Journal of Econometrics 1 (1973) 351–362.

THE PROBLEM OF IDENTIFICATION IN FINITE PARAMETER CONTINUOUS TIME MODELS*

P.C.B. PHILLIPS
University of Essex, Colchester, England

1. Introduction

One of the problems that arise in the estimation of the structural parameters of linear stochastic differential equations is that the function relating the parameters of the continuous time model to those of the corresponding discrete time model is not, in general, bijective. In a recent article Telser (1967) mentioned this point, but on the whole the problem has been neglected in most of the papers that have appeared on the subject of estimating continuous time models (Bartlett (1946), Phillips (1959), Durbin (1961)). The manifestation of the problem in the frequency domain is the aliasing phenomena, which is well known. If $f(\omega)$ is the spectral density matrix of the stationary random process $y(t)$, $-\infty < t < \infty$, then $f(\omega)$ cannot be uniquely determined from the folded spectrum

$$f^{(h)}(\omega) = f(\omega) + \sum_{n=1}^{\infty} \{f(\omega + 2n\pi/h) + f(\omega - 2n\pi/h)\}; \qquad |\omega| \leqq \pi/h, \quad h > 0,$$

in the absence of further information about $f(\omega)$ (such as $f(\omega) = 0$ for $|\omega| > \pi/h$). Since $f^{(h)}(\omega)$ is the spectrum of the sequence $\{y_t = y(th): t = \ldots, -1, 0, 1, \ldots\}$ this means that the observations $\{y_t : t = 1, \ldots, T\}$ taken at intervals of length h can be used to provide a consistent estimator of $f^{(h)}(\omega)$ but not $f(\omega)$. When the process $y(t)$ is generated by a system of linear stochastic differential equations the elements of $f(\omega)$ are rational functions which depend on the parameters of that system.

* I am very grateful to Professor J.D. Sargan and a referee for their helpful comments on an earlier version of this paper.

The fact that there is a whole class of different spectral density matrices $f(\omega)$ consistent with $f^{(h)}(\omega)$ accords with the fact that there is no bijection between the parameters of the continuous time model and its corresponding discrete time model.

To minimise the aliasing problem it has been suggested that h be taken sufficiently small so that the high frequency components outside $(-\pi/h, \pi/h)$ are of negligible importance, see Grenander and Rosenblatt (1957) ch. 1.9. Alternatively, it has been shown by Shapiro and Silverman (1960) that the spectral density of a continuous time process can be uniquely determined when the sampling is not periodic but 'additively random' and the characteristic function of the interval increments takes no value more than once over the real numbers. However, neither of these procedures will be of much use unless the sampling process can be controlled and in some fields of study like economics this is not usually the case. If we cannot control the sampling interval and proceed as if there were no high frequency components in the system then it seems that we will obtain biased estimators of both the spectral density and the parameters of the structural model, which may be of some importance if we are interested in forecasting a continuous time path of a variable.

Of course, the structural parameters of a stochastic differential equation system can always be estimated from an approximate discrete model and this is the usual procedure followed in practice. On this approach the reader may consult Bergstrom (1966 and 1967), Wymer (1972) and Sargan (1973). In the case where there are no restrictions on the structural parameters in the underlying continuous time model the non-recursive discrete approximation is just-identified. This last point is discussed by Bergstrom (1967) ch. 9.3. Although the problem of identification does not arise in the application of this method, the method does suffer from the specification error implicit in the approximate model, which means that estimators based on the discrete approximation are never consistent.

This is not to say that it is impossible to obtain consistent estimators of the parameters of a stochastic differential equation system when only discrete data sampled periodically are available. In an earlier paper by the author (1972) it was found that, at least in the special case of a first order three equation model, simple restrictions on the structural parameters can be sufficient to overcome the identification problem and then consistent estimators of the structural parameters in the differen-

tial equation system can be obtained from the exact discrete model by a non-linear regression.[1]

The object of the present paper is to consider whether parameter identification can be achieved by making use of a priori restrictions known from the specification of the model. Typical restrictions are similar to those in conventional econometric models where certain variables are known to be excluded from certain equations and the coefficients of different variables are related in some known way. Our discussion is in terms of a simple first order linear system[2] and we examine a set of conditions sufficient to ensure the identification of the structural parameters in this case, given certain linear restrictions.

2. Linear systems of equations

We shall consider the first order linear system

$$D\,y(t) = A\,y(t) + \zeta(t), \tag{1}$$

where $y(t)$ is an $n \times 1$ vector of random functions observable at discrete points in time, A is a parameter matrix whose elements are real numbers and the stochastic disturbance $\zeta(t)$ is taken to be pure noise. The covariance matrix of the integral of $\zeta(t)$ over a unit time interval is assumed to be the positive definite matrix Σ. We note in passing that (1) may be given a precise meaning by defining $\zeta(t)$ as a generalised (vector) random process, see Hannan (1970) pp. 23–26. The following additional assumption will be useful:

Assumption 1. The matrix A has distinct characteristic roots all of which have negative real parts. These roots do not differ by an integer multiple of $2\pi i/h$, where h is the time interval between observations of $y(t)$.

[1] Recently, Robinson (1973a and b) has suggested an interesting new approach which involves approximating the Fourier transform of the structural model by a model in which the Fourier transforms are replaced by normalised discrete Fourier transformations of the data. The approximate model is then estimated by a weighted non-linear regression. This method has the great advantage of treating cases where only very weak assumptions are made about the covariance structure of the disturbances. Its main limitations, however, are that the method is applicable only when exogenous variables occur in the system and that an aliasing condition is imposed so that the spectral density functions of the exogenous variables (at least) are essentially zero beyond the Nyquist frequency.

[2] A discussion of the higher order scalar model is given in Phillips (1973).

The discrete model which corresponds to (1) and which is satisfied by the observations $y_t = y(th)$ at the lattice points $\{th : t = 1, ..., T\}$ is given by

$$y_t = B y_{t-1} + \xi_t , \tag{2}$$

where $B = \exp(hA)$ and $\xi_t = \int_0^h \exp(sA)\, \zeta(th-s)\, ds$. The disturbance ξ_t in (2) is serially independent and under assumption 1 its covariance matrix

$$\Omega = \int_0^h \exp(sA)\, \Sigma \exp(sA')\, ds \tag{3}$$

is non-singular for positive h. From (3) we can deduce the alternative relationship

$$\exp(hA)\, \Sigma \exp(hA') - \Sigma = A\Omega + \Omega A',$$

which we write as

$$\operatorname{vec} \Omega = \{A \otimes I + I \otimes A\}^{-1} \{\exp(hA) \otimes \exp(hA) - I \otimes I\} \operatorname{vec} \Sigma, \tag{4}$$

where vec represents the vector formed by taking the direct sum of the rows of a matrix, and $\otimes$ denotes the right hand Kronecker product. From (4) it follows that Ω is unrestricted when there are no restrictions on Σ.

The difference eq. (2) is completely specified by the pair (B, Ω) and the distribution of $\int_{t-h}^{t} \zeta(s)\, ds$. Assuming that the distribution of $\int_{t-h}^{t} \zeta(s)\, ds$ is unknown except for its mean value which is zero, we may represent the structure of the model (1) by the pair (A, Σ). Clearly, (A, Σ) is identifiable in the model (2) if and only if the matrix A is identifiable in B. If the true structural coefficient matrix of (1) is A^0 we denote the true coefficient matrix of (2) by $B^0 = \exp(hA^0)$. However, when there are no restrictions on the matrix A, a countable infinity of systems such as (1) with different A matrices have the same discrete model (2) in which $B = B^0$, just as there are infinitely many solutions of the matrix equation

$$\exp(hA) = B^0, \tag{5}$$

see, for example, Coddington and Levinson (1955) ch. 3.1.

Since all theorems on the consistency of non-linear regressions depend ultimately on the assumption that the non-linear function (in this case $B = \exp(hA)$) is bijective at least in some neighbourhood of the true value – see, for instance, Malinvaud (1970a) – it would appear that we cannot obtain a consistent estimator of A by a non-linear regression on the model (2) with the discrete sample data $\{y_t : t = 1, ..., T\}$. One way of circumventing this problem in the estimation of A is to select that matrix whose characteristic roots have imaginary parts with modulus less than π/h. This method might be reasonable if the true matrix A^0 itself satisfied this criterion; but sometimes this may not be so and, in any case, since A^0 is unknown we may not be in a position to assume that it does. However, if the true system generates oscillations one or more of which have periods shorter than twice the interval between observations we cannot hope by this method anyway to estimate A^0 consistently.

On the other hand, the matrix A is usually not entirely unrestricted. We have already assumed that its elements are real numbers, and there may well be further restrictions which stem from the construction of the model. Under assumption 1 there exists a non-singular matrix T such that

$$T^{-1}A^0T = \operatorname{diag}(\lambda_1, ..., \lambda_n), \tag{6}$$

where the first $n - 2\rho$ (ρ integer and $0 \leqq \rho \leqq [n/2]$) components $\lambda_1, ..., \lambda_{n-2\rho}$ are real and the remainder occur in complex conjugate pairs as $\lambda_{n-2\rho+1}, ..., \lambda_{n-\rho}$, $\lambda_{n-\rho+1} = \bar{\lambda}_{n-2\rho+1}, ..., \lambda_n = \bar{\lambda}_{n-\rho}$. If the matrix A satisfies (5) then A is related to A^0 by the equation

$$T^{-1}AT = T^{-1}A^0T + \frac{2\pi i}{h}\begin{bmatrix} 0 & 0 & 0 \\ 0 & P & 0 \\ 0 & 0 & -P \end{bmatrix}, \tag{7}$$

where $P = \operatorname{diag}(m_1, ..., m_\rho)$ and the m_j are integers. This equation follows from the requirements in assumption 1 that the roots of A do not differ by an integer multiple of $2\pi i/h$. For this ensures that B^0 has distinct characteristic roots. Otherwise we would have the equation

$$U^{-1}T^{-1}ATU = T^{-1}A^0T + Q,$$

where U is an arbitrary matrix that commutes with the Jordan canonical

form of B^0 – Gantmacher (1959) vol. 1, pp. 239–241 – and Q is the second matrix on the right side of (23). We write (23) as

$$A = A^0 + TQT^{-1}, \tag{8}$$

and thus, in the absence of further restrictions on the coefficient matrix A, it is clear that (5) does not have a unique solution.

3. The effect of prior restrictions on the structural matrix

In formulating an econometric model such as (1) we will frequently wish to impose a number of prior restrictions on the matrix A. These restrictions arise naturally in the specification of the different equations of the model and in many cases will result from the fact that not all variables occur in every equation. Other restrictions which are often of the cross equation type arise when the elements of A are simple functions of a more basic set of economic parameters. Examples of disequilibrium models that fit this description are given in Bergstrom (1967).

We now assume that the restrictions on the structural matrix A can be expressed in terms of linear functions of the elements of A. If there are R such restrictions which we write as

$$f_r' A g_r = h_r; \qquad r = 1, ..., R,$$

where f_r and g_r are known real vectors and h_r is a known scalar, then the matrix A as given by (8) will be indistinguishable from A^0 in B^0 only if A satisfies these restrictions i.e. if

$$f_r' TQT^{-1} g_r = 0; \quad r = 1, ..., R. \tag{9}$$

We partition the matrices T and T^{-1} into their real, complex and complex conjugate blocks as follows: $T = [H : K : \bar{K}]$ and $T'^{-1} = [L' : S' : \bar{S}']$, where the bar indicates the matrix of complex conjugates. It follows that (9) leads to

$$f_r' \operatorname{Im}(KPS) g_r = 0; \qquad r = 1, ..., R, \tag{10}$$

where Im denotes the imaginary part. Since T is the transformation

matrix which diagonalises A^0 we have the relationship

$$A^0 = H\Lambda_{11} L + 2\,\mathrm{Re}(K\Lambda_{22} S),$$

where $\Lambda_{11} = \mathrm{diag}(\lambda_1, \ldots, \lambda_{n-2\rho})$, $\Lambda_{22} = \mathrm{diag}(\lambda_{n-2\rho+1}, \ldots, \lambda_{n-\rho})$ and Re denotes the real part. But A^0 satisfies the a priori restrictions so that

$$f_r' \{H\Lambda_{11} L + 2\,\mathrm{Re}(K\Lambda_{22} S)\} g_r = h_r; \qquad r = 1, \ldots, R, \tag{11}$$

and since $T T^{-1} = I$, we also have

$$f_r' \{HL + 2\,\mathrm{Re}(KS)\} g_r = f_r' g_r; \qquad r = 1, \ldots, R. \tag{12}$$

It is worth noting that (10) can be written

$$\sum_{j=1}^{\rho} m_j \,\mathrm{Im}\left(\sum_{i,k} f_{ri} k_{ij} s_{jk} g_{rk}\right) = 0; \qquad r = 1, \ldots, R, \tag{13}$$

where k_{ij} and s_{jk} are the (i, j)th and (j, k)th elements of K and S respectively. If R is at least as great as ρ (which is certainly true if $R \geqq [n/2]$) then we would expect the system (13) to imply, in general, that the integers m_j are zero and hence that Q is a zero matrix. The case $R < \rho$ is more complex because (13) may be approximately satisfied for some arbitrarily large m_j so that the true structural matrix may be difficult to identify in this case.[3]

It transpires that in some cases (10), (11) and (12) can be used to establish fairly easily that (5) has a unique solution of A^0 in the admissable parameter space and hence that A^0 is identifiable in the discrete model (2). We illustrate by taking a three equation model in which the true structural matrix $A^0 = (a_{ij}^0)$ has one real eigenvalue and it is known that $a_{13}^0 = 0$ so that the third variable does not enter the first equation. We write $T = (b, k, \bar{k})$, $T'^{-1} = (l, s, \bar{s})$ and $Q = 2\pi i\,\mathrm{diag}(0, m, -m)$ for some integer m. The time interval between observations (i.e. h) we have set equal to the time unit. Then (10), (11) and (12) lead to the following equations:

[3] However, if we restrict the admissable parameter space to be a compact set in the appropriate Euclidean space (a conventional assumption in non-linear regression) then this problem may not be very important. A rigorous theory in this case is a little difficult but it is hoped to continue this discussion in a later paper.

$$m \operatorname{Im}(k_1 s_3) = 0,$$

$$\lambda_1 b_1 l_3 + 2 \operatorname{Re}(\lambda_2 k_1 s_3) = 0,$$

$$b_1 l_3 + 2 \operatorname{Re}(k_1 s_3) = 0.$$

If $m \neq 0$, these equations imply that $\operatorname{Im}(k_1 s_3) = 0$ and $(\operatorname{Re}\lambda_2 - \lambda_1) \times \operatorname{Re}(k_1 s_3) = 0$ so that if the real parts of the roots of A are distinct it follows that either $k_1 = 0$ or $s_3 = 0$. We take the first case. Since

$$A^0 k = \lambda_2 k, \tag{14}$$

we have from the first row that $a_{12}^0 k_2 = 0$ and hence $k_2 = 0$ except for a set of structures with Lebesgue measure zero in $n^2 - 1$ (= 8) dimensional Euclidean space (i.e. those structures with $a_{12} = 0$). Then from the second row of (14) we obtain $k_3 = 0$ for almost all structures and therefore $k = 0$, which contradicts assumption 1. We would have found a similar result for s if we had taken $s_3 = 0$. Thus $m = 0$ and in this case a single zero restriction is sufficient to ensure that almost all structural matrices are identifiable in the discrete model.

4. Prior restrictions on particular rows and columns of the structural matrix

In this section we assume that each restriction is homogeneous and involves only elements from the same row or same column of A. Thus, if there are R_i restrictions on the ith row of A these can be represented by the system $\Phi_i a_i = 0$, where Φ_i is an $R_i \times n$ matrix and a_i' is the ith row of A. These restrictions correspond to the usual restrictions on the coefficient matrices of a simultaneous equations model in econometrics. Similarly, C_j restrictions on the jth column of A can be written $\Psi_j A_j = 0$, where Ψ_j is a $C_j \times n$ matrix and A_j is the jth column of A. We prove the following results.

Theorem 1. In the model (1) satisfying assumption 1 the ith row of the true structural matrix A^0 is identifiable if the matrix $\Phi_i A^{0\prime}$ has rank $n-1$. Similarly, the jth column of A^0 is identifiable if $\Psi_i A^0$ has rank $n-1$.

The case where row restrictions are present will be considered. Sup-

pose the matrix A satisfies (5) so that we can write $A = DA^0$, where D is the non-singular matrix $I + TQT^{-1}A^{0-1}$. If the ith row of A satisfies the a priori restrictions then

$$\Phi_i a_i = \Phi_i A^{0\prime} d_i = 0,$$

where d_i' is the ith row of D. It follows that d_i lies in the null space of $\Phi_i A^{0\prime}$. Since A^0 satisfies the a priori restrictions on the ith row we know that $\Phi_i A^{0\prime} f_i = 0$, where f_i is the ith column of the identity matrix of order n. Moreover, by the assumption of the theorem f_i spans the null space of $\Phi_i A^{0\prime}$, so that we can write $d_i = \alpha f_i$ for some real scalar α. But $d_i' = f_i' + t_i' QT^{-1}A^{0-1}$, where t_i' is the ith row of T and

$$T^{-1}A^{0-1} = \operatorname{diag}(1/\lambda_1, \ldots, 1/\lambda_n)T^{-1}.$$

Therefore

$$t_i' KT^{-1} = \gamma f_i', \tag{15}$$

where $\gamma = \alpha - 1$ and $K = Q \operatorname{diag}(1/\lambda_1, \ldots, 1/\lambda_n)$. From (15) we obtain the equation

$$t_i'(K - \gamma I) = 0. \tag{16}$$

Since t_i cannot be a zero vector (16) implies that the matrix $K - \gamma I$ is singular. But γ is real and the non-zero diagonal elements of K are complex with non-zero imaginary parts provided that the integers $m_j \neq 0$ for $j = 1, \ldots, \rho$. Hence, the singularity of $K - \gamma I$ implies that $\gamma = 0$ and therefore $\alpha = 1$. It follows that the ith row of TQT^{-1} is zero and thus the ith row of the structural matrix A^0 is identifiable. This proves the theorem. The corresponding result when column restrictions occur may be proved in the same way.

Theorem 2. If, in the model (1) satisfying assumption 1, any $n-1$ rows of the true structural matrix A^0 are identifiable the remaining row of A^0 is also identifiable, regardless of the number of a priori restrictions on the elements of that row.

If A satisfies (5) and we write $A = A^0 + TQT^{-1}$, it follows from the fact that $n-1$ rows of A^0 are identifiable that TQT^{-1} has only one non-zero row. The rank of TQT^{-1} must therefore be less than two. But

TQT^{-1} has the same rank as the matrix Q which will be at least two unless Q is a zero matrix. Hence, Q is zero and the remaining equation in the system is identifiable. Clearly the same result will be true if $n-1$ columns rather than rows of A^0 are identifiable.

The condition of theorem 1 is rather strong. It requires, in particular, that the total number of a priori restrictions R_i on a row of the structural matrix is not less than $n-1$ for that row (or equation) to be identifiable. In many cases this order condition may not be satisfied.[4] But the condition is not necessary and much weaker conditions will often be sufficient to ensure identifiability as the example given in the last section suggests. One interesting case that may be worthy of mention is that when (1) is the first order enlarged model corresponding to a higher order scalar model. From the form of the system matrix in this special case it follows immediately from theorems 1 and 2 that the unknown elements of the matrix are identifiable in the exact discrete model. But since the first $n-1$ derivatives of a variable are not usually observable this fact may not be of much practical use.

5. Final comments

In this paper the aliasing or identification problem that arises when we wish to estimate the characteristics of a continuous time series from a sequence of discrete observations has been discussed in the context of a simple linear model that depends on a finite number of parameters. The results obtained must be seen, therefore, as limited by this assumption. Instead of the simple parametric model (1), we could formulate a general system of continuously distributed lags such as those in Koopmans (1950) or Sims (1971a) and impose only weak conditions on the function space to which the lag distributions belong. However, this approach raises further problems, which are outside the scope of this paper, such as the extent to which finite parameter continuous time models are satisfactory approximations.[5]

[4] Since theorem 1 is similar in form to the classical theorem of identification in econometric theory (Malinvaud (1970 b) p. 658), it may be worth indicating why the implications of the two results differ. In our case A is a square matrix whereas in the classical case A is $n \times (n+m)$, where m is the number of predetermined variables. It is the fact that there are more degrees of freedom in each row of A in the classical case that makes the classical theorem less restrictive.

[5] The analogue of this problem in distributed lag models where the time parameter is discrete occurs when the parameter space is known to have infinite dimension and the effect of using finite parameter schemes as approximations in this case has been investigated by Sims (1971b) and (1972).

What it is hoped the present paper shows is that it is possible, in some cases, to use a priori information on the structure of a differential equation system to help overcome the aliasing problem we face when our data series are discrete. Although our results are preliminary, they do suggest that it is not always necessary for an investigator to assume that the spectral density matrix of the process in which he is interested is zero outside the Nyquist frequency.

References

Bartlett, M.S., 1946, On the theoretical specification and sampling properties of autocorrelated time series, Journal of the Royal Statistical Society (Supplement) 7, 27–41.

Bergstrom, A.R., 1966, Nonrecursive models as discrete approximations to systems of stochastic differential equations, Econometrica 34, 173–182.

Bergstrom, A.R., 1967, The Construction and Use of Economic Models (English Universities Press, London).

Coddington, E.A. and N. Levinson, 1955, Theory of Ordinary Differential Equations (McGraw-Hill, New York).

Durbin, J., 1961, Efficient fitting of linear models for continuous stationary time series from discrete data, Bulletin of the International Statistical Institute 38, 273–282.

Gantmacher, F.R., 1959, Theory of Matrices (Chelsea, New York).

Grenander, U. and M. Rosenblatt, 1957, Statistical analysis of stationary time series (Wiley, New York).

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

Koopmans, T.C., 1950, Models involving a continuous time variable, in: T.C. Koopmans, Statistical Inference in Dynamic Economic Models (Wiley, New York) 384–392.

Malinvaud, E., 1970a, The consistency of non-linear regressions, Annals of Mathematical Statistics 41, 956–969.

Malinvaud, E., 1970b, Statistical Methods of Econometrics (North-Holland, Amsterdam).

Phillips, A.W., 1959, The estimation of parameters in systems of stochastic differential equations, Biometrika 46, 67–76.

Phillips, P.C.B., 1972, The structural estimation of a stochastic differential equation system, Econometrica 40, 1021–1041.

Phillips, P.C.B., 1973, The problem of identification in finite parameter continuous time models, University of Essex Discussion Paper no. 48 (University of Essex, Colchester).

Robinson, P.M., 1973a, The estimation of linear differential equations with constant coefficients, to be published.

Robinson, P.M., 1973b, Continuous time regressions with discrete data, to be published.

Sargan, J.D., 1973, Some discrete approximations to continuous time stochastic models, Journal of the Royal Statistical Society, forthcoming.

Shapiro, H.S. and R.A. Silverman, 1960, Alias – free sampling of random noise, Journal of the Society of Industrial and Applied Mathematics 8, 225–248.

Sims, C.A., 1971a, Discrete approximations to continuous time distributed lags in econometrics, Econometrica 39, 545.

Sims, C.A., 1971b, Distributed lag estimation when the parameter space is explicity infinite dimensional, Annals of Mathematical Statistics 42, 1622–1636.

Sims, C.A., 1972, The role of approximate prior restrictions in distributed lag estimation, Journal of the American Statistical Association 67, 169–175.

Telser, L.G., 1967, Discrete samples and moving sums in stationary stochastic processes, Journal of the American Statistical Association 62, 484–499.
Wymer, C.R., 1972, Econometric estimation of stochastic differential equation systems, Econometrica 40, 565–577.