

Tilburg University

## Frequency Domain Gaussian Estimation of Temporally Aggregated Cointegrated Systems

Chambers, M.J.; McCrorie, J.R.

*Publication date:*  
2004

[Link to publication in Tilburg University Research Portal](#)

*Citation for published version (APA):*

Chambers, M. J., & McCrorie, J. R. (2004). *Frequency Domain Gaussian Estimation of Temporally Aggregated Cointegrated Systems*. (CentER Discussion Paper; Vol. 2004-40). Vakgroep CentER.

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Center



# Discussion Paper

No. 2004–40

## **FREQUENCY DOMAIN GAUSSIAN ESTIMATION OF TEMPORALLY AGGREGATED COINTEGRATED SYSTEMS**

By M.J. Chambers, J.R. McCrorie

April 2004

ISSN 0924-7815

# Frequency Domain Gaussian Estimation of Temporally Aggregated Cointegrated Systems

Marcus J. Chambers  
*University of Essex*

and

J. Roderick McCrorie  
*University of Essex*

March 2004

## Abstract

This paper discusses the joint estimation of the long run equilibrium coefficients and the parameters governing the short run dynamics of a fully parametric cointegrated system formulated in continuous time. The model allows the stationary disturbances to be generated by a stochastic differential equation system and for the variables to be a mixture of stocks and flows. We derive a precise form for the exact discrete analogue of the continuous time model in triangular error correction form, which acts as the basis for frequency domain Gaussian estimation of the unknown parameters using discrete time data. We formally establish the order of consistency and the asymptotic sampling properties of such an estimator. The function of the data that estimates the cointegrating parameters is shown to converge at the rate of the sample size to a mixed normal distribution, while that estimating the short run parameters converges at the rate of the square root of the sample size to a limiting normal distribution.

**JEL Nos.:** C32; C51.

**Key words:** temporal aggregation; cointegration; continuous time; frequency domain; Gaussian estimation.

**Acknowledgements:** We would like to thank Roy Bailey, Rex Bergstrom, an Editor, and an anonymous referee for helpful comments on this paper, without implicating them in any possible shortcomings. In particular we are grateful to an anonymous referee for providing us with detailed comments on an earlier version of this paper. The first author thanks the Economic and Social Research Council (grant number R000221818) and the Leverhulme Trust (Philip Leverhulme Prize) for financial support. The second author's research was partially performed as a visiting fellow at CentER, Tilburg University.

**Address for Correspondence:** Professor Marcus J. Chambers, Department of Economics, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, England. Tel: +44 1206 872756. Fax: +44 1206 872724. E-mail: mchamb@essex.ac.uk.

## 1. Introduction

While it is recognised that the frequency with which time series data are observed is seldom within the control of the econometrician, the consequences for estimation and inference are often ignored. A dynamic model that is naïvely specified in terms of the observation interval can suffer from severe misspecification, with estimates being contaminated by temporal aggregation bias. This can arise owing to economic agents making decisions in finer time intervals than the sampling interval and the attendant problems of not sampling frequently enough to capture the movements of the economic variables. As a consequence it can be difficult, in practice, to offer an economic interpretation of parameter estimates, rather than just an interpretation of the observations; see Christiano and Eichenbaum (1987). Another aspect of the problem was expounded by Weiss (1984) who showed that the aggregation of a discrete time autoregressive moving average (ARMA) process results in a model that depends on the frequency with which the underlying process is observed.

One remedy to the above problem is to formulate the econometric model in continuous time and indeed Phillips (1991a) established that, in a temporally aggregated (continuous time) cointegrated system, the long run parameters can be estimated *directly* from a corresponding error correction model formulated in discrete time.<sup>1</sup> While this result is powerful, it is really only pertinent when the focus is on estimating long run equilibria rather than dynamic adjustment mechanisms, for it does not apply in the context of jointly estimating short run and long run effects which is very much in the spirit of the literature on estimating cointegrating systems. This is because the problem of estimating the parameters governing the short run dynamics would be subject to temporal aggregation bias in the way described above.

The purpose of this paper is to provide an analysis of estimating the temporally aggregated cointegrated system that allows the long run and short run parameters to be treated together. Kessler and Rahbek (2001) have offered a theoretical discussion based on continuously recorded data but here we provide an analysis more appropriate for econometric time series data, which are observed discretely. We base the cointegrated system on the continuous time triangular representation of Phillips (1991a), although in contrast to his non-parametric approach we model the disturbances explicitly as a continuous time autoregressive process (in the form of a stochastic differential equation system). While we could, in principle, use as the basis of estimation exact discrete time representations by Bergstrom (1997) and Chambers (1999) that are applicable to cointegrated systems, we prefer to estimate the autoregressive parameters in conjunction with the cointegrating parameters by maximising a frequency domain Gaussian likelihood function. The advantages of such an

---

<sup>1</sup>See also Stock (1987).

approach in the context of stationary systems are outlined by Robinson (1993), and Phillips (1991a), for example, uses spectral regression methods in a multivariate cointegrated system context very similar to our own.

There are two main contributions contained in this paper. The first is the derivation of the discrete time triangular error correction model (ECM) representation of the continuous time system. The model allows for the variables to be a mixture of stocks and flows, and the triangular version of the ECM assigns the system dynamics to the stationary disturbance term. A time domain representation is provided that relates the discrete time disturbance vector to the stationary disturbances in the continuous time model. This is used mainly to establish an invariance principle for the discrete time disturbances based on certain assumptions concerning the continuous time disturbances. A filtering equation is also derived that depicts the same relationship, and which is used to derive the spectral density function of the discrete time process.

The second, more substantial, contribution is the derivation of the consistency and asymptotic sampling properties of the frequency domain Gaussian estimator. This is not, in fact, a trivial problem but we have found that recent work by Saikkonen (1995, 2001) has been helpful in this regard. Firstly, as in stationary systems, we have to confront the problem of establishing uniform convergence of the likelihood function over the parameter space owing to the fact that our estimator is defined implicitly as the maximum of a function.<sup>2</sup> Unlike the stationary case, however, the likelihood diverges at different rates in different directions of the parameter space. Based on techniques in Saikkonen (1995) we are able to establish the different orders of consistency of the the estimator of the short run and long run parameter vectors. Furthermore, it is not appropriate to use directly a mean value expansion of the score vector to establish the limiting distribution of the estimator because of the way the Hessian matrix behaves in our more general context. Saikkonen (1995) showed, however, that the usual Taylor series expansion can be used provided the order of consistency of the long run parameter estimator can be derived and the Hessian can be shown to satisfy a certain stochastic equicontinuity condition. Here, however, we follow the approach in Saikkonen (2001) and work directly with the normalised score vector, incorporating in an essential way the previously established results on the order of consistency of the estimator and thereby avoiding the need to verify the required stochastic equicontinuity conditions. Furthermore, the frequency domain Gaussian estimator of the cointegrating parameters falls within the class of optimal estimators defined by Phillips (1991b).

---

<sup>2</sup>This problem does not arise, for example, in Corradi (1997) who analyses ‘comovements’ between diffusion processes. While her paper is similar to ours in that it uses the framework of Phillips (1991b) as the basis of deriving a triangular error correction model, the allowable dynamics are constrained there to be driven by Brownian motion. Here, we are implicitly following Phillips (1991a) that allows our dynamics to be driven by processes whose paths, at least in principle, have appropriate degrees of differentiability.

The results obtained are applicable to cointegrated continuous time vector autoregressive (VAR) processes of any (finite) order, and can be regarded as continuous time counterparts of the discrete time VECM approach popularized by Johansen (1991) and extended by Pesaran and Shin (2002). They also allow the observable data vector to comprise both stock variables, observable at points in time, and flow variables, observable as the integral of the underlying rate of flow over the observation interval. The coefficient matrices in the continuous time cointegrated system are also allowed to be known functions of an underlying unknown parameter vector. As a result, the dynamic responses, as well as the cointegrating relationships, may contain nonlinear restrictions on the coefficients of the type that often arise in economics.

The organization of the paper is as follows. Section 2 defines the model and derives its discrete time triangular ECM representation. Section 3 defines the frequency domain likelihood function and establishes some limiting distributional results that are used in the asymptotic analysis of the estimator. The consistency of the estimator is established in section 4, while section 5 derives the limiting distribution. Some further discussion of the methods and results is provided in section 6, along with some concluding comments. An appendix contains the proofs of all the lemmas and theorems presented in the paper.

Finally, the following notation is used in the paper.  $I_k$  denotes an identity matrix of dimension  $k \times k$ ,  $\det(A)$  and  $\text{tr}(A)$  denote the determinant and trace of a square matrix  $A$ , respectively, while  $\|A\| = [\text{tr}(AA^*)]^{1/2}$  denotes the Euclidean norm of  $A$ , where  $A^*$  denotes the complex conjugate transpose of a complex-valued matrix  $A$ . The notation  $\text{vec}(A)$  denotes the column vector obtained by stacking the columns of  $A$  vertically on top of each other. The symbols  $\Rightarrow$ ,  $\xrightarrow{p}$  and  $\xrightarrow{d}$  are used to denote weak convergence of probability measures, convergence in probability, and convergence in distribution, respectively. The integrals  $\int_0^1 SS'$  and  $\int_0^1 SdS'$  denote, respectively, the stochastic integrals  $\int_0^1 S(r)S(r)'dr$  and  $\int_0^1 S(r)dS(r)'$ , where  $S(r)$  is vector Brownian motion. Finally,  $A > 0$  denotes that the matrix  $A$  is positive definite, and  $[x]$  denotes the integer part of the scalar  $x$ .

## 2. The model and the ECM representation

The continuous time model of cointegration is defined by

$$y_1(t) = B(\theta_1)y_2(t) + u_1(t), \quad t > 0, \quad (1)$$

$$dy_2(t) = u_2(t)dt, \quad t > 0, \quad (2)$$

where  $y_1(t)$  and  $y_2(t)$  are continuous time random vectors of dimensions  $m_1 \times 1$  and  $m_2 \times 1$

respectively,  $B(\theta_1)$  is an  $m_1 \times m_2$  coefficient matrix whose elements are known functions of a  $p_1 \times 1$  vector  $\theta_1$  of unknown cointegrating parameters ( $p_1 \leq m_1 m_2$ ) belonging to a parameter space  $\Theta_1$ , and  $u_1(t)$  and  $u_2(t)$  are stationary continuous time random disturbance vectors whose dimensions are conformable with  $y_1$  and  $y_2$  respectively. The initial conditions  $y_1(0)$  and  $y_2(0)$  are taken to be fixed and to be known to be zero vectors. The long run cointegrating relationships between  $y_1$  and  $y_2$  are depicted in (1), while the zero roots in the system (corresponding to unit roots in discrete time) are captured by (2). Note that the specification of the cointegrating relationships in (1) allows for the possibility that the cointegrating parameters in  $\theta_1$  enter the model nonlinearly.

The dynamics that drive the cointegrated system stem from the stationary disturbance vector  $u(t) = [u_1(t)', u_2(t)']'$ , which is of dimension  $m \times 1$ , where  $m = m_1 + m_2$ . The dynamics for  $u(t)$  are assumed to be governed by the stochastic differential equation system

$$C(D)u(t)dt = db(t), \quad -\infty < t < \infty, \quad (3)$$

where  $C(z) = z^q I_m + \sum_{j=0}^{q-1} C_j(\beta)z^j$ ,  $C_0, \dots, C_{q-1}$  are  $m \times m$  coefficient matrices whose elements are known functions of a  $p_\beta \times 1$  vector  $\beta$  of unknown parameters ( $p_\beta \leq qm^2$ ) belonging to a parameter space  $\mathcal{B}$ , and  $D$  is the mean square differential operator. It is assumed that all the roots of the equation  $\det[C(z)] = 0$  have negative real parts so that the stochastic differential equation system is stable. Furthermore,  $db(t)$  represents the increment in the  $m \times 1$  vector Brownian motion process<sup>3</sup>  $b(t)$ , so that  $db(t) \sim N(0, \Sigma(\mu)dt)$  and  $E[db(t_1)db(t_2)'] = 0$  for  $t_1 \neq t_2$ , where  $\Sigma(\mu)$  is a symmetric positive definite matrix whose elements are known functions of a  $p_\mu \times 1$  vector  $\mu$  of unknown parameters ( $p_\mu \leq m(m+1)/2$ ) belonging to the parameter space  $\mathcal{M} = \{\mu : \Sigma(\mu) > 0\}$ . The unknown parameters may be combined, for convenience, into the  $p \times 1$  vector  $\theta = (\theta_1', \theta_2)'$ , where  $\theta_2 = (\beta', \mu)'$  is a  $p_2 \times 1$  vector ( $p_2 = p_\beta + p_\mu$ ), and  $p = p_1 + p_2$ . Hence  $\theta_1$  contains the long run (cointegrating) parameters while  $\theta_2$  contains the parameters that govern the short run dynamics.

It will be assumed that the vectors  $y_1$  and  $y_2$  are comprised of both stock and flow variables, there being  $m_j^S$  stock variables and  $m_j^F$  flow variables in the vector  $y_j$ , and where  $m_j^S + m_j^F = m_j$  ( $j = 1, 2$ ). Without loss of generality, each vector will be organised with the stock variables first, followed by the flow variables, so that

$$y_1(t) = [y_1^S(t)', y_1^F(t)']' \quad \text{and} \quad y_2(t) = [y_2^S(t)', y_2^F(t)']',$$

---

<sup>3</sup>The Gaussian assumption was made at the suggestion of the Editor following an earlier version of the paper which attempted to allow for possibly non-Gaussian distributions.

the superscripts ‘S’ and ‘F’ denoting stocks and flows respectively. The observed vectors are

$$y_{1t} = \left[ y_1^S(t)', \int_0^1 y_1^F(t-r)' dr \right]' \quad \text{and} \quad y_{2t} = \left[ y_2^S(t)', \int_0^1 y_2^F(t-r)' dr \right]'$$

for  $t = 1, \dots, T$ , where  $T$  denotes sample size. Stock variables are therefore observed at (integer) points in time, while flows are observed as the integral of the underlying rate of flow over the unit interval.

Although a number of approaches could be implemented for deriving the Gaussian likelihood function, the approach adopted here is based on the analytically-appealing triangular ECM representation of cointegrated systems advanced by Phillips (1991b), which lends itself readily to frequency domain likelihood methods. In what follows, it is convenient to partition the coefficient matrix  $B(\theta_1)$ , and the stationary disturbance in the ECM ( $\xi_t$ ), as follows:

$$B(\theta_1) = \begin{bmatrix} B_{SS}(\theta_1) & B_{SF}(\theta_1) \\ B_{FS}(\theta_1) & B_{FF}(\theta_1) \end{bmatrix}, \quad \xi_t = \begin{bmatrix} \xi_{1t} \\ \xi_{2t} \end{bmatrix}, \quad \xi_{1t} = \begin{bmatrix} \xi_{1t}^S \\ \xi_{1t}^F \end{bmatrix}, \quad \xi_{2t} = \begin{bmatrix} \xi_{2t}^S \\ \xi_{2t}^F \end{bmatrix}.$$

The sub-matrix  $B_{SF}(\theta_1)$ , for example, is of dimension  $m_1^S \times m_2^F$ , while  $\xi_{1t}^F$  is of dimension  $m_1^F \times 1$ . The vector  $u(t)$  is also partitioned conformably with  $y(t)$ .

**Lemma 1.** *Let  $y(t) = [y_1(t)', y_2(t)']'$  be generated by (1) and (2). Then  $y_t = [y_{1t}', y_{2t}']'$  satisfies the triangular ECM given by*

$$\Delta y_t = -JA(\theta_1)y_{t-1} + \xi_t, \quad t = 1, \dots, T, \quad (4)$$

where  $J = [I_{m_1}, 0]'$ ,  $A(\theta_1) = [I_{m_1}, -B(\theta_1)]$ , and  $\xi_t$  is related to  $u(t)$  as follows:

$$\begin{aligned} \xi_{1t}^S &= u_1^S(t) + B_{SS}(\theta_1) \int_0^1 u_2^S(t-s) ds \\ &\quad + B_{SF}(\theta_1) \left[ \int_0^1 u_2^F(t-s) ds + \int_0^1 (1-s) u_2^F(t-1-s) ds \right], \\ \xi_{1t}^F &= \int_0^1 u_1^F(t-s) ds + B_{FS}(\theta_1) \left[ \int_0^1 \int_0^1 u_2^S(t-r-s) dr ds \right. \\ &\quad \left. - \int_0^1 (1-s) u_2^S(t-1-s) ds \right] + B_{FF}(\theta_1) \int_0^1 \int_0^1 u_2^F(t-r-s) dr ds, \\ \xi_{2t}^S &= \int_0^1 u_2^S(t-s) ds, \\ \xi_{2t}^F &= \int_0^1 \int_0^1 u_2^F(t-r-s) dr ds. \end{aligned}$$



An equivalent representation is given by the filtering equation  $\xi_t = M(D, \theta_1)u(t)$ , where the matrix filter function  $M(z, \theta_1)$  is defined by

$$M(z, \theta_1) = \begin{bmatrix} I_{m_1^S} & 0 & g(z)B_{SS}(\theta_1) & [g(z) + h(z)]B_{SF}(\theta_1) \\ 0 & g(z)I_{m_1^F} & [g(z)^2 - h(z)]B_{FS}(\theta_1) & g(z)^2B_{FF}(\theta_1) \\ 0 & 0 & g(z)I_{m_2^S} & 0 \\ 0 & 0 & 0 & g(z)^2I_{m_2^F} \end{bmatrix},$$

and where  $g(z) = (1 - e^{-z})/z$  and  $h(z) = e^{-z}[1 - g(z)]/z$ .

The triangular ECM defined in Lemma 1 forms the basis for the estimation of the unknown parameter vector  $\theta$ . The time domain equations relating  $\xi_t$  to  $u(t)$  are used to establish an invariance principle for  $\xi_t$  in the next section, while the filtering equation relating  $\xi_t$  to  $u(t)$  is particularly useful for deriving the spectral density function of  $\xi_t$ . From (3), the spectral density function of the continuous time process  $u(t)$  is given by

$$f_u(\lambda, \theta_2) = \frac{1}{2\pi} C(i\lambda)^{-1} \Sigma [C(-i\lambda)^{-1}]', \quad -\infty < \lambda < \infty,$$

where  $i = \sqrt{-1}$  and  $i\lambda$  denotes the frequency response of the operator  $D$ . The dependence of  $f_u(\cdot)$  on  $\theta_2$  arises because  $C(\cdot)$  is a function of  $\beta$  and  $\Sigma$  is a function of  $\mu$ . It follows that the spectral density for  $\xi_t = M(D, \theta_1)u(t)$ , regarded as a continuous time process, is therefore<sup>4</sup>  $f^c(\lambda, \theta) = M(i\lambda, \theta_1)f_u(\lambda, \theta_2)M(-i\lambda, \theta_1)'$ . The spectral density for the discrete time process  $\xi_t$  is then obtained by folding all the frequencies on the real line back into the interval  $(-\pi, \pi]$  using the formula  $f(\lambda, \theta) = \sum_{j=-\infty}^{\infty} f^c(\lambda + 2\pi j, \theta)$ , yielding

$$f(\lambda, \theta) = \sum_{j=-\infty}^{\infty} M(i(\lambda + 2\pi j), \theta_1) f_u(\lambda + 2\pi j, \theta_2) M(i(\lambda + 2\pi j), \theta_1)^*, \quad -\pi < \lambda \leq \pi. \quad (5)$$

Methods for accurately computing doubly infinite series of the type defining  $f(\lambda)$  are given in Robinson (1993) and have been applied to spectral density functions arising from differential-difference equations by Chambers (1998).

It should be noted that the triangular ECM in (4) is not the only possible representation of the discrete time vector  $y_t$ . An explicit vector ARMA representation can also be derived from the stochastic differential equation system obtained from (1), (2) and (3), using the techniques of Chambers (1999). An example of a discrete time vector ARMA model derived from a mixed first- and second-order stochastic differential equation system with unobservable stochastic trends is provided by Bergstrom (1997). Such discrete time representations

<sup>4</sup>Note that, when  $\lambda = 0$ , we define  $M(0, \theta_1)$  using the limits  $\lim_{\lambda \rightarrow 0} g(i\lambda) = 1$  and  $\lim_{\lambda \rightarrow 0} h(i\lambda) = 1/2$ ; see Lemma A5 of Chambers (2003).

provide an alternative way of constructing the Gaussian likelihood function in the time domain but would rely on the inversion of an  $mT \times mT$  covariance matrix to compute the likelihood function.

### 3. The Gaussian estimator and some asymptotic results

The frequency domain Gaussian likelihood function we consider is, for finite  $T$ , an approximation to (ignoring a constant) minus twice the negative of the logarithm of the true Gaussian likelihood function

$$\Lambda_T(\theta) = \ln \det(V_\xi) + \xi' V_\xi^{-1} \xi, \quad (6)$$

where  $\xi' = [\xi'_1, \dots, \xi'_T]'$  and  $V_\xi = E(\xi\xi')$ . It is motivated by the fact that the sub-blocks of the block Toeplitz covariance matrix  $V_\xi$  are of the form  $V_{\xi,j} = E(\xi_t \xi'_{t+j}) = \int_{-\pi}^{\pi} f(\lambda, \theta) e^{ij\lambda} d\lambda$ . Defining  $\lambda_j = 2\pi j/T$  and  $J_T = \{j : -T/2 < j \leq [T/2]\}$ , the discrete Whittle likelihood is given by

$$L_T^\xi(\theta) = \sum_{j \in J_T} \left\{ \ln \det[f(\lambda_j, \theta)] + \text{tr}[f(\lambda_j, \theta)^{-1} I_\xi(\lambda_j)] \right\}, \quad (7)$$

where  $f(\lambda, \theta)$  is defined in (5) and  $I_\xi(\lambda) = w_\xi(\lambda)w_\xi(\lambda)^*$  is the periodogram of  $\xi_t$  in which  $w_\xi(\lambda) = (2\pi T)^{-1/2} \sum_{t=1}^T \xi_t e^{-it\lambda}$  denotes the discrete Fourier transform (dFt) of  $\xi_t$ .

The problem with implementing (7) in practice is that  $\xi_t$  is not observed. However, noting that  $\xi_t = \Delta y_t + JA(\theta_1)y_{t-1}$  enables us to work with dFts of observable variables by using

$$w(\lambda, \theta_1) = w_\Delta(\lambda) + J[w_1(\lambda) - B(\theta_1)w_2(\lambda)],$$

where  $w_\Delta(\lambda)$ ,  $w_1(\lambda)$  and  $w_2(\lambda)$  denote the dFts of  $\Delta y_t$ ,  $y_{1,t-1}$  and  $y_{2,t-1}$  respectively. In view of  $w_\xi(\lambda) = w(\lambda, \theta_1)$  we replace  $I_\xi(\lambda)$  in (7) with  $I(\lambda, \theta_1) = w(\lambda, \theta_1)w(\lambda, \theta_1)^*$  and therefore consider the frequency domain Gaussian likelihood function

$$L_T(\theta) = \sum_{j \in J_T} \left\{ \ln \det[f(\lambda_j, \theta)] + \text{tr}[f(\lambda_j, \theta)^{-1} I(\lambda_j, \theta_1)] \right\}, \quad (8)$$

which converges to (6) as  $T \rightarrow \infty$ . The frequency domain Gaussian estimator is consequently defined as

$$\hat{\theta}_T = \left( \hat{\theta}'_{1T}, \hat{\theta}'_{2T} \right)' = \arg \min_{\theta \in \Theta} L_T(\theta),$$

where  $\Theta = \Theta_1 \times \Theta_2$  denotes the parameter space for  $\theta$  and where the subvectors  $\theta_1$  and  $\theta_2$  belong to the sets  $\Theta_1$  and  $\Theta_2 = \mathcal{B} \times \mathcal{M}$  respectively. In what follows, the true value of the

parameter vector is denoted  $\theta_0 = (\theta'_{10}, \theta'_{20})'$ . It is also convenient to define  $w_0(\lambda) = w(\lambda, \theta_0)$ , which is the dFt of the stationary process  $\xi_{0t} = [\xi'_{01,t}, \xi'_{02,t}]' = \Delta y_t + JA(\theta_{10})y_{t-1}$ .

It is useful, at this stage, to present a number of results that are utilised in establishing the consistency and asymptotic distribution of  $\hat{\theta}_T$ . The first of these establishes an invariance principle for the partial sums of  $\xi_{0t}$ .

**Lemma 2.** *Let  $\xi_{0t} = \Delta y_t + JA(\theta_{10})y_{t-1}$ . Then  $T^{-1/2} \sum_{j=1}^{[Tr]} \xi_{0j} \Rightarrow S(r)$  as  $T \rightarrow \infty$ , where  $r \in [0, 1]$  and  $S(r)$  is an  $m \times 1$  Brownian motion process with covariance matrix  $\Omega = 2\pi f(0, \theta_0)$ .*

The validity of Lemma 2 is verified by showing that the conditions of Corollary 2.2 of Phillips and Durlauf (1986) are satisfied, thereby ensuring that the invariance principle holds. The main role of Lemma 2 is to establish the limiting behaviour of various functions of  $\xi_{0t}$ . Noting, from the ECM representation of  $y_t$  in Lemma 1, that  $\Delta y_{2t} = \xi_{02,t}$ , it follows that  $y_{2t} = \sum_{j=1}^t \xi_{02,j}$ , and so the limiting behaviour of the sample moments of  $y_{2t}$  can be derived straightforwardly using Lemma 2. It is also convenient to partition the Brownian motion process as  $S(r) = [S_1(r)', S_2(r)']'$  and to partition  $\Omega$  conformably with  $S_1$  and  $S_2$ .

**Lemma 3.** *The following sample moments converge jointly as  $T \rightarrow \infty$ :*

$$\begin{aligned} (a) \quad & T^{-1} \sum_{t=1}^{T-s} \xi_{0t} \xi'_{0,t+s} \xrightarrow{p} \Gamma_s; \\ (b) \quad & T^{-1} \sum_{t=1}^{T-s} y_{2t} \xi'_{0,t+s} \Rightarrow \int_0^1 S_2 dS' + \sum_{j=s}^{\infty} \Gamma'_{2j}; \\ (c) \quad & T^{-2} \sum_{t=1}^T y_{2t} y'_{2t} \Rightarrow \int_0^1 S_2 S_2', \end{aligned}$$

where  $\Gamma_s = [\Gamma_{1s} : \Gamma_{2s}] = E[\xi_{00} \xi'_{0s}]$ .

The convergence of the sample moments depicted in Lemma 3 is now standard in the asymptotic theory of multivariate integrated processes; see, for example, Phillips and Durlauf (1986). These results are used here to derive the limiting distributions of various functions of such sample moments. In particular, we need to establish certain uniform convergence results for weighted sums of periodogram estimates. The precise uniform convergence results

that we require are presented in Lemma 4.<sup>5</sup>

**Lemma 4.** *Let  $I_{00,j} = w_0(\lambda_j)w_0(\lambda_j)^*$ ,  $I_{02,j} = w_0(\lambda_j)w_2(\lambda_j)^*$ ,  $I_{22,j} = w_2(\lambda_j)w_2(\lambda_j)^*$ , and partition  $f(\lambda, \theta)$  as  $f(\lambda, \theta) = [f_1(\lambda, \theta) : f_2(\lambda, \theta)]$ . Furthermore, let  $\phi(\lambda, \theta)$  denote a complex-valued matrix function that is continuous in  $\lambda$ , for every  $\theta \in \Theta$ , and continuous in  $\theta$ , for every  $\lambda \in (-\pi, \pi]$ , and define  $\phi_j(\theta) = \phi(\lambda_j, \theta)$ . Then, as  $T \rightarrow \infty$ :*

$$(a) \sup_{\theta \in \Theta} \left\| T^{-1} \sum_{j \in J_T} \phi_j(\theta) \text{vec} [I_{00,j}] - \frac{1}{2\pi} \int_{-\pi}^{\pi} \phi(\lambda, \theta) \text{vec} [f(\lambda, \theta_0)] d\lambda \right\| \xrightarrow{p} 0,$$

$$(b) \sup_{\theta \in \Theta} \left\| T^{-1} \sum_{j \in J_T} \phi_j(\theta) I_{02,j} - \frac{1}{2\pi} \phi(0, \theta) \int_0^1 dS S_2' - \bar{g}(\theta) \right\| \Rightarrow 0,$$

$$(c) \sup_{\theta \in \Theta} \left\| T^{-2} \sum_{j \in J_T} \left( J' \phi_j(\theta) J \otimes I_{22,j}' \right) - \left( \frac{1}{2\pi} J' \phi(0, \theta) J \otimes \int_0^1 S_2 S_2' \right) \right\| \Rightarrow 0,$$

where  $\bar{g}(\theta) = (1/2\pi) \int_{-\pi}^{\pi} \phi(\lambda, \theta) \sum_{q=1}^{\infty} e^{iq\lambda} f_2(\lambda, \theta_0) d\lambda$ .

Lemma 4 is a key result used primarily in establishing the limiting properties of components of  $I(\lambda, \theta_1)$  and its derivatives. Its proof is based on an important and fundamental result of Robinson (1976) whose Theorem 1 established almost sure convergence of similar quantities in the stationary case. Lemma 4 extends Robinson's result by allowing for a different mode of convergence and for integrated variables. Additional assumptions and results that are used in establishing consistency and the limiting distribution of  $\hat{\theta}_T$  will be presented as required in the following sections.

#### 4. Consistency of $\hat{\theta}_T$

Establishing the consistency of the frequency domain Gaussian estimator in cointegrated models is more difficult than in stationary models in which uniform convergence of the likelihood function over the parameter space is a key ingredient. As pointed out by Saikkonen (1995), whose method of proof we broadly follow, this is not a feature of the likelihood in cointegrated models, because the likelihood converges to limiting values at different rates in different directions of the parameter space (corresponding to  $\theta_1$  and  $\theta_2$ , the long run and short run parameters, respectively). In fact, we demonstrate that  $\hat{\theta}_{1T} - \theta_{10} = o_p(T^{-\gamma})$  for  $0 < \gamma < 1$  and that  $\hat{\theta}_{2T} - \theta_{20} = o_p(1)$ . Note that the requirement is somewhat stronger for the long run parameter vector  $\theta_1$  than for the vector of short run parameters  $\theta_2$ . Saikkonen

---

<sup>5</sup>Note that the dimension of the function  $\phi(\lambda, \theta)$  defined in Lemma 4 can vary across parts (a)–(c) to ensure conformability of the relevant products.

(1995) shows<sup>6</sup> that a sufficient condition for  $\hat{\theta}_{1T} - \theta_{10} = o_p(T^{-\gamma})$  is that, for every  $\delta > 0$ ,

$$\lim_{T \rightarrow \infty} \Pr \left\{ \inf_{\theta \in \bar{N}_{T,\gamma}(\theta_{10}, \delta) \times \Theta_2} [L_T(\theta) - L_T(\theta_0)] > 0 \right\} = 1, \quad (9)$$

where  $\bar{N}_{T,\gamma}(\theta_{10}, \delta) = \{\theta_1 \in \Theta_1 : \|\theta_1 - \theta_{10}\| \geq \delta/T^\gamma\}$  is the complement of an open ball in  $\mathcal{R}^{p_1}$  with centre  $\theta_{10}$  and radius  $\delta/T^\gamma$ . The order of consistency is therefore determined by the rate at which the radius of the ball tends to zero as  $T \rightarrow \infty$ . The consistency of  $\hat{\theta}_{2T}$  can be established by showing<sup>7</sup> that, for every  $\delta > 0$ ,

$$\lim_{T \rightarrow \infty} \Pr \left\{ \inf_{\theta \in \Theta_1 \times \bar{B}(\theta_{20}, \delta)} [L_T(\theta) - L_T(\theta_0)] > 0 \right\} = 1, \quad (10)$$

where  $\bar{B}(\theta_{20}, \delta) = \{\theta_2 \in \Theta_2 : \|\theta_2 - \theta_{20}\| \geq \delta\}$  denotes the complement of an open ball of radius  $\delta$  centered at  $\theta_{20}$ . In fact, if (9) is satisfied, it is sufficient to show that (10) holds with  $\Theta_1$  replaced with  $N_{T,\gamma}(\theta_{10}, \delta_1) = \{\theta_1 \in \Theta_1 : \|\theta_1 - \theta_{10}\| < \delta_1/T^\gamma\}$ , where  $\delta_1$  can be chosen freely; see Saikkonen (1995, p. 905). Conditions (9) and (10) will be demonstrated in turn.

In what follows, it is convenient to write  $f_j(\theta) = f(\lambda_j, \theta)$  and  $I_j(\theta_1) = I(\lambda_j, \theta_1)$ , so that the likelihood function becomes

$$L_T(\theta) = \sum_{j \in J_T} \left\{ \ln \det[f_j(\theta)] + \text{tr}[f_j(\theta)^{-1} I_j(\theta_1)] \right\}. \quad (11)$$

The difference  $L_T(\theta) - L_T(\theta_0)$  can then be written

$$\begin{aligned} L_T(\theta) - L_T(\theta_0) &= \sum_{j \in J_T} \{ \ln \det[f_j(\theta)] - \ln \det[f_j(\theta_0)] \} \\ &\quad + \sum_{j \in J_T} \left\{ \text{tr}[f_j(\theta)^{-1} I_j(\theta_1)] - \text{tr}[f_j(\theta_0)^{-1} I_j(\theta_{10})] \right\} \\ &= A_T(\theta, \theta_0) + B_T(\theta, \theta_{10}) + C_T(\theta, \theta_0), \end{aligned} \quad (12)$$

where, noting that  $I_j(\theta_{10}) = I_{00,j}$ ,

$$A_T(\theta, \theta_0) = \sum_{j \in J_T} \{ \ln \det[f_j(\theta)] - \ln \det[f_j(\theta_0)] \}, \quad (13)$$

$$B_T(\theta, \theta_{10}) = \sum_{j \in J_T} \text{tr} \left\{ f_j(\theta)^{-1} [I_j(\theta_1) - I_{00,j}] \right\}, \quad (14)$$

$$C_T(\theta, \theta_0) = \sum_{j \in J_T} \text{tr} \left\{ [f_j(\theta)^{-1} - f_j(\theta_0)^{-1}] I_{00,j} \right\}. \quad (15)$$

<sup>6</sup>See equation (26) of Saikkonen (1995).

<sup>7</sup>See equation (31) of Saikkonen (1995).

Furthermore, the difference  $I_j(\theta_1) - I_{00,j}$  in (14) has the convenient representation

$$\begin{aligned} I_j(\theta_1) - I_{00,j} &= -I_{02,j} [B(\theta_1) - B(\theta_{10})]' J' - J [B(\theta_1) - B(\theta_{10})] I_{02,j}^* \\ &\quad + J [B(\theta_1) - B(\theta_{10})] I_{22,j} [B(\theta_1) - B(\theta_{10})]' J'. \end{aligned} \quad (16)$$

In order to examine (13), (14) and (15) in more detail, the following further assumptions are also made.

**Assumption 1.** The parameter space  $\Theta$  is a compact subset of  $\mathcal{R}^p$  and  $\theta_0 \in \Theta$ .

**Assumption 2.** (a) The elements of the matrices  $C_j(\beta)$ ,  $\Sigma(\mu)$  and  $B(\theta_1)$  are continuously differentiable functions of  $\beta \in \mathcal{B}$ ,  $\mu \in \mathcal{M}$  and  $\theta_1 \in \Theta_1$  respectively. (b) For  $\theta \neq \theta_0$ ,  $f(\lambda, \theta) \neq f(\lambda, \theta_0)$  on a subset of  $(-\pi, \pi]$  having positive Lebesgue measure.

**Assumption 3.** Let  $F(\bar{\theta}_1)$  be the  $m_1 m_2 \times p$  matrix whose  $i$ 'th row is equal to the  $i$ 'th row of the matrix  $\partial \text{vec}[B(\theta_1)'] / \partial \theta_1'$  evaluated at  $\bar{\theta}_{1,i} = \alpha_i \theta_1 + (1 - \alpha_i) \theta_{10}$  for some  $0 < \alpha_i < 1$  ( $i = 1, \dots, m_1 m_2$ ) and  $\bar{\theta}_1 = [\bar{\theta}_{1,1}, \dots, \bar{\theta}_{1,m_1 m_2}]$ . Then  $F(\bar{\theta}_1)$  has full column rank  $p_1$  for all  $\bar{\theta}_1 \in \Theta_1^{m_1 m_2}$ , where  $\Theta_1^{m_1 m_2} = \Theta_1 \times \dots \times \Theta_1$  ( $m_1 m_2$  times).

Assumption 1 is a standard assumption in consistency proofs, while Assumption 2(a) ensures that the spectral density function  $f(\lambda, \theta)$  is differentiable throughout the parameter space. It is also positive definite on  $\Theta$  owing to the definition of the subspace  $\mathcal{M}$ . Assumption 2(b) is an identification requirement but one that is typically difficult to verify in continuous time models owing to aliasing effects; see Phillips (1973), Hansen and Sargent (1983) and McCrorie (2003). Note, however, that the elements of the matrix  $B(\theta_1)$  are not subject to the usual aliasing problem; see Phillips (1991a) for details. Finally, Assumption 3 imposes a rank condition on a matrix of partial derivatives that arises in a mean value expansion of  $\text{vec}[B(\theta_1)']$ . A similar assumption on the matrix of cointegrating vectors can be found in Assumption 1 of Saikkonen (2001).

**Theorem 1.** Under Assumptions 1–3, as  $T \rightarrow \infty$ ,  $\hat{\theta}_{1T} - \theta_{10} = o_p(T^{-\gamma})$  for  $0 < \gamma < 1$  and  $\hat{\theta}_{2T} - \theta_{20} = o_p(1)$ .

Theorem 1 establishes not only the consistency (in the usual sense) of the estimator  $\hat{\theta}_T$ , but also the order of consistency of  $\hat{\theta}_{1T}$ , the estimator of the long run cointegrating parameters. A key input into the proof of Theorem 1 for  $\hat{\theta}_{1T}$  concerns the asymptotics of the elements of the decomposition of the difference  $L_T(\theta) - L_T(\theta_0)$  given in (13), (14) and (15). This result is stated as Lemma B in the Appendix. The result for  $\hat{\theta}_{2T}$  is based on an alternative decomposition of the difference  $L_T(\theta) - L_T(\theta_0)$  that enables (10) to be verified. The orders of consistency of  $\hat{\theta}_{1T}$  and  $\hat{\theta}_{2T}$  are important in establishing the asymptotic distribution of the estimator, as will be seen in the next section.

## 5. Asymptotic distribution of $\hat{\theta}_T$

The usual approach to deriving the limiting distribution of a normalised optimisation estimator is based on a mean value expansion of the normalised score vector, defined by

$$s_T(\theta) = \begin{pmatrix} s_{T1}(\theta) \\ s_{T2}(\theta) \end{pmatrix} = \begin{pmatrix} (1/T)\partial L_T(\theta)/\partial\theta_1 \\ (1/\sqrt{T})\partial L_T(\theta)/\partial\theta_2 \end{pmatrix}.$$

The mean value expansion then yields an expression of the form

$$\left(T(\hat{\theta}_{1T} - \theta_{10})', \sqrt{T}(\hat{\theta}_{2T} - \theta_{20})'\right)' = J_T(\bar{\theta})^{-1} s_T(\theta_0), \quad (17)$$

where  $J_T(\bar{\theta})$  is the normalised Hessian matrix evaluated at the mean value points. Usually, the consistency of  $\hat{\theta}_T$  and the continuity of the Hessian ensure that  $J_T(\bar{\theta}) \xrightarrow{p} J(\theta_0)$  as  $T \rightarrow \infty$ , which, allied to the convergence of  $s_T(\theta_0)$  to  $s(\theta_0)$  and the establishment of the distribution of the latter vector, yields the limiting distribution of  $\hat{\theta}_T$ . In the current situation, however, the Hessian is not sufficiently smooth for the above arguments to be valid, although a similar approach can be employed, provided that the normalised Hessian satisfies a stochastic equicontinuity condition, which can be difficult to verify; see Saikkonen (1995) for details.

In this paper, an alternative approach is followed, based on the score vector directly and inspired by the techniques employed by Phillips (1991b) and Saikkonen (2001). The advantage of this approach is that it avoids the need to establish a certain stochastic equicontinuity condition for the Hessian, and relies more directly on the previously established orders of consistency of the estimators of the short-run and long-run parameters.

**Assumption 4.** The elements of the matrices  $C_j(\beta)$  ( $j = 0, \dots, q-1$ ),  $\Sigma(\mu)$  and  $B(\theta_1)$  are twice continuously differentiable functions of  $\beta \in \mathcal{B}$ ,  $\mu \in \mathcal{M}$  and  $\theta_1 \in \Theta_1$  respectively.

Assumption 4 extends Assumption 2 to second-order differentiability of the relevant matrices which ensures that the second derivatives of  $f(\lambda, \theta)$  exist. The limiting distribution of  $\hat{\theta}_T$  is presented in Theorem 2.

**Theorem 2.** *Under Assumptions 1–4, as  $T \rightarrow \infty$ ,*

$$\begin{aligned} T(\hat{\theta}_{1T} - \theta_{10}) &\Rightarrow \left[ \left( \frac{\partial \text{vec } B(\theta_{10})}{\partial \theta'_1} \right)' \left( \int_0^1 S_2 S_2' \otimes \Omega_{11.2}^{-1} \right) \left( \frac{\partial \text{vec } B(\theta_{10})}{\partial \theta'_1} \right) \right]^{-1} \\ &\quad \times \left( \frac{\partial \text{vec } B(\theta_{10})}{\partial \theta'_1} \right)' \left( I_{m_2} \otimes \Omega_{11.2}^{-1} \right) \left( \int_0^1 S_2 \otimes dS_{1.2} \right), \\ \sqrt{T}(\hat{\theta}_{2T} - \theta_{20}) &\xrightarrow{d} N \left( 0, 2V(\theta_0)^{-1} \right), \end{aligned}$$

where  $\Omega_{11.2} = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}$ ,  $\partial \text{vec } B(\theta_{10})/\partial \theta'_1$  denotes the matrix  $\partial \text{vec } B(\theta_1)/\partial \theta'_1$  evaluated at  $\theta_1 = \theta_{10}$ , and the  $(k, l)$  element of  $V(\theta)$  is given by

$$V_{kl}(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \text{tr} \left\{ f(\lambda, \theta)^{-1} \frac{\partial f(\lambda, \theta)}{\partial \theta_{2k}} f(\lambda, \theta)^{-1} \frac{\partial f(\lambda, \theta)}{\partial \theta_{2l}} \right\} d\lambda \quad (k, l = 1, \dots, p_2).$$

The limiting distribution of  $\hat{\theta}_{1T}$ , the estimator of the long run cointegrating parameters, is seen to be the familiar mixed normal distribution. It is identical to the limiting distribution of the spectral regression estimator of cointegrating parameters in continuous time systems given<sup>8</sup> in Phillips (1991a). It is also a member of the class of ‘optimal’ estimators as classified by Phillips (1991b), and is asymptotically efficient; see Saikkonen (1991) for details. Note that efficiency is obtained here by the correct parametric modelling of the dynamics, whereas the spectral regression estimators of Phillips (1991a) account for the dynamics nonparametrically in the frequency domain. The limiting distribution of  $\hat{\theta}_{2T}$ , the estimator of the short run dynamic parameters, corresponds to that of the Gaussian estimator of parameters in correctly specified parametric stationary time series models; see, for example, Dunsmuir (1979).

## 6. Discussion and concluding comments

Our concern in this paper has been with the derivation of the asymptotic properties of the frequency domain Gaussian estimator of the parameters in a temporally aggregated

---

<sup>8</sup>Note that Phillips (1991a) uses row, rather than column, vectorisation, so that the representation of the distribution in that paper is slightly different to that given here.



cointegrated system. The underlying model is written as a triangular system in continuous time, with the system dynamics driven by a continuous time VAR( $q$ ) in the form of a stochastic differential equation system of order  $q$ . In Lemma 1 we have shown that the discrete time observations also satisfy a triangular ECM, and the complicated form of the dynamics of the resulting disturbance vector is a key motivation behind our use of the frequency domain likelihood function, the dynamics effectively being represented by the spectral density function.

Alternative time domain approaches are possible. Combining (1) and (2) we may write  $dy(t) = -JAy(t)dt + w(t)dt$ , where the matrices  $J$  and  $A$  are defined in Lemma 1 and  $w(t)$  depends on  $u(t)$ ; see equation (3) of Chambers (2003). Writing this as  $(DI_m + JA)y(t)dt = w(t)dt$  and assuming that  $w(t)$  may be represented as a continuous time VAR( $q$ ) process of the form  $\Psi(D)w(t)dt = db(t)$ , with  $\Psi(z) = z^q I_m + \sum_{j=0}^{q-1} \Psi_j z^j$ , we obtain the system  $\Psi(D)(DI_m + JA)y(t)dt = db(t)$  which may be written

$$d[D^q y(t)] = \sum_{j=0}^q F_j(\theta) D^j y(t)dt + db(t),$$

where  $F_0 = -\Psi_0 JA$ ,  $F_j = -(\Psi_{j-1} + \Psi_j JA)$  ( $j = 1, \dots, q-1$ ), and  $F_q = -(\Psi_{q-1} + JA)$ . The parallels with discrete time cointegrated VARs are apparent. The reduced rank due to cointegration is evident in the matrix  $F_0$ . It is possible to show that the observed vector  $y_t$  satisfies the discrete time cointegrated VARMA system

$$\Delta y_t = \Phi_0 y_{t-1} + \sum_{j=1}^q \Phi_j \Delta y_{t-j} + \eta_t,$$

where  $\Phi_0$  has rank  $m_1$  and  $\eta_t$  is an MA( $q+1$ ) disturbance process. This discrete time VARMA system is extremely parsimonious compared to an unrestricted discrete time VARMA and is capable of producing a richer dynamic structure than a pure VAR in discrete time.

The results obtained in this paper are applicable more widely than to temporally aggregated cointegrated systems. They can also be applied (with suitable modification) to cointegrated models formulated directly in discrete time for which the triangular ECM representation is valid. It is also possible to exclude frequency bands, for example some seasonal frequencies, over which the model might not be felt to be entirely appropriate. The likelihood function would then be defined not over the entire set of frequencies  $J_T$  but over a restricted set  $B_T \subset J_T$ . Band-limited methods have been proposed by Hannan and Robinson (1973) and Robinson (1976) for stationary continuous time systems, by Phillips (1991a,1991c) for (discrete and continuous time) cointegrated systems, and by Corbae, Ouliaris and Phillips (2002) for stationary and nonstationary trending data. Subject to appropriate modifications, our results will continue to hold in this setup.

## Appendix

**Proof of Lemma 1.** As shown by Phillips (1991a), the partially observable vector  $y(t)$  generated by (1) and (2) satisfies the discrete time ECM  $\Delta y(t) = -JAy(t-1) + x(t)$ , where  $x(t) = [x_1(t)', x_2(t)']'$  is related to  $u(t)$  by the equations  $x_1(t) = u_1(t) + B \int_0^1 u_2(t-s)ds$  and  $x_2(t) = \int_0^1 u_2(t-s)ds$ ; see Lemma A1 of Chambers (2003) for details. Re-writing the ECM in more detail gives

$$\Delta y_1^S(t) = -y_1^S(t-1) + B_{SS}y_2^S(t-1) + B_{SF}y_2^F(t-1) + x_1^S(t), \quad (18)$$

$$\Delta y_1^F(t) = -y_1^F(t-1) + B_{FS}y_2^S(t-1) + B_{FF}y_2^F(t-1) + x_1^F(t), \quad (19)$$

$$\Delta y_2^S(t) = x_2^S(t), \quad (20)$$

$$\Delta y_2^F(t) = x_2^F(t). \quad (21)$$

In (18), the only unobservable variable (excluding the random disturbance) is  $y_2^F(t-1)$ . Adding and subtracting  $B_{SF}y_{2,t-1}^F$  yields  $\Delta y_{1t}^S = -y_{1,t-1}^S + B_{SS}y_{2,t-1}^S + B_{SF}y_{2,t-1}^F + \xi_{1t}^S$ , where  $\xi_{1t}^S = x_1^S(t) + B_{SF}[y_2^F(t-1) - y_{2,t-1}^F]$ . From the definition of  $x_1(t)$ , and noting from Lemma A2 of Chambers (2003) that  $y_2^F(t-1) - y_{2,t-1}^F = \int_0^1 (1-s)u_2^F(t-1-s)ds$ , yields the expression for  $\xi_{1t}^S$  in the lemma. Integrating (19) over  $[0, 1]$  gives

$$\begin{aligned} \Delta y_{1t}^F &= -y_{1,t-1}^F + B_{FS} \int_0^1 y_2^S(t-1-s)ds + B_{FF}y_{2,t-1}^F + \int_0^1 x_1^F(t-s)ds \\ &= -y_{1,t-1}^F + B_{FS}y_{2,t-1}^S + B_{FF}y_{2,t-1}^F + \xi_{1t}^F, \end{aligned}$$

the second expression being obtained by adding and subtracting  $B_{FS}y_2^S(t-1) = B_{FS}y_{2,t-1}^S$  and where  $\xi_{1t}^F = \int_0^1 x_1^F(t-s)ds + B_{FS} \left[ \int_0^1 y_2^S(t-1-s)ds - y_2^S(t-1) \right]$ . The expression for  $\xi_{1t}^F$  in the lemma follows from the definition of  $x_1(t)$  above and Lemma A2 in Chambers (2003). Equation (20) readily yields  $\Delta y_{2t}^S = \xi_{2t}^S$ , with  $\xi_{2t}^S = x_2^S(t)$  following from the definition of  $x_2(t)$ , while integrating (21) over  $[0, 1]$  yields  $\Delta y_{2t}^F = \xi_{2t}^F$ , where the expression for  $\xi_{2t}^F$  in the lemma comes from integrating  $x_2^F(t)$ . Finally, the filtering equation  $\xi_t = M(D)u(t)$  arises straightforwardly because  $\int_0^1 u(t-s)ds = g(D)u(t)$ ,  $\int_0^1 \int_0^1 u(t-r-s)drds = g(D)^2u(t)$ , and  $\int_0^1 (1-s)u(t-1-s)ds = h(D)u(t)$ ; see Lemma A4 of Chambers (2003).  $\square$

**Proof of Lemma 2.** The proof establishes that the conditions of Corollary 2.2 of Phillips and Durlauf (1986) are satisfied, which ensures that the stated invariance principle holds. First, note that (3) may be written as  $dv(t) = Cv(t)dt + db_v(t)$  ( $-\infty < t < \infty$ ), where  $v(t) = [u(t)', Du(t)', \dots, D^{q-1}u(t)']'$ ,  $C$  is the associated companion matrix whose eigenvalues have negative real parts, and  $db_v(t) = [0, \dots, 0, db(t)']'$ . Note that  $u(t) = S_u v(t)$ , where  $S_u = [I_m, 0, \dots, 0]$  is the selection matrix that picks out  $u(t)$  from  $v(t)$ . Since

$v(t) = \int_{-\infty}^t e^{C(t-r)} db_v(r)$ , where  $e^{rC} = \sum_{j=0}^{\infty} (rC)^j / j!$ , it follows that  $v(t)$ , and hence  $u(t)$ , are Gaussian. In fact,  $u(t) \sim N\left(0, S_u \int_0^{\infty} e^{Cr} \Sigma e^{C'r} dr S_u'\right)$ . Furthermore,

$$E[u(t)u(t+m)'] = S_u \int_0^{\infty} e^{Cr} \Sigma e^{C'r} dr e^{C'm} S_u', \quad m > 0,$$

which decays exponentially with  $m$  by virtue of  $C$  having eigenvalues with negative real parts. Now, Lemma 1 establishes that  $\xi_t$  is a measurable function of  $u(t)$  over a finite interval and hence it inherits the same mixing properties; this follows from Theorem 14.1 of Davidson (1994). Furthermore,  $\xi_t$  is Gaussian and inherits the exponential decay of autocorrelations depicted above. Then, from Rozanov(1967, pp.181 and 186), the maximal linear correlation coefficient of  $\xi_t$ , which is also equal to the maximal correlation coefficient due to the Gaussian nature of  $\xi_t$ , also decays exponentially. But the latter coefficient bounds the strong mixing numbers from above (see, for example, Davidson, 1994, p.209) and hence the mixing decay rate condition of Corollary 2.2 of Phillips and Durlauf (1986) is satisfied. The remaining conditions that need to be fulfilled are that  $E[\xi_t] = 0$  and  $E|\xi_{it}|^{\gamma} < \infty$  ( $i = 1, \dots, m$ ) for some  $2 \leq \gamma < \infty$ ; these are trivially satisfied because  $\xi_t$  is a zero mean Gaussian process.  $\square$

**Proof of Lemma 3.** The stated limiting properties follow from the now well-established asymptotic theory for multivariate integrated processes. See, for example, Phillips and Durlauf (1986).  $\square$

The proof of Lemma 4 relies on the following additional result.

**Lemma A.** *Let  $w_0(\lambda) = [w_{01}(\lambda)', w_{02}(\lambda)']'$ . Then*

$$w_2(\lambda) = \frac{1}{e^{i\lambda} - 1} w_{02}(\lambda) - \frac{1}{e^{i\lambda} - 1} \left( \frac{e^{-iT\lambda} y_{2,T} - y_{2,0}}{\sqrt{2\pi T}} \right).$$

**Proof.** Similar to Lemma B of Corbae, Ouliaris and Phillips (2002). Taking dFt's of the equation  $y_{2,t} = y_{2,t-1} + \xi_{02,t}$  gives  $(1/\sqrt{2\pi T}) \sum_{t=1}^T y_{2t} e^{-it\lambda} = w_2(\lambda) + w_{02}(\lambda)$ . Noting that  $(1/\sqrt{2\pi T}) \sum_{t=1}^T y_{2t} e^{-it\lambda} = e^{i\lambda} w_2(\lambda) + (e^{-iT\lambda} y_{2,T} - y_{2,0})/\sqrt{2\pi T}$  yields the result upon substitution and rearrangement.  $\square$

**Proof of Lemma 4.** The proof follows that of Theorem 1 of Robinson (1976) which itself extends results by Hannan and Robinson (1973) and Jennrich (1969). Part (a) follows immediately as an 'in probability' version of Robinson's result but for parts (b) and (c) we need to modify the proof slightly to account for the non-stationarity of the data. We begin by establishing pointwise convergence with  $\phi(\lambda, \theta)$  periodic of period  $2\pi$ . Pointwise convergence then applies under the more general hypothesis of the theorem by applying the technique used by Robinson (1976, pp.232–233). Finally, we demonstrate that the convergence is uniform.

(b) Let  $g_T(\theta, \phi) = T^{-1} \sum_j \phi_j(\theta) I_{02,j}$ . By Fejér's theorem, every continuous periodic function

$\phi : \mathcal{R} \rightarrow \mathcal{C}$  can be uniformly approximated by trigonometric polynomials and, in particular, by the Cesàro sum of its finite Fourier series representation. The  $M$ th first-order Cesàro mean, given by

$$\phi_M(\lambda, \theta) = (2\pi)^{-1} \int_{-\pi}^{\pi} \phi(\mu, \theta) F_M(\lambda - \mu) d\mu,$$

therefore converges uniformly in  $\lambda$  to  $\phi$ , where  $F_M(\lambda) = \sum_{|k| < M} (1 - |k|/M) e^{ik\lambda}$  denotes Fejér's kernel. Then, for given  $\epsilon > 0$ ,  $\sup_{\lambda} \|\phi(\lambda, \theta) - \phi_M(\lambda, \theta)\| < \epsilon$  for  $M$  sufficiently large. Defining  $w_{0j} = w(\lambda_j, \theta_0)$  and  $w_{2j} = w_2(\lambda_j)$ , it follows that

$$\|g_T(\theta, \phi) - g_T(\theta, \phi_M)\| < \frac{\epsilon}{T} \sum_j \|I_{02,j}\| \leq \frac{\epsilon}{T} \sum_j \|w_{0j}\| \|w_{2j}\| = O_p(\epsilon) \quad (22)$$

using Lemma A. Since  $\epsilon$  is arbitrary, we can replace  $\phi$  by  $\phi_M$ . Noting that

$$I_{02,j} = (2\pi T)^{-1} \sum_{m=1}^T \sum_{n=1}^T \xi_{0m} y'_{2,n-1} e^{i(n-m)\lambda_j},$$

$g_T(\theta, \phi_M)$  has the representation

$$\begin{aligned} & \frac{1}{2\pi T} \sum_k \sum_m \sum_n \sum_j \int_{-\pi}^{\pi} \left(1 - \frac{|k|}{M}\right) \phi(\lambda, \theta) \xi_{0m} y'_{2,n-1} e^{i(n-m+k)\lambda_j - ik\lambda} d\lambda \\ &= \left(\frac{1}{2\pi}\right)^2 \sum_k \left[ \int_{-\pi}^{\pi} \phi(\lambda, \theta) \left(1 - \frac{|k|}{M}\right) e^{-ik\lambda} d\lambda \right] \widehat{\Gamma}_{02,k+1} + o_p(1) \end{aligned} \quad (23)$$

for  $T > M$ , where  $\widehat{\Gamma}_{02,k} = T^{-1} \sum_{t=1}^{T-k} \xi_{0t} y'_{2,t+k}$ . By Lemma 3(b),  $g_T(\theta, \phi_M) \Rightarrow g(\theta, \phi_M)$ , where

$$g(\theta, \phi_M) = \frac{1}{2\pi} \phi_M(0, \theta) \int_0^1 dSS'_2 + \bar{g}(\theta, \phi_M), \quad (24)$$

and

$$\begin{aligned} \bar{g}(\theta, \phi_M) &= \left(\frac{1}{2\pi}\right)^2 \int_{-\pi}^{\pi} \phi(\lambda, \theta) \sum_k \left(1 - \frac{|k|}{M}\right) e^{-ik\lambda} d\lambda \sum_{j=k+1}^{\infty} \Gamma_{2j} \\ &= \left(\frac{1}{2\pi}\right)^2 \int_{-\pi}^{\pi} \phi(\lambda, \theta) \sum_k \left(1 - \frac{|k|}{M}\right) e^{-ik\lambda} d\lambda \sum_{j=k+1}^{\infty} \int_{-\pi}^{\pi} e^{ij\omega} f_2(\omega, \theta_0) d\omega \\ &= \left(\frac{1}{2\pi}\right)^2 \int_{-\pi}^{\pi} \phi(\lambda, \theta) \sum_k \left(1 - \frac{|k|}{M}\right) e^{-ik\lambda} d\lambda \sum_{q=1}^{\infty} \int_{-\pi}^{\pi} e^{i(q+k)\omega} f_2(\omega, \theta_0) d\omega \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[ \int_{-\pi}^{\pi} \phi(\lambda, \theta) F_M(\omega - \lambda) d\lambda \right] \sum_{q=1}^{\infty} e^{iq\omega} f_2(\omega, \theta_0) d\omega \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \phi_M(\omega, \theta) \sum_{q=1}^{\infty} e^{iq\omega} f_2(\omega, \theta_0) d\omega, \end{aligned}$$

which uses the fact that  $\Gamma_{2j} = E[\xi_{00}\xi'_{02,j}] = \int_{-\pi}^{\pi} e^{ij\omega} f_2(\omega, \theta_0) d\omega$ . But

$$\begin{aligned} \|g(\theta, \phi_M) - g(\theta, \phi)\| &\leq \left\| \frac{1}{2\pi} [\phi_M(0, \theta) - \phi(0, \theta)] \int_0^1 dSS'_2 \right\| \\ &\quad + \left\| \frac{1}{2\pi} \int_{-\pi}^{\pi} [\phi_M(\omega, \theta) - \phi(\omega, \theta)] \sum_q e^{iq\omega} f_2(\omega, \theta_0) d\omega \right\| \\ &< \frac{\epsilon}{2\pi} \left\| \int_0^1 dSS'_2 \right\| + \frac{\epsilon}{2\pi} \int_{-\pi}^{\pi} \left\| \sum_q e^{iq\omega} f_2(\omega, \theta_0) \right\| d\omega, \end{aligned} \quad (25)$$

which is  $O_p(\epsilon)$ . We have therefore demonstrated that, for any  $\theta \in \Theta$ ,  $g_T(\theta, \phi) \Rightarrow g(\theta, \phi)$ . The result extends to  $\phi$  not necessarily of period  $2\pi$  on applying the argument on pp.232–233 of Robinson (1976).

Uniformity of convergence is essentially implied generically by Theorem 1 of Jennrich (1969) and in our context (omitting the dependence on  $\phi$  for convenience)

$$\|g_T(\theta) - g(\theta)\| \leq \|g_T(\theta) - g_T(\bar{\theta})\| + \|g_T(\bar{\theta}) - g(\bar{\theta})\| + \|g(\bar{\theta}) - g(\theta)\|$$

for given  $\theta, \bar{\theta} \in \Theta$ . For given  $\epsilon > 0$  there exists a neighbourhood  $U$  of  $\bar{\theta}$ ,  $U \in \Theta$ , such that

$$\begin{aligned} \sup_{\theta \in U} \|g(\bar{\theta}) - g(\theta)\| &\leq \frac{1}{2\pi} \sup_{\theta \in U} \|\phi(0, \bar{\theta}) - \phi(0, \theta)\| \left\| \int_0^1 dSS'_2 \right\| \\ &\quad + \frac{1}{2\pi} \int_{-\pi}^{\pi} \sup_{\theta \in U} \|\phi(\lambda, \bar{\theta}) - \phi(\lambda, \theta)\| \left\| \sum_q e^{iq\lambda} f_2(\lambda, \theta_0) \right\| d\lambda = O_p(\epsilon). \end{aligned}$$

Moreover, for  $T$  sufficiently large,  $\|g_T(\bar{\theta}) - g(\bar{\theta})\| = O_p(\epsilon)$  and

$$\sup_{\theta \in U} \|g_T(\theta) - g_T(\bar{\theta})\| \leq \frac{1}{T} \sum_j \sup_{\theta \in U} \|\phi(\lambda_j, \theta) - \phi(\lambda_j, \bar{\theta})\| \|I_{02,j}\| = O_p(\epsilon),$$

thus implying that  $\sup_{\theta \in U} \|g_T(\theta) - g(\theta)\| = O_p(\epsilon)$ . Under Assumption 1, every open cover of  $\Theta$  has a finite sub-cover and so the above results hold uniformly in  $\theta \in \Theta$ . As  $\epsilon$  is arbitrary, it follows that (b) holds.

(c) The proof proceeds as in part (b) with the function of interest defined as

$$g_T(\theta, \phi) = \frac{1}{T^2} \sum_j J' \phi_j(\theta) J \otimes I'_{22,j}.$$

In place of (22) we have

$$\|g_T(\theta, \phi) - g_T(\theta, \phi_M)\| < \epsilon \|J\|^2 \frac{1}{T^2} \sum_j \|I_{22,j}\| \leq \epsilon m_1 \frac{1}{T^2} \sum_j \|w_{2j}\|^2 = o_p(\epsilon),$$

implying that  $\phi$  can be replaced by  $\phi_M$  as before. The function  $g_T(\theta, \phi_M)$  has the representation

$$\left(\frac{1}{2\pi}\right)^2 \sum_k J' \left[ \int_{-\pi}^{\pi} \phi(\lambda, \theta) \left(1 - \frac{|k|}{M}\right) e^{-ik\lambda} d\lambda \right] J \otimes \frac{1}{T} \widehat{\Gamma}'_{22,k} + o_p(1),$$

where  $\widehat{\Gamma}_{22,k} = T^{-1} \sum_{t=1}^{T-k} y_{2,t} y'_{2,t+k}$ . By Lemma 3(c),  $g_T(\theta, \phi_M) \Rightarrow g(\theta, \phi_M)$ , where

$$g(\theta, \phi_M) = \frac{1}{2\pi} J' \phi_M(0, \theta) J \otimes \int_0^1 S_2 S_2'.$$

But

$$\begin{aligned} \|g(\theta, \phi_M) - g(\theta, \phi)\| &\leq \frac{1}{2\pi} \|J\|^2 \|\phi_J(0, \theta) - \phi(0, \theta)\| \left\| \int_0^1 S_2 S_2' \right\| \\ &< \frac{\epsilon m_1}{2\pi} \left\| \int_0^1 S_2 S_2' \right\| = O_p(\epsilon). \end{aligned}$$

Hence pointwise convergence is established, and the uniformity follows from arguments identical to those in part (b).  $\square$

The proof of Theorem 1 relies on various mean value expansions that in turn establish Lemma B below. In particular,

$$\ln \det[f_j(\theta)] = \ln \det[f_j(\theta_0)] + h_j(\bar{\theta})(\theta - \theta_0),$$

where  $h_j(\bar{\theta}) = \partial \ln \det[f_j(\theta)] / \partial \theta'$  evaluated at  $\bar{\theta} = \alpha \theta + (1 - \alpha) \theta_0$  for some  $0 < \alpha < 1$ ;

$$\text{vec}[B(\theta_1)'] = \text{vec}[B(\theta_{10})'] + F(\bar{\theta}_1)(\theta_1 - \theta_{10}),$$

where the  $i$ 'th row of the  $m_1 m_2 \times p$  matrix  $F(\bar{\theta}_1)$  is equal to the  $i$ 'th row of the matrix  $\partial \text{vec}[B(\theta_1)'] / \partial \theta_1'$  evaluated at  $\bar{\theta}_{1,i} = \alpha_i \theta_1 + (1 - \alpha_i) \theta_{10}$  for some  $0 < \alpha_i < 1$  ( $i = 1, \dots, m_1 m_2$ ) and  $\bar{\theta}_1 = [\bar{\theta}_{1,1}, \dots, \bar{\theta}_{1,m_1 m_2}]$ ;

$$\text{vec}\{[f_j(\theta)^{-1}]\}' = \text{vec}\{[f_j(\theta_0)^{-1}]\}' + G_j(\tilde{\theta})(\theta - \theta_0),$$

where the  $i$ 'th row of  $G_j(\tilde{\theta})$  is equal to the  $i$ 'th row of the matrix  $\partial \text{vec}\{[f_j(\theta)^{-1}]\}' / \partial \theta'$  evaluated at  $\tilde{\theta}_i = \bar{\alpha}_i \theta + (1 - \bar{\alpha}_i) \theta_0$  for some  $0 < \bar{\alpha}_i < 1$  ( $i = 1, \dots, m^2$ ) and  $\tilde{\theta} = [\tilde{\theta}_1, \dots, \tilde{\theta}_{m^2}]$ .

**Lemma B.** *Let  $A_T(\theta, \theta_0)$ ,  $B_T(\theta, \theta_{10})$  and  $C_T(\theta, \theta_0)$  be defined as in (13), (14) and (15), respectively. Then, under Assumptions 1–3,*

$$(a) \quad A_T(\theta, \theta_0) \geq -T c_{A_T} \|\theta - \theta_0\|,$$

$$(b) \quad B_T(\theta, \theta_{10}) \geq -T c_{B_1} X_{1T} \|\theta_1 - \theta_{10}\| + T^2 c_{B_{2T}} \|\theta_1 - \theta_{10}\|^2,$$

$$(c) C_T(\theta, \theta_0) \geq -TX_{2T}\|\theta - \theta_0\|,$$

where  $c_{A_T} = O(1)$ ,  $c_{B_1}$  is a positive constant,  $X_{1T} = O_p(1)$ ,  $c_{B_{2T}}$  is positive with probability approaching 1, and  $X_{2T} = O_p(1)$ .

**Proof.** (a) Using the mean value expansion of  $\ln \det[f_j(\theta)]$  we obtain

$$A_T(\theta, \theta_0) = \sum_j h_j(\bar{\theta})(\theta - \theta_0) \geq -\sum_j \|h_j(\bar{\theta})\| \|\theta - \theta_0\| \geq -Tc_{A_T}\|\theta - \theta_0\|,$$

where  $c_{A_T} = T^{-1} \sum_j \sup_{\theta \in \Theta} \|h_j(\theta)\| = O(1)$  under Assumption 1 and the continuity of the  $h_j(\theta)$  which follows from Assumption 2.

(b) Using (16) we can write  $B_T(\theta, \theta_{10}) = B_{1T}(\theta, \theta_{10}) + B_{2T}(\theta, \theta_{10})$  where

$$\begin{aligned} B_{1T}(\theta, \theta_{10}) &= -2 \sum_j \text{tr} \left\{ f_j(\theta)^{-1} I_{02,j} (B - B_0)' J' \right\} \\ &= -2T \text{vec}(J')' \left( I_{m_1} \otimes \frac{1}{T} \sum_j f_j(\theta)^{-1} I_{02,j} \right) \text{vec}[(B - B_0)'] \\ &= -2T \text{vec}(J')' \left( I_{m_1} \otimes \frac{1}{T} \sum_j f_j(\theta)^{-1} I_{02,j} \right) F(\bar{\theta}_1) (\theta_1 - \theta_{10}) \\ &\geq -2T \|J\| \sqrt{m_1} \left\| \frac{1}{T} \sum_j f_j(\theta)^{-1} I_{02,j} \right\| \|F(\bar{\theta}_1)\| \|\theta_1 - \theta_{10}\| \\ &\geq -Tc_{B_1} X_{1T} \|\theta_1 - \theta_{10}\|, \end{aligned}$$

with  $c_{B_1} = 2m_1 \sup_{\bar{\theta}_1 \in \Theta_1^{m_1 m_2}} \|F(\bar{\theta}_1)\| < \infty$  and  $X_{1T} = \sup_{\theta \in \Theta} \left\| \frac{1}{T} \sum_j f_j(\theta)^{-1} I_{02,j} \right\| = O_p(1)$  by Lemma 4(b), and, defining  $Q_T(\theta) = F(\bar{\theta}_1)' \left( T^{-2} \sum_j J' f_j(\theta)^{-1} J \otimes I'_{22,j} \right) F(\bar{\theta}_1)$ ,

$$\begin{aligned} B_{2T}(\theta, \theta_{10}) &= \sum_j \text{tr} \left\{ f_j(\theta)^{-1} J (B - B_0) I_{22,j} (B - B_0)' J' \right\} \\ &= T^2 \text{vec}[(B - B_0)']' \left( \frac{1}{T^2} \sum_j J' f_j(\theta)^{-1} J \otimes I'_{22,j} \right) \text{vec}[(B - B_0)'] \\ &= T^2 (\theta_1 - \theta_{10})' Q_T(\theta) (\theta_1 - \theta_{10}). \end{aligned}$$

Hence  $B_{2T}(\theta, \theta_{10}) \geq T^2 c_{B_{2T}} \|\theta_1 - \theta_{10}\|^2$ , where  $c_{B_{2T}} = \inf_{\theta \in \Theta} \mu_1[Q_T(\theta)] > 0$  with probability approaching 1 under Assumption 3 and where  $\mu_1[Q_T(\theta)]$  denotes the smallest eigenvalue of the matrix  $Q_T(\theta)$ .

(c) Utilising the mean value expansion of  $\text{vec}\{[f_j(\theta)^{-1}]'\}$  yields

$$C_T(\theta, \theta_0) = \sum_j \text{tr} \left\{ [f_j(\theta)^{-1} - f_j(\theta_0)^{-1}] I_{00,j} \right\}$$

$$\begin{aligned}
&= \sum_j \text{vec} \left\{ \left[ f_j(\theta)^{-1} - f_j(\theta_0)^{-1} \right]' \right\}' \text{vec}(I_{00,j}) \\
&= T(\theta - \theta_0)' T^{-1} \sum_j G_j(\tilde{\theta}) \text{vec}(I_{00,j}) \\
&\geq -T X_{2T} \|\theta - \theta_0\|,
\end{aligned}$$

where  $X_{2T} = \sup_{\tilde{\theta} \in \Theta^{m^2}} \left\| T^{-1} \sum_j G_j(\tilde{\theta}) \text{vec}(I_{00,j}) \right\| = O_p(1)$  by Lemma 4(a).  $\square$

**Proof of Theorem 1.** Using Lemma B, (12) can be written

$$\begin{aligned}
L_T(\theta) - L_T(\theta_0) &\geq -T(c_{A_T} + X_{2T}) \|\theta - \theta_0\| - T c_{B_1} X_{1T} \|\theta_1 - \theta_{10}\| + T^2 c_{B_{2T}} \|\theta_1 - \theta_{10}\|^2 \\
&= T^2 c_{B_{2T}} \|\theta_1 - \theta_{10}\|^2 Z_T,
\end{aligned}$$

where

$$\begin{aligned}
Z_T &= 1 - \frac{c_{B_1} X_{1T}}{T c_{B_{2T}} \|\theta_1 - \theta_{10}\|} - \frac{(c_{A_T} + X_{2T}) (\|\theta_1 - \theta_{10}\|^2 + \|\theta_2 - \theta_{20}\|^2)^{1/2}}{T c_{B_{2T}} \|\theta_1 - \theta_{10}\|^2} \\
&= 1 - O_p(T^{-(1-\gamma)}) - O_p(T^{-(1-\gamma)}) = 1 - o_p(1)
\end{aligned}$$

since  $0 < \gamma < 1$ . Hence

$$\inf_{\theta \in \bar{N}_{T,\gamma}(\theta_{10}, \delta) \times \Theta_2} [L_T(\theta) - L_T(\theta_0)] \geq T^{2(1-\gamma)} \delta^2 c_{B_{2T}} > 0$$

with probability approaching 1, thereby satisfying (9).

Given that (9) holds, it is sufficient to show that (10) holds with the set  $\Theta_1$  replaced by  $N_{T,\gamma}(\theta_{10}, \delta_1)$  for arbitrary  $\delta_1$ , as noted in the text. We also need to take  $1/2 < \gamma < 1$ ; this is also required by Saikkonen for the consistency of  $\theta_2$ .<sup>9</sup> Note that

$$\begin{aligned}
&\Pr \left\{ \inf_{\theta \in N_{T,\gamma}(\theta_{10}, \delta_1) \times \bar{B}(\theta_{20}, \delta)} [L_T(\theta) - L_T(\theta_0)] > 0 \right\} \\
&\geq \Pr \left\{ \inf_{\theta \in N_{T,\gamma}(\theta_{10}, \delta_1) \times \bar{B}(\theta_{20}, \delta)} [L_T(\theta) - L_T(\theta_0)] > T\eta \right\}
\end{aligned}$$

for some  $\eta > 0$ . We therefore consider  $T^{-1} [L_T(\theta) - L_T(\theta_0)] = R(\theta, \theta_0) + U(\theta, \theta_0)$ , where  $R(\theta, \theta_0) = L(\theta) - L(\theta_0)$ ,  $U(\theta, \theta_0) = T^{-1} [L_T(\theta) - L_T(\theta_0)] - [L(\theta) - L(\theta_0)]$ , and

$$L(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left\{ \ln \det [f(\lambda, \theta)] + \text{tr} \left[ f(\lambda, \theta)^{-1} f(\lambda, \theta_0) \right] \right\} d\lambda.$$

---

<sup>9</sup>See Saikkonen (1995, pp.905 and 911).



Then we can write

$$\begin{aligned} \Pr \left\{ \inf T^{-1} [L_T(\theta) - L_T(\theta_0)] > \eta \right\} &= \Pr \{ \inf [R(\theta, \theta_0) + U(\theta, \theta_0)] > \eta \} \\ &\geq \Pr \{ \inf R(\theta, \theta_0) > \eta \} + \Pr \{ \sup |U(\theta, \theta_0)| \leq \eta \}, \end{aligned}$$

where the infima and supremum are taken over  $\theta \in N_{T,\gamma}(\theta_{10}, \delta_1) \times \bar{B}(\theta_{20}, \delta)$ . Hence (10) is satisfied if, for every  $\delta > 0$ , some  $\eta > 0$  and some  $\delta_1 > 0$ , the following two conditions hold:

$$\begin{aligned} \text{(a)} \quad \lim_{T \rightarrow \infty} \Pr \left\{ \inf_{\theta \in N_{T,\gamma}(\theta_{10}, \delta_1) \times \bar{B}(\theta_{20}, \delta)} R(\theta, \theta_0) > \eta \right\} &= 1, \\ \text{(b)} \quad \lim_{T \rightarrow \infty} \Pr \left\{ \sup_{\theta \in N_{T,\gamma}(\theta_{10}, \delta_1) \times \bar{B}(\theta_{20}, \delta)} |U(\theta, \theta_0)| \leq \eta \right\} &= 1. \end{aligned}$$

These conditions shall be demonstrated in turn.

(a) First, observe that, for  $i = 1, \dots, p$ ,

$$\begin{aligned} \frac{\partial R(\theta, \theta_0)}{\partial \theta_i} &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \text{tr} \left\{ \left[ f(\lambda, \theta)^{-1} \frac{\partial f(\lambda, \theta)}{\partial \theta_i} \right] - \left[ f(\lambda, \theta)^{-1} \frac{\partial f(\lambda, \theta)}{\partial \theta_i} f(\lambda, \theta)^{-1} f(\lambda, \theta_0) \right] \right\} d\lambda \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \text{tr} \left\{ f(\lambda, \theta)^{-1} \frac{\partial f(\lambda, \theta)}{\partial \theta_i} [I_p - f(\lambda, \theta)^{-1} f(\lambda, \theta_0)] \right\} d\lambda, \end{aligned}$$

so that  $R(\theta, \theta_0)$  is uniquely minimised at  $\theta = \theta_0$  under Assumption 2(b). Since  $\|\theta_2 - \theta_{20}\| > \delta$  on the set  $\bar{B}(\theta_{20}, \delta)$ , condition (a) then follows from the uniform continuity of  $R(\theta, \theta_0)$ .

(b) From the definition of  $U(\theta, \theta_0)$  we can write  $|U(\theta, \theta_0)| \leq |U_1(\theta)| + |U_1(\theta_0)|$ , where  $U_1(\theta) = T^{-1}L_T(\theta) - L(\theta)$ . Now

$$\begin{aligned} |U_1(\theta)| &= \left| \frac{1}{T} \sum_j \left\{ \ln \det[f_j(\theta)] + \text{tr} [f_j(\theta)^{-1} I_j(\theta_1)] \right\} \right. \\ &\quad \left. - \frac{1}{2\pi} \int_{-\pi}^{\pi} \left\{ \ln \det[f(\lambda, \theta)] + \text{tr} [f(\lambda, \theta)^{-1} f(\lambda, \theta_0)] \right\} d\lambda \right| \\ &\leq \left| \frac{1}{T} \sum_j \ln \det[f_j(\theta)] - \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln \det[f(\lambda, \theta)] d\lambda \right| \\ &\quad + \left| \frac{1}{T} \sum_j \text{tr} [f_j(\theta)^{-1} I_j(\theta_1)] - \frac{1}{2\pi} \int_{-\pi}^{\pi} \text{tr} [f(\lambda, \theta)^{-1} f(\lambda, \theta_0)] d\lambda \right|. \end{aligned}$$

Hence  $\sup_{\theta \in N_{T,\gamma} \times \bar{B}} |U_1(\theta)| \xrightarrow{p} 0$  if:

$$\text{(i)} \quad \sup_{\theta \in N_{T,\gamma} \times \bar{B}} \left| \frac{1}{T} \sum_j \ln \det[f_j(\theta)] - \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln \det[f(\lambda, \theta)] d\lambda \right| \rightarrow 0,$$

$$(ii) \quad \sup_{\theta \in N_{T,\gamma} \times \bar{B}} \left| \frac{1}{T} \sum_j \text{tr} [f_j(\theta)^{-1} I_j(\theta_1)] - \frac{1}{2\pi} \int_{-\pi}^{\pi} \text{tr} [f(\lambda, \theta)^{-1} f(\lambda, \theta_0)] d\lambda \right| \xrightarrow{p} 0.$$

Part (i) is satisfied by the uniform convergence of the Riemann sums. Let  $U_2(\theta)$  denote the function whose supremum is being considered in (ii). Then

$$\begin{aligned} |U_2(\theta)| &\leq \left| \frac{1}{T} \sum_j \text{tr} \{ f_j(\theta)^{-1} [I_j(\theta_1) - I_{00,j}] \} \right| \\ &\quad + \left| \frac{1}{T} \sum_j \text{tr} [f_j(\theta)^{-1} I_{00,j}] - \frac{1}{2\pi} \int_{-\pi}^{\pi} \text{tr} [f(\lambda, \theta)^{-1} f(\lambda, \theta_0)] d\lambda \right| \\ &= |U_{21}(\theta)| + |U_{22}(\theta)|. \end{aligned}$$

Lemma 4(a) ensures that  $\sup |U_{22}(\theta)| \xrightarrow{p} 0$ , while

$$\begin{aligned} |U_{21}(\theta)| &\leq \left| \frac{1}{T} \sum_j \text{tr} \{ f_j(\theta)^{-1} J[B - B_0] I_{22,j} [B - B_0]' J' \} \right| \\ &\quad + 2 \left| \frac{1}{T} \sum_j \text{tr} \{ f_j(\theta)^{-1} I_{02,j} (B - B_0)' J' \} \right| \\ &= \left| T \text{vec} [(B - B_0)']' \left( \frac{1}{T^2} \sum_j J' f_j(\theta)^{-1} J \otimes I'_{22,j} \right) \text{vec} [(B - B_0)'] \right| \\ &\quad + 2 \left| \text{vec}(J')' \left( I_{m_1} \otimes \frac{1}{T} \sum_j f_j(\theta)^{-1} I_{02,j} \right) \text{vec} [(B - B_0)'] \right| \\ &\leq T \|F(\bar{\theta}_1)\|^2 \|\theta_1 - \theta_{10}\|^2 \left\| \frac{1}{T^2} \sum_j J' f_j(\theta)^{-1} J \otimes I'_{22,j} \right\| \\ &\quad + 2m_1 \|F(\bar{\theta}_1)\| \|\theta_1 - \theta_{10}\| \left\| \frac{1}{T} \sum_j f_j(\theta)^{-1} I_{02,j} \right\|, \end{aligned}$$

which makes use of the mean value expansion of  $\text{vec} [(B - B_0)']$ . Hence

$$\begin{aligned} \sup_{\theta \in N_{T,\gamma} \times \bar{B}} |U_{21}(\theta)| &\leq T^{1-2\gamma} \delta_1^2 \sup_{\bar{\theta}_1 \in \Theta^{m_1 m_2}} \|F(\bar{\theta}_1)\|^2 \sup_{\theta \in \Theta} \left\| \frac{1}{T^2} \sum_j J' f_j(\theta)^{-1} J \otimes I'_{22,j} \right\| \\ &\quad + 2m_1 T^{-\gamma} \delta_1 \sup_{\bar{\theta}_1 \in \Theta^{m_1 m_2}} \|F(\bar{\theta}_1)\| \sup_{\theta \in \Theta} \left\| \frac{1}{T} \sum_j f_j(\theta)^{-1} I_{02,j} \right\| \xrightarrow{p} 0 \end{aligned}$$

because  $1/2 < \gamma < 1$ ,  $\sup_{\bar{\theta}_1 \in \Theta^{m_1 m_2}} \|F(\bar{\theta}_1)\| < \infty$ , and the two stochastic terms are each  $O_p(1)$  by Lemma 4(c) and 4(b) respectively. Finally, (ii) is established because  $\sup |U_1(\theta_0)| \xrightarrow{p} 0$  by the uniform convergence of Riemann sums and Lemma 4(a).  $\square$

**Proof of Theorem 2.** Consider, first, a typical element of the normalised score vector with regard to  $\theta_1$ , denoted

$$[s_{T1}(\theta)]_i = \frac{1}{T} \sum_j \text{tr} \Gamma_{ji}(\theta) - \frac{1}{T} \sum_j \text{tr} \Phi_{ji}(\theta) I_j(\theta_1) + \frac{1}{T} \sum_j \text{tr} f_j(\theta)^{-1} I_{ji}(\theta_1),$$

where  $i = 1, \dots, p_1$ , all summations are over  $j \in J_T$  and

$$\Gamma_{ji}(\theta) = f_j(\theta)^{-1} \frac{\partial f_j(\theta)}{\partial \theta_{1i}}, \quad \Phi_{ji}(\theta) = f_j(\theta)^{-1} \frac{\partial f_j(\theta)}{\partial \theta_{1i}} f_j(\theta)^{-1}, \quad \text{and} \quad I_{ji}(\theta_1) = \frac{\partial I_j(\theta_1)}{\partial \theta_{1i}}.$$

Evaluating the normalised score at  $\hat{\theta}_T$  and expanding yields

$$\begin{aligned} [s_{T1}(\hat{\theta}_T)]_i &= \frac{1}{T} \sum_j \text{tr} \Gamma_{ji}(\hat{\theta}_T) - \frac{1}{T} \sum_j \text{tr} \Phi_{ji}(\hat{\theta}_T) I_j(\hat{\theta}_{1T}) + \frac{1}{T} \sum_j \text{tr} f_j(\hat{\theta}_T)^{-1} I_{ji}(\hat{\theta}_{1T}) \\ &= \frac{1}{T} \sum_j \text{tr} \left\{ \Gamma_{ji}(\hat{\theta}_T) - \Phi_{ji}(\hat{\theta}_T) I_{00,j} \right\} + \frac{1}{T} \sum_j \text{tr} \Phi_{ji}(\hat{\theta}_T) \left[ I_{00,j} - I_j(\hat{\theta}_{1T}) \right] \\ &\quad + \frac{1}{T} \sum_j \text{tr} f_j(\hat{\theta}_T)^{-1} I_{ji}(\hat{\theta}_{1T}). \end{aligned} \tag{26}$$

The previously established consistency of  $\hat{\theta}_T$  implies that  $\hat{\theta}_T \in \Theta$  with probability approaching one, and so the first term in (26) is  $o_p(1)$  due to the uniform convergence of the Riemann sums and Lemma 4(a), yielding the limit

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \text{tr} \Gamma_i(\lambda, \theta_0) d\lambda - \frac{1}{2\pi} \int_{-\pi}^{\pi} \text{tr} \Phi_i(\lambda, \theta_0) f(\lambda, \theta_0) d\lambda = 0.$$

For the second term, recall that

$$I_{00,j} - I_j(\hat{\theta}_{1T}) = I_{02,j} [B_0 - \hat{B}]' J' + J [B_0 - \hat{B}] I_{02,j}^* + J [B_0 - \hat{B}] I_{22,j} [B_0 - \hat{B}]' J',$$

where  $B_0 = B(\theta_{10})$  and  $\hat{B} = B(\hat{\theta}_{1T})$ , and so

$$\begin{aligned} &\frac{1}{T} \sum_j \text{tr} \Phi_{ji}(\hat{\theta}_T) \left[ I_{00,j} - I_j(\hat{\theta}_{1T}) \right] \\ &= \frac{2}{T} \sum_j \text{tr} \Phi_{ji}(\hat{\theta}_T) I_{02,j} [B_0 - \hat{B}]' J' + \frac{1}{T} \sum_j \text{tr} \Phi_{ji}(\hat{\theta}_T) J [B_0 - \hat{B}] I_{22,j} [B_0 - \hat{B}]' J' \\ &= 2 \text{vec}(J)' \left( I \otimes \frac{1}{T} \sum_j \Phi_{ji}(\hat{\theta}_T) I_{02,j} \right) \text{vec}(B_0 - \hat{B}) \\ &\quad + \text{vec}(B_0 - \hat{B})' \frac{1}{T} \sum_j \left( J' \Phi_{ji}(\hat{\theta}_T) J \otimes I_{22,j}' \right) \text{vec} \left[ (B_0 - \hat{B})' \right]. \end{aligned}$$

The mean value expansion of  $\text{vec}(\hat{B})$  used in the proof of Lemma B indicates that  $\text{vec}(B_0 - \hat{B}) = -F(\bar{\theta}_1)(\hat{\theta}_{1T} - \theta_{10}) = o_p(T^{-\gamma})$  for  $0 < \gamma < 1$ . Using Lemma 4(b) the first term above

is  $2\text{vec}(J)'(I \otimes O_p(1))o_p(T^{-\gamma}) = o_p(1)$  while the second, using Lemma 4(c), can be seen to be  $o_p(T^{-\gamma})O_p(T)o_p(T^{-\gamma}) = o_p(T^{1-2\gamma}) = o_p(1)$  for  $\gamma > 1/2$ . Hence the second term in (26) is also  $o_p(1)$  and so we are led to concentrate on the third term, which can be written

$$\begin{aligned} \frac{1}{T} \sum_j \text{tr} f_j(\hat{\theta}_T)^{-1} I_{ji}(\hat{\theta}_{1T}) &= \frac{1}{T} \sum_j \text{tr} [f_j(\hat{\theta}_T)^{-1} - f_{0j}^{-1}] I_{ji}(\theta_{10}) \\ &+ \frac{1}{T} \sum_j \text{tr} f_j(\hat{\theta}_T)^{-1} [I_{ji}(\hat{\theta}_{1T}) - I_{ji}(\theta_{10})] + \frac{1}{T} \sum_j \text{tr} f_{0j}^{-1} I_{ji}(\theta_{10}), \end{aligned} \quad (27)$$

where  $f_{0j} = f_j(\theta_0)$ . Define  $B_i(\theta_1) = \partial B(\theta_1)/\partial \theta_{1i}$  and note that  $I_{ji}(\theta_{10}) = -JB_i(\theta_{10})I_{02,j}^* - I_{02,j}B_i(\theta_{10})'J'$ . Then the first term in (27) can be written

$$\begin{aligned} \frac{2}{T} \sum_j \text{tr} [f_j(\hat{\theta}_T)^{-1} - f_{0j}^{-1}] I_{02,j}B_i(\theta_{10})'J' \\ = 2\text{vec}(J)' \left( I \otimes \frac{1}{T} \sum_j [f_j(\hat{\theta}_T)^{-1} - f_{0j}^{-1}] I_{02,j} \right) \text{vec} [B_i(\theta_{10})] = o_p(1), \end{aligned}$$

using the consistency of  $\hat{\theta}_T$  and Lemma 4(b). For the second term, note that

$$\begin{aligned} I_{ji}(\hat{\theta}_{1T}) - I_{ji}(\theta_{10}) &= -J(\hat{B}_i - B_i^0)I_{02,j}^* - I_{02,j}(\hat{B}_i - B_i^0)'J' \\ &+ J\hat{B}_i I_{22,j}(B - B_0)'J' + J(\hat{B} - B_0)I_{22,j}\hat{B}_i'J', \end{aligned}$$

where  $\hat{B}_i = B_i(\hat{\theta}_{1T})$  and  $B_i^0 = B_i(\theta_{10})$ . Hence the second term is

$$\begin{aligned} \frac{1}{T} \sum_j \text{tr} f_j(\hat{\theta}_T)^{-1} [I_{ji}(\hat{\theta}_{1T}) - I_{ji}(\theta_{10})] \\ = -\frac{2}{T} \sum_j \text{tr} f_j(\hat{\theta}_T)^{-1} I_{02,j}(\hat{B}_i - B_i^0)'J' + \frac{2}{T} \sum_j \text{tr} f_j(\hat{\theta}_T)^{-1} J\hat{B}_i I_{22,j}(B - B_0)'J' \\ = -2\text{vec}(J)' \left( I \otimes \frac{1}{T} \sum_j f_j(\hat{\theta}_T)^{-1} I_{02,j} \right) \text{vec}[\hat{B}_i - B_i^0] \\ + 2\text{vec}(\hat{B} - B_0)' \frac{1}{T} \sum_j (Jf_j(\hat{\theta}_T)^{-1}J' \otimes I'_{22,j}) \text{vec}(\hat{B}_i'). \end{aligned}$$

The first component is  $-2\text{vec}(J)'O_p(1)o_p(1) = o_p(1)$ , using the consistency of  $\hat{\theta}_T$  and Lemma 4(b), while the second component is  $2o_p(T^{-\gamma})O_p(T)O_p(1) = o_p(T^{1-\gamma})$ , using the consistency of  $\hat{\theta}_{1T}$  and Lemma 4(c). Therefore the second component of the second term of (27) is not

asymptotically negligible, and it is convenient to write it as

$$\begin{aligned} & \frac{2}{T} \sum_j \text{tr} I_{22,j} [\hat{B}_i - B_i^0]' J' f_j(\hat{\theta}_T)^{-1} J \hat{B}_i \\ &= 2 \text{vec}(\hat{B}_i)' \frac{1}{T} \sum_j \left( I_{22,j} \otimes (J' f_j(\hat{\theta}_T)^{-1} J)' \right) \text{vec}[\hat{B}_i - B_i^0]. \end{aligned}$$

Using the mean value expansion of  $\text{vec}[\hat{B} - B_0]$ , the consistency of  $\hat{\theta}_T$ , and stacking the  $i = 1, \dots, p_2$  equations, yields

$$2 \left( \frac{\partial \text{vec} B(\theta_{10})}{\partial \theta'_1} \right)' \frac{1}{T^2} \sum_j \left( I_{22,j} \otimes (J' f_j(\hat{\theta}_T)^{-1} J)' \right) \left( \frac{\partial \text{vec} B(\theta_{10})}{\partial \theta'_1} \right) T(\hat{\theta}_{1T} - \theta_{10}) + o_p(1). \quad (28)$$

For the third term in (27),

$$\begin{aligned} \frac{1}{T} \sum_j \text{tr} f_{0j}^{-1} I_{ji}(\theta_{10}) &= -\frac{2}{T} \sum_j \text{tr} f_{0j}^{-1} J B_i^0 I_{02,j}^* \\ &= -2 \text{vec}(B_i^0)' (I_{m_2} \otimes J)' \text{vec} \left( \frac{1}{T} \sum_j f_{0j}^{-1} I_{02,j} \right), \end{aligned}$$

which is  $O_p(1)$  from Lemma 4(b). Stacking the  $i = 1, \dots, p_1$  equations yields

$$-2 \left( \frac{\partial \text{vec} B(\theta_{10})}{\partial \theta'_1} \right)' (I_{m_2} \otimes J)' \text{vec} \left( \frac{1}{T} \sum_j f_{0j}^{-1} I_{02,j} \right). \quad (29)$$

Now, since  $s_{T1}(\hat{\theta}_T) = 0$ , it follows, combining (28) and (29), that

$$\begin{aligned} T(\hat{\theta}_{1T} - \theta_{10}) &= \left[ \left( \frac{\partial \text{vec} B(\theta_{10})}{\partial \theta'_1} \right)' \frac{1}{T^2} \sum_j \left( I_{22,j} \otimes (J' f_j(\hat{\theta}_T)^{-1} J)' \right) \left( \frac{\partial \text{vec} B(\theta_{10})}{\partial \theta'_1} \right) \right]^{-1} \\ &\quad \times \left( \frac{\partial \text{vec} B(\theta_{10})}{\partial \theta'_1} \right)' (I_{m_2} \otimes J)' \text{vec} \left( \frac{1}{T} \sum_j f_{0j}^{-1} I_{02,j} \right) + o_p(1). \quad (30) \end{aligned}$$

From Lemma 4(b),

$$\frac{1}{T} \sum_j f_{0j}^{-1} I_{02,j} \Rightarrow \frac{1}{2\pi} f(0, \theta_0)^{-1} \int_0^1 dS S'_2 + \bar{g}(\theta_0),$$

where

$$\bar{g}(\theta_0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{q=1}^{\infty} e^{iq\lambda} f(\lambda, \theta_0)^{-1} f_2(\lambda, \theta_0) d\lambda = \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{q=1}^{\infty} e^{iq\lambda} d\lambda \begin{bmatrix} 0 \\ I_{m_2} \end{bmatrix},$$

because

$$f_2(\lambda, \theta_0) = f(\lambda, \theta_0) \begin{bmatrix} 0 \\ I_{m_2} \end{bmatrix}.$$

But

$$\int_{-\pi}^{\pi} \sum_{q=1}^{\infty} e^{iq\lambda} d\lambda = \sum_{q=1}^{\infty} \frac{e^{iq\pi} - e^{-iq\pi}}{iq} = 2 \sum_{q=1}^{\infty} \frac{\sin q\pi}{q} = 0,$$

since, from equation 1.441.1 of Gradshteyn and Ryzhik (1994),  $\sum_{j=1}^{\infty} \sin jx/x = (\pi - x)/2$ .

Hence  $\bar{g}(\theta_0) = 0$  and so

$$\text{vec} \left( \frac{1}{T} \sum_j f_{0j}^{-1} I_{02,j} \right) \Rightarrow \text{vec} \left( \Omega^{-1} \int_0^1 dSS'_2 \right), \quad (31)$$

where  $\Omega = 2\pi f(0, \theta_0)$ . Now,

$$\begin{aligned} (I \otimes J) \text{vec} \left( \Omega^{-1} \int_0^1 dSS'_2 \right) &= (I \otimes J)(I \otimes \Omega^{-1}) \text{vec} \left( \int_0^1 dSS'_2 \right) \\ &= (I \otimes J\Omega^{-1}) \left( \int_0^1 S_2 \otimes dS \right) \\ &= (I \otimes \Omega_{11.2}^{-1}) \left( \int_0^1 S_2 \otimes dS_{1.2} \right), \end{aligned} \quad (32)$$

since  $J'\Omega^{-1}S = \Omega_{11.2}^{-1}S_{1.2}$ . Meanwhile, from Lemma 4(c),

$$\begin{aligned} \frac{1}{T^2} \sum_j \left( I_{22,j} \otimes (J'f_j(\hat{\theta}_T)^{-1}J)' \right) &\Rightarrow \int_0^1 S_2 S'_2 \otimes \frac{1}{2\pi} J'f(0, \theta_0)^{-1}J \\ &= \int_0^1 S_2 S'_2 \otimes J'\Omega^{-1}J = \int_0^1 S_2 S'_2 \otimes \Omega_{11.2}^{-1}. \end{aligned} \quad (33)$$

Combining (33) and (32) in (30), we obtain the result stated in the Theorem for  $T(\hat{\theta}_{1T} - \theta_{10})$ .

Turning to the normalised score vector for  $\theta_2$ , a typical element can be written

$$[s_{T2}(\theta)]_i = \frac{1}{T^{1/2}} \sum_j \text{tr} \Lambda_{ji}(\theta) - \frac{1}{T^{1/2}} \sum_j \text{tr} \Psi_{ji}(\theta) I_j(\theta_1), \quad i = 1, \dots, p_2,$$

where all summations are over  $j \in J_T$  and

$$\Lambda_{ji}(\theta) = f_j(\theta)^{-1} \frac{\partial f_j(\theta)}{\partial \theta_{2i}} \quad \text{and} \quad \Psi_{ji}(\theta) = f_j(\theta)^{-1} \frac{\partial f_j(\theta)}{\partial \theta_{2i}} f_j(\theta)^{-1}.$$

Since  $\hat{\theta}_{2T}$  satisfies  $s_{T2}(\hat{\theta}_T) = 0$ , consider

$$[s_{T2}(\hat{\theta}_T)]_i = \frac{1}{T^{1/2}} \sum_j \text{tr} \Psi_{ji}(\hat{\theta}_T) \left[ f_j(\hat{\theta}_T) - I_j(\hat{\theta}_{1T}) \right]$$

$$\begin{aligned}
&= \frac{1}{T^{1/2}} \sum_j \text{tr} \Psi_{ji}(\hat{\theta}_T) \left\{ f_j(\hat{\theta}_T) - I_j(\hat{\theta}_{1T}) - [f_{0j} - I_{00,j}] \right\} \\
&\quad + \frac{1}{T^{1/2}} \sum_j \text{tr} [\Psi_{ji}(\hat{\theta}_T) - \Psi_{ji}(\theta_0)] [f_{0j} - I_{00,j}] \\
&\quad + \frac{1}{T^{1/2}} \sum_j \text{tr} \Psi_{ji}(\theta_0) [f_{0j} - I_{00,j}] \equiv A_T + B_T + C_T.
\end{aligned}$$

The first term can be written  $A_T = A_{1T} + A_{2T}$ , where

$$\begin{aligned}
A_{1T} &= \frac{1}{T^{1/2}} \sum_j \text{tr} \Psi_{ji}(\hat{\theta}_T) [I_{00,j} - I_j(\hat{\theta}_{1T})] \\
&= 2\text{vec}(J)' \left( I \otimes \frac{1}{T} \sum_j \Psi_{ji}(\hat{\theta}_T) I_{02,j} \right) T^{1/2} \text{vec}(B_0 - \hat{B}) \\
&\quad + T^{1/2} \text{vec}(B_0 - \hat{B})' \frac{1}{T} \sum_j \left( J' \Psi_{ji}(\hat{\theta}_T) J \otimes I'_{22,j} \right) \text{vec} [(B_0 - \hat{B})'].
\end{aligned}$$

The first component of  $A_{1T}$  is  $2\text{vec}(J)'(I \otimes O_p(1))o_p(T^{1/2-\gamma}) = o_p(1)$  if  $\gamma > 1/2$  while the second component is  $o_p(T^{1/2-\gamma})O_p(T)o_p(T^{-\gamma}) = o_p(1)$  if  $\gamma > 3/4$ . Turning to  $A_{2T}$ ,

$$\begin{aligned}
A_{2T} &= \frac{1}{T^{1/2}} \sum_j \text{tr} \Psi_{ji}(\hat{\theta}_T) [f_j(\hat{\theta}_T) - f_{0j}] \\
&= \frac{1}{T^{1/2}} \sum_j \text{vec}[\Psi_{ji}(\hat{\theta}_T)']' \text{vec}[f_j(\hat{\theta}_T) - f_{0j}] \\
&= \frac{1}{T} \sum_j \text{vec}[\Psi_{ji}(\hat{\theta}_T)']' \frac{\partial \text{vec} f_j(\bar{\theta})}{\partial \theta'_1} T^{1/2} (\hat{\theta}_{1T} - \theta_{10}) \\
&\quad + \frac{1}{T} \sum_j \text{vec}[\Psi_{ji}(\hat{\theta}_T)']' \frac{\partial \text{vec} f_j(\bar{\theta})}{\partial \theta'_2} T^{1/2} (\hat{\theta}_{2T} - \theta_{20}).
\end{aligned}$$

The first component of  $A_{2T}$  is  $O_p(1)o_p(T^{1/2-\gamma}) = o_p(1)$  if  $\gamma > 1/2$  while the second component is  $O_p(1)$  and hence is important in contributing to the asymptotics. It is convenient, for later reference, to stack the  $p_2$  equations in the second component, making use of the matrix

$$K_T(\hat{\theta}_T) = \begin{bmatrix} T^{-1} \sum_j \text{vec}[\Psi_{j1}(\hat{\theta}_T)']' \frac{\partial \text{vec} f_j(\bar{\theta})}{\partial \theta'_2} \\ \vdots \\ T^{-1} \sum_j \text{vec}[\Psi_{jp_2}(\hat{\theta}_T)']' \frac{\partial \text{vec} f_j(\bar{\theta})}{\partial \theta'_2} \end{bmatrix},$$

to give

$$K_T(\hat{\theta}_T) T^{1/2} (\hat{\theta}_{2T} - \theta_{20}). \tag{34}$$

We shall next consider  $C_T$ , which involves the periodogram,  $I_{00,j}$ , of the stationary process  $\xi_{0t}$ . Theorem 2.1 and Corollary 2.2 of Dunsmuir (1979) ensure that  $C_T = O_p(1)$  and its limiting distribution rests on establishing a central limit theorem for the elements  $c_{rs}(j)$  of the matrix

$$C(j) = \frac{1}{T^{1/2}} \sum_{t=1}^T \left[ \xi_{0t} \xi'_{0t+j} - E(\xi_{0t} \xi'_{0t+j}) \right].$$

From Hannan (1976), any finite set of the  $c_{rs}(j)$  is asymptotically jointly normal, provided that (i) the diagonal elements of  $f(\lambda, \theta_0)$  are square integrable, (ii) the  $\xi_{0t}$  are ergodic with square summable matrix norms in their Wold representation, and (iii) the first four conditional moments of the innovations in the Wold representation are constant. Condition (i) is clearly satisfied, so we need to demonstrate (ii) and (iii). Since  $f(\lambda, \theta_0)$  is continuous in both arguments and Hermitian positive definite and  $\xi_{0t}$  is a stationary sequence, Theorem 17.3.3 of Ibragimov and Linnik (1971) implies that  $\xi_{0t}$  is strong mixing. It is therefore, from Hannan (1970, p.202), also ergodic and has representation  $\xi_{0t} = \sum_{j=0}^{\infty} C_j \epsilon_{t-j}$  with  $\sum_{j=0}^{\infty} \|C_j\|^2 < \infty$  and  $\epsilon_t$  iid. Hence (ii) is satisfied, and (iii) follows from the Wold representation and the fact that  $\xi_{0t}$  is Gaussian. The results of Dunsmuir therefore apply, yielding

$$c_T(\theta_0) = \begin{bmatrix} T^{-1/2} \sum_j \text{tr} \Psi_{j1}(\theta_0) [f_{0j} - I_{00,j}] \\ \vdots \\ T^{-1/2} \sum_j \text{tr} \Psi_{jp2}(\theta_0) [f_{0j} - I_{00,j}] \end{bmatrix} \xrightarrow{d} N(0, 2V(\theta_0)). \quad (35)$$

Turning to  $B_T$ , we make use of the mean value expansion

$$\text{vec}[\Psi_{ji}(\hat{\theta}_T)'] = \text{vec}[\Psi_{ji}(\theta_0)'] + P_{ji}(\tilde{\theta})(\hat{\theta}_T - \theta_0),$$

where the  $k$ 'th row of  $P_{ji}(\tilde{\theta})$  is equal to the  $k$ 'th row of  $\partial \text{vec}[\Psi_{ji}(\theta)'] / \partial \theta'$  evaluated at  $\tilde{\theta}_k = \tilde{\alpha}_k \hat{\theta}_T + (1 - \tilde{\alpha}_k) \theta_0$  for some  $0 < \tilde{\alpha}_k < 1$  ( $k = 1, \dots, m^2$ ) and where  $\tilde{\theta} = [\tilde{\theta}_1, \dots, \tilde{\theta}_{m^2}]$ .

The existence of  $P_{ji}(\tilde{\theta})$  is ensured under Assumption 4. Then  $B_T$  can be written

$$\begin{aligned} B_T &= \frac{1}{T^{1/2}} \sum_j \text{tr} \left[ \Psi_{ji}(\hat{\theta}_T) - \Psi_{ji}(\theta_0) \right] [f_{0j} - I_{00,j}] \\ &= \frac{1}{T^{1/2}} \sum_j \text{vec} \left\{ [\Psi_{ji}(\hat{\theta}_T) - \Psi_{ji}(\theta_0)]' \right\}' \text{vec}[f_{0j} - I_{00,j}] \\ &= (\hat{\theta}_T - \theta_0)' \frac{1}{T^{1/2}} \sum_j \text{vec}[P_{ji}(\tilde{\theta})]' \text{vec}[f_{0j} - I_{00,j}], \end{aligned}$$

which is  $o_p(1)O_p(1) = o_p(1)$ . From the properties of  $A_T$ ,  $B_T$  and  $C_T$ , and utilising the fact



that  $s_{T2}(\hat{\theta}_T) = 0$ , we find, from (27) and (35), and noting that  $K_T(\hat{\theta}_T) \xrightarrow{p} K(\theta_0)$ , where

$$K(\theta_0) = \begin{bmatrix} (1/2\pi) \int_{-\pi}^{\pi} \text{vec}[\Psi_1(\lambda, \theta_0)]' \frac{\partial \text{vec} f_j(\theta_0)}{\partial \theta_2'} d\lambda \\ \vdots \\ (1/2\pi) \int_{-\pi}^{\pi} \text{vec}[\Psi_{p2}(\lambda, \theta_0)]' \frac{\partial \text{vec} f_j(\theta_0)}{\partial \theta_2'} d\lambda \end{bmatrix},$$

that  $T^{1/2}(\hat{\theta}_{2T} - \theta_{20})$  has the same limiting distribution as that of the vector  $-K(\theta_0)^{-1}c_T(\theta_0)$ , which is  $N(0, 2K(\theta_0)^{-1}V(\theta_0)(K(\theta_0)^{-1})')$ . Now, observe that

$$\frac{\partial \text{vec} f_j(\theta)}{\partial \theta_2'} = \left[ \text{vec} \left( \frac{\partial f_j(\theta)}{\partial \theta_{21}} \right), \dots, \text{vec} \left( \frac{\partial f_j(\theta)}{\partial \theta_{2p2}} \right) \right],$$

and so the  $(k, l)$  element of  $K(\theta_0)$  is

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \text{tr} \Psi_k(\lambda, \theta_0) \frac{\partial f(\lambda, \theta_0)}{\partial \theta_{2l}} d\lambda = \frac{1}{2\pi} \int_{-\pi}^{\pi} \text{tr} f(\lambda, \theta_0)^{-1} \frac{\partial f(\lambda, \theta_0)}{\partial \theta_{2k}} f(\lambda, \theta_0)^{-1} \frac{\partial f(\lambda, \theta_0)}{\partial \theta_{2l}} d\lambda.$$

Hence  $K(\theta_0) = V(\theta_0)$  and the distribution stated in the Theorem follows.  $\square$

## References

- Bergstrom, A.R., 1997. Gaussian estimation of mixed-order continuous-time dynamic models with unobservable stochastic trends from mixed stock and flow data. *Econometric Theory* 13, 467–505.
- Chambers, M.J., 1998. The estimation of systems of joint differential-difference equations. *Journal of Econometrics* 85, 1–31.
- Chambers, M.J., 1999. Discrete time representation of stationary and nonstationary continuous time systems. *Journal of Economic Dynamics and Control* 23, 619–639.
- Chambers, M.J., 2003. The asymptotic efficiency of cointegration estimators under temporal aggregation. *Econometric Theory* 19, 49–77.
- Christiano, L.J., Eichenbaum, M., 1987. Temporal aggregation and structural inference in macroeconomics. *Carnegie-Rochester Conference Series on Public Policy* 26, 63–130.
- Corbae, D., Ouliaris, S., Phillips, P.C.B., 2002. Band spectral regression with trending data. *Econometrica* 70, 1067–1109.
- Corradi, V., 1997. Comovements between diffusion processes: characterization, estimation and testing. *Econometric Theory* 13, 646–666.
- Davidson, J.E.H., 1994. *Stochastic Limit Theory*. Oxford University Press, Oxford.

- 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.
- Gradshteyn, I.S., Ryzhik, I.M., 1994. *Table of Integrals, Series, and Products (Fifth Edition)*. Academic Press, Boston.
- Hannan, E.J., 1970. *Multiple Time Series*. Wiley, New York.
- Hannan, E.J., 1976. The asymptotic distribution of serial covariances. *Annals of Statistics* 4, 396–399.
- Hannan, E.J., Robinson, P.M., 1973. Lagged regression with unknown lags. *Journal of the Royal Statistical Society (Series B)* 35, 252–267.
- Hansen, L.P., Sargent, T.J., 1983. The dimensionality of the aliasing problem in models with rational spectral densities. *Econometrica* 51, 377–387.
- Ibragimov, I.A., Linnik, Y.V., 1971. *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff, Groningen.
- Jennrich, R.I., 1969. Asymptotic properties of non-linear least squares estimators. *Annals of Mathematical Statistics* 40, 633–643.
- Johansen, S., 1991. Estimation and hypothesis testing of cointegration vectors in Gaussian autoregressive models. *Econometrica* 59, 1551–1580.
- Kessler, M, Rahbek, A., 2001. Asymptotic likelihood based inference for cointegrated homogeneous diffusions. *Scandinavian Journal of Statistics* 28, 455–470.
- McCrorie, J.R., 2003. The problem of aliasing in identifying finite parameter continuous time stochastic models. *Acta Applicandae Mathematicae* 79, 9–16.
- Pesaran, M.H., Shin, Y., 2002. Long run structural modelling. *Econometric Reviews* 21, 49–87.
- Phillips, P.C.B., 1973. The problem of identification in finite parameter continuous time models. *Journal of Econometrics* 1, 351–362.
- Phillips, P.C.B., 1991a. Error correction and long run equilibrium in continuous time. *Econometrica* 59, 967–980.
- Phillips, P.C.B., 1991b. Optimal inference in cointegrated systems. *Econometrica* 59, 283–306.

- Phillips, P.C.B., 1991c. Spectral regression for cointegrated time series. In: Barnett, W.A., Powell, J., Tauchen, G.E., (Eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*. Cambridge University Press, Cambridge.
- Phillips, P.C.B., Durlauf, S.N., 1986. Multiple time series regression with integrated processes. *Review of Economic Studies* 53, 473–495.
- Robinson, P.M., 1976. Fourier estimation of continuous time models. In: Bergstrom, A.R., (Ed.), *Statistical Inference in Continuous Time Economic Models*. North-Holland, Amsterdam.
- Robinson, P.M., 1993. Continuous time models in econometrics: closed and open systems, stocks and flows. In: Phillips, P.C.B., (Ed.), *Models, Methods and Applications of Econometrics: Essays in Honor of A.R. Bergstrom*. Blackwell, Oxford.
- Rozanov, Y.A., 1967. *Stationary Random Processes*. Holden Day, New York.
- Saikkonen, P., 1991. Asymptotically efficient estimation of cointegrating regressions. *Econometric Theory* 7, 1–21.
- Saikkonen, P., 1995. Problems with the asymptotic theory of maximum likelihood estimation in integrated and cointegrated systems. *Econometric Theory* 11, 888–911.
- Saikkonen, P., 2001. Statistical inference in cointegrated vector autoregressive models with nonlinear time trends in cointegrating relations. *Econometric Theory* 17, 327–356.
- Stock, J.H., 1987. Temporal aggregation and structural inference in macroeconomics: a comment. *Carnegie-Rochester Conference Series on Public Policy* 26, 131–140.
- Weiss, A.A., 1984. Systematic sampling and temporal aggregation in time series models. *Journal of Econometrics* 26, 271–281.