CHAPTER 15

The Treatment of Flow Data in the Estimation of Continuous Time Systems

P. C. B. Phillips*

1 Introduction

In most of the work that has recently been done in the field of estimating continuous time econometric models it has been assumed that point observations of the variables are available.[1] Although this is true of certain variables like wages, prices, interest rates and stocks (which we term instantaneous variables[2]), a large number of economic variables are observed as time integrals. For instance, in the case of aggregate consumption we use $C(t)$ to denote the rate of consumption expenditure at a particular point in time. If this rate were sustained for a whole unit time period then $C(t)$ also represents the flow of aggregate consumption over that period. More usually, the rate will vary over the time period so that the actual flow is the integral of $C(t)$ over time, and this is the quantity we observe.[3] A number of early writers emphasized this point (see Koopmans, 1950, and Phillips, 1956) and, in his study, Phillips indicated how the presence of flow data causes problems in the estimation of interdependent systems of continuously distributed lags.

Naturally, it is important to develop practical econometric methods that are designed to make use of data in the form in which it becomes available. Wymer (1971) has briefly discussed the problem of flow data in the context of estimating a system of stochastic differential equations from its non-recursive discrete approximation. If we centre observations approximately, then there is no difficulty in the construction of the discrete approximation. But estimators derived from the discrete approximation in this case are biased and the bias does not disappear as the sampling interval tends to zero.[4] Our main emphasis in the present paper, however, is on estimation via the exact discrete model where the presence of flow data causes the disturbance to become a first-order moving average. Our results may, therefore, be regarded as an extension of those in Phillips (1974a).

*University of Birmingham. This essay is a shorter version of a paper (Phillips, 1974b) read at the Grenoble meeting of the Econometric Society, 1974, and represents a development of Phillips (1974c, Chapter 7).

2 Linear Models with Flow Data

The structural system which we consider is

$$Dy(t) = Ay(t) + Bz(t) + \zeta(t) \tag{15.1}$$

where $y(t)$ is an $n \times 1$ vector of endogenous variables, $z(t)$ is an $m \times 1$ vector of exogenous variables, and $\zeta(t)$ is a vector of stochastic disturbances. The elements of the coefficient matrices A and B belong to the real number field and the eigenvalues of A are assumed to be distinct with negative real parts. D is the differential operator $\mathrm{d}/\mathrm{d}t$, taken in the mean square sense,[5] and the elements of $\zeta(t)$ are assumed to be pure noise so that the spectral density matrix of $\zeta(t)$ is a constant matrix, which we denoted by $\Sigma/2\pi$, over the whole real line. The latter assumption is certainly a strong one and useful methods are currently being developed to deal with systems such as (15.1) under much weaker assumptions about the residuals.[6] When the system is closed, however, these methods are not applicable and some assumption such as ours is necessary for a statistical treatment of the model. The methods we develop in this paper can still be used when exogenous variables are present, but an approximation along the lines of that suggested in Phillips (1974a) is generally needed for estimation purposes.

The discrete time system that corresponds to equation (15.1) we write as

$$\begin{aligned} y(t) = {} & \exp(hA)y(t-h) \\ & + \int_0^h \exp(sA)\,Bz(t-s)\,\mathrm{d}s \\ & + \int_0^h \exp(sA)\,\zeta(t-s)\,\mathrm{d}s \end{aligned} \tag{15.2}$$

We now define the time integrals

$$Y(t) = \int_{t-h}^{t} y(\tau)\,\mathrm{d}\tau \quad \text{and} \quad Z(t) = \int_{t-h}^{t} z(\tau)\,\mathrm{d}\tau \tag{15.3}$$

and integrating (15.2) we obtain

$$\begin{aligned} Y(t) = {} & \exp(hA)\,Y(t-h) \\ & + \int_0^h \exp(sA)\,BZ(t-s)\,\mathrm{d}s \\ & + \int_{t-h}^{t} \int_0^h \exp(sA)\,\zeta(\tau-s)\,\mathrm{d}s\,\mathrm{d}\tau \end{aligned} \tag{15.4}$$

If we use the notation $Y_t = Y(th)$ for integral t we can write equation (15.4) as

$$Y_t = \exp(hA)\,Y_{t-1} + \int_0^h \exp(sA)\,BZ(th-s) + \eta_t \tag{15.5}$$

where

$$\eta_t = h \int_{t-1}^{t} \int_0^h \exp(sA)\,\zeta(\tau h - s)\,\mathrm{d}s\,\mathrm{d}\tau$$

The disturbance η_t in equation (15.5) clearly has mean zero. To investigate the autocovariance properties of η_t we first define

$$\xi(\tau) = \int_0^h \exp(sA)\, \zeta(\tau h - s)\, ds$$

and note that

$$E[\xi(\tau)\, \xi(\rho)'] = \exp(\tau hA) \int_{h\max(\tau-1,\, \rho-1)}^{h\min(\tau,\rho)} \exp(-sA)\, \Sigma \exp(-sA') \, ds \exp(\rho hA') \tag{15.6}$$

Since

$$\eta_t = h \int_{t-1}^{t} \xi(\tau)\, d\tau$$

it follows from equation (15.6) that

$$E(\eta_t\, \eta_t') = h^2 \int_{t-1}^{t} \int_{t-1}^{t} \exp(\tau hA) \int_{h\max(\tau-1,\, \rho-1)}^{h\min(\tau,\rho)} \exp(-sA)\, \Sigma \exp(-sA') \, ds \exp(\rho hA')\, d\tau\, d\rho$$

which we may expand as

$$h^2 \int_{t-1}^{t} \int_{t-1}^{\tau} \exp(\tau hA) \int_{\tau h-h}^{\rho h} \exp(-sA)\, \Sigma \exp(-sA)\, ds \exp(\rho hA')\, d\rho\, d\tau$$
$$+ h^2 \int_{t-1}^{t} \int_{t-1}^{\rho} \exp(\tau hA) \int_{\rho h-h}^{\tau h} \exp(-sA)\, \Sigma \exp(-sA')\, ds \exp(\rho hA')\, d\tau\, d\rho \tag{15.7}$$

By carrying out a transformation of variables in the triple integral (15.7) we can show that (15.7) is equivalent to

$$h^2 \int_0^1 \int_0^p \exp(-qhA) \int_{hp}^{hq+h} \exp(sA)\, \Sigma \exp(sA')\, ds \exp(-phA')\, dq\, dp$$
$$+ h^2 \int_0^1 \int_0^q \exp(-qhA) \int_{hq}^{hp+h} \exp(sA)\, \Sigma \exp(sA')\, ds \exp(-phA')\, dq\, dp \tag{15.8}$$

which is independent of t.

We find also that

$$E(\eta_t\, \eta_{t-1}') = h^2 \int_{t-1}^{t} \int_{t-1}^{\rho} \exp(\tau hA) \int_{\tau h-h}^{\rho h-h} \exp(-sA)\, \Sigma \exp(-sA')\, ds \exp[(\rho h - h)A']\, d\tau\, d\rho$$

which, after a transformation of variables, becomes

$$h^2 \int_0^1 \int_0^q \exp(-qhA) \int_{ph+h}^{qh+h} \exp(sA)\, \Sigma \exp(sA')\, ds \exp[-(ph+h)A']\, dp\, dq \tag{15.9}$$

Under our assumptions, $E(\eta_t \eta'_{t-r})$ is a zero matrix for $r > 1$ so that the covariance generating function of η_t is

$$\Psi(z) = \sum_{\tau=-1}^{1} \phi_\eta(\tau) z^{-\tau} \tag{15.10}$$

where $\phi_\eta(\tau) = E(\eta_t \eta'_{t+\tau})$ and $\phi_\eta(\tau) = \phi_\eta(-\tau)'$. Since the determinant of $\psi(z)$ is not identically zero, we may factorize $\psi(z)$ uniquely as[7]

$$\psi(z) = G(z)\,\Omega G(z^{-1})' \tag{15.11}$$

where $G(z) = I + Fz$ and the zeros of $\det[G(z)]$ are on or outside the unit circle. The matrix Ω is positive definite and equation (15.11) implies the representation of η_t as the moving average

$$\eta_t = \epsilon_t + F\epsilon_{t-1} \tag{15.12}$$

where ϵ_t is a pure noise process in discrete time with $E(\epsilon_t \epsilon_t') = \Omega$. Equating coefficients of like powers of z in equations (15.10) and (15.11) we obtain the system

$$\phi_\eta(0) = \Omega + F\,\Omega\,F'$$
$$\phi_\eta(1) = \Omega\,F'$$

which relates the parameters of the moving average in equation (15.12) to the covariance structure of η_t.

In fact, if we assume that the roots of $\det[\psi(z)] = 0$ are all distinct and that no roots lie on the unit circle,[8] then it is a simple matter to construct the matrix F. We first let z_i, $i = 1, \ldots, n$, denote the n roots of $\det[\psi(z)] = 0$ which lie outside the unit circle and define the matrix

$$H = \operatorname{diag}\ \frac{1}{z_1}, \ldots, \frac{1}{z_n} \tag{15.13}$$

We then compute the adjugate matrices adj $[\psi(z_i)]$ at each of these roots. By assumption, adj $[\psi(z_i)]$ has unit rank and can therefore be written as the product

$$\operatorname{adj}\,[\psi(z_i)] = c_i r_i' \tag{15.14}$$

where c_i and r_i are non-zero $n \times 1$ vectors. An equation similar to (15.14) holds for each i and, if we let R be the matrix formed from the n rows $(r_i : i = 1, \ldots, n)$, the coefficient matrix F in equation (15.11) can then be derived from[9]

$$F = -R^{-1} HR \tag{15.15}$$

The implication of the present section is that if we wish to estimate the continuous system (15.1) with observable data such as (15.3) we can use the mixed autoregressive moving average model (with exogenous inputs) that is given by

$$Y_t = \exp(hA)\,Y_{t-1} + \int_0^h \exp(sA)\,BZ(th - s)\,ds + \epsilon_t + F\epsilon_{t-1} \tag{15.16}$$

The treatment of the exogenous variable component in equation (15.16) has been discussed earlier in Phillips (1974a). In principle, (15.16) can be handled by a number of different methods, but if the system is of a moderate size (ten to fifteen equations perhaps) and the moving average coefficient matrix F is taken to be unknown, then the application of standard maximum likelihood methods[10] to equation (15.16) will involve a very complicated non-linear regression. Moreover, since F is uniquely determined by A and Σ, the estimates obtained in this way will be inefficient to the extent that the procedure ignores the restrictions on F implied by equations (15.8) to (15.11) as well as the *a priori* restrictions on A from the formulation of the model. In the following section, therefore, we consider a model which approximates equation (15.16) and involves a moving average disturbance which we obtain by truncating the expansion of F in powers of h. The model can then be estimated by an iterative procedure in which we alternately estimate the pair $(\exp(hA), \Omega)$ and use this pair of estimates to revise our estimate of F.

3 An Approximate Moving Average and Its Misspecification Bias

Using the power series expansion of $\exp(A)$ in the integrals (15.8) and (15.9) which define $[\phi_\eta(\tau) : \tau = -1, 0]$, we find that

$$\phi_\eta(0) = \frac{2h^3 \Sigma}{3} + h^4 \frac{A\Sigma + \Sigma A'}{3} + 0(h^5) \tag{15.17}$$

$$\phi_\eta(-1) = \frac{\Sigma h^3}{6} + h^4 \left(\frac{A\Sigma}{8} + \frac{\Sigma A'}{24} \right) + 0(h^5) \tag{15.18}$$

We can now approximate the generating function $\psi(z)$ by taking the first terms in the expansions (15.17) and (15.18). We define

$$\begin{aligned}\Psi^*(z) &= \frac{h^3 \Sigma z^{-1}}{6} + \frac{2h^3 \Sigma}{3} + \frac{h^3 \Sigma z}{6} \\ &= h^3 \Sigma \frac{z^{-1} + 4 + z}{6}\end{aligned}$$

which we can write in the factorized form

$$\Psi^*(z) = (1 + \alpha z) \frac{h^3 \Sigma}{6\alpha} (1 + \alpha z^{-1}) \tag{15.19}$$

where $\alpha = 0.268$. The representation (15.19) leads us to consider as an approximation to (15.12) the simple moving average

$$\epsilon_t + \alpha\epsilon_{t-1} \qquad (\alpha = 0.268) \tag{15.20}$$

The model (15.16) can, therefore, be approximated by a model in which the moving average $\epsilon_t + F\epsilon_{t-1}$ is replaced by (15.20). Since the operator $1 + \alpha z$ has an inverse which can be expanded as a stable series of non-negative powers of z, this approximate model can be premultiplied by the operator $(1 + \alpha z)^{-1}$ so that the disturbance in the resulting model is pure noise. The approximate model then becomes

$$Y_t^* = \exp(hA)\, Y_{t-1}^* + \int_0^h \exp(sA)\, BZ^*(th - s)\, ds + \epsilon_t \tag{15.21}$$

where the asterisk denotes data transformations which are carried out before estimation[11] (for example $Y_t^* = \Sigma_{s=0}^{t}(-\alpha)^s Y_{t-s}$).

The approximate moving average (15.20) is the same as that suggested by Wymer (1972b) for the estimation of a second-order differential equation system from its discrete approximation. The reason for this equivalence is fairly clear: the discrete approximation of a second-order system can be derived by integrating the system over an interval of length h twice, giving a disturbance of the form $\int_{\tau-h}^{t} \int_{\tau-h}^{\tau} \zeta(s) ds\, d\tau$; the approximation we have just suggested in the case of flow variables is equivalent to replacing $\exp(sA)$ in the last term on the right-hand side of equation (15.4) by the identity matrix, giving a disturbance of the same form.

We now turn our attention to the specification error implicit in the use of (15.20) and establish the following result.

Theorem 15.1

The coefficient matrix F and the covariance matrix Ω in the moving average (15.12) *can be expanded in powers of h as*

$$F = \alpha I + \frac{\alpha h}{4}(A - \Sigma A' \Sigma^{-1}) + 0(h^2)$$

and

$$\Omega = \frac{h^3}{6\alpha}\left[\Sigma + \frac{h}{2}(A\Sigma + \Sigma A')\right] + 0(h^5)$$

Proof

We define

$$P(z) = \frac{z\Psi(z)}{h^3} \tag{15.22}$$

and, using the expansions (15.17) and (15.18) in the generating function $\psi(z)$ on the right-hand side of equation (15.22), we find that

$$P(z) = \frac{\Sigma}{6}(z^2 + 4z + 1) + \frac{hA\Sigma}{24}(3z^2 + 8z + 1) + \frac{h\Sigma A'}{24}(z^2 + 8z + 3) + 0(h^2) \tag{15.23}$$

The remainder after dividing $P(z)$ by the binomial $(1 + z/\alpha)I$ is, from the generalized Bézout theorem[12], just $P(-\alpha I)$. But $-\alpha$ is a root of the quadratic $z^2 + 4z + 1 = 0$ so that it is clear from equation (15.23) that this remainder reduces to

$$\frac{h}{24} A\,\Sigma(\alpha^2 - 8\alpha + 3) + \frac{h}{24} \Sigma A'(3\alpha^2 - 8\alpha + 1) + 0(h^2) \qquad (15.24)$$

Thus the remainder when we divide $P(z)$ by the binomial $(1 + z/\alpha)I$ is of $0(h)$.

Since $(1 + z/\alpha)I$ is a regular polynomial matrix, there exists a unique polynomial $Q(z)$ such that

$$P(z) = Q(z)\left(1 + \frac{z}{\alpha}\right)I + S \qquad (15.25)$$

where the remainder S is given by (15.24). We know also, from the unique factorization (15.11), that $P(z)$ can be expressed as

$$P(z) = \frac{(I + Fz)\Omega(zI + F')}{h^3}$$

so that $zI + F'$ is a right divisor of $P(z)$. It follows that $P(-F') = 0$ and, substituting in equation (15.25), we obtain

$$Q(-F')\left(I - \frac{F'}{\alpha}\right) + S = 0 \qquad (15.26)$$

In view of (15.24), we can rewrite equation (15.26) as

$$\frac{(F - \alpha I)Q(-F')'}{\alpha} = 0(h) \qquad (15.27)$$

and, returning to equation (15.23), we see that

$$Q(z) = \frac{\alpha\Sigma}{6}\left(z + \frac{1}{\alpha}\right) + 0(h)$$

Therefore,

$$Q(-F') = \frac{-\alpha\Sigma}{6}\left[F' - \frac{1}{\alpha} I\right] + 0(h) \qquad (15.28)$$

Using the canonical form given by (15.15), we write the right-hand side of equation (15.28) as

$$\frac{\alpha}{6}\Sigma R'\left[H + \frac{1}{\alpha} I\right]R'^{-1} + 0(h)$$

Since $|\alpha| < 1$ it follows from equation (15.13) that $H + (1/\alpha)I$ is non-singular and of $0(1)$ as h tends to zero. Hence, $Q(-F')$ is of $0(1)$ and tends to a non-singular matrix as h tends to zero. We then deduce from equation (15.27) that

$$F = \alpha I + 0(h) \qquad (15.29)$$

We now calculate the quotient $Q(z)$ in equation (15.25) by long division.[13] We obtain

$$Q(z) = \alpha\left(\frac{\Sigma}{6} + \frac{hA\Sigma}{8} + \frac{h\Sigma A'}{24}\right)z + \alpha\left(\frac{2\Sigma}{3} + \frac{hA\Sigma}{3} + \frac{h\Sigma A'}{3}\right)$$
$$- \alpha^2\left(\frac{\Sigma}{6} + \frac{hA\Sigma}{8} + \frac{h\Sigma A'}{24}\right) + 0(h^2) \quad (15.30)$$

Since $\alpha^2 - 4\alpha + 1 = 0$, we have $\alpha^2 - 8\alpha + 3 = 2 - 4\alpha$ and $3\alpha^2 - 8\alpha + 1 = 4\alpha - 2$. Thus, from (15.24) we find that

$$S = \frac{2\alpha - 1}{12} h(A\Sigma - \Sigma A') + 0(h^2) \quad (15.31)$$

Moreover, from equation (15.30) we have

$$Q(-F') = -\alpha\left(\frac{\Sigma}{6} + \frac{hA\Sigma}{8} + \frac{h\Sigma A'}{24}\right)F' + \alpha\left(\frac{2\Sigma}{3} + \frac{hA\Sigma}{3} + \frac{h\Sigma A'}{3}\right)$$
$$- \alpha^2\left(\frac{\Sigma}{6} + \frac{hA\Sigma}{8} + \frac{h\Sigma A'}{24}\right) + 0(h^2)$$

so that using equation (15.29) we find that

$$Q(-F') = \frac{1 - 2\alpha}{3}\Sigma + 0(h) \quad (15.32)$$

Thus, it follows from equations (15.27), (15.31), and (15.32) that

$$F = \alpha I + \alpha S'[Q(-F')']^{-1}$$
$$= \alpha I + \alpha\left[\frac{2\alpha - 1}{12}h(A\Sigma - \Sigma A')' + 0(h^2)\right]\left[\frac{3}{1 - 2\alpha}\Sigma^{-1} + 0(h)\right]$$
$$= \alpha I + \frac{\alpha(-h)}{4}(\Sigma A'\Sigma^{-1} - A) + 0(h^2)$$
$$= \alpha I + \frac{\alpha h}{4}(A - \Sigma A'\Sigma^{-1}) + 0(h^2)$$

which establishes the first part of the theorem.

Since $\phi_\eta(-1) = F\Omega$ we obtain from (15.18) the equation

$$F\Omega = \frac{h^3\Sigma}{6} + h^4\left(\frac{A\Sigma}{8} + \frac{\Sigma A'}{24}\right) + 0(h^5)$$

and since

$$F^{-1} = \frac{1}{\alpha}I - \frac{h}{4\alpha}(A - \Sigma A'\Sigma^{-1}) + 0(h^2)$$

we have

$$\Omega = \left[\frac{1}{\alpha}I - \frac{h}{4\alpha}(A - \Sigma A'\Sigma^{-1}) + 0(h^2)\right]\frac{h^3\Sigma}{6} + h^4\left(\frac{A\Sigma}{8} + \frac{\Sigma A'}{24}\right) + 0(h^5)$$

$$= \frac{h^3}{6\alpha}\left[\Sigma + \frac{h}{2}(A\Sigma + \Sigma A') + 0(h^5)\right]$$

End of Proof

Corollary

The coefficient matrix F in the moving average (15.12) *can alternatively be expanded as*

$$F = \alpha I + \frac{\alpha h}{4}(A - \Omega A'\Omega^{-1}) + 0(h^2)$$

and

$$F = \alpha I + \frac{\alpha}{4}[\exp(hA) - \Omega\exp(hA')\Omega^{-1}] + 0(h^2)$$

Proof Since $\Omega = h^3\Sigma/6\alpha + 0(h^4)$ we have $\Omega^{-1} = 6\alpha\Sigma^{-1}/h^3 + o(h^{-2})$ and the first alternative expansion follows. We also have

$$\exp(hA) - \Omega\exp(hA')\Omega^{-1} = I + hA - \Omega(I + hA')\Omega^{-1} + 0(h^2)$$
$$= h(A - \Omega A'\Omega^{-1}) + 0(h^2)$$

and this gives us the second alternative expansion.

End of Proof

According to Theorem 15.1 the misspecification bias involved in the use of the approximate moving average (15.20) is of $0(h)$ and therefore the bias disappears as $h \to 0$. But, in addition, by giving us the term of $0(h)$ in the expansion of F, Theorem 15.1 and its Corollary enable us to use an iterative procedure to estimate the model (15.16). In the next section we will develop this procedure and discuss the properties of the resulting estimators.

4 Estimation of the Flow Data Model

In this section we will confine our attention to the closed model in which the coefficient matrix B in (15.1) is null. Our discussion could be cast in terms of the more general model but would involve us in the complications that have been sorted out elsewhere (Phillips, 1974a, and Sargan, 1976). It seems unnecessary to repeat that work here and we will only indicate generalizations where appropriate.

When there are no exogenous variables the discrete system (15.16) becomes

$$Y_t = \exp(hA)Y_{t-1} + \epsilon_t + F\epsilon_{t-1} \tag{15.33}$$

It might be thought that a problem of identification arises in this model since the disturbances are not serially independent. But the coefficient matrices are identifiable in this model because the matrix polynomials $I - \exp(hA)z$ and $I + Fz$ have a greatest common left divisor which is the identity matrix and the matrix $\exp(hA)$ is non-singular. The latter ensures that the null spaces of $\exp(hA')$ and F' have null intersection and thus we can appeal to the theorem established by Hannan (1969).

Defining $Y' = [Y_1, \ldots, Y_T]$, $Y'_{-1} = [Y_0, \ldots, Y_{T-1}]$, $E' = [\epsilon_1, \ldots, \epsilon_T]$, and $E'_{-1} = [\epsilon_0, \ldots, \epsilon_{T-1}]$ we can write the complete system of equations (15.33) as

$$Y' = \exp(hA)Y'_{-1} + E' + FE'_{-1} \tag{15.34}$$

For estimation by maximum likelihood methods it is often found convenient in models such as (15.33) to treat ϵ_0 as a fixed vector.[14] We do this and set $\mu = F\epsilon_0$, writing equation (15.34) now as

$$Y' = BZ' + V' \tag{15.35}$$

where $B = [\exp(hA) \,\vdots\, \mu]$

$$Z' = \begin{bmatrix} Y_0 & Y_1 \ldots Y_{T-1} \\ 1 & 0 \ldots 0 \end{bmatrix}$$

and $V' = [\epsilon_1,\ \epsilon_2 + F\epsilon_1, \ldots, \epsilon_T + F\epsilon_{T-1}]$. Stacking the columns of equation (15.35) we have

$$y = W\beta + M\epsilon \tag{15.36}$$

where[15] $y = \text{vec}(Y')$, $W = Z \otimes I$, $\beta = \text{vec}(B)$, $\epsilon = \text{vec}(E')$, and

$$M = \begin{bmatrix} I & 0 & 0 \ldots 0 & 0 \\ F & I & 0 \ldots 0 & 0 \\ 0 & F & I \ldots 0 & 0 \\ \vdots & \vdots & \vdots \cdots \vdots & \vdots \\ 0 & 0 & 0 \ldots F & I \end{bmatrix}$$

Now

$$M^{-1} = \begin{bmatrix} I & 0 & 0 & \ldots 0 \\ -F & I & 0 & \ldots 0 \\ \vdots & \vdots & \vdots & \ldots \vdots \\ (-F)^{T-1} & (-F)^{T-2} & & \ldots I \end{bmatrix} = \begin{bmatrix} M'_1 \\ M'_2 \\ \vdots \\ M'_T \end{bmatrix}$$

and premultiplying equation (15.36) by M^{-1} we obtain

$$M'_t y = M'_t WB + \epsilon_t \qquad (t = 1, \ldots, T) \tag{15.37}$$

Thus, assuming that the vectors ϵ_t are normally distributed, we see from equation (15.37) that the joint frequency function of $y_1, \ldots, y_T$ conditional on the fixed vector ϵ_0 (and hence y_0) is given by

$$L = \frac{1}{(2\pi)^{nT/2} |\det M| (\det \Omega)^{T/2}} \exp\{-\tfrac{1}{2}(y - W\beta)' M'^{-1} (I \otimes \Omega^{-1}) M^{-1} (y - W\beta)\}$$

For a given sample data, L is a function of β and Ω. But since the elements of M depend on F which in turn are determined by A and Σ, it is clear the L depends on β indirectly through the matrix M. This causes great computational difficulties. One approach would be to replace F in the definition of M by the approximation

$$F^+ = \alpha I + \frac{\alpha h}{4}(A - \Omega A' \Omega^{-1})$$

suggested by Theorem 15.1. Even in this case, however, a preliminary concentration of L with respect to the unknown elements of A causes problems because of the dependence of F on Ω, and the computational burden of this approach is still considerable. In the remainder of this section and in Section 5, therefore, we will consider two alternatives to maximum likelihood.

The first is based on an iterative procedure and uses an approximate model derived from equation (15.34) by replacing F with F^+. We find it convenient to write equation (15.36) in the form

$$y = W\beta + M\epsilon + \psi_0 \tag{15.38}$$

where y, M, and ϵ are as in equation (15.36) but now $W = Y_{-1} \otimes I$, $\beta = \text{vec}[\exp(hA)]$ and $\psi_0{}' = [\epsilon_0' F' \vdots 0, \ldots, 0]$. Thus

$$M_t' y = M_t' W\beta + \epsilon_t + M_t' \psi_0 \tag{15.39}$$

The difference between equations (15.39) and (15.37) is that $F\epsilon_0$ is no longer incorporated in the parameter vector β and for estimation purposes we simply neglect the term $M_t'\psi_0$ in (15.39). As the sample size increases, it is clear that this term diminishes in importance and it will not affect our asymptotic theory.

The iteration we suggest is now as follows:

(a) Use $F^{(1)} = \alpha I$ as a first approximation to F and construct $M^{(1)}$ in the same way as M with F replaced by $F^{(1)}$. Writing $M^{(1)'-1} = [M_1^{(1)} \vdots \ldots \vdots M_T^{(1)}]$ we can then obtain the preliminary estimates of β and Ω given by

$$\beta^{(1)} = \left[\frac{1}{T} \sum_{t=1}^{T} W' M_t^{(1)} M_t^{(1)'} W\right]^{-1} \left[\frac{1}{T} \sum_{t=1}^{T} W' M_t^{(1)} M_t^{(1)'} y\right]$$

and

$$\Omega^{(1)} = \frac{1}{T} \sum_{t=1}^{T} M_t^{(1)'} [y - W\beta^{(1)}] [y - W\beta^{(1)}]' M_t^{(1)}$$

(b) Use $\beta^{(1)}$ and $\Omega^{(1)}$ to obtain a new estimate of F

$$F^{(2)} = \alpha I + \tfrac{\alpha}{4} [C^{(1)} - \Omega^{(1)} C^{(1)'} \Omega^{(1)-1}]$$

where $C^{(1)}$ is an estimate of $\exp(hA)$ constructed from $\beta^{(1)}$. From $F^{(2)}$ we derive a new estimate $M^{(2)}$ of M and then estimate A by minimizing with respect to the unknown elements of A the quadratic form

$$\frac{1}{T}\sum_{t=1}^{T}(y - W\beta)'M_t^{(2)}\Omega^{(1)-1}M_t^{(2)\prime}(y - W\beta)$$

$$= \frac{1}{T}(y - W\beta)'M^{(2)-1}[I \otimes \Omega^{(1)-1}]M^{(2)-1}(y - W\beta)$$

We denote the resulting estimate of A by $\hat{A}$

Remark 1

The iteration can be continued by revising the estimates of F and Ω after (b) and then proceeding as in (b) to reestimate A. Since the estimates of A in (b) satisfy the *a priori* restrictions there may be some advantage in such a procedure at least in small samples; but further iterations will not improve on the asymptotic bias of A.

Remark 2

The procedure that leads to $\beta^{(1)}$ in (a) is based on a regression with transformed data. For instance, $M_t^{(1)\prime}y = \Sigma_{s=1}^{t}(-\alpha)^{t-s}Y_s$ and $M_t^{(1)\prime}W$ involves a similar transformation of the rows of Y_{-1}. Thus, the regression in (a) closely resembles that involving the approximate model (15.21) discussed earlier. In fact, if we neglect the exogenous variable component in (15.21) the methods differ only in terms of end corrections (Y_T^* involves Y_0 while $M_t^{(1)\prime}y$ does not). This difference is clearly unimportant in an asymptotic theory,[16] and, in particular, the estimates of $\exp(hA)$ from the two models will have the same asymptotic bias.

Remark 3

$\beta^{(1)}$ is an unrestricted estimator in the sense that we do not take account of prior restrictions on A in the calculation of $\beta^{(1)}$. It would be possible to do this, of course, but such a step would increase the computational burden of the procedure.

An asymptotic theory can now be developed for the estimators derived at each stage in the iteration above. Although the proofs are rather lengthy (see Phillips, 1974b), we can, in particular, show that the asymptotic bias of $\beta^{(1)}$ as $T \to \infty$ is of $o(h)$ as $h \to 0$. Moreover, under a condition which ensures the identifiability of the unknown elements of A,[17] we can verify that the asymptotic bias of A as $T \to \infty$ is of $o(1)$ as $h \to 0$. Thus, the asymptotic bias of the structural estimates obtained from the iteration tends to zero with the sampling interval.[18]

5 Estimation by Instrumental Variables

If the sampling interval is not small then the specification error implicit in the use of the approximations $F^{(1)}$ and $F^{(2)}$ may not be inconsiderable. We now consider, therefore, an alternative method of estimation in which the moving average disturbance (15.12) is not approximated. This method requires more computational effort than the method described in the last section but is still much simpler from the point of view of the computations involved than the direct application of maximum likelihood methods to (15.16).

The idea behind the method is to use Y_{t-2} as an instrument for Y_{t-1} in the model (15.16). In the case where no exogenous variables occur, this procedure leads us to the estimating equations $\Sigma_{t=2}^{T} Y_t Y'_{t-2} = B\ \Sigma_{t=2}^{K} Y_{t-1} Y'_{t-2}$, from which we obtain, under the given assumptions, a consistent estimator B of $\exp(hA)$. The residuals $(v_t : t = 1, \ldots, T)$ from this regression are then used to construct the moment matrices $V_0 = (1/T)\ \Sigma_{t=2}^{T} v_t v'_t$ and $V_1 = (1/T)\ |\Sigma_t^{T-1} v_t v'_{t+1}$, which are consistent estimators of $\phi_\eta(0)$ and $\phi_\eta(1)$. Hence, we obtain the following estimate of the spectral density matrix of $\eta_t = \epsilon_t + F\epsilon_{t-1}$:

$$f^*_{\eta\eta}(\lambda) = \frac{1}{2\pi}(V'_1 e^{i\lambda} + V_0 + V_1 e^{-i\lambda}) \tag{15.40}$$

Assuming that $f^*_{\eta\eta}(\lambda)$ is positive definite, we may determine a unique estimate F^* of F from the factorization

$$f^*_{\eta\eta}(\lambda) = \frac{1}{2\pi}(I + F^* e^{i\lambda})\Omega^*(I + F^{*\prime} e^{-i\lambda}) \tag{15.41}$$

where the zeros of $\det(I + F^* z)$ are outside the unit circle.

Since $f^*_{\eta\eta}(\lambda)$ is a consistent estimator of the spectral density matrix of η_t, F^* converges in probability to F. However, F^* is likely to be a rather inefficient[19] estimator of F for several reasons. In the first place, if there are prior restrictions on the matrix A (which we can expect to be the usual situation in practice) then the residual moment matrices V_0 and V_1 may not be very good estimates of $\phi_\eta(0)$ and $\phi_\eta(1)$. Second, Y_{t-2} may not be a good instrument for Y_{t-1} if the sampling interval is large, so that even if there are no restrictions on A the moment matrices V_0 and V_1 may be rather inefficient estimators of $\phi_\eta(0)$ and $\phi_\eta(1)$. Finally, it is well known that the use of (15.40) and (15.41) to estimate the coefficients of a moving average may be an inefficient procedure.[20]

Nevertheless, when no exogenous variables enter the model, F^* is a consistent estimator of F so that, for large samples anyway, we would prefer F^* to the approximations $F^{(1)}$ and $F^{(2)}$; even in the case where exogenous variables are present F^* may provide a better estimate of F than these approximations. Once F^* has been calculated, we can proceed directly to stage (b) of the iteration described in the last section. We then estimate A by minimizing

$$\frac{1}{T}(y - W\beta)' M^{*\prime -1}(I \otimes \Omega^{*-1}) M^{*-1}(y - W\beta)$$

where M^* is constructed from F^* in the same way as M is from F, and Ω^* is given in the factorization (15.41). Since F^* and Ω^* are consistent estimates of F and Ω it follows that the resulting estimate of A is also consistent.

There may be one practical obstacle to the application of this method. This is the possibility that the matrix $f^*_{\eta\eta}(\lambda)$ may not be positive semidefinite for all λ and consequently will not admit the factorization (15.41). This situation is most likely to occur when the sample size is small. In this case, we can expect $f^*_{\eta\eta}(\lambda)$ (and hence F^* if $f^*_{\eta\eta}(\lambda)$ does factorize) to be a rather poor estimate and, thus, the iteration discussed in the last section is likely to be most useful when the data series are short.

6 Forecasting with a Flow Data Model

The presence of the moving average in equation (15.16) complicates somewhat the problem of forecasting future values of the endogenous variables. To forecast p periods ahead we would, in an instantaneous variable model, use the predictor[21]

$$Y^*_{T+p} = \exp(phA)Y_T + \int_0^{ph} \exp(sA)BZ(Th + ph - s)\mathrm{d}s \tag{15.42}$$

given the discrete observations $(Y_t : t = 1, \ldots, T)$ and the continuous time record $(Z(s) : Th \leq s \leq Th + ph]$. In the present case, however, where Y and Z are time integrals, equation (15.42) is no longer an optimum predictor since

$$Y_{T+p} - Y^*_{T+p} = h\int_{T-1}^{T} \int_0^{ph} \exp(sA)\zeta(\tau h + ph - s)\mathrm{d}s\,\mathrm{d}\tau$$

is correlated with Y^*_{T+p}.

Nevertheless, the theory of prediction for models similar in form to equation (15.16) is well developed.[22] We let M denote the Hilbert space spanned by the elements $[Y_t : -\infty < t < \infty]$ and M_t the closed linear manifold generated by the elements of $[Y_t : t \leq T]$. Then, when no exogenous variables occur in the model, the best linear predictor of Y_{T+p}. is given by the orthogonal projection of Y_{T+p} on M_T. If we denote this projection by $\hat{Y}_{T+p}$, then $\hat{Y}_{T+p}$ is known[23] to satisfy the equations

$$\hat{Y}_{T+p} = \exp(hA)\,\hat{Y}_{T+p-1} \qquad (p > 1) \tag{15.43}$$

$$\hat{Y}_{T+1} = \exp(hA)\,Y_T + F(Y_T - \hat{Y}_T) \tag{15.44}$$

where $\hat{Y}_T$ is the projection of Y_T on M_{T-1}. To find the optimum predictor $\hat{Y}_{T+1}$ we need the initial conditions $[\hat{Y}_T, \hat{Y}_{T-1}, \ldots]$ and through back-substitution in equation (15.44) we obtain

$$\begin{aligned}\hat{Y}_{T+1} = {} & [\exp(hA) + F]\; Y_T - F\,[\exp(hA) + F]\; Y_{T-1} \\ & + \ldots + (-1)^j F^j [\exp(hA) + F]\; Y_{T-j} \ldots\end{aligned} \tag{15.45}$$

For practical purposes we can truncate equation (15.45) after a finite number of terms[24] since, from equation (15.15) $F^j = (-1)^j R^{-1} H^j R$ and we can expect the diagonal elements of H^j to be small for moderate j if the sampling interval h is not too large. Once $\hat{Y}_{T+1}$ has been calculated we can use equation (15.43) interatively to obtain predictions for lead time ph where $p > 1$.

When exogenous variables are present in the model the formulae are similar. We must assume, however, that the exogenous variables are stochastically independent of the disturbances in equation (15.16). We first set

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

and then the optimum predictor of Y_{T+1} is given by

$$\hat{Y}_{T+1} = \{[\exp(hA) + F]\ Y_T - F\ [\exp(hA) + F]\ Y_{T-1} + F^2\ [\exp(hA) + F]\ Y_{T-2} - \ldots\} + [\phi_{T+1} - F\phi_T + F^2\phi_{T-1} - \ldots] \tag{15.46}$$

As before we truncate these expressions at the jth term when th[elements of F^j are small.

Of course, the optimum predictors (15.45) and (15.46) depend on the knowledge of the coefficient matrices A, F, and B as well as, in the case of (15.46), a continuous time record of the exogenous variables. Since this knowledge is not available in practice, we replace A, F, and B in equations (15.45) and (15.46) by their estimated values. The exogenous variable components $(\phi_t : t = T + 1, T, \ldots)$ in equation (15.46), which depend on continuous observations, can be replaced by the approximation given in Phillips (1974a) with the appropriate estimated coefficient matrices.

We should emphasize that the formulae (15.45) and (15.46) give us only point forecasts of the endogenous variables. We will, in general, be interested not only in obtaining point forecasts but also in estimating the forecast error covariance matrix.[25] However, when the expressions on the right-hand side of equations (15.45) and (15.46) have been truncated, this problem does not differ very greatly from the problem in conventional econometric models, and we will not discuss it in this paper.

7 Further Comments

Our discussion so far has been based on the assumption that all variables in the model (15.1), including exogenous variables, are flows and are observed in the form of time integrals. It has been convenient to make this assumption in order to develop the theory of Sections 2 to 6. Many practical models, however, can be expected to involve both flow variables and instantaneous variables. In these mixed variable models further approximations will usually be needed. One way of treating such models by the methods of this paper is to work with the system (15.16) and replace Y_{it} by $(h/2)\ [y_i(th) + y_i(th - h)]$ if the ith endogenous variable is a stock and the observable quantities are $[y_t(th) : t = 0, 1, \ldots]$ rather than the time integrals $[Y_{it} : t = 1, 2, \ldots]$. The same approximation can be used for the exogenous variables where necessary. We can expect the effects of such an approximation to be similar to the effects of replacing a differential equation system by a discrete approximation. The specification error induced by the latter[26] is of $0(h^2)$, so that it might be thought that in mixed variable models the additional approximation we have suggested is of less importance asymptotically as $h \to 0$ than the effect of the moving average approximations. However, a rigorous analysis is needed and is best reserved for a later paper.

Our discussion of the flow data problem in this paper has concentrated on the use of the exact discrete model. This is due to the fact that the problems posed by the presence of flow variables needed some systematic investigation in this context, especially since the exact model is currently being used for estimation purposes in empirical research.[27] But since a number of other methods are available for the estimation of continuous systems, we may well ask how the presence of flow data affects these other methods. As mentioned earlier in this paper, the treatment of flow data in the construction of the discrete approximation is discussed elsewhere (Phillips, 1974b, and Wymer, 1971) and it is known, in particular, that the asymptotic bias of estimators from the discrete approximation in a flow data model does not disappear as the sampling interval tends to zero. Thus, on the basis of asymptotic theory anyway, the procedures developed in this paper appear preferable.

The other main approach to the estimation of a continuous system is the interesting spectral method developed recently by Robinson (1975, 1976). In its most general form the method is designed for a model of continuously distributed lags, and if the model is transformed by the use of a linear filter such as (15.3) then the method can be applied to the transformed model and the resulting estimates will have good properties since the residuals in the transformed model are still stationary. Moreover, filters such as (15.3) concentrated the spectral mass of the exogenous series in the Nyquist frequency and, therefore, make the aliasing condition[28] in Robinson (1975 and 1976) more plausible. On the other hand, in mixed variable models the situation is more complicated and, as with the exact discrete model, warrants further investigation.

Notes

1. See Phillips (1974a, 1974c), Robinson (1973a, 1973b), Sargan (1976), and Wymer (1972a and 1972b).
2. Compare Wymer (1971).
3. A useful introductory discussion of stocks and flows is given in Allen (1968).
4. A complete discussion of this aspect of the problem is given in Phillips (1974b).
5. The mathematical problems associated with this interpretation when $\zeta(t)$ is a pure noise vector are discussed elsewhere (Doob, 1953; Phillips, 1974; Wymer, 1972b).
6. See Robinson (1975, 1976).
7. See Hannan (1970, p. 66).
8. This latter assumption is not strictly necessary but we make it now so that the operator $G(z)$ has an inverse which can be expanded as a stable realizable function.
9. For a similar derivation see Robinson (1967).
10. See Phillips (1966).

11. The exogenous variable component in (15.21) must also be approximated, in general, before estimation. We can use the procedure given in Phillips (1974a).
12. See Gantmacher (1959, Chapter IV, Section 3).
13. See Gantmacher (1959, p. 78).
14. Compare Phillips (1966).
15. Our definition of the vec operation is as follows. If $A = (a_1, \ldots, a_m)_{n \times m}$ then $\text{vec}(A) = (a_1', a_2', \ldots, a_m')'$.
16. Compare Anderson (1971, Section 5.5.2).
17. Compare Phillips (1973).
18. A detailed derivation of the results referred to here is given in Phillips (1974b).
19. By efficient we really mean asymptotically efficient in the usual sense (see, for instance, Hannan, 1970, Chapter VI, Section 5) so that an inefficient estimator is one whose asymptotic generalized variance is greater than that of the efficient estimator.
20. See Hannan (1970, p. 373).
21. Compare Sargan (1976).
22. See Doob (1953, Chapter XII) and Hannan (1970, Chapter III).
23. Compare Hannan (1970, p. 136).
24. Compare Box and Jenkins (1970, p. 130).
25. Given the form of the present model we have no guarantee of the existence of such a matrix, but the recent results of Sargan (1973) for models without lagged variables are suggestive.
26. See Sargan (1974).
27. See, in particular, Bergstrom and Wymer (1976).
28. This condition requires that the spectral mass of the exogenous series is zero outside the Nyquist frequency and is discussed fully in Robinson, (1976).

Bibliography

Anderson, T. W. (1971). *The Statistical Analysis of Time Series*, Wiley, New York.

Allen, R. G. D. (1968). *Macroeconomic Theory*, Macmillan, London.

Bergstrom, A. R., and C. R. Wymer (1976). 'A model of disequilibrium neoclassical growth and its application to the United Kingdom'. In A. R. Bergstrom (Ed.), *Statistical Inference in Continuous Time Economic Models*, North-Holland, Amsterdam.

Box, G. E. P., and G. M. Jenkins (1970). *Time Series Analysis Forecasting and Control*, Holden-Day, San Francisco.

Doob, J. L. (1953). *Stochastic Processes*, Wiley, New York.

Gantmacher, F. R. (1959). *Theory of Matrices*, Chelsea, London.

Hannan, E. J. (1969). 'The identification of vector mixed autoregressive-moving average systems,' *Biometrika*, **56**(1), March.

Hannan, E. J. (1970). *Multiple Time Series*, Wiley, New York.

Koopmans, T. C. (1950). 'Models involving a continuous time variable'. In T. C. Koopmans (Ed.), *Statistical Inference in Dynamic Economic Models*, Wiley, New York.

Phillips, A. W. (1956). 'Some notes on the estimation of time-forms of reactions in interdependent dynamic systems', *Economica*, n.s., 23(90), May.

Phillips, A. W. (1966). *Estimation of Systems of Difference Equations with Moving Average Disturbances*, Paper presented at the Econometric Society Meeting, San Francisco, and now published as Chapter II of this volume.

Phillips, P. C. B. (1973). 'The problem of identification in finite parameter continuous time models', *Journal of Econometrics*, 1(4), December.

Phillips, P. C. B. (1974a). 'The estimation of some continuous time models', *Econometrica*, **42**(5), September.

Phillips, P. C. B. (1974b). *The Treatment of Flow Data in the Estimation of Continuous Time Systems,* Paper presented at the Econometric Society Meeting, Grenoble, and available as a Discussion Paper, University of Essex.

Phillips, P. C. B. (1974c). *Problems in the Estimation of Continuous Time Models,* unpublished Ph.D thesis, University of London.

Robinson, E. A. (1967). *Multichannel Time Series Analysis with Digital Computer Programs,* Holden-Day, New York.

Robinson, P. M. (1976). The estimation of linear differential equations with constant coefficients, *Econometrica,* 44(4), July.

Robinson, P. M. (1975). Continuous time regressions with discrete data, *Annals of Statics,* 3, pp. 688–697.

Sargan, J. D. (1973). *The Tails of the FIML Estimates of the Reduced Form Coefficients,* mimeographed, London School of Economics.

Sargan, J. D. (1974). 'Some discrete approximations to continuous time stochastic models', *Journal of the Royal Statistical Society,* B, 36, (1).

Sargan, J. D. (1976). 'Some discrete approximations to continuous time stochastic models.' In A. R. Bergstrom (Ed.), *Statistical Inference in Continuous Time Economic Models*, North-Holland, Amsterdam.

Wymer, C. R. (1971). *Estimation of General Linear Differential Equation Systems,* mimeographed, London School of Economics.

Wymer, C. R. (1972a). 'A continuous disequilibrium adjustment model of United Kingdom financial markets'. In A. A. Powell and R. A. Williams (Eds.), *Economic Studies of Macroeconomic and Monetary Relations,* North-Holland, Amsterdam.

Wymer, C. R. (1972b). 'Econometric estimation of stochastic differential equation systems', *Econometrica,* **40**(3), May.