Cowles Foundation Paper 796

CHAPTER 16

Spectral regression for cointegrated time series

P. C. B. Phillips

1 Introduction

Efficient techniques for estimating the coefficients in a multiple system of linear equations by spectral methods were introduced by Hannan (1963). These techniques, which are related to work by Whittle (1951) on Gaussian likelihood estimation, provide the basis for a regression analysis in the frequency domain. Their principal advantage is that they permit a nonparametric treatment of regression errors so that it is not necessary for an investigator to be explicit about the generating mechanism for the errors other than to assume stationarity. In addition, the techniques make it possible to focus attention in a regression on the most relevant frequency, thereby offering a selective approach that has become known as band spectrum regression - see Hannan and Robinson (1973) and Engle (1974). They have also been extended to nonlinear models under conditions that parallel those of nonlinear regression theory - see Hannan (1971) and Robinson (1972). Most recently the methods have been refined to accommodate automatic, data-driven bandwidth selectors - see Robinson (1988). All of the theory has been developed for models where the time series are stationary or where the regressors are amenable to a generalized harmonic analysis.

The objective of the present chapter is to show how spectral methods may also be usefully employed in regressions for certain nonstationary time series such as integrated processes. Indeed, there are good reasons

This chapter was written in March 1988 while the author was a visiting professor at the University of Auckland and revised in January 1989. Two referees made helpful comments on the presentation and Mark Watson kindly pointed out a missing factor in the variance matrix V_{T0} given in Remark (f). My thanks also go to Glena Ames for her skill and effort in keyboarding the manuscript and to the NSF for support under Grants SES 8519191 and SES 8821180.

Reprinted from *Nonparametric and semiparametric methods in econometrics and statistics: Proceedings of the fifth international symposium in economic theory and econometrics*, edited by William A. Barnett, James Powell and George E. Tauchen.

why their use may even be more appealing in this context than in regressions for stationary series. The model I have in mind is a multivariate system of cointegrated time series. Such systems have been the object of study in many recent papers - see Engle and Granger (1987), Phillips and Durlauf (1986), Stock (1987), Park and Phillips (1988, 1989), and the special issues of the *Oxford Bulletin of Economics and Statistics* (1986) and the *Journal of Economic Dynamics and Control* (1988).

The approach in this chapter follows that of some of my other ongoing work - see Phillips (1988). This research focuses attention on full information estimation of cointegrated systems and gives strong arguments for the use of full maximum likelihood estimation of the system in error correction model (ECM) format. It is shown that such estimation brings the problem of inference within the locally asymptotically mixed normal (LAMN) family of Jeganathan (1980, 1982). This means that the cointegrating coefficient estimates are asymptotically median unbiased and symmetrically distributed, that an optimal theory of inference applies, and that hypothesis tests may be conducted using standard asymptotic chi-squared tests.

This chapter shows that similar advantages are enjoyed by system spectral methods. Moreover, these methods have the additional advantage over classical maximum likelihood that they permit a nonparametric treatment of the regression errors. In other words, full system specification and estimation (as in maximum likelihood) is not required. Indeed, the system spectral methods given here involve linear estimating equations and result in simply computed explicit formulas. These features mean that the methods avoid what can be awkward methodological problems of dynamic specification and that they focus entirely on what is the central problem of cointegrating regression theory - the estimation of long-run equilibrium relationships.

The analysis in this chapter employs the block triangular ECM representation of a cointegrated system that was given in my earlier work (1988). The triangular structure is especially appealing in the context of spectral regression. The reason is simple: In a simultaneous equations model with serially independent errors and a triangular structural coefficient matrix, it is well known that maximum likelihood is equivalent to generalized least squares (GLS) - see Lahiri and Schmidt (1978). When the errors are serially dependent, the ECM model can be transformed into the frequency domain, retaining the triangular structure but inducing errors that are asymptotically independent across frequency. Then efficient estimation of the ECM requires GLS in the frequency domain, which is popularly known as "Hannan efficient" spectral regression. Interestingly, my results show that in cointegrated systems full frequency band regression is unnecessary for efficient estimation in large samples, although it may well be helpful in finite samples. The explanation is that the time-domain regressors in the ECM are integrated processes and in the frequency domain their spectra have dominant behavior at the origin. This means that in constructing efficient estimates by GLS methods the asymptotic theory calls only for GLS weights that take account of the covariance structure at the zero frequency or what we call the long-run error covariance matrix. This simplification is of great significance. In effect, efficient estimation of a cointegrating regression and hence the parameters of a long-run equilibrium relationship do not require full maximum likelihood estimation as in Phillips (1988) or Johansen (1988). Individual parameters that govern the short-run dynamics of the model are in large part irrelevant to the optimal estimation of the long-run coefficients. All that is needed for the latter is the composite effects that are embodied in the long-run error covariance matrix, which is simply the arithmetic sum of the serial covariances for all lags.

The following notation is used throughout the chapter. The symbol "$\Rightarrow$" signifies weak convergence, the symbol "$\equiv$" signifies equality in distribution, and the inequality ">0" signifies positive definite when applied to matrices. Stochastic processes such as the Brownian motion $W(r)$ on $[0,1]$ are frequently written as W to achieve notational economy. Similarly, I write integrals with respect to Lebesgue measure such as $\int_0^1 W(s)\,ds$ more simply as $\int_0^1 W$. Vector Brownian motion with covariance matrix Ω is written $BM(\Omega)$. I use $\|A\|$ to represent the Euclidean norm $\mathrm{tr}(A'A)^{1/2}$ of the matrix A, A^* to signify its complex conjugate transpose, $[x]$ to denote the smallest integer $\leq x$, and $I(1)$ and $I(0)$ to signify time series that are integrated of order 1 and 0, respectively. All limits given in the chapter are taken as the sample size $T\to\infty$ unless otherwise stated. Proofs of theorems are given in the chapter appendix.

2 Model and estimators

The model is the cointegrated system

$$y_{1t}=\beta' y_{2t}+u_{1t}, \tag{1}$$

$$\Delta y_{2t}=u_{2t}, \tag{2}$$

where

$$y_t=\begin{bmatrix} y_{1t} \\ y_{2t}\end{bmatrix}\begin{matrix}1\\ m\end{matrix} \equiv I(1)$$

is an integrated n-vector process $(n=m+1)$ and

$$u_t = \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}^{1}_{m} \equiv I(0)$$

is stationary with continuous spectral density matrix $f_{uu}(\lambda) > 0$ over $-\pi < \lambda \leq \pi$. As formulated, (1) is a single equation cointegrating regression with cointegrating vector $\alpha' = (1, -\beta')$. This approach may be extended to multiple equation cointegrating regressions, in which case β is a matrix of coefficients. The required extensions are straightforward and involve nothing that is fundamentally new, so they will not be given here.

We shall assume that the partial sum process $\underset{\sim}{P}_t = \Sigma_1^t u_j$ satisfies the multivariate invariance principle

$$T^{-1/2}\underset{\sim}{P}_{[Tr]} \Rightarrow \underset{\sim}{B}(r) \equiv BM(\underset{\sim}{\Omega}), \quad 0 < r \leq 1, \tag{3}$$

where $\underset{\sim}{\Omega} = 2\pi f_{uu}(0)$. We decompose the "long-run" covariance matrix $\underset{\sim}{\Omega}$ as follows:

$$\underset{\sim}{\Omega} = \underset{\sim}{\Sigma} + \underset{\sim}{\Lambda} + \underset{\sim}{\Lambda}',$$

where

$$\underset{\sim}{\Sigma} = E(u_0 u_0'), \qquad \underset{\sim}{\Lambda} = \sum_{k=1}^{\infty} E(u_0 u_k').$$

Also, we define

$$\underset{\sim}{\Delta} = \underset{\sim}{\Sigma} + \underset{\sim}{\Lambda}.$$

In addition to (3) we assume weak convergence of the stochastic process constructed from the sample covariance between $\underset{\sim}{P}_t$ and u_t, namely,

$$T^{-1}\sum_{1}^{[Tr]} \underset{\sim}{P}_t u_t' \Rightarrow \int_0^r \underset{\sim}{B}\, d\underset{\sim}{B}' + r\underset{\sim}{\Delta}. \tag{4}$$

Explicit conditions under which (3) and (4) hold are discussed in earlier work, and the reader is referred to Phillips (1987) for references and for a review. Suffice it to say here that they are general enough to include a wide class of weakly dependent processes $\{u_t\}$ under mild moment conditions.

It is convenient to partition the Brownian motion $\underset{\sim}{B}$ and the matrices $\underset{\sim}{\Omega}, \underset{\sim}{\Sigma}, \underset{\sim}{\Lambda}, \underset{\sim}{\Delta}$ conformably with the vector y_t. For example, we shall write

$$\underset{\sim}{B} = \begin{bmatrix} \underset{\sim}{B}_1 \\ \underset{\sim}{B}_2 \end{bmatrix}, \qquad \underset{\sim}{\Omega} = \begin{bmatrix} \underset{\sim}{\omega}_{11} & \underset{\sim}{\omega}_{21}' \\ \underset{\sim}{\omega}_{21} & \underset{\sim}{\Omega}_{22} \end{bmatrix}, \qquad \underset{\sim}{\Sigma} = \begin{bmatrix} \underset{\sim}{\sigma}_{11} & \underset{\sim}{\sigma}_{21}' \\ \underset{\sim}{\sigma}_{21} & \underset{\sim}{\Sigma}_{22} \end{bmatrix},$$

and so on. We also define $\underset{\sim}{\omega}_{11\cdot 2} = \underset{\sim}{\omega}_{11} - \underset{\sim}{\omega}_{21}' \underset{\sim}{\Omega}_{22}^{-1} \underset{\sim}{\omega}_{21}$.

The cointegrated system (1) and (2) has the following ECM representation:

$$\Delta y_t = \gamma \alpha' y_{t-1} + v_t, \tag{5}$$

where

$$\gamma' = (-1, 0), \qquad \alpha' = (1, -\beta'),$$

$$v_t = \begin{bmatrix} 1 & \beta' \\ 0 & I \end{bmatrix} u_t = Du_t.$$

This is the triangular system ECM representation derived in Phillips (1988). It is this system that we now propose to estimate using spectral methods.

Note that the error process v_t in (5) is stationary with spectral matrix $f_{vv}(\lambda) = Df_{uu}(\lambda)D' > 0$. We write $\Omega = 2\pi f_{vv}(0)$, $B(r) = D\underline{B}(r) \equiv BM(\Omega)$, $P_t = D\underline{P}_t$ and similarly define $\Sigma = D\underline{\Sigma}D'$, $\Lambda = D\underline{\Lambda}D'$, $\Delta = D\underline{\Delta}D'$. These matrices and vectors are partitioned conformably with y_t just as their counterparts without the sub bar. Corresponding to (3) and (4) we have

$$T^{-1/2}P_{[Tr]} \Rightarrow B(r), \tag{6}$$

$$T^{-1}\sum_{1}^{[Tr]} P_t v_t' \Rightarrow \int_0^r B\,dB + r\Delta. \tag{7}$$

We make use of the efficient method of estimation introduced by Hannan (1963) for linear systems and later extended by Hannan (1971) and Robinson (1972) to nonlinear regression equations. To this end we introduce the finite Fourier transforms

$$w_\Delta(\lambda) = (2\pi T)^{-1/2}\sum_{t=1}^{T} \Delta y_t e^{it\lambda}$$

$$w_*(\lambda) = (2\pi T)^{-1/2}\sum_{t=1}^{T} y_{*t} e^{it\lambda}$$

$$w_y(\lambda) = (2\pi T)^{-1/2}\sum_{t=1}^{T} y_{t-1} e^{it\lambda}$$

$$w_v(\lambda) = (2\pi T)^{-1/2}\sum_{t=1}^{T} v_t e^{it\lambda}$$

for $\lambda \in [-\pi, \pi]$, $y_{*t}' = (y_{1t}, \Delta y_{2t}')$ and we transform (5) accordingly as

$$w_\Delta(\lambda) = \gamma\alpha' w_y(\lambda) + w_v(\lambda). \tag{8}$$

We partition $w_y(\lambda)$ conformably with y_t using the notation

$$w_y(\lambda)' = (w_1(\lambda), w_2(\lambda)').$$

Following Hannan (1971) and Robinson (1972), a class of nonlinear weighted least squares estimates of α may be obtained by minimization of the Hermitian form

$$\sum_{\mathcal{B}} \operatorname{tr}\{[w_\Delta(\lambda_s) - \gamma\alpha' w_y(\lambda_s)][w_\Delta(\lambda_s) - \gamma\alpha' w_y(\lambda_s)]^* \Phi(\lambda_s)\} \tag{9}$$

with respect to α, where Φ is a given positive definite Hermitian matrix, $\lambda_s = 2\pi s/T$ and s is integral with values in the interval $-[T/2] < s \leq [T/2]$. The summation in (9) is over $\lambda_s \in \mathcal{B}$, which is a subset of $(-\pi, \pi)$ such that if $\lambda \in \mathcal{B}$ then $-\lambda \in \mathcal{B}$ also. The use of $\mathcal{B}$ permits the restriction of the regression to a set of frequency bands in $(-\pi, \pi)$ and is inspired by the idea that the model, when formulated in the frequency domain as in (8), may be more appropriate for λ in certain bands than in others. In practice, therefore, the regression may be confined to what seem to be the relevant frequency bands and as such is known as band spectrum regression. The reader is referred to Hannan (1963, 1970), Hannan and Robinson (1973), and Robinson (1972) for further details and discussion, and to Engle (1974) for an econometric application of these ideas.

In the present context, because the cointegrating vector α defines a long-run relationship between the components of the time series, one possibility would be to confine the regression to a band around the origin so that low-frequency elements in the series are emphasized. In this application to a U.S. aggregate expenditure relationship, Engle (1974) in one case eliminated high-frequency elements from the regression, using the argument that these are associated with transitory components of the two variables - expenditure and income - in the regression.

In conventional spectral regression, choice of the weight function $\Phi(\cdot)$ that appears in (9) involves only efficiency considerations. Indeed, the criterion (9) would be proportional to the exponent in the Gaussian likelihood of (8) if the $w_v(\lambda_s)$ were independent (complex) normal random vectors with covariance matrix $\Phi(\lambda_s)^{-1}$. In the Hannan efficient procedure $\Phi(\cdot)$ is selected in such a way that this is achieved asymptotically. This approach, which originates in the work of Whittle (1951), relies on the fact that under rather general conditions on $\{v_t\}$ and for λ_s in a band around ω (so that $\lambda_s \to \omega$ as $T \to \infty$), we find

$$w_v(\lambda_s) \Rightarrow N^c(0, f_{vv}(\omega)), \quad \omega \neq 0, \pi \tag{10}$$

(for example, Brillinger 1974, theorem 4.4.1), where N^c signifies the complex normal distribution. In designing an efficient procedure we may then select $\Phi(\lambda_s) = \hat{f}_{vv}(\omega)^{-1}$ for λ_s in a band centered on ω and for some suitable choice of consistent spectral estimate $\hat{f}_{vv}$. Details of the construction are given by Hannan (1963).

I envisage a straightforward application of these ideas in the present context. However, unlike the conventional spectral regression model, the regressors y_{t-1} in (5) (and hence $w_y(\lambda)$ in (8)) are in general coherent with the errors $v_t(w_v(\lambda))$. The regressors are also $I(1)$, not $I(0)$, processes. These features of the present model make the choice of weight function Φ critical. As I will show, a nonefficient choice of Φ induces a (second-order) bias effect in estimation as well as a loss of efficiency. As a result, fully efficient procedures have much more to recommend them in the present application.

A simple way to estimate $f_{vv}(\lambda)$ is to use the residuals from an initial least squares regression on (1). Writing $\hat{v}_t = \Delta y_t - \gamma\hat{\alpha}'y_{t-1}$, we may now compute the smoothed periodogram estimate

$$\hat{f}_{vv}(\omega_j) = \frac{2M}{T}\sum_{\mathcal{B}_j}[w_\Delta(\lambda_s) - \gamma\hat{\alpha}'w_y(\lambda_s)][w_\Delta(\lambda_s) - \gamma\hat{\alpha}'w_y(\lambda_s)]^*, \tag{11}$$

where the summation is over

$$\lambda_s \in \mathcal{B}_j = \left(\omega_j - \frac{\pi}{2M} < \lambda \leq \omega_j + \frac{\pi}{2M}\right),$$

that is, a frequency band of width π/M centered on

$$\omega_j = \frac{\pi j}{M}, \quad j = -M+1, \ldots, M$$

for M integer. Setting $m = [T/2M]$, we are now in effect averaging m neighboring periodogram ordinates around the frequency ω_j to obtain $\hat{f}_{vv}(\omega_j)$. As usual, we require $M \to \infty$ in such a way that $M/T \to 0$ (so that $m \to \infty$). In fact, it is convenient for the proofs to require that $M = o(T^{1/2})$, as in Hannan (1970, p. 489). Because $\hat{\alpha}$ is consistent (Phillips and Durlauf 1986; Stock 1987), we find that when $\omega_j \to \omega$ we have

$$\hat{f}_{vv}(\omega_j) \xrightarrow{p} f_{vv}(\omega) \quad \text{as } T \to \infty.$$

A further consideration is that, since $\gamma' = (-1, 0)$ is known by virtue of the construction of (5), nonlinear methods are not required. Indeed, minimization of (9) with the following choice of weight function

$$\Phi(\lambda_s) = \hat{f}_{vv}(\omega_j)^{-1} \quad \text{for all } \lambda_s \in \mathcal{B}_j$$

leads directly to the estimator

$$\begin{aligned}\tilde{\beta} &= -\left[\frac{1}{2M}\sum_{j=-M+1}^{M}\gamma'\hat{f}_{vv}^{-1}(\omega_j)\gamma\hat{f}_{22}'(\omega_j)\right]^{-1} \\ &\quad \times \left[\frac{1}{2M}\sum_{j=-M+1}^{M}\hat{f}_{2*}(\omega_j)\hat{f}_{vv}^{-1}(\omega_j)\gamma\right] \\ &= \left[\frac{1}{2M}\sum_{j=-M+1}^{M}e'\hat{f}_{vv}^{-1}(\omega_j)e\hat{f}_{22}'(\omega_j)\right]^{-1} \\ &\quad \times \left[\frac{1}{2M}\sum_{j=-M+1}^{M}\hat{f}_{2*}(\omega_j)\hat{f}_{vv}^{-1}(\omega_j)e\right],\end{aligned} \tag{12}$$

where

$$\hat{f}_{22}(\omega_j) = \frac{1}{m} \sum_{\mathcal{B}_j} w_2(\lambda_s) w_2(\lambda_s)^*, \tag{13}$$

$$\hat{f}_{2*}(\omega_j) = \frac{1}{m} \sum_{\mathcal{B}_j} w_2(\lambda_s) w_*(\lambda_s)^*, \tag{14}$$

and $e' = -\gamma' = (1, 0, \ldots, 0)$ is the first unit n-vector.

Since our focus of interest is the (long-run) cointegrating vector $\alpha' = (1, -\beta')$, alternative estimators might be considered that are based on low-frequency averages. One such possibility is

$$\tilde{\beta}_{(0)} = \frac{-\hat{f}_{22}(0)^{-1}\hat{f}_{2*}(0)\hat{f}_{vv}^{-1}(0)\gamma}{\gamma'\hat{f}_{vv}^{-1}(0)\gamma} = \frac{\hat{f}_{22}(0)^{-1}\hat{f}_{2*}(0)\hat{f}_{vv}^{-1}(0)e}{e'\hat{f}_{vv}^{-1}(0)e}, \tag{15}$$

which relies only on spectral estimates at the origin. The information that is neglected in the formation of $\tilde{\beta}_{(0)}$ (in relation to $\tilde{\beta}$) turns out to be unimportant at least asymptotically, as we shall show in Section 3.

In formulas (12) and (15) above, $\tilde{\beta}$ and $\tilde{\beta}_{(0)}$ have been constructed from the smoothed periodogram spectral estimates (11), (13), and (14). We observe that other conventional choices of spectral estimates may be employed in these formulas without affecting the asymptotic theory obtained below. Note also that the periodogram at frequency zero is itself zero if data have been derived from their sample means. This does not cause any difficulty in the computation of $\tilde{\beta}_{(0)}$ because smoothed periodogram estimates involve ordinates at points other than zero. As in the work of Park and Phillips (1988, 1989), however, the use of demeaned or detrended data does affect the limit distribution theory, typically by the replacement of Brownian motion in the limit functionals by respective demeaned or detrended processes.

3 Asymptotic theory

The main result is the following:

Theorem 3.1.

(a) $T(\tilde{\beta} - \beta) \Rightarrow (\int_0^1 B_2 B_2')^{-1}(\int_0^1 B_2\, dB_{1\cdot2})$,
(b) $T(\tilde{\beta}_{(0)} - \beta) \Rightarrow (\int_0^1 B_2 B_2')^{-1}(\int_0^1 B_2\, dB_{1\cdot2})$,

where

$$\begin{bmatrix} B_{1\cdot2} \\ B_2 \end{bmatrix}_m^1 \equiv BM\left(\begin{bmatrix} \omega_{11\cdot2} & 0 \\ 0 & \Omega_{22} \end{bmatrix}\right)$$

and

$$\omega_{11\cdot2} = \omega_{11} - \omega_{21}'\Omega_{22}^{-1}\omega_{21}.$$

The common limit distribution in (a) *and* (b) *is given explicitly in mixed normal form by the integral*

$$\int_{g>0} N(0, g\omega_{11\cdot 2}\Omega_{22}^{-1})\, dP(g), \tag{16}$$

where

$$g = e_1'\left(\int_0^1 W_2 W_2'\right)^{-1} e_1, \tag{17}$$

e_1 *is the first unit m-vector, and* $W_2 \equiv BM(I_m)$.

Remark (a): We see from the result stated that $\tilde{\beta}$ and $\tilde{\beta}_{(0)}$ are asymptotically equivalent. Information in the component spectral estimates at the origin is all that is relevant in the limit distribution, and this is all that is used in the construction of $\tilde{\beta}_{(0)}$. In empirical work, of course, $\tilde{\beta}$ and $\tilde{\beta}_{(0)}$ will differ. However, since much of the spectral power is concentrated in an immediate neighborhood of the origin for most aggregate economic time series it seems likely that this difference between the estimates will not be great, at least in those practical applications where the sample size is large.

The asymptotic equivalence of $\tilde{\beta}$ and $\tilde{\beta}_{(0)}$ shows that full frequency band regression is unnecessary for efficient estimation in large samples. As pointed out in the introduction, the reason for this is that the regressors in (5) are integrated processes, whose spectral power is concentrated at the origin. This means that in constructing GLS estimates in the frequency domain, the contribution of the zero frequency ordinate dominates and all that is required for the optimal GLS weighting asymptotically is the error covariance structure at the zero frequency.

Remark (b): The representation (16) shows that the limit distribution is a continuous mixture of normals. The mixing variate is the scalar (17). If we partition the m-vector standard Brownian motion W_2 as

$$W_2' = [\overset{1}{W_{21}}\quad \overset{m-1}{W_{22}'}],$$

then we can also write (17) in the form

$$g = \left\{\int_0^1 W_{21}^2 - \int_0^1 W_{21}W_{22}'\left(\int_0^1 W_{22}W_{22}'\right)^{-1}\int_0^1 W_{22}W_{21}\right\}^{-1}.$$

Remark (c): The limit distribution (16) is the same as that of the full maximum likelihood estimator of β in (5) when an explicit parametric model is assumed for the data-generating mechanism of the innovation vector v_t. When v_t is generated by an ARMA process this estimator is obtained by constructing the full (Gaussian) likelihood by a method such as the innovations algorithm (see, for example, Brockwell and Davis 1987). The properties of this maximum likelihood estimator of β are explored in Phillips (1988). The result shows that spectral regression offers a simple alternative to maximum likelihood that has several advantages:

(i) The method leads to explicit easily calculated formulas;
(ii) it offers the additional generality of stationary (rather than ARMA) errors in (5);
(iii) it avoids the methodological problems that are involved in the specification of short-run dynamics through what is, in effect, a nonparametric treatment of the errors.

Remark (d): As remarked in Section 2, the choice of an efficient estimator is critical to the above result. Suppose, for example, a general weight function $\Phi(\cdot)$ were employed in (9). This would lead to the following estimator in place of $\tilde{\beta}$:

$$\tilde{\beta}_\Phi = -\left[\frac{1}{2M}\sum_{j=-M+1}^{M}\gamma'\Phi(\omega_j)\gamma\hat{f}_{22}'(\omega_j)\right]^{-1}\left[\frac{1}{2M}\sum_{j=-M+1}^{M}\hat{f}_{2*}(\omega_j)\Phi(\omega_j)\gamma\right].$$

The asymptotics for this estimator are given by

Theorem 3.2.

$$T(\tilde{\beta}_\Phi-\beta) \Rightarrow\left(\int_0^1 B_2B_2'\right)^{-1}\left(\int_0^1 B_2\,dB'\Phi(0)e+\sum_{g=-\infty}^{\infty}\Delta_2(g+1)F_ge\right)\Big/e'\Phi(0)e, \tag{18}$$

where $\Phi(\lambda)$ has the following Fourier series representation:

$$\Phi(\lambda)=\frac{1}{2\pi}\sum_{g=-\infty}^{\infty}F_ge^{ig\lambda} \tag{19}$$

and where

$$\Delta_2(g)=\sum_{j=0}^{\infty}E(u_{20}v_{j+g}').$$

The limit distribution (18) is no longer mixed normal. The distribution is, in fact, miscentered by a second-order bias that arises from two sources: the term $\sum_g \Delta_2(g+1)F_ge$ and the fact that the Brownian motion $B_2(r)$ is in general correlated with the Brownian motion $B'(r)\Phi(0)e$. Second and perhaps more important, the limit distribution (18) involves nuisance parameters that inhibit statistical inference. These nuisance parameters involve both the bias effects and the covariance matrix of the Brownian motion $B(r)$. They cannot be easily eliminated and their presence in the limit distribution renders (18) effectively impotent for inferential purposes.

Remark (e): The limit results given in Theorem 3.1 belong to the LAMN theory of Jeganathan (1980, 1982), LeCam (1986), and Davies (1986). As pointed out earlier, the criterion function (9) is asymptotically proportional to the exponent of the Gaussian likelihood of the model (5). This Gaussian likelihood belongs to the LAMN family of Jeganathan (1980) (see Phillips 1988 for details). The estimators $\tilde{\beta}$ and $\tilde{\beta}_{(0)}$ may therefore be regarded as spectral versions of maximum likelihood. As such they have all of the advantages of the latter, namely,

- (i) they are asymptotically median unbiased and symmetrically distributed;
- (ii) the nuisance parameters that appear in the limit distribution (16) involve only scale effects and are readily eliminated to facilitate inference;
- (iii) an optimal theory of inference applies (from LeCam 1986);
- (iv) hypothesis testing may be conducted using conventional asymptotic chi-squared criteria.

Remark (f): To pursue points (ii) and (iv), suppose we wish to test the following hypotheses about the cointegration space

$$H_0 : h(\beta) = 0, \qquad H_1 : h(\beta) \neq 0,$$

where $h(\ )$ is a twice continuously differentiable q-vector function of restrictions on β. We assume that $H = \partial h(\beta)/\partial \beta'$ has rank $q < m$.

To test H_0 against H_1 we may employ the Wald statistic in its usual form. Thus for the estimator $\tilde{\beta}$ we set up

$$M_1 = h(\tilde{\beta})'[\tilde{H} V_T \tilde{H}']^{-1} h(\tilde{\beta}),$$

where $\tilde{H} = H(\tilde{\beta})$ and

$$V_T = \frac{1}{T}\left[\frac{1}{2M}\sum_{j=-M+1}^{M} \gamma' \hat{f}_{vv}^{-1}(\omega_j)\gamma \hat{f}_{22}'(\omega_j)\right]^{-1}.$$

Here V_T is the conventional estimate of the asymptotic variance matrix of $\tilde{\beta}$ from spectral regression theory (see Hannan 1970, p. 442).

Similarly for $\tilde{\beta}_{(0)}$ we construct

$$M_2 = h(\tilde{\beta}_0)'[\tilde{H}_0 V_{T0} \tilde{H}_0']^{-1} h(\tilde{\beta}_0),$$

where

$$\tilde{H}_0 = H(\tilde{\beta}_0), \qquad V_{T0} = \frac{MK}{T}[\gamma'\hat{f}_{vv}(0)^{-1}\gamma\hat{f}_{22}(0)]^{-1}, \quad \text{and}$$

$$K = \int_{-1}^{1} k(s)\,ds.$$

In this variance formula $k(\cdot)$ is the lag window used in the construction of the spectral estimates that appear in (15). Thus, for the smoothed periodogram estimates given in (11), (13), and (14) we have $k(s) = \sin(\frac{1}{2}\pi s)/(\frac{1}{2}\pi s)$ and $K = 2$ (e.g., Hannan 1970, pp. 275–6).

The asymptotic theory for the test statistics M_1 and M_2 is as follows:

Theorem 3.3.

$$M_1, M_2 \Rightarrow \chi^2_q.$$

Thus, statistical tests of H_0 may be conducted in the usual fashion of asymptotic chi-squared tests. Interestingly, no modification to the conventional formulas from spectral regression theory are required in the case of the test M_1. The test based on M_2 involves a scale factor of $1/MK$ relative to the conventional test based on a band spectral regression for stationary time series. This is because the asymptotic behavior of the spectral estimate $\hat{f}_{22}(0)$ in V_{T0} is as follows:

$$\frac{1}{MT}\hat{f}_{22}(0) \Rightarrow \frac{1}{2\pi}K\int_0^1 B_2B_2'$$

(see (A.11) in the appendix). This limit involves the scale factor $K = \int_{-1}^{1} k(s)\,ds$ that measures the weight contributed by the lag window $k(\cdot)$ to the spectral estimate.

Remark (g): Single equation spectral regression methods do not have the same advantages as the systems estimators $\tilde{\beta}$ and $\tilde{\beta}_0$. To see this it is helpful to consider the following estimate, which is the analogue of $\tilde{\beta}$ for the first equation of (5):

$$\beta^* = \left[\frac{1}{2M}\sum_{j=-M+1}^{M}\hat{f}_{v_1v_1}^{-1}(\omega_j)\hat{f}_{22}'(\omega_j)\right]^{-1}\left[\frac{1}{2M}\sum_{j=-M+1}^{M}\hat{f}_{21}(\omega_j)\hat{f}_{v_1v_1}^{-1}(\omega_j)\right],$$

where $\hat{f}_{21}(\lambda)$ is an estimate of the cross spectrum between y_{2t-1} and y_{1t}. The estimator β^* is the Hannan (1963) efficient estimator of β in the equation

$$y_{1t} = \beta' y_{2t-1} + v_{1t}. \tag{20}$$

After minor modifications to adjust for the lag in (20), this is just the standard spectral regression estimator of β in the cointegrating regression equation (1). The asymptotic theory for β^* is given by

Theorem 3.4.

$$T(\beta^*-\beta) \Rightarrow \left(\int_0^1 B_2 B_2'\right)^{-1}\left(\int_0^1 B_2\, dB_1+\delta\right), \tag{21}$$

where

$$\delta = \left(\sum_{g=-\infty}^{\infty} \Delta_{21}(g+1)d_g\right)\Big/\left(\sum_{g=-\infty}^{\infty} d_g\right)$$

and

$$\Delta_{21}(g) = \sum_{j=0}^{\infty} E(u_{20}v_{1j+g}).$$

The limit distribution (21) involves second-order bias effects and nuisance parameters arising from the presence of δ in the second factor of (21) and the correlation between the Brownian motions B_1 and B_2. As in the case of (18), these problems severely inhibit the usefulness of the estimator β^* for inferential purposes.

Note that by decomposing B_1 as follows:

$$B_1(r) = \omega_{21}'\Omega_{22}^{-1}B_2(r) + \omega_{11\cdot 2}^{1/2}W_1(r),$$

where $W_1(r)$ is standard Brownian motion, that is, $BM(1)$, and W_1 is independent of B_2, we deduce an alternative representation of (21) in the form

$$\left(\int_0^1 B_2 B_2'\right)^{-1}\left(\int_0^1 B_2\, dB_2'\Omega_{22}^{-1}\omega_{21}+\delta\right)+\omega_{11\cdot 2}^{1/2}\left(\int_0^1 B_2 B_2\right)^{-1}\int_0^1 B_2\, dW_1. \tag{22}$$

The first term of (22) involves the "unit root" distribution

$$\left(\int_0^1 B_2 B_2'\right)^{-1}\left(\int_0^1 B_2\, dB_2\right)$$

and the "bias effects" from the factor $\Omega_{22}^{-1}\omega_{21}$ and δ. The second term of (22) is mixed normal with the same distribution as (16).

The decomposition (22) highlights the differences between single-equation and systems spectral regressions in the model (5). Single-equation methods neglect the prior information of the m unit roots in (5) and ignore the joint dependence of y_{1t} and y_{2t}. As a result, these methods implicitly involve the estimation of unit roots and this is responsible for the presence of the unit root distribution in the first term of (22). In addition, we see that the neglect of the rest of the system in (5) imports a second-order bias effect through the term δ. The magnitude of this term depends on the extent of the contemporaneous and serial correlation between u_{2t} and v_{1t}.

Finally, we observe that Theorem 3.4 gives the asymptotic theory for the (full band) spectral estimator used by Engle (1974) in his application of spectral regression to the aggregate consumption function with quarterly U.S. data on money income and consumption. Our results suggest that the estimates of the propensity to consume obtained by Engle in this study are likely to be biased and that conventional tests are not validated, at least by the asymptotic theory given here. It would seem worthwhile to reanalyze this data set using the systems estimator $\tilde{\beta}$ (and $\tilde{\beta}_{(0)}$) and associated test statistics such as M_1 (and M_2). It is also of some interest to determine the extent of the bias effects in finite samples through simulation experiments.

Remark (h): The estimators $\tilde{\beta}$ and $\tilde{\beta}_{(0)}$ both rely on first-stage estimates of the residual spectrum. This in turn depends on a first-stage estimate of the cointegrating vector, which can be delivered by least squares. However, since the latter estimate has a second-order bias - manifested in a miscentered limit distribution (see Stock 1987 and Phillips and Durlauf 1986) - it would appear that some improvement may be expected by the use of iteration. This would entail the reestimation of the residual spectrum using residuals calculated from the regression coefficients $\tilde{\beta}$ or $\tilde{\beta}_{(0)}$ and then reestimation of β by (12) or (15). Again, simulations would be useful in advising about the merits of such iteration in finite samples.

4 Conclusion

This chapter provides a frequency domain extension of the results in Phillips (1988) on the maximum likelihood estimation of cointegrated systems. Indeed, full system spectral regression in an ECM is asymptotically equivalent to maximum likelihood and shares with it the advantages of belonging to the LAMN family. But spectral regression techniques seem to have more appeal in the context of cointegrated time series. This is true for several reasons.

1. They involve only linear estimating equations and thereby avoid the nonlinear optimization methods that are typically called for in the application of maximum likelihood (for instance, when there are ARMA error processes).
2. The nonparametric treatment of regression errors that is inherently involved in spectral methods avoids the methodological difficulties that are encountered with the need to completely specify the data generating mechanism of the errors before maximum likelihood is applied.
3. The nonparametric approach brings with it additional generality concerning the error processes at what seems to be little or no extra cost.

4. Even simpler methods are available - for example, the systems band spectral regression estimator given by (15) - and such estimators continue to enjoy the same asymptotic properties as the full system estimator (12).
5. Standard systems spectral regressions may be used with no modifications being necessary to deal with the regressor endogeneity that is characteristic of cointegrated systems.

It is important to emphasize that it is the systems spectral estimators given by (12) and (15) that have these advantages. Single-equation or subsystem spectral regressions have quite different asymptotic properties as shown in Theorem 3.4. In particular, they suffer from bias and nuisance parameter dependencies that inhibit their use for inference. Thus systems estimation brings with it considerably more than the usual efficiency gains we have come to expect from traditional asymptotic theory. In view of these apparent advantages of systems spectral estimators over direct maximum likelihood and single equation spectral methods, it would seem worthwhile to investigate their performance in sampling experiments and in empirical work.

In developing his alternative strategy for empirical macroeconomics via unrestricted vector autoregressions, Sims (1980) set out to emulate the spectral regression approach. He argued as follows:

> The style I am suggesting we emulate is that of frequency-domain time series theory (although it will be clear I am not suggesting we use frequency-domain methods themselves), in which what is being estimated (e.g. the spectral density) is implicitly part of an infinite dimensional parameter space, and the finite parameter models we actually use are justified as part of a procedure in which the number of parameters is explicitly a function of sample size or the data.

The present chapter puts forward some new arguments for the explicit use of frequency-domain methods in empirical macroeconomics. It is argued that if the parameters of a model can be separated sensibly into the coefficients of long-run relationships on the one hand and short-run dynamics on the other, then frequency domain regression provides a natural method for the efficient estimation of the long-run coefficients. If the short-run dynamic adjustment coefficients are best thought of in terms of an infinite dimensional parameter space, then spectral regression offers a convenient mechanism for dealing with this complexity. Interestingly, efficient estimation in the frequency domain does not necessarily involve the full spectral density. As we have seen, all that is important, at least in the model studied here, is the composite effect of serial covariances at all lags, that is, what is here termed the long-run covariance matrix and what is measured by the value of the spectral matrix at the origin.

Appendix

Proof of Theorem 3.1: (a) From (12) we find that

$$T(\tilde{\beta}-\beta)=\left[\frac{1}{2MT}\sum_{j=-M+1}^{M} e'\hat{f}_{vv}^{-1}(\omega_j)e\hat{f}_{22}'(\omega_j)\right]^{-1}$$
$$\times\left[\frac{1}{2M}\sum_{j=-M+1}^{M}\hat{f}_{2v}(\omega_j)\hat{f}_{vv}^{-1}(\omega_j)e\right]. \tag{A.1}$$

Our approach follows Hannan (1963) in general outline, with the main differences arising from the treatment of the nonstationary elements.

It is convenient to work with spectral estimates in (A.1) of the same general form, say

$$\hat{f}_{ab}(\lambda)=\frac{1}{2\pi}\sum_{n=-M}^{M} k\left(\frac{n}{M}\right)c_{ab}(n)e^{-in\lambda},$$

where

$$c_{ab}(n)=T^{-1}\sum_{1}^{T} a_t b_{t+n}, \quad 1\leq t+n\leq T$$

and where the lag window $k(\ )$ is a bounded even function defined on $[-1,1]$ with $k(0)=1$. For example, when $k(n/M)=1-|n|/M$, $\hat{f}_{ab}(\lambda)$ is the Bartlett estimator (e.g., Hannan 1970, p. 278). We may also replace $k(n/M)$ by $k(n/M)(1-|n|/T)$ in the formula without affecting the arguments that follow. Other spectral estimates may also be employed, but the formula above helps to simplify derivations and avoid repetition.

As in Hannan (1963) we have

$$\max_{\lambda}\|\hat{f}_{vv}(\lambda)-f_{vv}(\lambda)\|\xrightarrow{p}0$$

as $T\to\infty$ and then the limit behavior of (A.1) is equivalent to that of the same expression but with $f_{vv}(\omega_j)$ replacing $\hat{f}_{vv}(\omega_j)$ in both factors on the right-hand side. We take each of these in turn.

Using the Fourier series

$$f_{vv}^{-1}(\lambda)=\frac{1}{2\pi}\sum_{g=-\infty}^{\infty} D_g e^{ig\lambda},$$

we have

$$\frac{1}{2MT}\sum_{j=-M+1}^{M} e'f_{vv}^{-1}(\omega_j)e\hat{f}_{22}'(\omega_j)$$
$$=\frac{1}{2\pi T}\sum_{g=-\infty}^{\infty} e'D_g e\frac{1}{2M}\sum_{j=-M+1}^{M} e^{ig\pi j/M}\hat{f}_{22}'\left(\frac{\pi j}{M}\right)$$
$$=\left(\frac{1}{2\pi}\right)^2\frac{1}{T}\sum_{g=-\infty}^{\infty} e'D_g ec_{22}'(\underline{g})k\left(\frac{g}{M}\right), \tag{A.2}$$

where

$$\underline{g}+2lM=g, \quad -M+1\le \underline{g}\le M$$

for some integer l and where

$$c_{22}(n)=T^{-1}\sum_{1}^{T} y_{2t}y'_{2t+n}, \quad 1\le t+n\le T.$$

The next step is to determine the asymptotic behavior of (A.2). We start by defining the random elements

$$X_T(r)=T^{-1/2}P_{[Tr]}, \qquad Z_{Ti}(r)=T^{-1}\sum_{1}^{[Tr]} P_t v'_{t+i}.$$

In view of (6) and (7) we have the weak convergence

$$X_T(r)\Rightarrow X_\infty(r)=B(r)$$
$$Z_{Ti}(r)\Rightarrow Z_{\infty i}(r)=\int_0^r B\,dB+r\Delta(i),$$

where

$$\Delta(i)=\sum_{j=0}^{\infty} E(v_0v'_{j+i}).$$

Using the Skorohod construction, we now employ a new probability space with random elements $\{(X_T^*, Z_{Ti}^*)\}$, $(X_\infty^*, Z_{\infty i}^*)$ for which

$$X_T^* \xrightarrow{\text{a.s}} X_\infty^*, \qquad Z_{Ti}^* \xrightarrow{\text{a.s.}} Z_{\infty i}^* \tag{A.3}$$

and where

$$X_T\equiv X_T^*, \qquad Z_{Ti}\equiv Z_{Ti}^*$$

(with "$\equiv$" as usual representing equivalence in distribution). It will be convenient to use a superscript "2" on these random elements to signify subelements of matrices and vectors that correspond to the component u_{2t} of v_t. Thus we write

$$X_\infty^{(2)}=B_2, \qquad Z_{\infty i}^{(2)}=\int_0^r B_2\,dB'+r\Delta_2(i),$$
$$Z_{\infty i}^{(22)}=\int_0^r B_2\,dB_2'+r\Delta_{22}(i),$$

and so on. With this notation we have, for $0\le n\le M$ and up to a term of $O_p(T^{-1})$,

$$T^{-1}c_{22}(n)=T^{-2}\sum_{1}^{T} y_{2t}y'_{2t+n}$$
$$=\int_0^1 X_T^{(2)}X_T^{(2)\prime}+T^{-1}\{Z_{T1}^{(22)}(1)+\cdots+Z_{Tn}^{(22)}(1)\}$$

$$\equiv \int_0^1 X_T^{*(2)} X_T^{*(2)\prime} + T^{-1}\{Z_{T1}^{*(22)}(1) + \cdots + Z_{Tn}^{*(22)}(1)\} \quad (:= T^{-1}c_{22}^*(n), \text{ say})$$

$$\xrightarrow{\text{a.s.}} \int_0^1 B_2^* B_2^{*\prime}.$$

In view of (A.3) the final convergence takes place almost surely and uniformly in $|n| \leq M$ as $T \to \infty$. The same result also applies when $-M \leq n \leq 0$.

Because $k(g/M) \to 1$ for all fixed g as T (and hence M) $\to \infty$, we deduce that

$$\left(\frac{1}{2\pi}\right)^2 \frac{1}{T} \sum_{g=-\infty}^{\infty} e' D_g e c_{22}^{*\prime}(g) k\left(\frac{g}{M}\right) \xrightarrow{\text{a.s.}} \left(\frac{1}{2\pi}\right)^2 \left(\sum_{g=-\infty}^{\infty} e' D_g e\right) \int_0^1 B_2^* B_2^{*\prime}.$$

This, of course, implies that

$$\left(\frac{1}{2\pi}\right)^2 \frac{1}{T} \sum_{g=-\infty}^{\infty} e' D_g e c_{22}^{*\prime}(g) k\left(\frac{g}{M}\right) \Rightarrow \left(\frac{1}{2\pi}\right)^2 \left(\sum_{g=-\infty}^{\infty} e' D_g e\right) \int_0^1 B_2^* B_2^{*\prime}.$$

However,

$$c_{22}(g) \equiv c_{22}^*(g) \quad \text{for all } g$$

and

$$B_2(r) \equiv B_2^*(r),$$

so that by a simple modification of the Skorohod-Dudley-Wichura theorem (e.g., Shorack and Wellner 1986, p. 47) we deduce that

$$\left(\frac{1}{2\pi}\right)^2 \frac{1}{T} \sum_{g=-\infty}^{\infty} e' D_g e c_{22}'(g) k\left(\frac{g}{M}\right)$$
$$\Rightarrow \left(\frac{1}{2\pi}\right)^2 \left(\sum_{g=-\infty}^{\infty} e' D_g e\right) \int_0^1 B_2 B_2'. \tag{A.4}$$

Next observe that

$$\frac{1}{2\pi} \sum_{g=-\infty}^{\infty} e' D_g e = e' f_{vv}^{-1}(0) e$$

so that the right-hand side of (A.4) is simply

$$e'\Omega^{-1} e \int_0^1 B_2 B_2' = \frac{1}{\omega_{11\cdot 2}} \int_0^1 B_2 B_2'. \tag{A.5}$$

It remains to consider the second factor in the right-hand element of (A.1). Replacing $\hat{f}_{vv}$ by f_{vv} for the reason given earlier, we have

$$\frac{1}{2M} \sum_{j=-M+1}^{M} \hat{f}_{2v}(\omega_j) f_{vv}^{-1}(\omega_j) e$$

$$= \frac{1}{2\pi} \sum_{g=-\infty}^{\infty} \left(\frac{1}{2M} \sum_{j=-M+1}^{M} \hat{f}_{2v}(\omega_j) e^{ig\omega_j} \right) D_g e$$
$$= \left(\frac{1}{2\pi} \right)^2 \sum_{g=-\infty}^{\infty} c_{2v}(\underline{g}) D_g e k\left(\frac{g}{M} \right).$$

Now

$$\begin{aligned} c_{2v}(n) &= T^{-1} \sum_{1}^{T} y_{2t-1} v_{t+n}, \quad 1 \le t+n \le T \\ &= Z_{Tn}^{(2)}(1) \\ &\equiv Z_{Tn}^{*(2)}(1) \quad (:= c_{2v}^{*}(n), \text{ say}) \\ &\xrightarrow{\text{a.s.}} Z_{\infty n}^{*(2)}(1) = \int_0^1 B_2^* \, dB^{*\prime} + \Delta_2(n+1). \end{aligned}$$

We deduce that

$$\left(\frac{1}{2\pi} \right)^2 \sum_{g=-\infty}^{\infty} c_{2v}^{*}(\underline{g}) D_g e k\left(\frac{g}{M} \right) \xrightarrow{\text{a.s.}} \left(\frac{1}{2\pi} \right)^2 \int_0^1 B_2^* \, dB^{*\prime} \left(\sum_{g=-\infty}^{\infty} D_g e \right) + \left(\frac{1}{2\pi} \right)^2 \sum_{g=-\infty}^{\infty} \Delta_2(g+1) D_g e.$$

Using the Skorohod-Dudley-Wichura theorem as before, we obtain

$$\left(\frac{1}{2\pi} \right)^2 \sum_{g=-\infty}^{\infty} c_{2v}^{*}(\underline{g}) D_g e k\left(\frac{g}{M} \right) \Rightarrow \left(\frac{1}{2\pi} \right)^2 \int_0^1 B_2 \, dB' \left(\sum_{g=-\infty}^{\infty} D_g e \right) + \left(\frac{1}{2\pi} \right)^2 \sum_{g=-\infty}^{\infty} \Delta_2(g+1) D_g e. \tag{A.6}$$

Note that

$$e' \left(\frac{1}{2\pi} \right)^2 \left(\sum_g D_g \right) B(r) = e' \Omega^{-1} B(r) := B_e(r) \equiv BM\left(\frac{1}{\omega_{11 \cdot 2}} \right). \tag{A.7}$$

Next we define

$$\underline{v}_j = \sum_{g=-\infty}^{\infty} v'_{g+j+1} D_g e,$$

from which we deduce

$$\sum_g \Delta_2(g+1) D_g e = \sum_{j=0}^{\infty} E(u_{20} \underline{v}_j) = 0 \tag{A.8}$$

because

$$E(u_{20} \underline{v}_j) = \int_{-\pi}^{\pi} e^{ij\lambda} f_{2\underline{v}}(\lambda) \, d\lambda = 0 \quad \text{for all } j.$$

The latter follows in view of the fact that

$$f_{2v}(\lambda)=[0\ \ I]f_{vv}(\lambda)\left(\sum_{g=-\infty}^{\infty}D_g e^{ig\lambda}\right)e$$
$$=2\pi[0\ \ I]f_{vv}(\lambda)f_{vv}^{-1}(\lambda)e$$
$$=0$$

for all $\lambda\in(-\pi,\pi]$.

From (A.6)–(A.8) we obtain

$$\left(\frac{1}{2\pi}\right)^2\sum_{g=-\infty}^{\infty}c_{2v}(g)D_g e k\left(\frac{g}{M}\right)\Rightarrow\int_0^1 B_2\,dB_e. \tag{A.9}$$

Combining (A.9) with (A.4), (A.5), and (A.1), we deduce that

$$T(\tilde{\beta}-\beta)\Rightarrow\left[\frac{1}{\omega_{11\cdot 2}}\int_0^1 B_2B_2'\right]^{-1}\left[\int_0^1 B_2\,dB_e\right].$$

Because

$$\omega_{11\cdot 2}B_e(r)\equiv B_{1\cdot 2}(r)\equiv BM(\omega_{11\cdot 2}),$$

the stated result (a) follows immediately.

The proof of part (b) follows similar lines. For the reasons advanced earlier, $\hat{f}_{vv}(0)$ may be replaced by $f_{vv}(0)$ in the formula for $\tilde{\beta}_{(0)}$, giving

$$T(\tilde{\beta}_{(0)}-\beta)\sim\left\{\frac{1}{TM}\hat{f}_{22}(0)\right\}^{-1}\left\{\frac{1}{M}\hat{f}_{2v}(0)f_{vv}^{-1}(0)e\right\}\Big/e'f_{vv}^{-1}(0)e. \tag{A.10}$$

Using the Skorohod construction given in the proof of part (a), we have

$$\frac{1}{MT}\hat{f}_{22}(0)\equiv\frac{1}{2\pi M}\sum_{n=-M}^{M}k\left(\frac{n}{M}\right)\frac{1}{T}c_{22}^{*}(n)$$
$$\xrightarrow{\text{a.s.}}\nu\int_0^1 B_2^{*}B_2^{*\prime},\qquad \nu=\frac{1}{2\pi}\int_{-1}^{1}k(s)\,ds$$

as $T\to\infty$, from which we deduce that

$$\frac{1}{MT}\hat{f}_{22}(0)\Rightarrow\nu\int_0^1 B_2B_2'. \tag{A.11}$$

Similarly we find

$$\frac{1}{M}\hat{f}_{2v}^{*}(0)=\frac{1}{2\pi M}\sum_{n=-M}^{M}k\left(\frac{n}{M}\right)c_{2v}^{*}(n)$$
$$\xrightarrow{\text{a.s.}}\nu\int_0^1 B_2^{*}\,dB^{*\prime}+\frac{1}{2\pi}\Delta_2, \tag{A.12}$$

where

$$\Delta_2=\sum_{j=-\infty}^{\infty}E(u_{20}v_j').$$

To see this, note that for fixed n we have

$$c_{2v}^*(n) \xrightarrow{\text{a.s.}} \int_0^1 B_2^* \, dB^{*\prime} + \Delta_2(n+1),$$

where

$$\Delta_2(n) = \sum_{j=0}^{\infty} E(u_{20} v_{j+n}') = \sum_{l=n}^{\infty} E(u_{20} v_l').$$

Because $\Delta_2(n) \to 0$ as $n \to \infty$ and $\Delta_2(n) \to \Delta_2$ as $n \to -\infty$, we find that the Cesaro sum

$$\frac{1}{M} \sum_{n=-M}^{M} \Delta_2(n) \to \Delta_2, \quad M \to \infty,$$

giving (A.12). Hence

$$\frac{1}{M} \hat{f}_{2v}(0) \Rightarrow \nu \int_0^1 B_2 \, dB' + \frac{1}{2\pi} \Delta_2. \tag{A.13}$$

Next we observe that

$$f_{vv}^{-1}(0) e = 2\pi \Omega^{-1} e \tag{A.14}$$

and

$$\begin{aligned} \Delta_2 \Omega^{-1} e &= [0 \ \ I_m] \left\{ \sum_{j=-\infty}^{\infty} E(v_0 v_j') \right\} \Omega^{-1} e \\ &= [0 \ \ I_m] \Omega \Omega^{-1} e \\ &= [0 \ \ I_m] e = 0. \end{aligned} \tag{A.15}$$

It then follows from (A.10), (A.11), and (A.13)–(A.15) that

$$\begin{aligned} T(\tilde{\beta}_{(0)} - \beta) &\Rightarrow \left(\nu \int_0^1 B_2 B_2' \right)^{-1} \left(\nu \int_0^1 B_2 \, dB' \Omega^{-1} e \Big/ e' \Omega^{-1} e \right) \\ &= \left(\int_0^1 B_2 B_2' \right)^{-1} \int_0^1 B_2 \, dB_{1\cdot 2} \end{aligned}$$

as required. **Q.E.D.**

Proof of Theorem 3.2: This follows in the same way as the proof of part (a) of Theorem 3.1. We simply use (19) in place of the Fourier series for $f_{vv}^{-1}(\lambda)$. **Q.E.D.**

Proof of Theorem 3.3: This is the same as the proof of Theorem 4.1 of Phillips (1988). **Q.E.D.**

Proof of Theorem 3.4: This follows the same lines as the proof of part (a) of Theorem 3.1 above. We use the Fourier series

$$f_{v_1v_1}^{-1}(\lambda) = \frac{1}{2\pi} \sum_{g=-\infty}^{\infty} d_g e^{ig\lambda},$$

and then

$$\frac{1}{2MT} \sum_{j=-M+1}^{M} \hat{f}_{v_1v_1}^{-1}(\lambda)\hat{f}_{22}(\omega_j) \Rightarrow \left(\frac{1}{2\pi}\right)^2 \left(\sum_{g=-\infty}^{\infty} d_g\right) \int_0^1 B_2 B_2' = \frac{1}{\omega_{11}} \int_0^1 B_2 B_2'$$

because

$$\left(\frac{1}{2\pi}\right)^2 \sum_g d_g = \frac{1}{2\pi} f_{v_1v_1}^{-1}(0) = \frac{1}{\omega_{11}}.$$

Next,

$$\frac{1}{2M} \sum_{j=-M+1}^{M} \hat{f}_{2v_1}(\omega_j)\hat{f}_{v_1v_1}^{-1}(\omega_j)$$
$$\Rightarrow \left(\frac{1}{2\pi}\right)^2 \int_0^1 B_2\, dB_1 \left(\sum_g d_g\right) + \left(\frac{1}{2\pi}\right)^2 \sum_g \Delta_{21}(g+1) d_g$$
$$= \frac{1}{\omega_{11}} \int_0^1 B_2\, dB_1 + \delta_1.$$

The result stated now follows with $\delta = \omega_{11}\delta_1$. Q.E.D.

REFERENCES

Brillinger, D. R. (1974), *Time Series: Data Analysis and Theory* (New York: Holt, Rinehart and Winston).

Brockwell, P. J., and R. A. Davis (1987), *Time Series: Theory and Methods* (New York: Springer).

Davies, R. B. (1986), "Asymptotic Inference When the Amount of Information Is Random," in L. M. LeCam and R. A. Olshen (eds.), *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer,* Vol. II (Wadsworth).

Engle, R. F. (1974), "Band Spectrum Regression," *International Economic Review,* 15, 1-11.

Engle, R. F., and C. W. J. Granger (1987), "Cointegration and Error Correction: Representation, Estimation and Testing," *Econometrica,* 55, no. 87, 251-76.

Hannan, E. J. (1963), "Regression for Time Series," in M. Rosenblatt (ed.), *Time Series Analysis* (New York: Wiley).

(1970), *Multiple Time Series* (New York: Wiley).

(1971), "Nonlinear Time Series Regression," *Journal of Applied Probability,* 8, 767-80.

Hannan, E. J., and P. M. Robinson (1973), "Lagged Regression with Unknown Lags," *Journal of the Royal Statistical Society,* Series B, 35, 252-67.

Jeganathan, P. (1980), "An Extension of a Result of L. LeCam Concerning Asymptotic Normality," *Sankhya Series A,* 42, 146-60.

(1982), "On the Asymptotic Theory of Estimation When the Limit of the Log-Likelihood Ratios is Mixed Normal," *Sankhya Series A,* 44, 173-212.

Johansen, S. (1988), "Statistical Analysis of Cointegrating Vectors," *Journal of Economic Dynamics and Control*, 12, 231–54.

Journal of Economic Dynamics and Control (1988), 12, special issue.

Lahiri, K., and P. Schmidt (1978), "On the Estimation of Triangular Structural Systems," *Econometrica*, 46, 1217–22.

LeCam, L. (1986), *Asymptotic Methods in Statistical Decision Theory* (New York: Springer).

Oxford Bulletin of Economics and Statistics (1986), 48, special issue.

Park, J. Y., and P. C. B. Phillips (1988), "Statistical Inference in Regressions with Integrated Processes: Part 1," *Econometric Theory*, 4, 468–98.

(1989), "Statistical Inference in Regressions with Integrated Processes: Part 2," *Econometric Theory*, 5, 95–132.

Phillips, P. C. B. (1987), "Multiple Regression with Integrated Processes," in N. U. Prabhu (ed.), *Statistical Inference from Stochastic Processes, Contemporary Mathematics*, 80, 79–106.

(1988), "Optimal Inference in Cointegrated Systems," Cowles Foundation Discussion Paper No. 866, Yale University, to appear in *Econometrica*, 1991.

Phillips, P. C. B., and S. N. Durlauf (1986), "Multiple Time Series with Integrated Variables," *Review of Economic Studies*, 53, 473–96.

Robinson, P. M. (1972), "Nonlinear Vector Time Series Regression," *Journal of Applied Probability*, 9.

(1988), "Automatic Generalized Least Squares" (mimeod, LSE).

Shorack, G. R., and J. A. Wellner (1986), *Empirical Processes with Applications to Statistics* (New York: Wiley).

Sims, C. A. (1980), "Macroeconomics and Reality," *Econometrica*, 48, 1–48.

Stock, J. H. (1987), "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors," *Econometrica*, 55, 1035–56.

Whittle, P. (1951), *Hypothesis Testing in Time Series Analysis* (Uppsala, Sweden: Almqvist and Wicksell).