frac{Y'_{-1}Z_{-1}}{T} = O(h). \tag{19}$$

If we assume that the first derivatives of the elements of $z(t)$ are uniformly bounded,[12] (18) is straightforward because

$$\begin{aligned} &\left| \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} z_{it}z_{jt} - \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} z_{it}z_{jt-1} \right| \\ &\qquad = \lim_{T\to\infty} T^{-1} \left| \sum_{t=1}^{T} z_{it}(z_{jt} - z_{jt-1}) \right| \\ &\qquad \leqq \limsup_{T\to\infty} T^{-1} \sum_{t=1}^{T} |z_{it}| \, |z_j^{(1)}(\theta)| h, \quad th - h < \theta < th, \end{aligned}$$

which, by assumption 2, is of $O(h)$. To verify (19) we first write $y(th) = \int_0^\infty \exp(sA)Bz(th-s)\,ds + \int_0^\infty \exp(sA)\zeta(th-s)\,ds$, so that

$$\begin{aligned} \operatorname*{plim}_{T\to\infty} \frac{Y'_{-1}Z}{T} &= \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \int_0^\infty \exp(sA)Bz(th-h-s)z(th)'\,ds \\ &\quad + \operatorname*{plim}_{T\to\infty} T^{-1} \sum_{t=1}^{T} \int_0^\infty \exp(sA)\zeta(th-h-s)z(th)'\,ds \\ &= \int_0^\infty \exp(sA)B \lim_{T\to\infty} \left\{ T^{-1} \sum_{t=1}^{T} z(th-h-s)z(th)' \right\} ds. \end{aligned}$$

[12] If $z(t)$ were a stationary ergodic stochastic process it would be sufficient to assume that its autocovariance matrix be differentiable except at the origin where only the left and right derivatives need to exist [c.f., assumption B′(ii) in appendix A].

Taking the norm $\|D\|$ of a matrix $D = [(d_{ij})]$ to be the sum $\sum_{i,j}|d_{ij}|$, we have

$$\left\| \int_0^{\infty} \exp(sA)B \lim_{T\to\infty} \left[T^{-1} \sum_{t=1}^{T} z(th-h-s)\{z(th)-z(th-h)\}' \right] ds \right\|$$

$$\leqq h \int_0^{\infty} \|\exp(sA)\| \, \|B\| \lim_{T\to\infty} T^{-1} \left\{ \sum_{t=1}^{T} \|z(th-h-s)\| \, \|z^{(1)}(\theta)'\| \right\} ds,$$

$$th - h < \theta < th,$$

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

Returning to (17) we deduce that

$$\left[E_1^* - E_1^0 \,\vdots\, \sum_{i=2}^{4} (E_i^* - E_i^0) \right] \begin{bmatrix} S_{11} & i' \otimes S_{12} \\ S_{21} & i' \otimes S_{22} \end{bmatrix} = O(h^4),$$

and postmultiplying this system by J we find that

$$\left[E_1^* - E_1^0 \,\vdots\, \sum_{i=2}^{4} (E_i^* - E_i^0) \right] J'\overline{M}_{22}J = O(h^4),$$

where

$$J'\overline{M}_{22}J = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}.$$

The theorem is now proved because, by assumption 1, $J'\overline{M}_{22}J$ is positive definite.

We now turn to the QML estimators of A and B in the model (10) and prove the following result.

Theorem 2.

Under the additional assumptions 1, 2 and 5, the limits in probability $\hat{A}$ and $\hat{B}$ of the QML estimators of A^0 and B^0 in the model (10) satisfy

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

Proof

From equations (12) and (13) we obtain

$$(G^* - G^0)M_{22}(G^* - G^0)' = M_{\phi 2}M_{22}^{-1}M_{2\phi},$$

where $M_{\phi 2} = \text{plim}_{T\to\infty} T^{-1}\Sigma_t \phi_t x_t'$ and $M_{2\phi} = M_{\phi 2}'$. We now introduce the non-singular matrix K for which

$$K'M_{22}K = I. \tag{20}$$

The existence of K is assured by assumption 1 for positive h and we can, in particular, write K as

$$K = T\Lambda^{-\frac{1}{2}},$$

where T is the orthogonal matrix which diagonalises M_{22} and $\Lambda^{-\frac{1}{2}}$ is the inverse of the matrix whose diagonal elements are the square roots of the eigenvalues of M_{22}. Clearly, the elements of K depend on h, but for positive h we have

$$\begin{aligned}(G^* - G^0)M_{22}(G^* - G^0)' &= M_{\phi 2}KK'M_{2\phi}\\ &= \operatorname*{plim}_{T\to\infty}\left\{\left(\frac{\phi' X}{T}\right)KK'\left(\frac{X'\phi}{T}\right)\right\},\end{aligned}$$

where $\phi' = [\phi_1, \ldots, \phi_T]$. We denote the i'th row of ϕ' by ϕ_i', and then by the Cauchy inequality we have

$$\left(\frac{\phi_i' X}{T}\right)KK'\left(\frac{X'\phi_i}{T}\right) \leq \left(\frac{\phi_i'\phi_i}{T}\right)\text{tr}\left\{K'\left(\frac{X'X}{T}\right)K\right\}, \tag{21}$$

and

$$\left|\left(\frac{\phi_i' X}{T}\right)KK'\left(\frac{X'\phi_j}{T}\right)\right| \leq \left\{\left(\frac{\phi_i'\phi_i}{T}\right)\left(\frac{\phi_j'\phi_j}{T}\right)\right\}^{\frac{1}{2}}\text{tr}\left\{K'\left(\frac{X'X}{T}\right)K\right\}, \tag{22}$$

for $i \neq j$. Taking the limits in probability of (21) and (22) we find from (20) that

$$[M_{\phi 2}KK'M_{2\phi}]_{ii} \leq (n + 3m)\operatorname*{plim}_{T\to\infty}\left(\frac{\phi_i'\phi_i}{T}\right),$$

and

$$\left|[M_{\phi 2}KK'M_{2\phi}]_{ij}\right| \leq (n + 3m)\left\{\operatorname*{plim}_{T\to\infty}\left(\frac{\phi_i'\phi_i}{T}\right)\operatorname*{plim}_{T\to\infty}\left(\frac{\phi_j'\phi_j}{T}\right)\right\}^{\frac{1}{2}}.$$

Hence, it follows from assumption 2(iii) that

$$(G^* - G^0)M_{22}(G^* - G^0)' = O(h^8). \tag{23}$$

We now define $\hat{G} = G(\hat{A}, \hat{B})$, $H = [I \vdots -G]$, $\hat{H} = [I \vdots -G]$, $H^0 = [I \vdots -G^0]$ and $H^* = [I \vdots -G^*]$. Then, from the definition of G^*, we obtain the relationship

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

which, according to (23) can be written as

$$H^0 M H^{0\prime} = H^* M H^{*\prime} + O(h^8). \tag{24}$$

From (10) and assumption 2 we know that

$$M_{11} = G^0 M_{22} G^{0\prime} + \Omega + O(h^4),$$

where $\Omega = \int_0^h \exp(sA)\Sigma \exp(sA')\, ds$ is the covariance matrix of ξ_t. Therefore

$$H^0 M H^{0\prime} = \Omega + O(h^4),$$

and by expanding $\exp(sA)$ as a power series in the expression for Ω we find that $\Omega = h\Sigma + O(h^2)$ so that $H^0 M H^{0\prime}/h$ is of $O(1)$ and tends to a positive definite limit as h tends to zero.

Since $\hat{A}$ and $\hat{B}$ minimise (11) we know that $\det(H^0 M H^{0\prime}) \geqq \det(\hat{H} M \hat{H}')$, and from the relationship

$$\hat{H} M \hat{H}' = H^* M H^{*\prime} + (\hat{G} - G^*)M_{22}(\hat{G} - G^*)', \tag{25}$$

it then follows that

$$\begin{aligned} 0 &\leqq \det(\hat{H} M \hat{H}') - \det(H^* M H^{*\prime}) \\ &\leqq \det(H^0 M H^{0\prime}) - \det(H^* M H^{*\prime}). \end{aligned} \tag{26}$$

Hence, from (24) and (26) we obtain

$$\det \frac{\hat{H} M \hat{H}'}{h} - \det \frac{H^* M H^{*\prime}}{h} = O(h^7). \tag{27}$$

Using a lemma proved by Sargan [6], we deduce from (25) and (27) that

$$(\hat{G} - G^*)M_{22}(\hat{G} - G^*)' = O(h^8). \tag{28}$$

But

$$(\hat{G} - G^0)M_{22}(\hat{G} - G^0)' = (\hat{G} - G^*)M_{22}(\hat{G} - G^*)' + (G^* - G^0)M_{22}(G^* - G^0)' + (\hat{G} - G^*)M_{22}(G^* - G^0)' + (G^* - G^0)M_{22}(\hat{G} - G^*)'. \quad (29)$$

By (23) and (28) the first two terms on the right side of (29) are of $O(h^8)$ and, using the Cauchy inequality, we can readily show that

$$(\hat{G} - G^*)M_{22}(G^* - G^0)' = O(h^8),$$

and, therefore,

$$(\hat{G} - G^0)M_{22}(\hat{G} - G^0) = O(h^8). \quad (30)$$

Given $\varepsilon > 0$, we know from the lemma above that for small enough h,

$$M_{22} - (1 - \varepsilon)\overline{M}_{22}$$

is positive semi-definite. Hence, it follows from (30) that

$$(\hat{G} - G^0)M_{22}(\hat{G} - G^0)' = O(h^8),$$

which implies that

$$\left[\hat{E}_1 - E_1^0 : \sum_{i=2}^{4} (\hat{E}_i - E_i^0)\right] J'\overline{M}_{22}J \begin{bmatrix} (\hat{E}_1 - E_1^0)' \\ \sum_{i=2}^{4} (E_i - E_i^0)' \end{bmatrix} = O(h^8).$$

It follows from assumption 1 that

$$\hat{E}_1 - E_1^0 = O(h^4) \quad \text{and} \quad \sum_{i=2}^{4} (\hat{E}_i - E_i^0) = O(h^4). \quad (31)$$

The last part of our proof involves the step from (31) to the asymptotic bias of $\hat{A}$ and $\hat{B}$. By assumption 5, the equation $\exp(hA) = \exp(hA^0)$ has a unique solution A^0 in ψ. Moreover, the function $\exp(hA)$ is continuous in the elements of A so that, given $\varepsilon > 0$, there exists $\eta > 0$ such that

$$h^{-1} \|\exp(hA) - \exp(hA^0)\| < \eta$$

implies

$$\|A - A^0\| < \varepsilon,$$

in ψ (see, for instance, the lemma on p. 669 of Malinvaud [5]). From (31) we know that

$$\lim_{h \to \infty} h^{-1} \| \exp(h\hat{A}) - \exp(hA^0) \| = 0.$$

Hence, given a sequence of positive numbers $\{\varepsilon^{(r)}: r = 1, 2, \ldots\}$ for which $\lim_{r\to\infty} \varepsilon^{(r)} = 0$, we can find a corresponding sequence of positive numbers $\{h^{(r)}: r = 1, 2, \ldots\}$ such that $\lim_{r\to\infty} h^{(r)} = 0$ and

$$\| \hat{A} - A^0 \| < \varepsilon^{(r)},$$

for all $h < h^{(r)}$. It follows that

$$\hat{A} = A^0 + o(1). \tag{32}$$

However, we can be more precise about the asymptotic bias of $\hat{A}$ than (32). We introduce the vector a formed by taking the direct sum of the non-zero (or unrestricted) elements in each row of A. Then, since the elements of E_1 are differentiable with respect to a, we have

$$\frac{\partial \operatorname{vec}(E_1)}{\partial a} = hS_A + O(h^2), \tag{33}$$

where S_A is a $p \times n^2$ matrix whose rows are linearly independent unit vectors, and p is the number of components in the vector a. Clearly, S_A has full rank.

Thus, we can write

$$h^{-1}\{\operatorname{vec}(E_1) - \operatorname{vec}(E_1^0)\} = h^{-1}\left(\frac{\partial \operatorname{vec} E_1(\tilde{a})}{\partial a}\right)'(\hat{a} - a^0), \tag{34}$$

where $\tilde{a}$ lies on the line segment joining $\tilde{a}$ and a^0. In view of (33) we know that the matrix

$$h^{-1} \frac{\partial \operatorname{vec} E_1(\tilde{a})}{\partial a}$$

has full rank for small enough h so that, since (31) and (34) imply that

$$h^{-1}\left(\frac{\partial \operatorname{vec} E_1(\hat{a})}{\partial a}\right)'(\hat{a} - a^0) = O(h^3),$$

it follows directly that

$$\hat{A} - A^0 = O(h^3). \tag{35}$$

To establish the asymptotic bias of $\hat{B}$ we first introduce the matrix function

$$F(A, B) = E_2 + E_3 + E_4 = A^{-1}\{\exp(hA) - I\}B, \tag{36}$$

so that from (31) we have

$$F(\hat{A}, \hat{B}) - F(A^0, B^0) = O(h^4). \tag{37}$$

Since the elements of F are differentiable with respect to A and B we obtain from (36) and (37) the equation

$$\left(\frac{\partial \operatorname{vec} F(\tilde{a}, \tilde{b})}{\partial a}\right)'(\hat{a} - a^0) + \left(\frac{\partial \operatorname{vec} F(\tilde{a}, \tilde{b})}{\partial b}\right)'(\hat{b} - b^0) = O(h^4), \tag{38}$$

where b is the vector formed by taking the direct sum of the non-zero (or unrestricted) elements in each row of B, and the elements of $\tilde{a}$ and $\tilde{b}$ lie between those of $\hat{a}$ and a^0, $\hat{b}$ and b^0, respectively. Clearly, $\partial \operatorname{vec} F/\partial a$ is of $O(h)$ as h tends to zero, and

$$\begin{aligned}\left(\frac{\partial \operatorname{vec} F}{\partial b}\right)' = \frac{\partial \operatorname{vec} F}{\partial b'} &= [A^{-1}\{\exp(hA) - I\} \otimes I]\frac{\partial \operatorname{vec} B}{\partial b'}\\ &= [A^{-1}\{\exp(hA) - I\} \otimes I]S_B',\end{aligned}$$

where S_B is a $q \times nm$ matrix whose rows are linearly independent unit vectors, and q is the number of components in the vector b. For h small enough, it follows that

$$h^{-1}\frac{\partial \operatorname{vec} F}{\partial b'}$$

has full rank. In view of (35) we can now write (38) as

$$h^{-1}\frac{\partial \operatorname{vec} F(\tilde{a}, \tilde{b})}{\partial b'}(\hat{b} - b^0) = O(h^3).$$

Thus

$$\hat{B} - B_0 = O(h^3). \tag{39}$$

and the theorem is proved.

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 (1).[13] This suggests that the use of the approximate model (7) for estimation purposes may be worthwhile in spite of the computational difficulties. We should emphasize, however, that the final results (35) and (39) are conditional on the assumptions we have made earlier. When these assumptions [particularly assumption 2(ii) and 2(iii)] are not satisfied, the order of magnitude of the asymptotic bias of our estimators may be larger than that given by (35) and (39). We take up this problem in appendix A. We show there, inter alia, that if the first derivatives 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 $O(h^3)$ but is of $O(h)$. Thus, when our assumptions are not satisfied, asymptotic theory does not lead us to prefer the approximate model (7) rather than models derived from simpler approximations such as the discrete approximation to (1). We now turn to discuss the limiting distribution of the estimators.

6. *The limiting distribution of the QML estimators*

Since the exogenous variables are non-random we know that the disturbance η_t in (7) has the same covariance matrix 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 $O(h^{-1})$ as h tends to zero.[14]

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

$$\operatorname{trace}\,(V^{-1}\,\mathrm{d}V) = 0, \tag{40}$$

where

$$V = T^{-1}\sum_{t=1}^{T}(y_t - Gx_t)(y_t - Gx_t)'.$$

[13] Sargan [6] has shown that the QML estimators of the coefficient matrices A and B obtained from the discrete approximation to (1) have an asymptotic bias of $O(h^2)$.

[14] 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 $O(h^{-1})$ and not $O(1)$.

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 (40) can be written

$$H(c) = T^{-1} \sum_{t=1}^{T} W_t' V^{-1}(y_t - Gx_t) = 0, \tag{41}$$

where W_t' is the matrix whose (i,j)'th element is $w_{ijt} = \sum_{k=1}^{n+3m} (\partial g_{jk}/\partial c_i) x_{kt}$, 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.[15]

The i'th row of (41) has the following limited expansion[16] in the neighbourhood 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, $\hat{c}$ satisfies

$$H(c^0) + Q_T(\hat{d} - c^0) = 0, \tag{42}$$

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

$$\partial H_i(c^0)/\partial c_j + \tfrac{1}{2}(\hat{c} - c^0)' \partial^2 H_i(c^1)/\partial c \partial c_j. \tag{43}$$

Under the assumptions we have made, (43) 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

$$-\operatorname*{plim}_{T\to\infty} T^{-1} \sum_{t=1}^{T} W_t^{0\prime} N^{-1} W_t^0, \tag{44}$$

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

$$N = \operatorname*{plim}_{T\to\infty} V^0 = \Omega + \lim_{T\to\infty} T^{-1} \sum^{T} \phi_t \phi_t'.$$

From an earlier remark it follows that N^{-1} is of $O(h^{-1})$, so that (44) is of $O(h)$. Assumption 4 suffices to ensure that (44) is non-singular. Hence, if we write $\bar{Q} = \operatorname{plim}_{T\to\infty} Q_T$, $\bar{Q}^{-1}$ is of $O(h^{-1})$.

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

[16] It may be worth mentioning that in the system (41) the elements of W_t', V and G are all functions of c.

From (42) we obtain

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

so that $T^{\frac{1}{2}}(\hat{c} - \text{plim}\,\hat{c})$ has the same asymptotic distribution as

$$-\bar{Q}^{-1}T^{-\frac{1}{2}}\sum_{t=1}^{T} W_t^{0\prime}N^{-1}\xi_t - \bar{Q}^{-1}T^{\frac{1}{2}} \times \left[T^{-1}\sum_{t=1}^{T} W_t^{0\prime}N^{-1}\phi_t - \text{plim}\,T^{-1}\sum_{t=1}^{T} W_t^{0\prime}N^{-1}\phi_t\right]. \tag{45}$$

The first term of (45) has a limiting normal distribution. For, if we denote the derivative $\partial g_{jk}/\partial c_i$ by l_{ijk}, we may write the i'th element of the vector $T^{-\frac{1}{2}}\sum_{t=1}^{T} W_t^{0\prime}N^{-1}\xi_t$ as

$$\sum_{j,k=1}^{n}\sum_{r=1}^{n+3m} l_{ijr}^{0}n^{jk}T^{-\frac{1}{2}}\sum_{t=1}^{T} x_{rt}\xi_{kt}, \tag{46}$$

where n^{jk} is the (j, k)'th element of N^{-1}. Under the assumptions we have made it is clear that (46) has a limiting normal distribution.[17] Moreover, the asymptotic covariance matrix of the first term of (45) is

$$\bar{Q}^{-1}\left\{\underset{T\to\infty}{\text{plim}}\,T^{-1}\sum_{t=1}^{T} W_t^{0\prime}N^{-1}\Omega N^{-1}W_t^{0}\right\}\bar{Q}^{\prime-1}, \tag{47}$$

and this matrix is of $O(h^{-1})$ as h tends to zero.

Since

$$y_t = \sum_{s=0}^{\infty}\exp(shA)\xi_{t-s} + \mu_t,$$

where

$$\mu_t = \sum_{s=0}^{\infty}\exp(shA)\chi_{t-s},$$

and

$$\chi_t = \int_0^h \exp(sA)Bz(th - s)\,\mathrm{d}s,$$

[17] (46) is a finite linear combination of elements such as $T^{-\frac{1}{2}}\sum_{t=1}^{T} x_{rt}\,\xi_{kt}$ 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 [7].

the limiting distribution of the second term of (45) depends on that of

$$\sum_{j,k=1}^{n} \sum_{r=1}^{n} \sum_{s=0}^{\infty} \sum_{p=1}^{n} l_{ijr}^{0} n^{jk} [\exp(shA)]_{rp} T^{-\frac{1}{2}} \sum_{t=1}^{T} \phi_{kt} \xi_{p\ t-1-s},$$

for $i = 1, 2, \ldots, n$. But the elements of ϕ_t are bounded, so that $T^{-\frac{1}{2}} \sum_{t=1}^{T} \phi_{kt} \xi_{p\ t-1-s}$ has an asymptotic normal distribution with mean zero and variance $[\Omega]_{pp} \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \phi_{kt}^2$. We can deduce that the limiting distribution of the second term of (45) is normal with mean zero and covariance matrix

$$\bar{Q}^{-1} K \bar{Q}'^{-1}, \tag{48}$$

where

$$K = \sum_{s_1, s_2 = 0}^{\infty} \sum_{i,j=1}^{n} L_i' \exp(s_1 hA) \Omega \exp(s_2 hA') L_j N^{i'} \times \left\{ \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \phi_t \phi_{t-s_1+s_2} \right\} N^j.$$

L_i' is the matrix whose (r, s)'th element is l_{ris}^0 and $N^{i'}$ denotes the i'th row of N^{-1}. From assumption 2 and the orders of magnitude of Ω, N^{-1} and the elements l_{ris}^0, it follows that K is of $O(h^4)$. The matrix (48) is then of $O(h^2)$ as h tends to zero.

We can similarly show that the asymptotic covariance between the first and second terms of (45) is of $O(h^2)$. Hence, the dominant term as h tends to zero, of the asymptotic covariance matrix of $T^{\frac{1}{2}}(\hat{c} - \operatorname{plim} \hat{c})$ is the matrix (47). We have now proved the following result:

Theorem 3

The limiting distribution of $(hT)^{\frac{1}{2}}(\hat{c} - \operatorname{plim} \hat{c})$ 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

$$\begin{aligned} \lim_{h\to 0} h \left\{ \operatorname*{plim}_{T\to\infty} \left[T^{-1} \sum_{t=1}^{T} W_t^{0\prime} \Omega^{-1} W_t^0 \right]^{-1} \right\} &= [(\lim_{h\to 0} h^{-2} \operatorname{trace}\{(\partial G^{0\prime}/\partial c_i) \Sigma^{-1} (\partial G^0/\partial c_j) \bar{M}_{22}\})_{ij}]^{-1} \\ &= [S(\Sigma^{-1} \otimes J' \bar{M}_{22} J) S']^{-1}, \end{aligned}$$

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.

The fact that the asymptotic covariance matrix of $T^{\frac{1}{2}}(\hat{c} - \operatorname{plim} \hat{c})$ is of $O(h^{-1})$ is not entirely unexpected. In the first place, the QML estimator of c obtained from the discrete approximation to (1) has an asymptotic covariance matrix of the same order.[18] Furthermore, in the simple case of a single-equation model with no exogenous variables, say $Dy(t) = \alpha y(t) + \zeta(t)$, the result is fairly obvious. For the estimator we have been using becomes, in this case, the least squares estimator which satisfies

$$e^{h\hat{\alpha}} = \sum_{t=1}^{T} y_t y_{t-1} \Big/ \sum_{t=1}^{T} y_{t-1}^2 .$$

The limiting distribution of $T^{\frac{1}{2}}(e^{h\hat{\alpha}} - e^{h\alpha})$ is normal with mean zero and variance $1 - e^{2h\alpha}$. We readily deduce that the asymptotic variance of $T^{\frac{1}{2}}(\hat{\alpha} - \alpha)$ is of $O(h^{-1})$.

7. Concluding remarks

In this paper we have shown that the presence of exogenous variables in systems of stochastic differential equations can cause serious complications. However, we found in section 3 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 the 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 favour the discrete time model rather than the discrete approximation for estimation purposes, but only in certain cases. This qualification depends on the results of appendix A to this paper where it is shown that the crucial assumptions 1, 2(ii) and 2(iii) will be satisfied when the exogenous variables are, essentially, either smooth non-random functions of time or

[18] See Sargan [6].

stationary stochastic processes that are differentiable in mean square to the third order.

Appendix A: Some conditions on the exogenous variables

The purpose of this appendix is to indicate how assumptions 1 and 2 in section 5 of the paper can be derived from more basic assumptions about the exogenous variables. In treating the discrete approximate model Sargan [6] 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

A(i) The elements of $z(t)$ are non-random continuous functions of time that are bounded uniformly in t.

A(ii) The first three derivatives of $z(t)$ exist and are bounded uniformly in t.

A(iii) The matrix $\lim_{N\to\infty} N^{-1} \int_0^N z(\tau)z(\tau - r)' \, d\tau$ exists for all real r and when $r = 0$ it is positive definite.

A(iv)[19] The matrix $\lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} z(th - s)z(th - r)'$ exists for all real s and r.

A(v) The submatrix

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

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

In view of A(ii) there exist finite positive quantities d_0 and d_1 for which $\|z(t)\| < d_0$ and $\|z^{(1)}(t)\| < d_1$ for all t. Then, $\|z(t) - z(t - s)\| < d_1 s$, for $s > 0$, and we can show that

$$\left\| h^{-1} \int_{(t-1)h}^{th} z(\tau)z(\tau)' \, d\tau - z(th)z(th)' \right\| < d_0 d_1 h + d_1^2 h^2/3.$$

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

If we now let $N = Th$, where T is an integer representing the number of sample observations, we have

$$\left\| N^{-1}\int_0^N z(\tau)z(\tau)'\,\mathrm{d}\tau - T^{-1}\sum_{t=1}^{T} z(th)z(th)' \right\|$$
$$\leqq T^{-1}\sum_{t=1}^{T}\left\| h^{-1}\int_{(t-1)h}^{th} z(\tau)\ (\tau)'\,\mathrm{d}\tau - z(th)z(th)' \right\|$$
$$< d_0 d_1 h + d_1^2 h^2/3. \tag{49}$$

It follows from A(iv) and (49) that

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

which is, by A(iii), positive definite. We know that

$$T^{-1}\sum_{t=1}^{T} y_t y_t' = T^{-1}\sum_{t=1}^{T}\int_0^\infty\int_0^\infty \exp(sA)Bz(th-s)\,z(th-r)'B'$$
$$\times \exp(rA')\,\mathrm{d}s\,\mathrm{d}r$$
$$+ T^{-1}\sum_{t=1}^{T}\int_0^\infty\int_0^\infty \exp(sA)\zeta(th-s)\zeta(th-r)'$$
$$\times \exp(rA')\,\mathrm{d}s\,\mathrm{d}r, \tag{50}$$

and since the second term on the right side of (50) converges in probability to

$$\Omega^* = \int_0^\infty \exp(sA)\Sigma\exp(sA')\,\mathrm{d}s, \tag{51}$$

which is positive definite, it follows that

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

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

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

which is positive definite in view of (51) 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 chapter is clearly satisfied. But 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 assume:

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 chapter 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 (7) has no real advantage over simpler approximate models.

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

A′(iii) The matrix

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

where $D_N = \operatorname{diag}(\sqrt{d_{11}^N}, \ldots, \sqrt{d_{mm}^N})$ and $d_{ii}^N = \int_0^N z_i(\tau)^2\,\mathrm{d}\tau$ $(i = 1, \ldots, m)$, 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 normalised. Although the exogenous variables are not uniformly bounded when they are polynomial functions of time the interpolation formula (4) 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 normalised.

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

Assumption B

B(i) $z(t)$ is a strictly stationary ergodic process that is stochastically independent of the process $\zeta(t)$.

B(ii) The autocovariance matrix $R_{zz}(\tau) = E\{z(t)z(t-\tau)'\}$ has continuous derivatives up to the sixth order and $R_{zz}(0)$ is positive definite.

B(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 1.

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

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

and

$$\begin{aligned} R_{yy}(\tau) &= E\{y(t)y(t-\tau)'\} \\ &= \int_0^\infty \int_0^\infty \exp(sA)BR_{zz}(\tau - s + r)B'\exp(sA')\,\mathrm{d}s\,\mathrm{d}r \\ &\quad + \int_\tau^\infty \exp(sA)\Sigma\exp\{(s-\tau)A'\}\,\mathrm{d}s. \end{aligned} \tag{52}$$

We denote the first matrix on the right side of (52) by $R_{yy}^{**}(\tau)$. It follows from B(i) that

$$M_{22} = \begin{bmatrix} R_{yy}^{**}(0) & R_{yz}(0) & R_{yz}(h) & R_{yz}(2h) \\ R_{yz}'(0) & R_{zz}(0) & R_{zz}(h) & R_{zz}(2h) \\ R_{yz}'(h) & R_{zz}'(h) & R_{zz}(0) & R_{zz}(h) \\ R_{yz}'(2h) & R_{zz}'(2h) & R_{zz}'(h) & R_{zz}(0) \end{bmatrix} + \begin{bmatrix} \Omega^* & 0 \\ 0 & 0 \end{bmatrix}, \tag{53}$$

where Ω^* is given by (51). 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 clear from (51) 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

$$\operatorname*{plim}_{T\to\infty} T^{-1} \sum_{t=1}^{T} x_t \phi'_{t-r} = E(x_t \phi'_{t-r}), \tag{54}$$

and

$$\operatorname*{plim}_{T\to\infty} T^{-1} \sum_{t=1}^{T} \phi_t \phi'_{t-r} = E(\phi_t \phi'_{t-r}), \tag{55}$$

for all integral r. To find the order of magnitude of (54) when $r = 0$ we consider first the mean value

$$E(z_t \phi'_t) = \int_0^h E\{z_t \psi(th - s)'\} B' \exp(sA') \, ds. \tag{56}$$

From (4) and (5) we obtain

$$\begin{aligned} E\{z_t \psi(th - s)'\} = R_{zz}(s) - [&R_{zz}(0) - s\{R_{zz}(2h) - 4R_{zz}(h) \\ &+ 3R_{zz}(0)\}/2h + s^2\{R_{zz}(0) - 2R_{zz}(h) \\ &+ R_{zz}(2h)\}/2h^2]. \end{aligned} \tag{57}$$

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 (57) 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 $0 < \theta_1 < s$, $0 < \theta_2 < h$, $0 < \theta_3 < 2h$. 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 (56) that $E(z_t \phi'_t)$ is of $O(h^4)$. We can similarly show that the other elements of (54) are of $O(h^4)$ when $r = 0$. Thus, assumption 2(ii) of the paper is satisfied.

We now turn to (55). If we put $r = 0$ we have

$$\begin{aligned} E(\phi_t \phi'_t) = \int_0^h \int_0^h &\exp(s_1 A) B E\{\psi(th - s_1)\psi(th - s_2)'\} B' \\ &\times \exp(s_2 A') \, ds_1 \, ds_2, \end{aligned} \tag{58}$$

and

$$E\{\psi(th - s_1)\psi(th - s_2)'\} = E[z(th - s_1)\{z(th - s_2) - \hat{z}(th - s_2)\}'] - E[\hat{z}(th - s_1) \times \{z(th - s_2) - \hat{z}(th - s_2)\}']. \quad (59)$$

From the definition of $\hat{z}$ in (4) it follows that the right side of (59) is (omitting the z subscripts in R_{zz})

$$\begin{aligned}
&R(s_2 - s_1) - \Big[R(s_2) - \frac{s_1}{2h}\{R(s_2 - 2h) - 4R(s_2 - h) + 3R(s_2)\} \\
&\quad + \frac{s_1^2}{2h^2}\{R(s_2) - 2R(s_2 - h) + R(s_2 - 2h)\}\Big] \\
&\quad - \Big\{R(-s_1) - \Big[R(0) - \frac{s_1}{2h}\{R(-2h) - 4R(-h) \\
&\quad + 3R(0)\} + \frac{s_1^2}{2h^2}\{R(0) - 2R(-h) + R(-2h)\}\Big]\Big\} \\
&\quad + \frac{s_2}{2h}\Big[\Big\{R(2h - s_1) - \Big[R(2h) - \frac{s_1}{2h}\{R(0) - 4R(h) \\
&\quad + 3R(2h)\} + \frac{s_1^2}{2h^2}\{R(2h) - 2R(h) + R(0)\}\Big]\Big\} \\
&\quad - 4\Big\{R(h - s_1) - \Big[R(h) - \frac{s_1}{2h}\{R(-h) - 4R(0) \\
&\quad + 3R(h)\} + \frac{s_1^2}{2h^2}\{R(h) - 2R(0) + R(-h)\}\Big]\Big\} \\
&\quad + 3\Big\{R(-s_1) - \Big[R(0) - \frac{s_1}{2h}\{R(-2h) - 4R(-h) \\
&\quad + 3R(0)\} + \frac{s_1^2}{2h^2}\{R(0) - 2R(-h) + R(-2h)\}\Big]\Big\}\Big]
\end{aligned}$$

$$- \frac{s_2^2}{2h^2}\left[\left\{R(-s_1) - \left[R(0) - \frac{s_1}{2h}\{R(-2h) - 4R(-h)\right.\right.\right.$$

$$\left.\left. + 3R(0)\} + \frac{s_1^2}{2h^2}\{R(0) - 2R(-h) + R(-2h)\}\right]\right\}$$

$$- 2\left\{R(h - s_1) - \left[R(h) - \frac{s_1}{2h}\{R(-h) - 4R(0)\right.\right.$$

$$\left.\left. + 3R(h)\} + \frac{s_1^2}{2h^2}\{R(h) - 2R(0) + R(-h)\}\right]\right\}$$

$$+ \left\{R(2h - s_1) - \left[R(2h) - \frac{s_1}{2h}\{R(0) - 4R(h)\right.\right.$$

$$\left.\left.\left. + 3R(2h)\} + \frac{s_1^2}{2h^2}\{R(2h) - 2R(h) + R(0)\}\right]\right\}\right]. \tag{60}$$

We now expand each of $R(s_2 - s_1)$, $R(s_2)$, $R(s_2 - 2h)$, $R(s_2 - h)$, $R(-s_1)$, $R(h - s_1)$, $R(2h - s_1)$, $R(-h)$, $R(-2h)$, $R(h)$ and $R(2h)$ in a Taylor series to the sixth order about the origin and, collecting terms, we find that (60) reduces to

$$R^{(6)}(\theta_1)\frac{(s_2 - s_1)^6}{720} - R^{(6)}(\theta_2)\left(\frac{s_2^6}{720} - \frac{s_1 s_2^6}{480h} + \frac{s_1^2 s_2^6}{1440h^2}\right)$$

$$+ R^{(6)}(\theta_3)\left\{\frac{s_1^2(s_2 - h)^6}{720h^2} - \frac{s_1(s_2 - h)^6}{360h}\right\}$$

$$+ R^{(6)}(\theta_4)\left\{\frac{s_1(s_2 - 2h)^6}{1440h} - \frac{s_1^2(s_2 - 2h)^6}{1440h^2}\right\}$$

$$- R^{(6)}(\theta_5)\left(\frac{s_1^6}{720} - \frac{s_2 s_1^6}{480h} + \frac{s_2^2 s_1^6}{1440h^2}\right)$$

$$+ R^{(6)}(\theta_6)\left\{\frac{s_2^2(h - s_1)^6}{720h^2} - \frac{s^2(h - s_1)^6}{360h}\right\}$$

$$+ R^{(6)}(\theta_7)\left\{\frac{s_2(2h - s_1)^6}{1440h} - \frac{s_2^2(2h - 2_1)^6}{1440h^2}\right\}$$

$$
\begin{aligned}
&+ R^{(6)}(\theta_8)\left\{\frac{s_1h^5}{360} - \frac{s_1^2h^4}{720} - \frac{s_1s_2h^4}{180} + \frac{s_2s_1^2h^3}{288}\right. \\
&\qquad \left. + \frac{s_1s_2^2h^3}{480} - \frac{s_1^2s_2^2h^2}{720}\right\} \\
&+ R^{(6)}(\theta_9)\left(-\frac{2s_1h^5}{45} + \frac{2s_1^2h^4}{45} + \frac{s_1s_2h^4}{15}\right. \\
&\qquad \left. - \frac{s_1^2s_2h^3}{15} - \frac{s_1s_2^2h^3}{45} + \frac{s_1^2s_2^2h^2}{45}\right) \\
&+ R^{(6)}(\theta_{10})\left(+\frac{s_2h^5}{380} - \frac{s_1s_2h^4}{180} + \frac{s_1^2s_2h^3}{480}\right. \\
&\qquad \left. - \frac{s_2^2h^4}{720} + \frac{s_1s_2^2h^3}{288} - \frac{s_1^2s_2^2h^2}{720}\right) \\
&+ R^{(6)}(\theta_{11})\left(-\frac{2s_2h^5}{45} + \frac{s_1s_2h^4}{15} - \frac{s_1^2s_2h^3}{45}\right. \\
&\qquad \left. + \frac{2s_2^2h^5}{45} - \frac{s_1s_2^2h^3}{15} + \frac{s_1^2s_2^2h^2}{45}\right),
\end{aligned}
\tag{61}
$$

where the θ_i lie in the range $-2h < \theta_i < 2h$. From (58) and (61) it now follows that

$$E(\phi_t\phi_t') = O(h^8),$$

and assumption 2(iii) 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 somewhat restrictive. For, if $F(\lambda)$ is the spectral distribution matrix of the process $z(t)$, the existence of the sixth derivative of $R_{zz}(\tau)$ at the origin implies that

$$\operatorname{trace}\left\{\int_{-\infty}^{\infty} \lambda^{2r}\, \mathrm{d}F(\lambda)\right\} < \infty, \qquad r = 1, 2, 3.$$

Thus, the process $z(t)$ is differentiable in mean square to the third order.[20] 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.

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

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

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

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

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 $\operatorname{plim}_{T\to\infty} T^{-1}\sum_{t=1}^{T} x_t\phi_t'$ in assumption 2(ii) of the paper is generally of $O(h^2)$ as h tends to zero.

Moreover, under B′(ii), assumption 2(iii) of the paper is no longer true. We now find that for $h \geq s_2 \geq s_1 \geq 0$,

$$E\{\psi(th-s_1)\psi(th-s_2)'\} = \left(-s_1 + \frac{s_1 s_2}{2h} + \frac{s_1^2 s_2}{2h^2} + \frac{s_1 s_2^2}{2h^2} - \frac{s_1^2 s_2^2}{2h^3}\right) \times \{R^{(1)}(0+) - R^{(1)}(0-)\} + O(h^2),$$

[20] See, for instance, Hannan [2].

and, by dividing the domain of integration in (58) up into the region $s_2 > s_1$ and the rest (as in Sargan [6]) we obtain

$$E(\phi_t \phi_t') = \frac{-7}{72} h^3 B\{R^{(1)}(0+) - R^{(1)}(0-)\} B' + O(h^4).$$

Thus, assumption 2(iii) is not satisfied.

The effect of replacing B(ii) by the weaker assumption B′(ii), therefore, is to increase the asymptotic bias of the estimators in section 5. In the case of the QML estimators of A and B we can deduce, by following the line of argument in theorem 2, only that the asymptotic bias is of $O(h^{\frac{1}{2}})$. 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 and the model is estimated by a conventional non-linear regression.

Appendix B: An alternative procedure based on instrumental variables[21]

When the exogenous variables are stochastic and not differentiable in mean square, the asymptotic bias of the estimators we have been considering is somewhat disappointing. The main reason for this is that in estimating (7) we use as regressors the set of variables $\{y_{t-1}, z_t, z_{t-1}, z_{t-2}\}$ and as we have seen in the last appendix, $E(y_{t-1}\phi_t')$ and $E(z_{t-1}\phi_t')$ are of $O(h^2)$ rather than of $O(h^4)$ in this case.

One way of overcoming this problem is to replace y_{t-1} and z_{t-1} in the regressor set by the instrumental variables y_{t-2} and z_{t-2}. We can then estimate (7) with the set of instrumental variables $\{y_{t-2}, z_t, z_{t-2}, z_{t-3}\}$. The procedure we propose is as follows:

(i) Estimate the coefficient matrix G by an unrestricted least squares regression and construct the residual moment matrix Ω^* from this regression as an estimator of Ω.

(ii) Estimate A and B by minimising the quadratic form

$$\text{trace}\,\{\Omega^{*-1}(Y' - GX')Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime}(Y - XG')\},$$

[21] I am very grateful to Professor Sargan who first suggested to me that instrumental variables would be useful in this context.

with respect to the unknown elements of A and B, where $Z^{*\prime} = [z_1^*, \ldots, b_T^*]$ and $z_t^{*\prime}$ is the vector of instrumental variables $(y'_{t-2}, z'_t, z'_{t-2}, z'_{t-3})$.

We let $\tilde{A}$ and $\tilde{B}$ be the limits in probability of the instrumental variable estimators of A and B obtained via (i) and (ii). To establish the asymptotic bias of $\tilde{A}$ and $\tilde{B}$ we find it convenient to modify assumption B(iii) somewhat and introduce B(iv) as follows:

Assumption

B′(iii) There exist no vectors a, b, d and d, not all zero, for which

$$a'z(t) + b'z(t-\tau) + c'z(t-2\tau) + d'z(t-3\tau) = 0, \qquad \tau > 0,$$

with probability 1.

B(iv) The matrix $M_{xz}^* = E(x_t z_t^{*\prime})$ is non-singular for positive h.

Then we have:

Theorem 4

If assumptions B(i), B′(ii), B′(iii), B(iv) and 5 are satisfied, then the asymptotic bias of $\tilde{A}$ and $\tilde{B}$ is given by

$$\tilde{A} - A^0 = O(h^3) \quad \text{and} \quad \tilde{B} - B^0 = O(h^3).$$

Proof

$\tilde{A}$ and $\tilde{B}$ minimise

$$\operatorname{tr}\{N^{-1}(M_{yz}^* - GM_{xz}^*)M_{zz}^{*-1}(M_{zy}^* - M_{zx}^*G')\}, \tag{62}$$

where

$$N = \operatorname*{plim}_{T\to\infty} T^{-1}(Y' - GX')(Y - GX) = \Omega + E(\phi_t\phi_t'),$$

$$M_{yz}^* = \operatorname*{plim}_{T\to\infty} T^{-1}Y'Z^* = E(y_t z_t^{*\prime}),$$

and

$$M_{zz}^* = \operatorname*{plim}_{T\to\infty} T^{-1}Z^{*\prime}Z^* = E(z_t^* z_t^{*\prime}).$$

The first-order conditions for a minimum of (62) are

$$N^{-1}M_{yz}^*M_{zz}^{*-1}M_{zx}^* - N^{-1}\tilde{G}M_{xz}^*M_{zz}^{*-1}M_{zx}^* = 0, \tag{63}$$

where $\tilde{G} = G(\tilde{A}, \tilde{B})$. But, since $M_{yz}^* = G^0M_{xz}^* + M_{\phi z}^*$, where $M_{\phi z}^* = \operatorname{plim}_{T\to\infty} T^{-1}\Sigma_t\phi_t z_t^{*\prime}$, we derive from (63) the equation

$$(G^0 - \tilde{G})M^*_{xz}M^{*-1}_{zz}M^*_{zx} = -M^*_{\phi z}M^{*-1}_{zz}M^*_{zx}. \tag{64}$$

Now $M^*_{\phi z} = E(\phi_t z^{*\prime}_t)$ and this matrix has elements which are of $O(h^4)$ as h tends to zero. For, $E(\phi_t z'_t)$, $E(\phi_t z'_{t-2})$ and $E(\phi_t z'_{t-3})$ are all of $O(h^4)$ as in the last appendix, and

$$E(\phi_t y'_{t-2}) = \int_0^h \exp(sA)BE\{\psi(th - s)y'_{t-2}\}\,ds.$$

But

$$\begin{aligned} E\{\psi(th - s)y'_{t-2}\} = R_{zz}(2h - s) - \Bigg[& R_{zy}(2h) \\ & - \frac{s}{2h}\{R_{zy}(0) - 4R_{zy}(h) + 3R_{zy}(2h)\} \\ & + \frac{s^2}{2h^2}\{R_{zy}(2h) - 2R_{zy}(h) + R_{zy}(0)\}\Bigg], \end{aligned}$$

and, expanding $R_{zy}(2h - s)$, $R_{zy}(2h)$ and $R_{zy}(h)$ in a Taylor series to the third order on the positive side of the origin,[22] we find that $E\{\psi(th - s)y'_{t-2}\}$ is of $O(h^3)$ when $0 \le s \le h$ so that $E(\phi_t y'_{t-2})$ is of $O(h^4)$. Now

$$\begin{aligned} M^*_{zz} &= \begin{bmatrix} R_{yy} & R_{yz}(-2h) & R_{yz}(0) & R_{yz}(h) \\ R_{zy}(2h) & R_{zz}(0) & R_{zz}(2h) & R_{zz}(3h) \\ R_{zy}(0) & R_{zz}(-2h) & R_{zz}(0) & R_{zz}(h) \\ R_{zy}(-h) & R_{zz}(-3h) & R_{zz}(-h) & R_{zz}(0) \end{bmatrix} \\ &= \begin{bmatrix} R_{y*y*}(0) & M_{34} \\ M_{43} & M_{44} \end{bmatrix} + \begin{bmatrix} \Omega^* & 0 \\ 0 & 0 \end{bmatrix}, \end{aligned}$$

where Ω^* and $R_{y*y*}(0)$ are defined in (51) and (52), respectively, and

$$M_{34} = [R_{yz}(-2h) : R_{yz}(0) : R_{yz}(h)],$$

and

$$M_{44} = \begin{bmatrix} R_{zz}(0) & R_{zz}(2h) & R_{zz}(3h) \\ R_{zz}(-2h) & R_{zz}(0) & R_{zz}(h) \\ R_{zz}(-3h) & R_{zz}(-h) & R_{zz}(0) \end{bmatrix}.$$

[22] Such an expansion is possible because $R_{zy}(\tau) = \int_0^\infty R_{zz}(\tau + s)B' \exp(sA')\,ds$, $R_{zz}(\tau + s)$ has continuous derivatives to the third order for $s > 0$, $\tau > 0$ and the matrix A has stable roots.

It follows from B′(iii) that M_{44} is positive definite so that, since Ω^* is positive definite, M^*_{zz} is non-singular for positive h. Thus, (64) is well defined for $h > 0$.

From equation (64) and the order of magnitude of $M^*_{\phi z}$ we now obtain

$$(G^0 - \tilde{G})M^*_{xz}M^*_{zx}(G^0 - \tilde{G})' = O(h^8). \tag{65}$$

The matrix $M^*_{xz}M^*_{zx}$ is positive definite when $h > 0$ and converges as h tends to zero to the limit matrix $\overline{M}^*_{xz}\overline{M}'^*_{zx}$, where

$$\overline{M}^*_{xz} = \begin{bmatrix} R_{yy}(0) & i' \otimes R_{yz}(0) \\ i \otimes R_{zy}(0) & ii' \otimes R_{zz}(0) \end{bmatrix}.$$

Using the lemma given in section 5 and following the line of argument in the proof of theorem 1 we find from (65) that

$$(G^0 - \tilde{G})\overline{M}^*_{xz}\overline{M}_{zx}(G^0 - \tilde{G})' = O(h^8),$$

which in turn implies that

$$E^0_1 - \tilde{E}_1 = O(h^4) \quad \text{and} \quad \sum_{i=2}^{4} (E^0_i - \tilde{E}_i) = O(h^4).$$

In view of assumption 5, we can now establish in the same way as in the proof of theorem 2 that

$$\tilde{A} - A^0 = O(h^3) \quad \text{and} \quad \tilde{B} - B^0 = O(h^3).$$

There seems to be no reason why the set of instrumental variables should be confined to those we have considered here. When h is small, there may be some advantage to be gained in small samples of including extra variables such as y_{t-3} in Z^*. On the other hand, since the $n \times (n + 3m)$ matrix G depends only on the unknown elements of A and B we need only use the instrumental variables y_{t-2} and z_t in Z^*. However, in this latter case the asymptotic bias of the resulting estimators is not assured by the proof of the above theorem because the rank of M^*_{xz} in this case will be less than $n + 3m$.

Appendix C: Computational note

When we come to estimate the model (7) in practice, we will frequently wish to use the derivatives of E_1, E_2, E_3 and E_4 with respect to the unknown elements of A and B in the computation. The expressions given earlier for the matrix functions E_i ($i = 2, 3, 4$) have complicated derivatives which can lead to serious rounding errors and it is useful to simplify these expressions first by expanding $\exp(hA)$ in a power series so that

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

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

and

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

where $C = \sum_{r=0}^{\infty}(hA)^r/(r+3)!$

These expressions lead to simpler derivatives that do not, in particular, involve powers of A^{-1}.

References

[1] Doob, J. L., 1953, Stochastic processes (Wiley, New York).
[2] Hannan, E. J., 1970, Multiple time series (New York).
[3] Malinvaud, E., 1970, Statistical methods of econometrics, 2nd ed. (North-Holland, Amsterdam).
[4] Phillips, P. C. B., 1974, The estimation of some continuous time models, Econometrica 42, pp. 803–824.
[5] Phillips, P. C. B., 1974, Problems in the estimation of continuous time models, Ph.D. thesis (University of London).
[6] Sargan, J. D., 1974, Some discrete approximations to continuous time stochastic models, Journal of the Royal Statistical Society, Series B, 36, pp. 74–90.
[7] Schonfeld, P., 1971,A useful central limit theorem for m-dependent variables, Metrika 15, pp. 116–128.
[8] Wong, E., 1974, Stochastic processes in information and dynamical systems (McGraw-Hill, New York).