Econometric Theory, **22**, 2006, 483–498. Printed in the United States of America.
DOI: 10.1017/S0266466606060233

ESTIMATION OF DIFFERENTIAL-DIFFERENCE EQUATION SYSTEMS WITH UNKNOWN LAG PARAMETERS

JOANNE S. ERCOLANI
University of Birmingham

MARCUS J. CHAMBERS
University of Essex

This paper considers the estimation of the parameters of general systems of stochastic differential-difference equations in which the lag parameters themselves are treated as unknown and are not restricted to be integers and therefore form part of the parameter vector to be estimated. The asymptotic properties of an infeasible frequency domain maximum likelihood estimator are established in addition to those of a feasible version based on truncating an infinite series that arises in the computation of the spectral density function of the observed discrete time series. Precise conditions that the truncation parameter must satisfy for the asymptotic results to hold are provided.

1. INTRODUCTION

The continuous time econometrics literature to date is primarily concerned with the estimation of the parameters governing stochastic differential equations. As with many branches of econometrics, the field has grown and diversified over time, and part of this growth and diversification can be attributed to the stimulus provided by developments in economics and finance. A review of the development of linear continuous time systems and their applications in economics can be found in Bergstrom (1990), and an extensive survey of continuous time methods in finance, including the estimation of nonlinear diffusions, can be found in Sundaresan (2000).

A number of continuous time models finding recent application in economics can be classified as being forms of differential-difference equations (DDEs).

We are grateful to Gordon Kemp, Andrew Harvey, the Editor, and two anonymous referees for helpful comments on an earlier version of this paper. Any remaining errors are the sole responsibility of the authors. The first author thanks the Economic and Social Research Council for financial support under grant number R00429434216, and the second author thanks the Leverhulme Trust for financial support in the form of a Philip Leverhulme Prize. Address correspondence to Dr. J.S. Ercolani, Department of Economics, Ashley Building, University of Birmingham, Edgbaston, Birmingham B15 2TT, England; e-mail: J.S.Ercolani@bham.ac.uk.

© 2006 Cambridge University Press 0266-4666/06 $12.00

https://doi.org/10.1017/S0266466606060233 Published online by Cambridge University Press

Such models relate time derivatives of a variable to the levels and derivatives of the variable in current *and* previous time periods, unlike pure differential equations, in which the influence of lagged values is neglected in continuous time. An early and widely cited example of a DDE in economics is found in the work of Kalecki (1935), who, in his analysis of business cycles, derived an equation of the form $(d/dt)y(t) = \alpha y(t) + \beta y(t - \theta)$, where $y(t)$ denotes investment, α and β are constants, and θ denotes the lag parameter (often called the delay). A similar equation, but also involving the lagged derivative $(d/dt)y(t - \theta)$ on the right-hand side, was obtained by Ioannides and Taub (1992) in the context of time-to-build production technology and was further developed by Asea and Zak (1999).

More recent examples of DDEs in economics have enriched the specification in a number of directions. Boucekkine, de la Croix, and Licandro (2002), in an overlapping generations model of endogenous growth, derive a DDE for detrended human capital that (i) is of *second order*, i.e., the term on the left-hand side of the DDE is $(d^2/dt^2)y(t)$; (ii) contains *multiple lags* of $y(t)$ and $(d/dt)y(t)$ on the right-hand side; and (iii) contains *nonlinear parameter restrictions* on the coefficients that also involve the lag parameters themselves. Benhabib (2004) analyzes monetary policy in a model in which nominal interest rates are set by a central bank in response to a lagged inflation measure. The resulting model is of second order in inflation but has the novel feature of containing *leads* of variables and derivatives on the right-hand side, a similar equation for the price level being obtained by von Kalckreuth and Schröder (2002) in an analysis of monetary transmission mechanisms. De Cesare and Sportelli (2005) consider an IS-LM macroeconomic model involving delayed tax revenues and obtain a *system* of DDEs (in output, the interest rate and money supply), the coefficient matrices of which involve nonlinear parameter restrictions.

The estimation of DDEs has only recently been given serious consideration in the econometrics literature, although a significant literature exists in statistics on the related problem of time delay estimation.[1] Chambers (1998) developed methods for estimating the parameters of first-order DDE systems in which the lag parameters are required to be known integers, whereas Chambers and McGarry (2002) focused on a univariate first-order DDE containing a single (not necessarily integer) lag within a continuous time unobserved components model. Neither of these approaches is, therefore, capable of estimating DDEs of relevance in the recent economics literature, in which higher order DDEs, multiple lags, leads, complete systems, and nonlinear parameter restrictions are much in evidence. Hence our concern here is to develop methods of estimation of the parameters of a much more general system of DDEs, in which all of the preceding features are treated *simultaneously*. We also stress that the lag (or lead) parameters are treated as unknown and therefore constitute elements of the parameter vector to be estimated. Allowing the lags to be nonintegers enables the dynamic responses to be free of the interval at which $y(t)$ is sampled. The ability to estimate such noninteger lags is therefore a considerable advantage,

https://doi.org/10.1017/S0266466606060233 Published online by Cambridge University Press

and it can lead to more accurate estimates of cycle durations and time-to-build effects than might be obtained if the lags were restricted to coincide with the sampling interval.[2] The results governing single equation specifications and models with fixed lag parameter values emerge straightforwardly as special cases of the model considered here.

The paper is organized as follows. Section 2 defines the most general system of equations with which we are concerned and derives the spectral density function of the discrete time vector of observations. The model allows for a system of higher order DDEs with unknown lag or lead parameters in which the variables may be a mixture of stocks and flows. Section 3 derives the asymptotic properties of the frequency domain maximum likelihood estimator of the unknown system parameters. We refer to this estimator as "infeasible" because it relies on the computation of a doubly infinite summation defining the spectral density of the observed discrete time series. We therefore also propose a feasible alternative estimator that truncates the infinite series, and we provide conditions on the truncation point that ensure that the desirable asymptotic properties of consistency and asymptotic normality continue to hold. Section 4 concludes. Throughout the paper, λ denotes frequency,[3] A^* denotes the complex conjugate transpose of a matrix A, $\mathrm{tr}[A]$ denotes the trace of the matrix A, and $\|A\| = \sqrt{\mathrm{tr}[AA^*]}$. Proofs of theorems can be found in the Appendix.

2. THE DDE SYSTEM AND ITS SPECTRAL PROPERTIES

To deal simultaneously with the general features detailed in the previous section, we shall consider a real-valued stationary continuous time q-dimensional random process $\{y(t)\}_{t=-\infty}^{\infty}$ generated according to the system of DDEs defined by

$$d[D^{n-1}y(t)] = \left[\sum_{l=0}^{n}\sum_{k=1}^{s_l} A_{lk}(\theta)D^l y(t-\nu_{lk})\right]dt + \zeta(dt), \qquad -\infty < t < \infty, \tag{1}$$

where $n \geq 1$ is an integer, D denotes the mean square differential operator, $s_0,\ldots,s_n$ are finite integers, and $A_{lk}(\theta)$ $(k = 1,\ldots,s_l; l = 0,\ldots,n)$ are $q \times q$ matrices whose elements are known functions of an unknown $p \times 1$ parameter vector θ, where $p \leq (s_0 + \cdots + s_n)q^2$. The lag (or lead) terms ν_{lk} are finite real numbers satisfying $\nu_{l1} < \nu_{l2} < \cdots < \nu_{ls_l}$, where ν_{lk} may equal zero for $l \neq n$ but ν_{nk} is restricted to be nonzero for all k. Although we shall refer to the ν_{lk} as lag parameters we stress that leads can appear in (1) if the relevant ν_{lk} are negative. The term $\zeta(dt)$ is a $q \times 1$ vector of random measures satisfying $E[\zeta(dt)] = 0$, $E[\zeta(dt)\zeta(dt)'] = \Omega dt$, and $E[\zeta(\Delta_1)\zeta(\Delta_2)'] = 0$ for disjoint sets (intervals) Δ_1 and Δ_2 on the real line and where Ω is an unknown $q \times q$ symmetric positive definite covariance matrix. The *differential order* of the sys-

https://doi.org/10.1017/S0266466606060233 Published online by Cambridge University Press

tem (the highest order derivative) is equal to n, and the *difference order* of the system (the largest lag) is equal to $\max(|\nu_{lk}|)$.

The system of DDEs in (1) clearly nests the specifications that have arisen recently in economics, as described in the previous section. It also nests the model considered in Chambers (1998), in which $n = 1$ and all lags are integer valued, in addition to the univariate specification in Chambers and McGarry (2002), although in the latter case the DDE was utlized to capture the cyclical movement of a series in an unobserved components model.

The unknown parameters in (1) are the p elements of θ, the $s = \sum_{j=0}^{n} s_j$ lag parameters ν_{lk} $(l = 0,\ldots,n; k = 1,\ldots,s_l)$ contained in the $s \times 1$ vector ν, and the $q(q+1)/2$ distinct elements of the symmetric covariance matrix Ω, which are contained in the vector ω. We shall assume that the vectors θ, ν, and ω belong to suitable parameter spaces Θ, $\mathcal{V}$, and $\mathcal{W}$, respectively, where $\mathcal{W} = \{\omega : \Omega \text{ is a symmetric positive definite matrix}\} \subset \mathcal{R}^q$ and where Θ and $\mathcal{V}$ are defined in Assumption 1, which follows. It is convenient to combine the $m = p + s + q(q+1)/2$ unknown parameters into the single $m \times 1$ parameter vector $\beta \equiv (\theta', \nu', \omega')'$, which belongs to the composite parameter space $\mathcal{B} = \Theta \times \mathcal{V} \times \mathcal{W} \subset \mathcal{R}^m$.

For reasons that shall become apparent, we propose a frequency domain maximum likelihood estimator of β. Central to formulating the frequency domain likelihood function is the spectral density function of the observed discrete time process generated by (1), which can be derived from the spectral density of the continuous time process $y(t)$. First note that (1) has the representation $\alpha(D)y(t)\,dt = \zeta(dt)$, where

$$\alpha(z) = z^n I_q - \sum_{l=0}^{n} \sum_{k=0}^{s_l} A_{lk}(\theta) z^l e^{-z\nu_{lk}} \tag{2}$$

denotes the characteristic function of (1) and whose dependence on θ and ν has been suppressed for notational convenience. The appropriate behavior of $\alpha(z)$ is ensured by restricting the parameter spaces Θ and $\mathcal{V}$ as follows.

Assumption 1. The parameter spaces Θ and $\mathcal{V}$ are restricted so that the product space $\Theta \times \mathcal{V} = \{\theta, \nu :$ the roots of the equation $\det[\alpha(z)] = 0$ have negative real parts and the matrix $\alpha(i\lambda)$ is nonsingular for all $-\infty < \lambda < \infty\}$.

The condition governing the roots of the determinantal equation ensures that $y(t)$ is a stationary process, whereas the inversion condition ensures that the spectral density matrix function of the continuous time process $y(t)$ is defined for all frequencies. Proofs of the existence and uniqueness of solutions to systems like (1) are identical to Chambers (1998), and the spectral density matrix of the continuous time process $y(t)$ is given by

$$f(\lambda, \beta) = \frac{1}{2\pi}\, \alpha(i\lambda)^{-1} \Omega \alpha^*(i\lambda)^{-1}, \qquad -\infty < \lambda < \infty, \tag{3}$$

https://doi.org/10.1017/S0266466606060233 Published online by Cambridge University Press

which follows because the spectral density matrix of the white noise vector $\zeta(dt)$ is $(\frac{1}{2}\pi)\Omega$ and the frequency response function of the operator D is $i\lambda$. We have emphasized the dependence of the spectral density function on the parameter vector β for purposes of clarity.

The underlying time series is allowed to comprise both stock and flow variables, such that $y(t) = [y^s(t)', y^f(t)']'$, where the vector of stock variables (y^s) has dimension $q^s \times 1$ and the vector of flows (y^f) has dimension $q^f \times 1$ and where $q^s + q^f = q$. The observed discrete time vector is then given by $y_t = [y^s(t)', \int_{t-1}^{t} y^f(r)\, dr']'$ for $t = 1, \ldots, T$, where T denotes the sample size. The frequency response function of the time invariant linear filter that transforms $y(t)$ into y_t is given by the matrix

$$H(\lambda) = \begin{bmatrix} I_{q^s} & 0 \\ 0 & h(i\lambda) I_{q^f} \end{bmatrix},$$

where $h(i\lambda) = (1 - e^{-i\lambda})/i\lambda$ is the filter for flows.[4] The relationship between the spectral density function of a discrete time process and that of the underlying continuous time process is given by[5]

$$F(\lambda, \beta) = \sum_{j=-\infty}^{\infty} \Phi(\lambda + 2\pi j, \beta), \qquad -\pi < \lambda \leq \pi, \tag{4}$$

where

$$\Phi(\lambda, \beta) = H(\lambda) f(\lambda, \beta) H^*(\lambda) = \frac{1}{2\pi} H(\lambda) \alpha (i\lambda)^{-1} \Omega \alpha^* (i\lambda)^{-1} H^*(\lambda).$$

That $f(\lambda, \beta)$ is Hermitian positive definite and integrable[6] and $F(\lambda, \beta)$ is a convergent series can be proved in a similar fashion to the proofs given in Chambers (1998).

The aliasing phenomenon is, of course, relevant in this setting. It is impossible to detect variation at high frequencies in the continuous time series because it appears as variation below the Nyquist frequency $\pi/\Delta t$ in the discrete time series, where Δt is the observation frequency. Using high-frequency data is one solution so that the continuous time spectral density becomes negligible outside $[-\pi/\Delta t, \pi/\Delta t]$, but such an option is typically not available for the majority of economic time series. For conditions on the identification of the parameters in continuous time models the reader is referred to Hansen and Sargent (1983) and Phillips (1973).

The difficulty for estimation arises from the computation of the discrete time spectral density function (4). When the lags are integers it is possible to calculate (4) exactly. Chambers (1998) shows how to achieve this by exploiting the periodicity of the exponential terms[7] in the characteristic equation such that its determinant has q roots (which are also poles of $\Phi(\lambda + 2\pi j, \beta)$). Using Cauchy's

https://doi.org/10.1017/S0266466606060233 Published online by Cambridge University Press

residue theorem it is then possible to reduce the infinite series given in (4) to the sum of a finite number of residues. For the system (1), however, this procedure is not at all useful. The number of roots to the equation $\det[\alpha(\lambda;j)] = 0$ will typically be infinite, where $\alpha(\lambda;j) = \alpha(\lambda + 2\pi j)$ is given by

$$\alpha(\lambda;j) = (i(\lambda + 2\pi j))^n I_q - \sum_{l=0}^{n} \sum_{k=1}^{s_l} A_{lk}(\theta)(i(\lambda + 2\pi j))^l e^{-i(\lambda+2\pi j)\nu_{lk}}. \tag{5}$$

Only if ν_{lk} were integer could the periodicity of the exponential terms be exploited to reduce them to be of the form $e^{-i\lambda\nu_{lk}}$.

One solution to the problem of computing (4), and the procedure used throughout this paper, is to simply truncate the summation at a finite number M, i.e., to use

$$F^M(\lambda,\beta) = \sum_{j=-M}^{M} \Phi(\lambda + 2\pi j, \beta). \tag{6}$$

Clearly (6) is only an approximation to the true discrete time spectral density function, but if the continuous time spectral density is sufficiently small after some finite point, i.e., if $\sum_{j\notin\mathcal{M}} \Phi(\lambda + 2\pi j, \beta)$ is negligible, where $\mathcal{M} = \{-M,\ldots,M\}$, then the approximation should be sufficiently accurate. We do at least know that the continuous time spectral density matrix $f(\lambda,\beta)$ behaves like λ^{-2n} uniformly in $\beta \in \mathcal{B}$ as $\lambda \to \infty$. Although the choice of truncation point M may seem somewhat arbitrary, we provide in the next section conditions on M that ensure the consistency and asymptotic normality of the frequency domain estimator.

The error arising from the use of (6) in place of (4) is given by

$$R^M(\lambda,\beta) = \sum_{j=M+1}^{\infty} [\Phi(\lambda + 2\pi j, \beta) + \Phi(\lambda - 2\pi j, \beta)]. \tag{7}$$

A useful property of the error, and one that will be called upon when investigating the asymptotic properties of the estimator, is that $R^M(\lambda,\beta) = O(M^{1-2n})$ uniformly in (λ,β).[8] Hence for a given truncation point, the error will decrease as the order of derivative of the model increases. This is because the spectral density itself decreases with λ at a faster rate as n increases.

3. FREQUENCY DOMAIN MAXIMUM LIKELIHOOD ESTIMATION

A commonly used frequency domain estimator is the minimizer of the negative of $(2/T)$ times the frequency domain approximation to the Gaussian log-likelihood function, given by

$$L_T(\beta) = \frac{1}{T} \sum_{k\in\mathcal{K}_T} \{\ln\det[F(\lambda_k,\beta)] + \operatorname{tr}[F(\lambda_k,\beta)^{-1} I(\lambda_k)]\}, \tag{8}$$

https://doi.org/10.1017/S0266466606060233 Published online by Cambridge University Press

where $\mathcal{K}_T = \{k : -T/2 < k \leq [T/2]\}$, $I(\lambda) = w(\lambda)w^*(\lambda)$ is the periodogram, $w(\lambda) = (1/\sqrt{2\pi T})\sum_{t=1}^{T} y_t e^{it\lambda}$ is the discrete Fourier transform of y_t, and $\lambda_k = 2\pi k/T$ for $k \in \mathcal{K}_T$. However, for the reasons stated earlier, it is not possible to accurately compute the spectral density function $F(\lambda, \beta)$. We therefore consider a feasible version of (8), replacing $F(\lambda, \beta)$ with $F^M(\lambda, \beta)$, resulting in

$$L_T^M(\beta) = \frac{1}{T} \sum_{k \in \mathcal{K}_T} \{\ln \det[F^M(\lambda_k, \beta)] + \operatorname{tr}[F^M(\lambda_k, \beta)^{-1} I(\lambda_k)]\}. \tag{9}$$

The likelihood function in (9) contains a bias term that may persist even asymptotically if the truncation is independent of the sample size. We therefore allow M to increase with T and prescribe appropriate rates of increase for the minimizer of (9) to be consistent and asymptotically normally distributed. In fact, we consider the asymptotic properties of two estimators, defined by

$$\hat{\beta}_T = \arg\min_{\beta \in \mathcal{B}} L_T(\beta) \quad \text{with } \hat{\beta}_T^M = \arg\min_{\beta \in \mathcal{B}} L_T^M(\beta),$$

where $\mathcal{B}$ denotes the parameter space. We shall refer to $\hat{\beta}_T$ as being an *infeasible* estimator of β because the spectral density function $F(\lambda, \beta)$ in (8) cannot be computed accurately. Conversely we shall refer to $\hat{\beta}_T^M$ as being a *feasible* estimator of β because the truncated spectral density $F^M(\lambda, \beta)$ can be computed exactly for finite M. We shall require the following additional assumptions to establish the consistency of both estimators.

Assumption 2. The elements of the matrices $A_{lk}(\theta)$ $(l = 0, \ldots, n; k = 1, \ldots, s_l)$ are continuous functions of $\theta \in \Theta$.

Assumption 3. (a) $\beta_0 \in \operatorname{int}(\mathcal{B})$, where β_0 is the true value of β and $\mathcal{B}$ is a compact subset of $\mathcal{R}^m$. (b) For $\beta \neq \beta_0$, $F(\lambda, \beta) \neq F(\lambda, \beta_0)$ on a set of positive Lebesgue measure.

Assumption 4. The random measure $\zeta(dt)$ is Gaussian.

Assumption 5. $M = O(T^\delta)$ for some $\delta > 0$.

Assumption 2 is a standard continuity assumption that ensures, in particular, that the spectral density function $F(\lambda, \beta)$ is continuous in $\beta \in \mathcal{B}$. Assumption 3(a) is also standard in consistency proofs of optimization estimators, whereas 3(b) is an identification condition. Assumption 4 is equivalent to defining $\zeta(dt)$ as the increment of a Brownian motion process and ensures that the processes $y(t)$ in continuous time and y_t in discrete time are also Gaussian. This assumption could undoubtedly be relaxed so that non-Gaussian processes could be considered but at the cost of complicating the proofs. One consequence of doing this would be that the asymptotic covariance matrix in Theorem 2, which follows, would become more complicated and would depend on the higher moments of the process. Finally, Assumption 5 requires the truncation point M to increase

https://doi.org/10.1017/S0266466606060233 Published online by Cambridge University Press

at an arbitrary rate with T, so that asymptotically the approximation error is eliminated. The consistency result is as follows.

THEOREM 1. *Under Assumptions 1–5,* $\hat{\beta}_T \xrightarrow{p} \beta_0$ *and* $\hat{\beta}_T^M \xrightarrow{p} \beta_0$ *as* $T \to \infty$.

The proof of Theorem 1 first establishes the consistency of the infeasible estimator $\hat{\beta}_T$ by demonstrating that $L_T(\beta)$ converges uniformly over $\beta \in \mathcal{B}$ to a function $L(\beta)$ that is uniquely minimized at $\beta = \beta_0$. The consistency of the feasible estimator $\hat{\beta}_T^M$ is then shown by establishing the uniform convergence of $L_T^M(\beta)$ to $L_T(\beta)$. The feasible estimator, based on the truncated spectral density, therefore shares the consistency property of the infeasible estimator, provided that the truncation point increases with sample size.

To derive the asymptotic distributions of the two estimators, two additional assumptions are required.

Assumption 5′. $M = O(T^\delta)$ for some $\delta > 1/(4n - 2)$.

Assumption 6. The elements of the matrices $A_{lk}(\theta)$ $(l = 0, \ldots, n; k = 1, \ldots, s_l)$ are twice continuously differentiable functions of $\theta \in \Theta$.

Assumption 5′ represents a strengthening of Assumption 5 and is required for the asymptotic distribution theory. Note that δ is a decreasing function of the differential order n, so that in models with higher order derivatives, a slower rate of increase of M with T will suffice. The intuition behind this is fairly simple because we know that the error of truncation (R^M) is inversely related to the order of derivative (n) and therefore as the order of derivative increases, the same magnitude for R^M can be obtained with a smaller truncation point. In models containing just a first derivative the condition reduces to $\delta > 0.5$ as in Chambers and McGarry (2002). Assumption 6 is common in the asymptotic analysis of optimization estimators and ensures that the objective functions (8) and (9) can be differentiated to second order.

THEOREM 2. *Under Assumptions 1–4, 5′, and 6,* $\sqrt{T}(\hat{\beta}_T - \beta_0) \xrightarrow{d} N(0, 2S(\beta_0)^{-1})$ *and* $\sqrt{T}(\hat{\beta}_T^M - \beta_0) \xrightarrow{d} N(0, 2S(\beta_0)^{-1})$ *as* $T \to \infty$, *where the* (i, j) *element of* $S(\beta_0)$ *is given by*

$$S_{ij}(\beta_0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \operatorname{tr}\left[F(\lambda, \beta_0)^{-1} \frac{\partial F(\lambda, \beta_0)}{\partial \beta_i} F(\lambda, \beta_0)^{-1} \frac{\partial F(\lambda, \beta_0)}{\partial \beta_j} \right] d\lambda \qquad \textbf{(10)}$$

for $i, j = 1, \ldots, m$, *and* $\partial F(\lambda, \beta_0)/\partial \beta_i$ *denotes the matrix* $\partial F(\lambda, \beta)/\partial \beta_i$ *evaluated at* $\beta = \beta_0$.

Both estimators share the same normal limiting distribution, provided the truncation point for the infeasible estimator increases at the appropriate rate with sample size, as required in Assumption 5′. The proof of Theorem 2 relies on standard arguments for the infeasible estimator $\hat{\theta}_T$, based on a mean value

https://doi.org/10.1017/S0266466606060233 Published online by Cambridge University Press

expansion of the score vector, whereas the result for $\hat{\theta}_T^M$ is obtained by establishing that $\sqrt{T}(\hat{\beta}_T - \hat{\beta}_T^M) = o_p(1)$ as $T \to \infty$. It is worth noting that the elements of the matrix $S(\beta_0)$ can be consistently estimated using $\hat{S}(\hat{\beta}_T)$ and $\hat{S}(\hat{\beta}_T^M)$, where the (i,j) element of the latter is defined by the computationally more convenient expression

$$\hat{S}_{ij}(\hat{\beta}_T^M) = \frac{1}{T} \sum_{k \in \mathcal{K}_T} F(\lambda_k, \hat{\beta}_T^M)^{-1} \frac{\partial F(\lambda_k, \hat{\beta}_T^M)}{\partial \beta_i} F(\lambda_k, \hat{\beta}_T^M)^{-1} \frac{\partial F(\lambda_k, \hat{\beta}_T^M)}{\partial \beta_j},$$

where $\partial F(\lambda, \hat{\beta}_T^M)/\partial \beta_i$ denotes the matrix $\partial F(\lambda, \beta)/\partial \beta_i$ evaluated at $\beta = \hat{\beta}_T^M$. The matrix $\hat{S}(\hat{\beta}_T)$ is obtained in the same way with $\hat{\beta}_T^M$ replaced by $\hat{\beta}_T$.

The estimators $\hat{\beta}_T^M$ and $\hat{\beta}_T$ are both asymptotically efficient under the stated assumptions. Departures from these assumptions, however, such as model misspecification or non-Gaussianity, can be expected to lead to inefficiencies, even asymptotically. Appropriately designed sampling experiments could be utilized to assess the extent of such inefficiencies in finite samples, but such an investigation remains beyond the scope of the present paper.

The asymptotic normality of the feasible estimator and the likelihood-based approach to estimation both mean that tests of hypotheses concerning β can be conducted using the standard trinity of Wald, likelihood ratio, and Lagrange multiplier test statistics. Under the appropriate null hypothesis standard techniques can be used to show that these statistics will be asymptotically distributed as chi-squared with degrees of freedom equal to the number of restrictions being tested. For example, tests of the nonlinear restrictions $h(\beta) = 0$, where $h : \mathcal{R}^m \to \mathcal{R}^c$ $(c \leq m)$ is twice continuously differentiable and the matrix $\partial h/\partial \beta'$ evaluated at $\beta = \beta_0$ has full rank, have asymptotic χ_c^2 distributions under the null, but other hypotheses are also of interest. Prominent among these is a test of equality of lag parameters across the elements of $y(t)$. In a simple setting, suppose

$$dy_i(t) = [a_{0i}\, y_i(t) + a_{1i}\, y_i(t - \nu_i)]\, dt + \zeta_i(dt), \qquad i = 1, \ldots, q,$$

in which case the null of interest would be $H_0 : \nu_1 = \nu_2 = \cdots = \nu_q$. This might correspond to a test of whether the time-to-build factors in a group of categories of investment expenditures were equal. Within the same setting another hypothesis might be that the ν_i were all equal to a particular integer value, e.g., unity. If this null hypothesis were found to be consistent with the data then its imposition would enable the spectral density function to be computed exactly using the techniques of Chambers (1998).

4. CONCLUDING COMMENTS

In this paper we have considered the problem of estimating the unknown parameters, including lag and lead parameters, in a system of DDEs of sufficient gen-

https://doi.org/10.1017/S0266466606060233 Published online by Cambridge University Press

erality that nests the important recent DDE specifications that have appeared in economics, thereby expanding the scope for estimation of DDEs beyond those that have appeared in the econometrics literature to date. The main technical results of this paper concern the derivation of the asymptotic properties of the (infeasible) frequency domain maximum likelihood estimator and its feasible counterpart based on an approximation to the discrete time spectral density function. The feasible frequency domain maximum likelihood estimator, based on the truncated spectral density, is shown to share the same desirable asymptotic properties as the infeasible estimator, based on the exact spectral density, provided the truncation parameter increases at an appropriate rate with the sample size (which is given by Assumption 5 for consistency and by Assumption 5′ for asymptotic normality). It is therefore possible to obtain consistent and asymptotically normal estimates of the lag (or lead) parameters themselves in addition to the other parameters in the DDE system and to perform asymptotic chi-squared inference on hypotheses concerning those parameters.

NOTES

1. References to some of this alternative literature can be found in Chambers and McGarry (2002).

2. In some circumstances, it may be possible to sample at a sufficiently high frequency so as to render all lags integer valued. However, for most applications in economics and, indeed, finance, such flexibility in the frequency with which a process is sampled is simply not available. It also presupposes that the true lag values are known, which is contrary to the objective of this paper, namely, the estimation of the unknown lag values.

3. This is actually the angular frequency representing the number of radians per unit of time. The number of cycles per time unit is then given by $f = \lambda/2\pi$, and the period of a cycle is $1/f$.

4. This follows from the relationship $y_t^f = h(D)y^f(t)$ and the fact that the frequency response of the operator D is $i\lambda$. See, e.g., Priestley (1981, p. 175) for details.

5. For a formal proof see Priestley (1981, Sect. 7.1.1).

6. The Hermitian property requires $f(\lambda,\beta) = f^*(\lambda,\beta)$. Integrability is important as it ensures that $y(t)$ has finite variance.

7. The exponential term $e^{i(\lambda+2\pi j)k} = e^{i\lambda k}$ for j and k integer.

8. A proof can be found in Lemma 1 in the Appendix.

REFERENCES

Amemiya, T. (1985) *Advanced Econometrics*. Blackwell.

Asea, P.K. & P.J. Zak (1999) Time-to-build and cycles. *Journal of Economic Dynamics and Control* 23, 1155–1175.

Benhabib, J. (2004) Interest rate policy in continuous time with discrete delays. *Journal of Money, Credit and Banking* 36, 1–16.

Bergstrom, A.R. (1990) *Continuous Time Econometric Modelling*. Oxford University Press.

Boucekkine, R., D. de la Croix, & O. Licandro (2002) Vintage human capital, demographic trends, and endogenous growth. *Journal of Economic Theory* 104, 340–375.

Chambers, M.J. (1998) The estimation of systems of joint differential-difference equations. *Journal of Econometrics* 85, 1–31.

Chambers, M.J. & J.S. McGarry (2002) Modeling cyclical behavior with differential-difference equations in an unobserved components framework. *Econometric Theory* 18, 387–419.

https://doi.org/10.1017/S0266466606060233 Published online by Cambridge University Press

De Cesare, L. & M. Sportelli (2005) A dynamic IS-LM model with delayed taxation revenues. *Chaos, Solitons and Fractals* 25, 233–244.
Dunsmuir, W. (1979) A central limit theorem for parameter estimation in stationary vector time series and its application to models for a signal observed with noise. *Annals of Statistics* 7, 490–506.
Dunsmuir, W. & E.J. Hannan (1976) Vector linear time series models. *Advances in Applied Probability* 8, 339–364.
Hannan, E.J. (1970) *Multiple Time Series*. Wiley.
Hansen, L.P. & T.J. Sargent (1983) The dimensionality of the aliasing problem in models with rational spectral densities. *Econometrica* 51, 377–387.
Ibragimov, I.A. & Y.V. Linnik (1971) *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff.
Ioannides, Y.M. & B. Taub (1992) On dynamics with time-to-build investment technology and non-time-separable leisure. *Journal of Economic Dynamics and Control* 16, 225–241.
Kalecki, M. (1935) A macro-dynamic theory of business cycles. *Econometrica* 3, 327–344.
Marcus, M. & H. Minc (1965) *Introduction to Linear Algebra*. Macmillan.
Phillips, P.C.B. (1973) The problem of identification in finite parameter continuous time models. *Journal of Econometrics* 1, 351–362.
Priestley, M.B. (1981) *Spectral Analysis and Time Series*. Academic Press.
Robinson, P.M. (1976) Fourier estimation of continuous time models. In A.R. Bergstrom (ed.), *Statistical Inference in Continuous Time Economic Models*, pp. 215–266. North-Holland.
Sundaresan, S.M. (2000) Continuous-time methods in finance: A review and an assessment. *Journal of Finance* 55, 1569–1622.
von Kalckreuth, U. & J. Schröder (2002) Monetary Transmission in the New Economy: Service Life of Capital, Transmission Channels and the Speed of Adjustment. Discussion paper 16/02, Economics Research Centre of the Deutsche Bundesbank.
Zygmund, A. (1959) *Trigonometric Series*, 2nd ed. Cambridge University Press.

APPENDIX

LEMMA 1. *Let $R^M(\lambda,\beta)$ denote the error from truncating the spectral density after M terms, as defined in (7). Then $R^M(\lambda,\beta) = O(M^{1-2n})$ uniformly in (λ,β).*

Proof. Recall, from (7), that

$$R^M(\lambda,\beta) = \sum_{j=M+1}^{\infty} [\Phi(\lambda + 2\pi j,\beta) + \Phi(\lambda - 2\pi j,\beta)],$$

where $\Phi(\lambda,\beta) = H(\lambda)f(\lambda,\beta)H^*(\lambda)$. The elements of $\Phi(\lambda \pm 2\pi j,\beta)$ can be shown to be $O(j^{-2n})$ uniformly in (λ,β) as $j \to \infty$ by noting that $f(\lambda \pm 2\pi j,\beta) = O(j^{-2n})$ uniformly in (λ,β) and $h(\lambda \pm 2\pi j) = (1 - e^{i\lambda})/[i(\lambda \pm 2\pi j)] = O(j^{-1})$ uniformly in λ. Hence, for $k,l = 1,\ldots,q$,

$$\left| \sum_{j=M+1}^{\infty} \Phi_{kl}(\lambda + 2\pi j,\beta) \right| \leq \sum_{j=M+1}^{\infty} |\Phi_{kl}(\lambda + 2\pi j,\beta)| \leq K \sum_{j=M+1}^{\infty} j^{-2n} = O(M^{1-2n})$$

uniformly in (λ,β) for $2n > 1$ and where K is a generic constant. ■

https://doi.org/10.1017/S0266466606060233 Published online by Cambridge University Press

Proof of Theorem 1. The consistency of $\hat{\beta}_T$ is established first by demonstrating that $\sup_{\beta\in\mathcal{B}}|L_T(\beta) - L(\beta)| \xrightarrow{p} 0$ and that $L(\beta)$ is uniquely minimized at $\beta = \beta_0$, where

$$L(\beta) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\{\ln\det[F(\lambda,\beta)] + \operatorname{tr}[F(\lambda,\beta)^{-1}F(\lambda,\beta_0)]\}\,d\lambda.$$

From the definition of $L_T(\beta)$ we have $\sup_{\beta\in\mathcal{B}}|L_T(\beta) - L(\beta)| \le \Lambda_{1T} + \Lambda_{2T}$, where

$$\Lambda_{1T} = \sup_{\beta\in\mathcal{B}}\left|\frac{1}{T}\sum_{k\in\mathcal{K}_T}\ln\det[F(\lambda_k,\beta)] - \frac{1}{2\pi}\int_{-\pi}^{\pi}\ln\det[F(\lambda,\beta)]\,d\lambda\right| \to 0$$

by the uniform convergence of the Riemann sums and the continuity of $F(\lambda,\beta)$ under Assumptions 1 and 2 and

$$\Lambda_{2T} = \sup_{\beta\in\mathcal{B}}\left|\frac{1}{T}\sum_{k\in\mathcal{K}_T}\operatorname{tr}[F(\lambda_k,\beta)^{-1}I(\lambda_k)] - \frac{1}{2\pi}\int_{-\pi}^{\pi}\operatorname{tr}[F(\lambda,\beta)^{-1}F(\lambda,\beta_0)]\,d\lambda\right| \xrightarrow{p} 0$$

by Theorem 1 of Robinson (1976), which requires that

$$\frac{1}{T}\sum_{t=1}^{T-k} y_t y_{t+k}' \xrightarrow{a.s.} E(y_0 y_k') = \int_{-\pi}^{\pi} e^{ik\lambda}F(\lambda,\beta_0)\,d\lambda.$$

This follows from Theorem 2 of Hannan (1970, p. 203) because y_t is a stationary Gaussian process with continuous spectral distribution function under Assumptions 1, 2, and 5. Hence $L_T(\beta)$ converges to $L(\beta)$, uniformly in $\beta \in \mathcal{B}$, as $T \to \infty$. That $L(\beta)$ is uniquely minimized at $\beta = \beta_0$ can be demonstrated following the argument in Dunsmuir and Hannan (1976).

To establish the consistency of $\hat{\beta}_T^M$, note that

$$\sup_{\beta\in\mathcal{B}}|L_T^M(\beta) - L(\beta)| \le \sup_{\beta\in\mathcal{B}}|L_T^M(\beta) - L_T(\beta)| + \sup_{\beta\in\mathcal{B}}|L_T(\beta) - L(\beta)|.$$

We have shown already that the second term converges to 0 in probability, so we consider $|L_T^M(\beta) - L_T(\beta)| \le B_{1T}(\beta) + B_{2T}(\beta)$, where

$$B_{1T}(\beta) = \left|\frac{1}{T}\sum_{k\in\mathcal{K}_T}\{\ln\det[F^M(\lambda_k,\beta)] - \ln\det[F(\lambda_k,\beta)]\}\right|,$$

$$B_{2T}(\beta) = \left|\frac{1}{T}\sum_{k\in\mathcal{K}_T}\operatorname{tr}\{[F^M(\lambda_k,\beta)^{-1} - F(\lambda_k,\beta)^{-1}]I(\lambda_k)\}\right|.$$

Fundamental to much that follows is the relationship

$$F^M(\lambda,\beta) = F(\lambda,\beta) - R^M(\lambda,\beta) = F(\lambda,\beta)[I - F(\lambda,\beta)^{-1}R^M(\lambda,\beta)].$$

Considering $B_{1T}(\beta)$ first, the preceding decomposition implies that

$$\ln\det[F^M(\lambda_k,\beta)] - \ln\det[F(\lambda_k,\beta)] = \ln\det[A^M(\lambda_k,\beta)],$$

https://doi.org/10.1017/S0266466606060233 Published online by Cambridge University Press

where $A^M(\lambda,\beta) = I - F(\lambda,\beta)^{-1}R^M(\lambda,\beta)$. The diagonal elements of $A^M(\lambda_k,\beta)$ are of the form $1 + O(M^{1-2n})$, whereas the off-diagonal elements are $O(M^{1-2n})$, both uniformly in (λ,β). Hence as $T \to \infty$,

$$\sup_{\beta\in\mathcal{B}} B_{1T}(\beta) = \sup_{\beta\in\mathcal{B}} \left| \frac{1}{T} \sum_{k\in\mathcal{K}_t} \ln\det[A^M(\lambda_k,\beta)] \right| \to 0.$$

To deal with $B_{2T}(\beta)$, define $D^M(\lambda,\beta) = F^M(\lambda,\beta)^{-1} - F(\lambda,\beta)^{-1} = A^M(\lambda,\beta)^{-1} F(\lambda,\beta)^{-1} - F(\lambda,\beta)^{-1}$. Let $\{\mu_i\}_{i=1}^q$ denote the eigenvalues of $F(\lambda,\beta)^{-1}R^M(\lambda,\beta)$, suppressing their dependence on M, λ, and β for notational convenience. From Theorem 3.7 of Marcus and Minc (1965), $|\mu_i| \le q \max_{i,j}|\phi_{ij}|$, where ϕ_{ij} denotes the (i,j) element of $F(\lambda,\beta)^{-1}R^M(\lambda,\beta)$. Hence $|\mu_i| = O(M^{1-2n})$ uniformly in (λ,β), and so, for large enough M, $\sup_{\lambda,\beta}|\mu_i| < 1$ $(i = 1,\ldots,q)$. Hence $A^M(\lambda,\beta)^{-1}$ can be written as a geometric progression of the form

$$A^M(\lambda,\beta)^{-1} = I + \sum_{l=1}^{\infty} [F(\lambda,\beta)^{-1}R^M(\lambda,\beta)]^l = I + G^M(\lambda,\beta),$$

where $G^M(\lambda,\beta) = O(M^{1-2n})$, uniformly in (λ,β). Hence

$$D^M(\lambda,\beta) = [I + G^M(\lambda,\beta)]F(\lambda,\beta)^{-1} - F(\lambda,\beta)^{-1} = G^M(\lambda,\beta)F(\lambda,\beta)^{-1} = O(M^{1-2n})$$

uniformly in (λ,β), and so

$$\sup_{\beta\in\mathcal{B}} B_{2T}(\beta) \le \sup_{\lambda_k,\beta} \|D^M(\lambda_k,\beta)\| \frac{1}{T} \sum_{k\in\mathcal{K}_T} \|I(\lambda_k)\|$$

$$\le KM^{1-2n} \frac{1}{T} \sum_{k\in\mathcal{K}_T} \|I(\lambda_k)\| = O_p(M^{1-2n}),$$

where K is a constant and because $T^{-1}\sum_k \|I(\lambda_k)\| = O_p(1)$. Under Assumption 5, the preceding expression is $O_p(T^{\delta(1-2n)}) = o_p(1)$. Hence $\sup_{\beta\in\mathcal{B}}|L_T^M(\beta) - L_T(\beta)| = o_p(1)$, and so $\sup_{\beta\in\mathcal{B}}|L_T^M(\beta) - L(\beta)| = o_p(1)$, thereby implying that $\hat{\beta}_T^M \xrightarrow{p} \beta_0$. ■

Proof of Theorem 2. Expanding the score vector $s_T(\hat{\beta}_T)$ around β_0 gives, for T large enough,

$$s_T(\hat{\beta}_T) = s_T(\beta_0) + S_T(\bar{\beta})(\hat{\beta}_T - \beta_0),$$

where $S_T(\bar{\beta})$ denotes the Hessian matrix whose ith row is equal to the ith row of the matrix $\partial s_T(\beta)/\partial\beta' = \partial^2 L_T(\beta)/\partial\beta\partial\beta'$ evaluated at $\beta = \bar{\beta}_i$, where $\bar{\beta}_i$ lies on the line segment joining $\hat{\beta}_T$ and β_0 and where $\bar{\beta} = [\bar{\beta}_1,\ldots,\bar{\beta}_m]$. Hence, because $s_T(\hat{\beta}_T) = 0$,

$$\sqrt{T}(\hat{\beta}_T - \beta_0) = -S_T(\bar{\beta})^{-1}[\sqrt{T}s_T(\beta_0)].$$

We shall demonstrate that, as $T \to \infty$, (a) $\sqrt{T}s_T(\beta_0) \xrightarrow{d} N(0,2S(\beta_0))$ and (b) $S_T(\bar{\beta}) \xrightarrow{p} S(\beta_0)$, from which the assertion in the theorem concerning $\hat{\beta}_T$ follows.

https://doi.org/10.1017/S0266466606060233 Published online by Cambridge University Press

(a) From the definition of the likelihood, we obtain, for $i = 1, \ldots, m$,

$$
\begin{aligned}
s_{T,i}(\beta) &= \frac{1}{T} \sum_{k \in \mathcal{K}_T} \left[\frac{\partial \ln \det[F_k(\beta)]}{\partial \beta_i} + \frac{\partial \operatorname{tr}[F_k(\beta)^{-1} I_k]}{\partial \beta_i} \right] \\
&= \frac{1}{T} \sum_{k \in \mathcal{K}_T} \left\{ \operatorname{tr}\left[F_k(\beta)^{-1} \frac{\partial F_k(\beta)}{\partial \beta_i} \right] - \operatorname{tr}\left[F_k(\beta)^{-1} \frac{\partial F_k(\beta)}{\partial \beta_i} F_k(\beta)^{-1} I_k \right] \right\} \\
&= \frac{1}{T} \sum_{k \in \mathcal{K}_T} \operatorname{tr}[G_{k,i}(\beta)(F_k(\beta) - I_k)], \qquad \textbf{(A.1)}
\end{aligned}
$$

where $G_{k,i}(\beta) = F_k(\beta)^{-1}[\partial F_k(\beta)/\partial \beta_i] F_k(\beta)^{-1}$, $F_k(\beta) = F(\lambda_k, \beta)$, and $I_k = I(\lambda_k)$ for notational convenience. The required asymptotic normality of $\sqrt{T} s_T(\beta_0)$ then holds if conditions C2.1–C2.3 on p. 492 of Dunsmuir (1979) can be shown to be satisfied. (These conditions are required for Theorem 2.1 of Dunsmuir, 1979, to be valid, which concerns the asymptotic properties of the minimizer of a version of the likelihood that incorporates integrals of $F(\lambda)$ and $I(\lambda)$, rather than the discrete sums considered here. However, Dunsmuir's Corollary 2.2 and his Remark 1 [on p. 498] demonstrate that these same conditions yield identical asymptotic properties for the minimizer of the discrete version of the likelihood employed here.) Condition C2.1 is satisfied in view of the definition of $F(\lambda, \beta)$ and Assumption 6. The differentiability of $F(\lambda, \beta)$ (with respect to λ) implies that $F(\lambda, \beta)$ satisfies a Lipschitz condition of order 1 (see, e.g., Zygmund, 1959, p. 43), and hence C2.2 is satisfied. Because $F(\lambda, \beta)$ is continuous in both arguments and Hermitian positive definite, and y_t is a stationary Gaussian sequence, Theorem 17.3.3 of Ibragimov and Linnik (1971) implies that y_t is strong mixing. It is therefore, from Hannan (1970, pp. 202–204), also ergodic and has representation $y_t = \sum_{j=0}^{\infty} C_j \epsilon_{t-j}$ with $\sum_{j=0}^{\infty} \|C_j\|^2 < \infty$ and where the ϵ_t are independently normally distributed. This ensures that C2.3 is satisfied, and hence $\sqrt{T} s_T(\beta_0) \xrightarrow{d} N(0, 2S(\beta_0))$.

(b) Differentiating (A.1) with respect to β_j and evaluating at $\beta = \bar{\beta}_i$ yields

$$
\begin{aligned}
S_{T,ij}(\bar{\beta}_i) = &- \frac{1}{T} \sum_{k \in \mathcal{K}_T} \operatorname{tr}\left[F_k(\bar{\beta}_i)^{-1} \frac{\partial F_k(\bar{\beta}_i)}{\partial \beta_j} F_k(\bar{\beta}_i)^{-1} \frac{\partial F_k(\bar{\beta}_i)}{\partial \beta_i} \right] \\
&+ \frac{1}{T} \sum_{k \in \mathcal{K}_T} \operatorname{tr}\left[F_k(\bar{\beta}_i)^{-1} \frac{\partial^2 F_k(\bar{\beta}_i)}{\partial \beta_i \partial \beta_j} \right] \\
&+ \frac{1}{T} \sum_{k \in \mathcal{K}_T} \operatorname{tr}\left[F_k(\bar{\beta}_i)^{-1} \frac{\partial F_k(\bar{\beta}_i)}{\partial \beta_j} F_k(\bar{\beta}_i)^{-1} \frac{\partial F_k(\bar{\beta}_i)}{\partial \beta_i} F_k(\bar{\beta}_i)^{-1} I_k \right] \\
&+ \frac{1}{T} \sum_{k \in \mathcal{K}_T} \operatorname{tr}\left[F_k(\bar{\beta}_i)^{-1} \frac{\partial F_k(\bar{\beta}_i)}{\partial \beta_i} F_k(\bar{\beta}_i)^{-1} \frac{\partial F_k(\bar{\beta}_i)}{\partial \beta_j} F_k(\bar{\beta}_i)^{-1} I_k \right] \\
&- \frac{1}{T} \sum_{k \in \mathcal{K}_T} \operatorname{tr}\left[F_k(\bar{\beta}_i)^{-1} \frac{\partial^2 F_k(\bar{\beta}_i)}{\partial \beta_i \partial \beta_j} F_k(\bar{\beta}_i)^{-1} I_k \right],
\end{aligned}
$$

where, e.g., $\partial F_k(\bar{\beta}_i)/\partial \beta_i$ denotes $\partial F_k(\beta)/\partial \beta_i$ evaluated at $\beta = \bar{\beta}_i$. The first two terms are of the generic form $T^{-1} \sum_k \operatorname{tr}[\Phi_k(\bar{\beta}_i)]$, whereas the remaining three are of the generic

https://doi.org/10.1017/S0266466606060233 Published online by Cambridge University Press

form $T^{-1}\sum_k \mathrm{tr}[\Phi_k(\bar{\beta}_i)I_k]$ with the matrix $\Phi_k(\beta)$ defined appropriately in each case. The terms satisfy, respectively, $\sup_{\beta\in\mathcal{B}}|T^{-1}\sum_k \mathrm{tr}[\Phi_k(\beta)] - (2\pi)^{-1}\int_{-\pi}^{\pi}\mathrm{tr}[\Phi(\lambda,\beta)]\,d\lambda| \to 0$ and $\sup_{\beta\in\mathcal{B}}|T^{-1}\sum_k \mathrm{tr}[\Phi_k(\beta)I_k] - (2\pi)^{-1}\int_{-\pi}^{\pi}\mathrm{tr}[\Phi(\lambda,\beta)F(\lambda,\beta_0)]\,d\lambda| \xrightarrow{p} 0$ as $T\to\infty$, the latter following from Theorem 1 of Robinson (1976). This uniform convergence, the fact that $\|\bar{\beta}_i - \beta_0\| \le \|\hat{\beta}_T - \beta_0\|$, and the consistency of $\hat{\beta}_T$ imply that the first two terms converge to

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\mathrm{tr}\left[F(\lambda,\beta_0)^{-1}\frac{\partial^2 F(\lambda,\beta_0)}{\partial\beta_i\partial\beta_j} - F(\lambda,\beta_0)^{-1}\frac{\partial F(\lambda,\beta_0)}{\partial\beta_j}F(\lambda,\beta_0)^{-1}\frac{\partial F(\lambda,\beta_0)}{\partial\beta_i}\right]d\lambda,$$

whereas Theorem 4.1.5 of Amemiya (1985) ensures that the final three terms converge in probability to

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\mathrm{tr}\left[F(\lambda,\beta_0)^{-1}\frac{\partial F(\lambda,\beta_0)}{\partial\beta_j}F(\lambda,\beta_0)^{-1}\frac{\partial F(\lambda,\beta_0)}{\partial\beta_i} - F(\lambda,\beta_0)^{-1}\frac{\partial^2 F(\lambda,\beta_0)}{\partial\beta_i\partial\beta_j}\right]d\lambda$$

$$+\frac{1}{2\pi}\int_{-\pi}^{\pi}\mathrm{tr}\left[F(\lambda,\beta_0)^{-1}\frac{\partial F(\lambda,\beta_0)}{\partial\beta_i}F(\lambda,\beta_0)^{-1}\frac{\partial F(\lambda,\beta_0)}{\partial\beta_j}\right]d\lambda.$$

Combining these limits we find that $S_{T,ij}(\bar{\beta}_i) \xrightarrow{p} S_{ij}(\beta_0)$ with $S_{ij}(\beta_0)$ given by the final term in the preceding expression, and hence (b) is proved. The claim in the theorem concerning the distribution of $\hat{\beta}_T$ then follows by combining the results of (a) and (b).

To establish that $\sqrt{T}(\hat{\beta}_T^M - \beta_0) \xrightarrow{d} N(0, 2S(\beta_0)^{-1})$ also, it is sufficient to establish that $\sqrt{T}(\hat{\beta}_T - \hat{\beta}_T^M) = o_p(1)$. From the mean value expansion of the score vector of $\hat{\beta}_T^M$, it can be shown that, for T large enough,

$$\sqrt{T}(\hat{\beta}_T^M - \beta_0) = -S_T^M(\tilde{\beta})^{-1}\sqrt{T}s_T^M(\beta_0), \tag{A.2}$$

where $S_T^M(\tilde{\beta})$ denotes the Hessian matrix whose ith row is equal to the ith row of the matrix $\partial s_T^M(\beta)/\partial\beta' = \partial^2 L_T^M(\beta)/\partial\beta\partial\beta'$ evaluated at $\beta = \tilde{\beta}_i$ where $\tilde{\beta}_i$ lies on the line segment joining $\hat{\beta}_T$ and β_0 and where $\tilde{\beta} = [\tilde{\beta}_1,\ldots,\tilde{\beta}_m]$. It can be shown therefore that

$$\begin{aligned}\sqrt{T}(\hat{\beta}_T - \hat{\beta}_T^M) &= S_T^M(\tilde{\beta})^{-1}\sqrt{T}s_T^M(\beta_0) - S_T(\bar{\beta})^{-1}\sqrt{T}s_T(\beta_0)\\ &= [S_T^M(\tilde{\beta})^{-1} - S_T(\bar{\beta})^{-1}]\sqrt{T}s_T^M(\beta_0)\\ &\quad + S_T(\bar{\beta})^{-1}\sqrt{T}[s_T^M(\beta_0) - s_T(\beta_0)].\end{aligned}$$

We have already established that $S_T(\bar{\beta})^{-1} = O_p(1)$ and that $\sqrt{T}s_T(\beta_0) = O_p(1)$. To show that $\sqrt{T}(\hat{\beta}_T - \hat{\beta}_T^M) = o_p(1)$, it is sufficient to show that (i) $\sqrt{T}[s_T^M(\beta_0) - s_T(\beta_0)] = o_p(1)$ and (ii) $S_T^M(\tilde{\beta})^{-1} - S_T(\bar{\beta})^{-1} = o_p(1)$. We look at each of these in turn.

(i) The score vector for $\hat{\beta}_T^M$ is given by

$$s_{T,i}^M(\beta) = \frac{1}{T}\sum_{k\in\mathcal{K}_T}\left[\frac{\partial \ln\det[F_k^M(\beta)]}{\partial\beta_i} + \frac{\partial\,\mathrm{tr}[F_k^M(\beta)^{-1}I_k]}{\partial\beta_i}\right], \qquad (i=1,\ldots,m),$$

https://doi.org/10.1017/S0266466606060233 Published online by Cambridge University Press

where $F_k^M(\beta) = F^M(\lambda_k, \beta)$ for notational economy. We know that $\ln\det[F_k^M(\beta)] = \ln\det[F_k(\beta)] + \gamma_{1k}^M$ where $\gamma_{1k}^M = A^M(\lambda_k, \beta) = O(M^{1-2n})$ uniformly in (λ_k, β). Hence

$$\frac{\partial \ln\det[F_k^M(\beta)]}{\partial \beta_i} = \frac{\partial \ln\det[F_k(\beta)]}{\partial \beta_i} + \gamma_{2k}^M, \tag{A.3}$$

where $\gamma_{2k}^M = O(M^{1-2n})$ uniformly in (λ_k, β). Note that $\mathrm{tr}[(F_k^M(\beta)^{-1} - F_k(\beta)^{-1})I_k]$ can be written in the form $\mathrm{tr}[D(\lambda_k, \beta)I_k] = O_p(M^{1-2n})$ uniformly in (λ_k, β). It therefore follows that $\mathrm{tr}[F_k^M(\beta)^{-1}I_k] = \mathrm{tr}[F_k(\beta)^{-1}I_k] + \gamma_{3k}^M$ where $\gamma_{3k}^M = O_p(M^{1-2n})$ uniformly in (λ_k, β), and so

$$\frac{\partial \,\mathrm{tr}[F_k^M(\beta)^{-1}I_k]}{\partial \beta_i} = \frac{\partial \,\mathrm{tr}[F_k(\beta)^{-1}I_k]}{\partial \beta_i} + \gamma_{4k}^M, \tag{A.4}$$

where $\gamma_{4k}^M = O_p(M^{1-2n})$ uniformly in (λ_k, β). Using (A.3) and (A.4) yields

$$\sqrt{T}s_T^M(\beta) = \frac{1}{\sqrt{T}} \sum_{k \in \mathcal{K}_T} \left[\frac{\partial \ln\det[F_k(\beta)]}{\partial \beta_i} + \frac{\partial \,\mathrm{tr}[F_k(\beta)^{-1}I_k]}{\partial \beta_i} + \gamma_{2k}^M + \gamma_{4k}^M \right],$$

implying that $\sqrt{T}s_T^M(\beta) = \sqrt{T}s_T(\beta) + \sqrt{T}O_p(M^{1-2n})$ uniformly in $\beta \in \mathcal{B}$. This gives the required result that $\sqrt{T}(s_T^M(\beta_0) - s_T(\beta_0)) = \sqrt{T}O_p(M^{1-2n}) = O_p(T^{1/2+\delta(1-2n)}) = o_p(1)$ for $\delta > 1/(4n-2)$, justifying the condition on δ that was given in Assumption 5′ in Section 3.

(ii) Differentiating equation (A.3) again with respect to β_j yields

$$\frac{\partial^2 \ln\det[F_k^M(\beta)]}{\partial \beta_i \partial \beta_j} = \frac{\partial^2 \ln\det[F_k(\beta)]}{\partial \beta_i \partial \beta_j} + \psi_{1k}^M,$$

where $\psi_{1k}^M = O(M^{1-2n})$ uniformly in (λ_k, β). Differentiating equation (A.4) again with respect to β_j yields

$$\frac{\partial^2 \mathrm{tr}[F_k^M(\beta)^{-1}I_k]}{\partial \beta_i \partial \beta_j} = \frac{\partial^2 \mathrm{tr}[F_k(\beta)^{-1}I_k]}{\partial \beta_i \partial \beta_j} + \psi_{2k}^M,$$

where $\psi_{2k}^M = O_p(M^{1-2n})$ uniformly in (λ_k, β). Therefore

$$S_{T,ij}^M(\beta) = \frac{1}{T} \sum_{k \in \mathcal{K}_T} \left[\frac{\partial^2 \ln\det[F_k(\beta)]}{\partial \beta_i \partial \beta_j} + \frac{\partial^2 \mathrm{tr}[F_k(\beta)^{-1}I_k]}{\partial \beta_i \partial \beta_j} + \psi_{1k}^M + \psi_{2k}^M \right],$$

and so $S_{T,ij}^M(\beta) - S_{T,ij}(\beta) = O_p(M^{1-2n}) = O_p(T^{\delta(1-2n)}) = o_p(1)$ uniformly in $\beta \in \mathcal{B}$. Theorem 4.1.5 of Amemiya (1985) and the consistency of $\hat{\beta}_T$ and $\hat{\beta}_T^M$ then imply that $S_{T,ij}^M(\tilde{\beta})^{-1} - S_{T,ij}(\bar{\beta})^{-1} = o_p(1)$. ■

https://doi.org/10.1017/S0266466606060233 Published online by Cambridge University Press