Efficient estimation and testing of cointegrating vectors in the presence of deterministic trends*

Bruce E. Hansen

University of Rochester, Rochester, NY 14627, USA

Estimation and inference in cointegrated models is examined in the presence of deterministic trends in the data. It is suggested that trends be excluded in the levels regression for maximal efficiency. Fully modified test statistics are asymptotically chi-square. A chi-square test for the validity of trend exclusion is presented. The asymptotic distributions of standard cointegration test statistics are shown to depend both upon regressor trends and estimation detrending methods.

1. Introduction

A vector of random variables is said to be cointegrated if a linear combination of the variables has a stationary distribution, yet considered individually, each element is integrated of order one [$I(1)$]. The latter processes are nonstationary in levels, but stationary after differencing. A new body of statistical theory has developed for the estimation of these cointegrating vectors; see, for example, Engle and Granger (1987), Stock (1987), Johansen (1988a, b), Park and Phillips (1988, 1989), Phillips (1988, 1991), Phillips and Hansen (1990), and Sims, Stock, and Watson (1990). Much of this literature has abstracted away from the presence of deterministic trends in the regressors. Many macroeconomic variables which are commonly described as $I(1)$ such as GNP, consumption, or the price level are actually best

thought of as '$I(1)$ with drift', which is the sum of an $I(1)$ process with zero-mean increments and a linear trend. The theoretical literature and econometric practice has been to detrend the data, often by inclusion of a linear trend in the levels regression. This has the advantage of rendering estimates of the cointegrating vector invariant to the presence of trends and simplifies the asymptotic theory.

Some exceptions should be noted. West (1988) explicitly examines the case of a single nonstationary regressor without detrending. His analysis, however, does not generalize to multiple regression, as pointed out by Park and Phillips (1988). These authors provide a preliminary analysis of the effects of failing to detrend when the regressors have drift (their theorems 3.5 and 3.6), but emphasize the degenerate nature of their limiting representations and fail to consider inference based upon these estimates or contrast the relative efficiency of these estimates versus those obtained under detrending. Johansen (1988b) sets up a Gaussian vector autoregression (VAR) with possible drift estimated by maximum likelihood without detrending and finds that likelihood ratio and Wald test statistics are asymptotically chi-squared. Sims, Stock, and Watson (1990) allow for trends in a VAR framework and achieve similar results.

This paper examines regression estimation methods, focusing on the semi-parametric corrections proposed in Phillips and Hansen (1990) and the residual-based tests for cointegration proposed in Engle and Granger (1987) and Phillips and Ouliaris (1990). The central result is that *detrending has adverse effects upon the precision of estimation*, while the asymptotic distributions of the Phillips–Hansen *fully modified test statistics* are chi-square regardless of detrending procedures. The maintained assumption is that the cointegrating error is stationary, as might be expected from a theory of long-run equilibrium. General trend processes are allowed for the regressors. A second result uncovered is that *cointegration tests* are *not* invariant to the actual trends in the data, if the data is not detrended. The distributions under the null and alternative hypotheses are sensitive to the true trend processes and detrending procedures.

Section 2 introduces the model. Section 3 outlines a theory of least squares estimation. Section 4 develops an asymptotic theory of fully modified estimation. Section 5 develops a theory of inference for tests on the cointegrating vector and for the presence of a trend in the cointegrating relationship. Section 6 examines residual-based tests of the hypothesis of no cointegration. Some extensions are discussed in section 7. Proofs of the theorems are left to the appendix.

The notation follows that of Phillips and Hansen (1990). The symbol $\Rightarrow$ denotes weak convergence, $\equiv$ denotes equality in distribution, and $[\cdot]$ denotes 'integer part'. Brownian motion $B(r)$ on $[0,1]$ is frequently written as B to achieve notational economy. Similarly, the integral $\int_0^1 B(s)\,\mathrm{d}s$ is written

more simply as $\int_0^1 B$. Vector Brownian motion with covariance matrix Ω is written $BM(\Omega)$. We use $\|A\|$ to represent the Euclidean norm $\operatorname{tr}(A'A)^{1/2}$ of the matrix A. Finally, all limits given in the paper are as the sample size $T \to \infty$ unless otherwise stated.

2. Model

We shall be working with an $(n + 1)$-dimensional time series $\{y_t\}_0^\infty$ partitioned as

$$y_t' = \left(\underset{1}{y_{1t}}, \underset{n}{y_{2t}'}\right) \tag{1}$$

and generated by the system

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

$$y_{2t} = \Pi k_t + S_{2t}, \tag{3}$$

$$\Delta S_{2t} = u_{2t}. \tag{4}$$

The initialization of this system is at $t = 0$ and y_0 may be any random variable. Our first assumption characterizes the innovation vector $u_t = (u_{1t}, u_{2t}')'$.

> The random sequence $\{u_t\}$ is mean zero and α-mixing
> with mixing coefficients α_m such that $\sum_{m=1}^\infty \alpha_m^{(\nu-2)/4\nu} < \infty$
> and $\sup_t E|u_t' u_t|^\nu < \infty$ for some $\nu > 2$. In addition,
> $\Omega = \lim_{T \to \infty}(1/T)E(S_T S_T') > 0.$ (U)

Condition (U) permits quite general forms of temporal dependence, heterogeneity, and endogeneity. Note that $\Omega > 0$ requires that S_{2t} is $I(1)$, but excludes cointegration among the elements of S_{2t} and 'multicointegration' [see Granger and Lee (1990)]. (U) implies that the partial sum process $S_t = \sum_1^t u_j$ satisfies the multivariate invariance principle

$$T^{-1/2}S_{[Tr]} \Rightarrow B(r) \equiv BM(\Omega), \qquad 0 \le r \le 1. \tag{5}$$

The univariate result was shown by Herrndorf (1984) and was extended to

the multivariate case by Phillips and Durlauf (1986). (U) also implies convergence to the matrix stochastic integral

$$T^{-1} \sum_1^T S_t u_t' \Rightarrow \int_0^1 B \, \mathrm{d}B' + \Lambda, \qquad \Lambda = \sum_0^\infty \mathrm{E}(u_0 u_k'),$$

as shown by Hansen (1990, theorem 4.1). We partition B, Λ, and Ω conformably with u_t:

$$B(r) = \begin{pmatrix} B_1(r) \\ B_2(r) \end{pmatrix} \begin{matrix} 1 \\ n \end{matrix}, \qquad \Omega = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \Omega_{22} \end{pmatrix} \begin{matrix} 1 \\ n \end{matrix}. \tag{6}$$

The restriction to univariate models (y_{1t} is scalar) is unimportant. The generalization to arbitrary dimensions is straightforward and omitted. Also, an intercept is excluded from (2) only in order to ease exposition of the main argument. It should be emphasized that our results in no way depend upon this abstraction.

The m elements of k_t in (3) are positive integer powers of time. That is,

$$k_t = (t^{p_1} \ t^{p_2} \ \cdots \ t^{p_m})', \qquad 1 \le p_1 < \cdots < p_m. \tag{K}$$

In most applications k_t will be either a time trend ($p = 1$) or a quadratic ($p_1 = 1$, $p_2 = 2$). For expositional ease we will refer to k_t as the 'trend', and variables which are projected orthogonal to k_t as being 'detrended'. This is not meant to restrict attention to linear trends, and the extra generality which is allowed will remain implicit. Set

$$\delta_T = \mathrm{diag}\{T^{p_1}, T^{p_2}, \ldots, T^{p_m}\},$$

so that

$$\delta_T^{-1} k_{[Tr]} \to k(r) = (r^{p_1} \ r^{p_2} \ \cdots \ r^{p_m})',$$

uniformly over $r \in [0, 1]$. Indeed for $k_t = t^p$,

$$\sup_r \left| \delta_T^{-1} k_{[Tr]} - k(r) \right| = \sup_r \left| \left(\frac{[Tr]}{T} \right)^p - r^p \right| = \left| 1 - \left(1 - \frac{1}{T} \right)^p \right| \to 0.$$

Note as well that under (K), $\int_0^1 kk' > 0$. For example, if $k_t = (t\ t^2)'$, then

$$\int_0^1 kk' = \begin{pmatrix} \frac{1}{3} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{5} \end{pmatrix}.$$

In Phillips and Hansen (1990) and Phillips and Ouliaris (1990), the distributional theory was derived under the assumption that $\Pi = 0$. In this paper we take the polar case in which Π is full rank,

$$\text{rank}(\Pi) = m \le n. \tag{R}$$

3. Least squares estimation

Although (2) does not include k_t as a regressor, Park and Phillips (1988, 1989) and Phillips and Hansen (1990) primarily consider estimation of the unrestricted model

$$y_{1t} = \alpha' y_{2t} + \beta' k_t + u_{1t} = \gamma' x_t + u_{1t}, \tag{7}$$

where $\gamma = (\alpha', \beta')'$, $x_t = (y_{2t}', k_t')'$. Denote the least squares estimate of γ in (7) by $\hat{\gamma}$ and the estimated residuals by $\hat{u}_{1t}$. Even if we maintain $\beta = 0$, unrestricted estimation of (7) is advantageous as the estimate $\hat{\alpha}$ is invariant to the presence of drift in the regressors, i.e., to the values of Π. The covariate k_t in (7) detrends y_{2t}, leaving the stochastic trend S_{2t}. This device is an important component of the asymptotic theory developed in earlier work.

The alternative is to estimate (2) by least squares; equivalently, estimate (7) subject to the constraint

$$\beta = 0. \tag{8}$$

Denote the restricted estimate of α by $\tilde{\alpha}$ and the residuals from the restricted estimates by $\tilde{u}_{1t}$. Throughout this paper, we will use hats ($\hat{\ }$) and tildes ($\tilde{\ }$) on γ and α to refer to the unrestricted and restricted regressions, respectively. In the restricted regression, y_{2t} is not detrended by k_t, so its behavior is driven in part by the asymptotic behavior of $k(r)$.

In order to develop an asymptotic theory, we have to find an appropriate weighting matrix which allows the reweighted regressors to converge weakly to a full-ranked process. Phillips and Hansen (1990) have shown [remark (e)

following their theorem 3.2] that if we set

$$
\hat{W}_T = \begin{pmatrix} \Omega_{22}^{1/2} T^{-1/2} & -\Omega_{22}^{1/2} \Pi T^{-1/2} \\ 0 & \delta_T^{-1} \end{pmatrix}^{-1} = \begin{pmatrix} \Omega_{22}^{-1/2} \sqrt{T} & \Pi \delta_T \\ 0 & \delta_T \end{pmatrix}, \quad (9)
$$

then

$$
\hat{W}_T^{-1} x_{[Tr]} \Rightarrow \begin{pmatrix} W_n(r) \\ k(r) \end{pmatrix} \equiv J_U(r), \tag{10}
$$

where $W_n(r) = \Omega_{22}^{-1/2} B_2(r) \equiv BM(I_n)$. J_U is a full-ranked process, in the sense that $\int_0^1 J_U J_U' > 0$ a.s. [see Phillips and Hansen (1990, lemma A2)].

A similar result is needed for y_{2t} alone. This is complicated by the fact that if we define π_m as the right-most column of Π, then

$$
T^{-(1+2p_m)} \sum_1^T y_{2t} y_{2t}' \to_p \pi_m \pi_m',
$$

and this limit $n \times n$ moment matrix has rank one. [Note that $\pi_m \neq 0$ under (R).]

If $n > m$, we need to develop a sequence of weights which yield a nondegenerate design matrix in the limit. Construct an $n \times (n - m)$ matrix Π^* which spans the null space of Π and

$$
C = [C_1, C_2] = \left(\Pi^* (\Pi^{*\prime} \Omega_{22} \Pi^*)^{-1/2}, \Pi (\Pi' \Pi)^{-1} \right). \tag{11}
$$

Under (U), $\Pi^{*\prime} \Omega_{22} \Pi^* > 0$ so C is well-defined. Note that

$$
C' y_{2t} = \begin{pmatrix} C_1' S_{2t} \\ C_2' S_{2t} + k_t \end{pmatrix}.
$$

Defining the weight matrix $\bar{W}_T = \mathrm{diag}\{I_{n-m}\sqrt{T}, \delta_T\}$, we can see

$$
\bar{W}_T^{-1} C' y_{2[Tr]} = \begin{pmatrix} T^{-1/2} C_1' S_{2[Tr]} \\ \delta_T^{-1} C_2' S_{2[Tr]} + \delta_T^{-1} k_{[Tr]} \end{pmatrix} \Rightarrow \begin{pmatrix} W_{n-m}(r) \\ k(r) \end{pmatrix} \equiv J_R(r),
$$

where

$$W_{n-m}(r) = C_2' B_2(r) = (\Pi^{*'} \Omega_{22} \Pi^*)^{-1/2} \Pi^{*'} B_2(r) \equiv BM(I_{n-m}).$$

$$(12)$$

It is clear that the limit process J_R is full-ranked; that is, $\int_0^1 J_R J_R' > 0$ a.s. This allows the derivation of a nondegenerate asymptotic theory.

Theorem 1.   (a) *If* (U) *and* (K) *hold, then*

$$T^{1/2} \hat{W}_T' (\hat{\gamma} - \gamma) \Rightarrow \left(\int_0^1 J_U J_U' \right)^{-1} \left[\int_0^1 J_U \, dB_1 + \begin{pmatrix} \lambda_{21} \\ 0 \end{pmatrix} \right].$$

(b) *If in addition* (R) *holds, then*

$$T^{1/2} \tilde{W}_T C^{-1} (\tilde{\alpha} - \alpha) \Rightarrow \left(\int_0^1 J_R J_R' \right)^{-1} \left[\int_0^1 J_R \, dB_1 + \begin{pmatrix} C_1' \lambda_{21} \\ 0 \end{pmatrix} \right],$$

where $\lambda_{21} = \sum_{j=0}^{\infty} E(u_{20} u_{1j})$ *and* J_U *and* J_R *are given in* (10) *and* (12).

The distributions in Theorem 1 are plagued by a number of nuisance parameters; particularly, λ_{21} and ω_{21} which describe the nature of the serial and long-run dependence of u_{2t} and u_{1t}. This implies that unadjusted least squares does not lead to a parameter-invariant theory of inference. Two alternatives are available: inclusion of stationary covariates to orthogonalize the error term or the semi-parametric correction of Phillips and Hansen (1990), which they call *fully modified estimation*. Both procedures can yield mixture normal asymptotics. In the next section we will consider the Phillips–Hansen approach, but the main results will apply to both procedures. See Phillips (1988, 1991) or Johansen (1988a, b) for discussions of alternative parametric approaches.

4. Fully modified estimation

The fully modified estimates proposed in Phillips and Hansen (1990) make use of first-stage estimates of the long-run covariance matrix Ω. While any consistent estimate of Ω will produce the same asymptotic distributions, we discuss a specific class of kernel estimates for concreteness. Set $\hat{u}_t = (\hat{u}_{1t}, \hat{u}_{2t}')'$, where $\hat{u}_{1t}$ is the least squares residual as defined before and $\hat{u}_{2t}$ is the residual from a regression of Δy_{2t} upon Δk_t. This requires the knowledge of the drifts Δk_t which drive y_{2t}. Frequently k_t is simply a time trend, in which

case Δk_t is a constant and $\hat{u}_{2t} = u_{2t} - \bar{u}_2$. If k_t is a quadratic in time, then $\Delta k'_t = (1, t)$ and $\hat{u}_{2t}$ is obtained by detrending Δy_{2t}.

The class of positive semidefinite kernel estimates of Ω we consider is given by

$$\hat{\Omega} = \sum_{j=-T}^{T} w(j/M)T^{-1}\sum_t \hat{u}_{t-j}\hat{u}'_t,$$

where the kernel weights $w(\cdot)$ satisfy

For all $x \in \mathbb{R}$, $|w(x)| \leq 1$ and $w(x) = w(-x)$; $w(0) = 1$;

$w(x)$ is continuous at zero and for almost all $x \in \mathbb{R}$;

$\int_{\mathbb{R}}|w(x)|\,\mathrm{d}x < \infty$; for all $\lambda \in \mathbb{R}$, $\int_{-\infty}^{\infty} w(x)\mathrm{e}^{-ix\lambda} \geq 0$.   (W)

This class of estimates has its origin in the literature on spectral density estimation. (When u_t is weakly stationary, Ω is proportional to its spectral density matrix at frequency zero.)

Kernels which satisfy (W) include

Bartlett:

$$w(x) = \begin{cases} 1 - |x| & \text{for} \quad |x| \leq 1, \\ 0 & \text{otherwise}; \end{cases}$$

Parzen:

$$w(x) = \begin{cases} 1 - 6x^2 + 6|x|^3 & \text{for} \quad 0 \leq |x| \leq \frac{1}{2}, \\ 2(1 - |x|)^3 & \text{for} \quad \frac{1}{2} \leq |x| \leq 1, \\ 0 & \text{otherwise}; \end{cases}$$

Quadratic Spectral:

$$w(x) = \frac{25}{12\pi^2 x^2}\left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\right).$$

In a recent study of kernel estimates of covariance matrices, Andrews (1991) shows that the quadratic spectral kernel has the best performance with respect to asymptotic truncated mean square error (MSE). In applications, however, it is frequently found that the choice of kernel is of secondary importance to the choice of the bandwidth parameter M, which is sometimes

called the lag truncation number. We assume:

$$M \to \infty \quad \text{as} \quad T \to \infty \quad \text{such that} \quad M/T^{1/2} \to 0. \tag{M}$$

Assumptions such as (M) which are used in consistency proofs are not particularly helpful for selecting M in applications. It should be noted that conditions such as $M = o(T^{1/4})$ or $M = o(T^{1/2})$ are frequently imposed as technical artifacts of the proof method, and should not be interpreted as guides for selection of M in practice. The choice of bandwidth parameter may, however, have important influences upon the estimates and test statistics; hence appropriate care should be exercised in their selection. One reasonable approach, dating at least back to Parzen (1957), is to choose M to minimize the MSE of $\hat{\Omega}$. A recent exposition of this approach is provided by Andrews (1991). Optimal selection rules are given and automatic bandwidth estimators proposed. Applied researchers who wish to adopt an automatic (or plug-in) estimator of this form are recommended to consult Andrews' paper.

We also define $\tilde{\Omega}$ analogously to $\hat{\Omega}$ using the residuals $\bar{u}_{1t}$ in place of $\hat{u}_{1t}$. Partition $\hat{\Omega}$ and $\tilde{\Omega}$ conformably with Ω.

Theorem 2. (a) *If* (U), (K), (W), *and* (M) *hold, then* $\hat{\Omega} \to_p \Omega$. (b) *If, in addition,* (R) *holds, then* $\tilde{\Omega} \to_p \Omega$.

Now define the corrected residual,

$$u_t^+ = u_{1t} - \omega_{12}\Omega_{22}^{-1}u_{2t},$$

which has zero coherence at the origin with u_{2t}. Define the nuisance parameter,

$$\lambda^+ = \sum_{j=0}^{\infty} \mathrm{E}(u_{20}u_j^+),$$

and an estimate of λ^+,

$$\hat{\lambda}^+ = \sum_{j=0}^{T} w(j/M)T^{-1}\sum_t \hat{u}_{2t-j}\hat{u}_t^+,$$

where

$$\hat{u}_t^+ = \hat{u}_{1t} - \hat{\omega}_{12}\hat{\Omega}_{22}^{-1}\hat{u}_{2t}.$$

Similarly, define $\tilde{\lambda}^+$ using $\bar{u}_{1t}$ in place of $\hat{u}_{1t}$. A similar argument to the

proof of Theorem 2 yields

$$\hat{\lambda}^+ = \sum_{j=0}^{T} w(j/m)T^{-1}\sum_t \hat{u}_{2t-j}\hat{u}_{1t} + \sum_{j=0}^{T} w(j/M)T^{-1}\sum_t \hat{u}_{2t-j}\hat{u}'_{2t}\hat{\Omega}_{22}^{-1}\hat{\omega}_{21}$$

$$\to_p \sum_{j=0}^{\infty} \mathrm{E}(u_{20}u_{1j}) + \sum_{j=0}^{\infty} \mathrm{E}(u_{20}u'_{2j})\Omega_{22}^{-1}\omega_{21} = \lambda^+.$$

The fully modified estimate for the unrestricted model is

$$\hat{\gamma}^+ = \begin{pmatrix} \hat{\alpha}^+ \\ \hat{\beta}^+ \end{pmatrix} = \left(\sum_1^T x_t x'_t\right)^{-1}\left[\sum_1^T x_t y_t^+ - T\begin{pmatrix} \hat{\lambda}^+ \\ 0 \end{pmatrix}\right], \tag{13}$$

and the fully modified estimate for the restricted model is

$$\tilde{\alpha}^+ = \left(\sum_1^T y_{2t} y'_{2t}\right)^{-1}\left[\sum_1^T y_{2t} y_t^+ - T\tilde{\lambda}^+\right]. \tag{14}$$

In (13), $y_t^+ = y_{1t} - \hat{\omega}_{12}\hat{\Omega}_{22}^{-1}\hat{u}_{2t}$. In (14), $y_t^+ = y_{1t} - \tilde{\omega}_{12}\tilde{\Omega}_{22}^{-1}\tilde{u}_{2t}$.

Theorem 3. (a) *If* (U), (K), (W), *and* (M) *hold, then*

$$T^{1/2}\hat{W}'_T(\hat{\gamma}^+ - \gamma) \Rightarrow \left(\int_0^1 J_U J'_U\right)^{-1}\int_0^1 J_U \,\mathrm{d}B_{1\cdot 2}.$$

(b) *If, in addition,* (R) *holds, then*

$$T^{1/2}\tilde{W}_T C^{-1}(\tilde{\alpha}^+ - \alpha) \Rightarrow \left(\int_0^1 J_R J'_R\right)^{-1}\int_0^1 J_R \,\mathrm{d}B_{1\cdot 2}.$$

In (a) and (b), $B_{1\cdot 2}(r) = B_1(r) - \omega_{12}\Omega_{22}^{-1}W_2(r) \equiv BM(\omega_{1\cdot 2})$, $\omega_{1\cdot 2} = \omega_{11} - \omega_{12}\Omega_{22}^{-1}\omega_{21}$, *and* $B_{1\cdot 2}$ *is independent of* J_U *and* J_R.

The distributions in Theorem 3 are full ranked, median unbiased, and mixture normal. Both $\tilde{\alpha}^+$ and $\hat{\alpha}^+$ are consistent and their limiting distributions are free of nuisance dependencies. As is well known, restricted coefficient estimation is more efficient than unrestricted estimation. Thus, if economic theory suggests that $\{y_{1t} - y'_{2t}\alpha\}$ should have an equilibrium relationship (the discrepancy should not contain a trend), then $\beta = 0$ and this restriction should be imposed in the estimation process. Estimation of the unrestricted model in earlier work has been motivated primarily to render a

tractable asymptotic theory of inference, and not because the restrictions were believed to not hold. An asymptotic theory of inference is developed in the next section which fills this gap.

5. Hypothesis testing

Hypothesis tests of interest will take two forms: (i) restrictions on the cointegrating vector α and (ii) restrictions on β in (7) (most commonly, $\beta = 0$). The latter test will be of primary importance as a verification that α can be estimated without inclusion of a trend term in the regression. We show that hypothesis tests constructed using the fully modified estimators have asymptotic chi-squared distributions.

5.1. Tests on the cointegrating vector

We consider test of linear hypotheses

$$H_0: \quad Q'\alpha = c, \qquad \text{rank}(Q) = q. \tag{15}$$

Define the Wald statistic constructed from $\bar{\alpha} = \hat{\alpha}$ or $\tilde{\alpha}$:

$$G_Q(\bar{\alpha}, \bar{V}) = [Q'\bar{\alpha} - c]'[Q'\bar{V}Q]^{-1}[Q'\bar{\alpha} - c], \tag{16}$$

and covariance matrix estimators

$$\hat{V}_T = \hat{\omega}_{1 \cdot 2}(I_n \ 0)\left(\sum_1^T x_t x_t'\right)^{-1}\begin{pmatrix} I_n \\ 0 \end{pmatrix},$$

$$\tilde{V}_T = \tilde{\omega}_{1 \cdot 2}\left(\sum_1^T y_{2t} y_{2t}'\right)^{-1},$$

where

$$\hat{\omega}_{1 \cdot 2} = \hat{\omega}_{11} - \hat{\omega}_{12}\hat{\Omega}_{22}^{-1}\hat{\omega}_{21},$$

and $\tilde{\omega}_{1 \cdot 2}$ is defined similarly to $\hat{\omega}_{1 \cdot 2}$.

Analysis of the distribution of Wald statistics for hypotheses on α using the unrestricted estimates $\hat{a}^+$ is fairly straightforward and is carried out in Phillips and Hansen (1990).

Theorem 4.   If (15), (U), (K), (W), and (M) hold, then

$$G_Q\left(\hat{\alpha}^+, \hat{V}_T\right) \to_d \chi_q^2.$$

The derivation of Theorem 4 is facilitated by the fact that $T(\hat{\alpha}^+ - \alpha)$ has a nondegenerate limiting distribution, so all linear combinations of the coefficient estimates converge at the same rate. In contrast, differing linear combinations of the restricted estimates, $\tilde{\alpha}^+$, converge at differing rates. This problem has not been studied in a general setting in the previous literature. The solution is to note that linear combinations of estimates which converge at different rates only converge as fast as the slowest elements. Careful handling of rank conditions enables the derivation of an asymptotic theory.

Our derivations will be facilitated by the following lemma:

Lemma.   If Q is $n \times q$, $rank(Q) = q$, and for any partition $\{n_1, n_2, \ldots, n_k:$ $\sum_1^k n_j = n\}$ there exists an $n \times q$ matrix B and an $n \times n$ matrix D such that

$$QB' = D, \qquad rank(D) = q,$$

where D is block lower diagonal, in that

$$D = \begin{pmatrix} D_{11} & 0 & 0 & \cdots & \\ D_{21} & D_{22} & 0 & & \\ \hdotsfor{5} \\ D_{k1} & D_{k2} & & \cdots & D_{kk} \end{pmatrix}$$

and the $\{D_{ii}: i = 1, \ldots, k\}$ are diagonal matrices of ones and zeros.

Note that under the null,

$$Q'\tilde{\alpha}^+ - c = Q'(\tilde{\alpha}^+ - \alpha) = \overline{Q}'C^{-1}(\tilde{\alpha}^+ - \alpha),$$

where $\overline{Q} = C'Q$ is of rank q and C is taken from (11). By the lemma, there exists an $n \times q$ full rank matrix B and a lower triangular matrix D such that $B\overline{Q}' = D$ and D has only ones and zeros on the diagonal. Set $\overline{D}$ as the diagonal matrix with the diagonal elements of D, and S as the $n \times q$ matrix consisting of the nonzero columns of $\overline{D}$ (which implies that $\overline{D}S = S$). Finally,

set $D_T = S'\tilde{W}_T B$. We then have

$$\sqrt{T}\,D_T(Q'\tilde{\alpha}^+ - c) = \sqrt{T}\,S'\tilde{W}_T B\bar{Q}'C^{-1}(\tilde{\alpha}^+ - \alpha)$$

$$= \sqrt{T}\,S'\tilde{W}_T D'C^{-1}(\tilde{\alpha}^+ - \alpha)$$

$$= \sqrt{T}\,S'\bar{D}\tilde{W}_T C^{-1}(\tilde{\alpha}^+ - \alpha) + o_p(1)$$

$$= \sqrt{T}\,S'\tilde{W}_T C^{-1}(\tilde{\alpha}^+ - \alpha) + o_p(1)$$

$$\Rightarrow S'\left(\int_0^1 J_R J_R'\right)^{-1}\int_0^1 J_R \, dB_{1\cdot 2}. \tag{17}$$

The above construction has shown that with appropriate reweighting, $Q'(\tilde{\alpha}^+ - \alpha)$ has a full ranked, mixture normal asymptotic distribution. It immediately follows that Wald tests of linear hypotheses have limiting chi-square distributions. Its implications are given in the following useful theorem:

Theorem 5.   If (U), (K), (R), (W), and (M) hold, then

(a) $\sqrt{T}\,D_T(Q'\tilde{\alpha}^+ - c) \Rightarrow S'\left(\int_0^1 J_R J_R'\right)^{-1}\int_0^1 J_R \, dB_{1\cdot 2},$

(b) $G_Q(\tilde{\alpha}^+, \bar{V}_T) \to_d \chi_q^2.$

Theorem 5 shows that the differing rates of convergence of the elements of $\tilde{\alpha}^+$ do not invalidate the use of the Wald test for standard linear hypotheses. Extension of this result to nonlinear hypotheses, however, does not directly follow from the above analysis. This useful extension is left to future research.

5.2. Tests on the regression trend

The efficiency gains can be realized only if $\beta = 0$ in eq. (7). This suggests that we may wish to test the hypothesis

$$H_0: \quad \beta = 0, \tag{18}$$

using $\hat{\beta}^+$ from (13). Phillips and Hansen (1990) have derived the limiting distribution of this test statistic under the assumption $\Pi = 0$. We consider

here the generic case $\Pi \neq 0$. (18) may be written as

$$H_0: \quad R'\gamma = 0, \qquad R = \begin{pmatrix} 0 \\ I_m \end{pmatrix}. \tag{19}$$

The Wald statistic is

$$G_R(\hat{\beta}^+) = \hat{\beta}^{+\prime} \left(R' \left(\sum_1^T x_t x_t' \right)^{-1} R \hat{\omega}_{1 \cdot 2} \right)^{-1} \hat{\beta}^+. \tag{20}$$

Theorem 6.   If (19), (U), (K), (W), and (M) hold, then

$$G_R(\hat{\beta}^+) \to_d \chi_m^2.$$

Theorem 6 allows for general drift in the regressors, yet yields an asymptotic chi-square test. This enables applied researchers to start with the unrestricted specification, test for $\beta = 0$, and then estimate the restricted equation if it is compatible with the data. The standard caveat about sequential hypothesis tests should be noted. In a context of low power, mere lack of rejection of an hypothesis does not mean that it is in fact true. Implementation of such tests should be handled with appropriate caution.

6. Testing for cointegration

An important application of our theory is to tests of the hypothesis of no cointegration. Engle and Granger (1987) suggested that the residuals from OLS estimation of the cointegrating regression be examined for the presence of a unit root in the autoregressive representation. They suggested several such tests, the most popular probably being the augmented Dickey–Fuller (ADF) test, originally suggested by Said and Dickey (1984) in the context of univariate unit root testing. A first attempt at a distributional theory was provided by Engle and Yoo (1987) and recently extended and generalized by Phillips and Ouliaris (1990).

Phillips and Ouliaris examined both the ADF test and the Phillips' $Z(\alpha)$ and $Z(t)$ tests suggested by Phillips (1987) in the context of univariate unit root testing. They derived the asymptotic distributions of the test statistics under the assumption that no deterministic trends are present in the regressors or the regression equation. The tabulated distributions they present imply that they understood their results to generalize to the presence of trends. It is not clear exactly how they meant this generalization to be

allowed [i.e., allowing for trends in the regressors only if trends are included in the levels regression as in Park and Phillips (1988) and Phillips and Hansen (1990)]. It turns out that the actual limiting distributions are quite sensitive to both the actual trends in the regressors as well as the detrending procedures, as we now show. Explicitly examined are the $Z(\alpha)$ and $Z(t)$ tests. The results will apply as well to the ADF test.

Since the null hypothesis is no cointegration, (2) is invalid. We instead make the alternative assumption that y_{1t} is generated by

$$y_{1t} = \alpha' y_{2t} + v_t, \tag{21}$$

$$v_t = \rho v_{t-1} + u_{1t}, \qquad \rho = 1, \tag{22}$$

and y_{2t} is generated as before.

We consider two procedures. The first is the unrestricted OLS regression

$$y_{1t} = \hat{\gamma}' x_t + \hat{v}_t,$$

where $x_t = (y_{2t}', k_t)$. The second is the restricted OLS regression

$$y_{1t} = \tilde{\alpha}' y_{2t} + \tilde{v}_t.$$

The Phillips' tests for the unrestricted regression are constructed from the OLS residuals $\hat{v}_t$ as follows. Calculate the first-order serial correlation coefficient and residuals

$$\hat{\rho} = \frac{\sum_t \hat{v}_t \hat{v}_{t+1}}{\sum_t \hat{v}_t^2}, \qquad \hat{\varepsilon}_t = \hat{v}_t - \hat{\rho} \hat{v}_{t-1},$$

and the bias correction

$$\hat{\lambda} = \sum_{j=1}^{T} w(j/M) T^{-1} \sum_t \hat{\varepsilon}_{t-j} \hat{\varepsilon}_t,$$

where the kernel weights $w(\cdot)$ satisfy (W). As discussed earlier, finite sample performance of the test may depend critically upon selection of kernel and the bandwidth parameter M. See section 4 for discussion and references.

The 'bias-corrected' serial correlation coefficient estimate is given by

$$\hat{\rho}^* = \frac{\sum\limits_t \left(\hat{v}_t \hat{v}_{t+1} - \hat{\lambda} \right)}{\sum\limits_t \hat{v}_t^2},$$

yielding the test statistics:

$$Z(\hat{\rho}^*) = T(\hat{\rho}^* - 1),$$

$$Z(\hat{t}^*) = \frac{\hat{\rho}^* - 1}{\left(\sum\limits_t \hat{v}_t^2 \right)^{-1/2} \hat{\omega}_{11}^{1/2}}, \qquad \hat{\omega}_{11} = \hat{\sigma}_\varepsilon^2 + 2\hat{\lambda}, \quad \hat{\sigma}_\varepsilon^2 = \frac{1}{T} \sum\limits_1^T \hat{\varepsilon}_t^2.$$

The statistics for the restricted regression are constructed from $\tilde{v}_t$ in the same way.

Theorem 7.   If (21), (22), (U), (K), (W), and (M) hold, then

(a) $Z(\hat{\rho}^*) \Rightarrow \int^1 Q(W_1, J_U) \, \mathrm{d}Q(W_1, J_U),$

(b) $Z(\hat{t}^*) \Rightarrow \int_0^1 Q(W_1, J_U) \, \mathrm{d}S(W_1, J_U),$

where

$$Q(W, X) = W_X \bigg/ \left(\int_0^1 W_X^2 \right)^{1/2},$$

$$S(X) = W_X \big/ (K_X' K_X)^{1/2},$$

$$W_X(r) = W(r) - \int_0^1 WX' \left(\int_0^1 XX' \right)^{-1} X(r),$$

$$K_X = \left(1, -\int_0^1 W_1 X' \left(\int_0^1 XX' \right)^{-1} \binom{I_d}{0} \right), \qquad d = \dim(X),$$

and $W_1 \equiv BM(1)$ is independent of J_U and J_R, which are defined in (10) and

(12), respectively. If, in addition, (R) holds, then

(c) $$Z(\tilde{\rho}^*) \Rightarrow \int_0^1 Q(W_1, J_R) \, \mathrm{d}Q(W_1, J_R),$$

(d) $$Z(\tilde{t}^*) \Rightarrow \int_0^1 Q(W_1, J_R) \, \mathrm{d}S(W_1, J_R).$$

Theorem 7 is important because it shows that the deterministic trends in the data affect the limiting distribution of the test statistics *whether or not we detrend the data*. The distributions of the test statistics from the unrestricted regression depend upon $J_U = (W_n' \quad k')$, while the distributions of the test statistics from the restricted regression depend upon $J_R = (W_{n-m}' \quad k')'$. The detrending procedure implicit in the unrestricted regression affects the limiting distribution by increasing the number of stochastic trends (from $n - m$ to n) entering the limiting representation. This has a tendency to shift the limiting distribution away from the origin [see the tables in Phillips and Ouliaris (1990)], increasing the absolute value of the critical value needed for rejection, and thus requiring a smaller estimated serial correlation coefficient to reject the hypothesis of a unit root. It seems reasonable that excess detrending will reduce the test's power. Careful consideration of the nature of the trends in the regressors must be given if tests of the cointegration hypothesis are to have the correct size and optimal power.

Theorem 7 states that the asymptotic distribution of the test statistics depends upon the number and nature of the nonstationary variables in the regression. Consider the common case $k_t = t$, which implies $k(r) = r$. The asymptotic distributions of the test statistics from the unrestricted regression depend upon $J_U = (W_n' \quad r)'$, and thus upon the detrending procedure and the number of regressors (n). In this case, tables Ic and IIc in Phillips and Ouliaris (1990) are appropriate. On the other hand, the asymptotic distributions of the statistics from the restricted regression depend upon the rank of Π. If $\Pi = 0$, then Phillips and Ouliaris have shown that the distribution depends upon W_n, and their tables Ib and IIb provide the appropriate critical values. If $\Pi \neq 0$, Theorem 7 above shows that the distribution of the test statistics depends upon $J_R = (W_{n-1}' \quad r')'$, and thus the appropriate critical values are found from tables Ic and IIc in Phillips and Ouliaris (1990), but using $n - 1$ instead of n. If $n = 1$, then the appropriate critical values are found in the third section of tables 8.5.1 and 8.5.2 of Fuller (1976). I have reproduced in tables 1 and 2 the asymptotic critical values for the $Z(\tilde{\rho}^*)$ and $Z(\tilde{t}^*)$ statistics from the restricted regression. The absolute value of the critical values are reported for one-tailed tests of size 10%, 5%, and 1%, for the cases of $n = 1$, 2, and 3. The data is assumed to be demeaned. To obtain

Table 1

$Z(\tilde{\rho}^*)$ critical values, $k_t = t$, restricted regression.[a]

		10%	5%	1%
$n = 1$:	$\Pi \neq 0$	18.3	21.8	29.5
	$\Pi = 0$	17.0	20.5	28.3
$n = 2$:	$\Pi \neq 0$	23.2	27.1	35.4
	$\Pi = 0$	22.2	26.1	34.2
$n = 3$:	$\Pi \neq 0$	27.8	32.2	40.3
	$\Pi = 0$	27.6	32.1	41.1

[a]*Source:* Fuller (1976) and Phillips and Ouliaris (1990).

Table 2

$Z(\tilde{t}^*)$ and ADF critical values, $k_t = t$, restricted regression.[a]

		10%	5%	1%
$n = 1$:	$\Pi \neq 0$	3.12	3.41	3.96
	$\Pi = 0$	3.07	3.37	3.96
$n = 2$:	$\Pi \neq 0$	3.52	3.80	4.36
	$\Pi = 0$	3.45	3.77	4.31
$n = 3$:	$\Pi \neq 0$	3.84	4.16	4.65
	$\Pi = 0$	3.83	4.11	4.73

[a]*Source:* Fuller (1976) and Phillips and Ouliaris (1990).

the critical values for the unrestricted regression, use Phillips and Ouliaris (1990), tables Ic and IIc, or tables 1 and 2 here with $\Pi \neq 0$ and $n + 1$.

The most striking fact about tables 1 and 2 is that the asymptotic critical values depend very little upon the nature of the regressors. Essentially, the critical values only depend upon the number of nonstationary regressors (counting the deterministic trends if included). This suggests a useful practical device.[1] Since it is advantageous to have a test procedure which does not depend upon nuisance parameters (such as knowledge about the true nature of trends in the data), it is reasonable to use the larger of the two critical values tabulated (if critical values were tabulated also incorporating a quadratic in time, then this could be considered in some cases as well). Thus, if two regressors are included as independent variables, then the practical critical value for the $Z(\tilde{\rho}^*)$ statistic would be 27.1, and for the $Z(\tilde{t}^*)$ statistic it would be 3.80.

As mentioned above, it seems reasonable that cointegration tests constructed using a smaller number of regressors would be more powerful

[1] I owe this suggestion to an anonymous referee.

against the alternative of cointegration. Specifically, exclusion of deterministic trends from equilibrium relationships will increase the likelihood of rejection of the null when the data is cointegrated.

This can be illustrated with a simple Monte Carlo exercise. The model is

$$y_{1t} = \alpha y_{2t} + u_t, \qquad u_t = \rho u_{t-1} + \varepsilon_{1t},$$

$$y_{2t} = y_{2t-1} + \pi + \varepsilon_{2t},$$

and $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})'$ are independent $N(0, I_2)$ draws. The test statistics are invariant to α, leaving one nuisance parameter, π, which is set at two values, zero and one. The null hypothesis of no cointegration holds when $\rho = 1$, and the alternative holds for $\rho < 1$.

We estimate the unrestricted model,

$$y_{1t} = \hat{\mu} + \hat{\alpha} y_{2t} + \hat{\pi}_2 t + \hat{u}_t,$$

and the restricted model,

$$y_{1t} = \tilde{\mu} + \tilde{\alpha}_{2t} + \tilde{u}_t.$$

The Phillips $Z(\hat{\rho}^*)$ test using a Bartlett kernel with four lags is mounted on the estimated residuals from the two regressions. (A simple Dickey–Fuller test would be appropriate in this model, but it seems more reasonable to make the correction for possible serial correlation as would be done in practice.) Sample size is set at 50, to emphasize small sample considerations, and 5,000 replications are used. To assess size distortion, tables 3 and 4 report the null rejection percentages using the asymptotic critical values (for size 10%, 5%, and 1%) under the parameter settings $\pi = 0$ and $\pi = 1$. The first line is the test based on the unrestricted regression, which uses the critical values reported in table 1, $n = 2$, $\Pi \neq 0$. The second and third lines are the tests based upon the restricted regression. There are two choices for critical values from table 1, $n = 1$: $\Pi \neq 0$ and $\Pi = 0$. The first is appropriate when $\pi = 1$, and the second is appropriate when $\pi = 0$. Tests based on both critical values are reported in the second and third lines of the tables.

The size distortion shown in tables 3 and 4 is fairly mild (especially noting that the sample size is only 50). The distortion is generally conservative (underrejection).

We have conjectured that cointegration tests based on the restricted regression are more powerful than cointegration tests based on the unrestricted regression. Figs. 1 and 2 show the size-adjusted Monte Carlo power functions of the two procedures (using the exact finite-sample critical values for tests of size 5%). Rejection frequencies were evaluated at ten values of

Table 3

Case A: $\pi = 0$, percentage rejections under the null ($\rho = 1$), standard errors in parentheses.

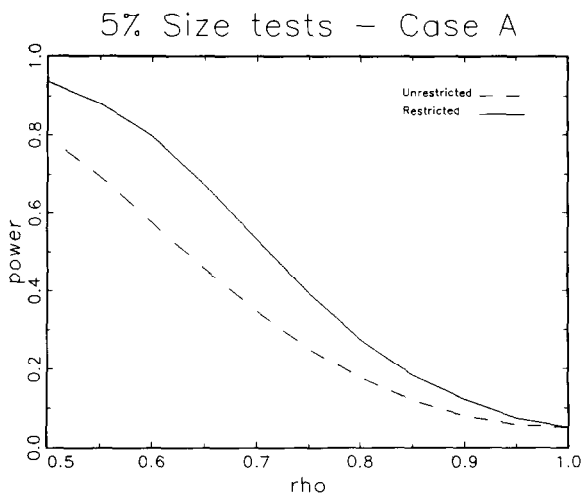|	10%	5%	1%
Unrestricted[a]	6.2%	2.3%	0.3%
	(0.3)	(0.2)	(0.1)
Restricted[b]	7.0%	2.3%	0.4%
	(0.4)	(0.2)	(0.1)
Restricted[c]	9.2%	3.6%	0.5%
	(0.4)	(0.2)	(0.1)

[a] $Z(\hat{\rho}^*)$ test using critical values from table 1, $n = 2$, $\Pi \neq 0$.
[b] $Z(\tilde{\rho}^*)$ test using critical values from table 1, $n = 1$, $\Pi \neq 0$.
[c] $Z(\tilde{\rho}^*)$ test using critical values from table 1, $n = 1$, $\Pi = 0$.

Table 4

Case B: $\pi = 1$, percentage rejections under the null ($\rho = 1$), standard errors in parentheses.

	10%	5%	1%
Unrestricted[a]	6.2%	2.2%	0.1%
	(0.3)	(0.2)	(0.1)
Restricted[b]	9.0%	3.6%	0.3%
	(0.4)	(0.2)	(0.1)
Restricted[c]	13.1%	5.3%	0.5%
	(0.5)	(0.3)	(0.1)

[a] $Z(\hat{\rho}^*)$ test using critical values from table 1, $n = 2$, $\Pi \neq 0$.
[b] $Z(\tilde{\rho}^*)$ test using critical values from table 1, $n = 1$, $\Pi \neq 0$.
[c] $Z(\tilde{\rho}^*)$ test using critical values from table 1, $n = 1$, $\Pi = 0$.
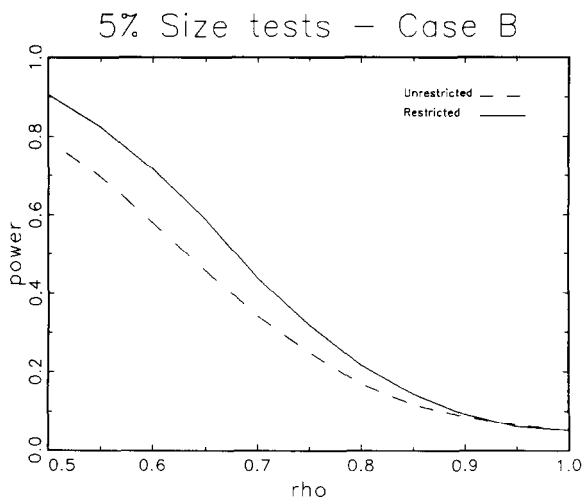

Fig. 1

Fig. 2

the alternative hypothesis, for $\rho = 0.5$ to $\rho = 0.95$ in steps of 0.05. The figures show as expected that the power function of the tests based on the restricted residuals are uniformly more powerful than the tests based on the unrestricted residuals.

7. Extensions

Without going through the details, some extensions of the results obtained in the previous sections are straightforward.

(a) The asymptotic distribution of full information maximum likelihood estimation in the presence of trends [see Phillips (1991) for the analysis without trends] will have the same form as in Theorems 3, 5, and 6.

(b) The analysis could be extended to include models with some trends maintained in the linear specification and some excluded, i.e.,

$$y_{1t} = \alpha' y_{2t} + \pi_1' k_{1t} + u_{1t},$$

$$y_{2t} = S_{2t} + \Pi_2' k_{2t}.$$

The asymptotic analysis of sections 3, 4, and 5 generalizes in the obvious way. That is, coefficient estimates are consistent and fully modified test statistics have asymptotic chi-square distributions. The analysis of section 6 also generalizes. For example, if $n = 2$, $k_{2t} = (t \ t^2)'$, and $k_{1t} = t$, the asymptotic distribution of the tests for cointegration based on the restricted regression

(y_{1t} on y_{2t} and k_{1t}) is given in Theorem 7 with $J_R(r)' = (W_1(r), r, r^2)$, $W_1 \equiv BM(1)$.

(c) As mentioned in section 6, Theorem 7 (b) and (d) could be extended along the lines of Phillips and Ouliaris (1990, theorem 4.2) to encompass the Augmented Dickey–Fuller test statistic.

Appendix

Proof of Theorem 1

To show (a),

$$\sqrt{T}\,\hat{W}_T'(\hat{\gamma} - \gamma) = \left(T^{-1}\hat{W}^{-1}\sum_1^T x_t x_t' \hat{W}_T^{-1\prime}\right)^{-1}\left(T^{-1/2}\hat{W}^{-1}\sum_1^T x_t u_{1t}\right)$$

$$\Rightarrow \left(\int_0^1 J_U J_U'\right)^{-1}\left(\int_0^1 J_U\, \mathrm{d}B_1 + \binom{\lambda_{21}}{0}\right),$$

by weak convergence of the moment matrices and the Continuous Mapping Theorem (CMT). Similarly, (b) follows by the CMT:

$$\sqrt{T}\,\tilde{W}_T C^{-1}(\tilde{\alpha} - \alpha)$$

$$= \left(T^{-1}\tilde{W}^{-1}C'\sum_1^T y_{2t} y_{2t}' C \tilde{W}_T^{-1}\right)^{-1}\left(T^{-1/2}\tilde{W}^{-1}C'\sum_1^T y_{2t} u_{1t}\right)$$

$$\Rightarrow \left(\int_0^1 J_R J_R'\right)^{-1}\left(\int_0^1 J_R\, \mathrm{d}B_1 + \binom{H_1'\lambda_{21}}{0}\right). \quad \square$$

Proof of Theorem 2

The proofs that the elements of $\hat{\Omega}$ and $\tilde{\Omega}$ s7are consistent are similar. We prove that $\hat{\omega}_{21} \to_p \omega_{21}$. First note that

$$\hat{u}_{2t} = u_{2t} - (\hat{\Pi} - \Pi)\, \Delta k_t,$$

where

$$\hat{\Pi} = \left(\sum_1^T \Delta k_t\, \Delta k_t'\right)^{-1}\sum_1^T \Delta k_t\, \Delta y_{2t}.$$

Under (K),

$$T\delta_T^{-1}\, \Delta k_{[Tr]} \to \mathrm{d}k(r) \quad \text{uniformly in } r, \tag{A.1}$$

where

$$\mathrm{d}k(r) = \mathrm{d}k(r)/\mathrm{d}r = (r^{p_1-1}, \dots, r^{p_m-1})'.$$

Thus,

$$T^{-1/2}\delta_T(\hat{\Pi} - \Pi)$$

$$= \left(\frac{1}{T}(T\delta_T^{-1}) \sum_1^T \Delta k_t\, \Delta k_t'(T\delta_T^{-1}) \right)^{-1} \frac{1}{\sqrt{T}}(T\delta_T^{-1}) \sum_1^T \Delta k_t u_{2t}$$

$$\Rightarrow \left(\int_0^1 (\mathrm{d}k)(\mathrm{d}k)' \right)^{-1} \int_0^1 (\mathrm{d}k)\, \mathrm{d}B_2. \tag{A.2}$$

Now note that

$$\hat{\omega}_{21} = \sum_j w(j/M)T^{-1} \sum_t \hat{u}_{2t+j}\hat{u}_{1t}$$

$$= \sum_j w(j/M)T^{-1} \sum_t u_{2t+j}u_{1t}$$

$$\quad - \sum_j w(j/M)T^{-1} \sum_t u_{2t+j}x_t'(\hat{\gamma} - \gamma)$$

$$\quad - \sum_j w(j/M)T^{-1} \sum_t u_{1t}\, \Delta k_{t+j}'(\hat{\Pi} - \Pi)$$

$$\quad + \sum_j w(j/M)(\hat{\Pi} - \Pi)'T^{-1} \sum_t \Delta k_{t+j}x_t'(\hat{\gamma} - \gamma)$$

$$= B_{1T} - B_{2T} - B_{3T} + B_{4T}, \quad \text{say}. \tag{A.3}$$

Condition (U) is stronger than the conditions of lemma 1 of Andrews (1991), which is sufficient for his proposition 1, which implies

$$\frac{\sqrt{T}}{M}(B_{1T} - \omega_{21}) \to_p 0. \tag{A.4}$$

Now

$$\left\| \frac{\sqrt{T}}{M} B_{2T} \right\| \le M^{-1} \sum_{j} |w(j/M)| \left\| T^{-1} \sum_{t} u_{2t+j} x_t' (\hat{\gamma} - \gamma) \sqrt{T} \right\|$$

$$\le \left(\int_{-\infty}^{\infty} |w(x)| \, dx \right) \left(T^{-1} \sum_{1}^{T} u_{2t}' u_{2t} \right)^{1/2} G_{xT}^{1/2}$$

$$= O_p(1), \tag{A.5}$$

where

$$G_{xT} = \sqrt{T} (\hat{\gamma} - \gamma)' \hat{W}_T T^{-1} \sum_{1}^{T} \hat{W}_T^{-1} x_t x_t' \hat{W}_T^{-1'} \sqrt{T} \, \hat{W}_T' (\hat{\gamma} - \gamma)$$

$$= O_p(1).$$

Also,

$$\left\| \frac{\sqrt{T}}{M} B_{3T} \right\| \le M^{-1} \sum_{j} |w(j/M)| \left\| T^{-1} \sum_{t} u_{1t} \, \Delta k_{t+j}' (\hat{\Pi} - \Pi) \sqrt{T} \right\|$$

$$\le \left(\int_{-\infty}^{\infty} |w(x)| \, dx \right) \left(T^{-1} \sum_{1}^{T} u_{1t}' u_{1t} \right)^{1/2} G_{kT}^{1/2}$$

$$= O_p(1), \tag{A.6}$$

where

$$G_{kT} = T^{-1/2} (\hat{\Pi} - \Pi)' \delta_T T^{-1}$$

$$\times \sum_{1}^{T} (T \delta_T^{-1}) \, \Delta k_t \, \Delta k_t' (T \delta_T^{-1}) T^{-1/2} \delta_T (\hat{\Pi} - \Pi)$$

$$= O_p(1),$$

using (A.1) and (A.2). Finally,

$$\left\| \frac{\sqrt{T}}{M} B_{4T} \right\| \le \frac{1}{\sqrt{TM}} \sum_j |w(j/M)|$$

$$\times \left\| \sqrt{T}(\hat{\Pi} - \Pi)T^{-1} \sum_t \Delta k_{t+j} x_t'(\hat{\gamma} - \gamma)\sqrt{T} \right\|$$

$$\le T^{-1/2} \left(\int_{-\infty}^{\infty} |w(x)| \, dx \right) G_{kt}^{1/2} G_{xT}^{1/2} \to_p 0. \tag{A.7}$$

(A.4)–(A.7) together imply that

$$\frac{\sqrt{T}}{M} (\hat{\omega}_{21} - \omega_{21}) = O_p(1).$$

Combined with condition (M), we find $\hat{\omega}_{21} \to_p \omega_{21}$. □

Proof of Theorem 3

Part (a) follows from the analysis of Phillips and Hansen (1990) and part (a) of Theorem 1. Part (b) will follow from Theorem 1 if we show

$$T^{-1/2}\bar{W}^{-1}C'\left(\sum_1^T y_{2t}u_t^+ - T\tilde{\lambda}^+ \right) \Rightarrow \int_0^1 J_R \, dB_{1\cdot2}. \tag{A.8}$$

From Phillips and Hansen (1990) and Theorem 1 we know that

$$T^{-1/2}\bar{W}^{-1}C'\sum_1^T y_{2t}u_t^+ \Rightarrow \int_0^1 J_R \, dB_{1\cdot2} + \begin{pmatrix} C_1'\lambda^+ \\ 0 \end{pmatrix}. \tag{A.9}$$

By the consistency of $\tilde{\lambda}^+$ we have

$$\sqrt{T}\,\bar{W}^{-1}C'\tilde{\lambda}^+ = \begin{pmatrix} C_1' \\ \delta_T^{-1}\sqrt{T}\,C_2' \end{pmatrix}\tilde{\lambda}^+ \to_p \begin{pmatrix} C_1'\lambda^+ \\ 0 \end{pmatrix}. \tag{A.10}$$

(A.9) and (A.10) yield (A.8). □

Proof of the Lemma

Without loss of generality we will prove the Lemma for the case that the submatrices D_{ii} are scalar. Partition $Q' = (Q_1, Q_2, \ldots, Q_n)$. We construct $B' = (B_1, B_2, \ldots, B_n)$ using the following modification of the Gram–Schmidt procedure. Set $Q_i^* = [Q_1, \ldots, Q_{i-1}]$ and $B_i^* = [B_1, \ldots, B_{i-1}]$.

(i)   If $Q_{11} = Q_1'Q_1 = 0$, set $B_1 = 0$.
      If $Q_{11} > 0$, set $B_1 = Q_1/Q_{11}$.

(ii) If $Q_{22} = Q_2'Q_2 - Q_2'Q_1B_1'Q_2 = 0$, set $B_1 = 0$.
      If $Q_{22} > 0$, set $B_2 = (Q_2 - Q_1B_1'Q_2)/Q_{22}$.

(iii) If $Q_{ii} = Q_i'Q_i - Q_i'Q_i^*B_i^{*\prime}Q_i = 0$, set $B_i = 0$.
      If $Q_{ii} > 0$, set $B_i = [Q_i - Q_i^*B_i^{*\prime}Q_i]/Q_{ii}$.

This construction yields $\{B_i\}$ which satisfy $Q_i'B_j = 0$, $i < j$, and $Q_i'B_i$ equals either zero or one for each i. Thus

$$QB' = (Q_i'B_j) = (D_{ij}), \quad \text{as required.} \quad \square$$

Proof of Theorem 5

Part (a) is proved in the text. That is,

$$\sqrt{T}\,D_T(Q'\tilde{\alpha}^+ - c) \Rightarrow S'\left(\int_0^1 J_R J_R'\right)^{-1}\int_0^1 J_R\,\mathrm{d}B_{1\cdot 2}$$

$$\equiv \left(S'\left(\int_0^1 J_R J_R'\right)^{-1}S\right)^{1/2}\cdot\eta, \qquad (A.11)$$

where $\eta \equiv N(0, \omega_{1\cdot 2}I_q)$, and is independent of $V \equiv S'(\int_0^1 J_R J_R')^{-1}S$. This represents the limit distribution as a variance mixture of normals.

Similar analysis to that in the text yields

$$TD_T Q'\left(\sum_1^T y_{2t}y_{2t}'\right)^{-1}QD_T' = D_T\overline{Q}'\left(T^{-1}C'\sum_1^T y_{2t}y_{2t}'C\right)^{-1}\overline{Q}D_T'$$

$$\Rightarrow S'\left(\int_0^1 J_R J_R'\right)^{-1}S. \qquad (A.12)$$

Combining (A.11) and (A.12) with the fact that $\bar{\omega}_{1\cdot2} \to_p \omega_{1\cdot2}$ and the CMT gives

$$G_Q(\tilde{\alpha}^+, \tilde{V}_T) = (Q'\tilde{\alpha}^+ - c)' \left(Q' \left(\sum_1^T y_{2t} y_{2t}' \right)^{-1} Q \right)^{-1} (Q'\tilde{\alpha}^+ - c) / \bar{\omega}_{1\cdot2}$$

$$= \left(\sqrt{T} D_T \overline{Q}' C^{-1} (\tilde{\alpha}^+ - \alpha) \right)'$$

$$\times \left(T D_T Q' \left(C' \sum_1^T y_{2t} y_{2t}' C \right)^{-1} Q D_T' \right)^{-1}$$

$$\times \left(\sqrt{T} D_T \overline{Q}' C^{-1} (\tilde{\alpha}^+ - \alpha) \right) / \bar{\omega}_{1\cdot2}$$

$$\Rightarrow \eta' \left(S' \left(\int_0^1 J_R J_R' \right)^{-1} S \right)^{1/2} \left(S' \left(\int_0^1 J_R J_R' \right)^{-1} S \right)^{-1}$$

$$\times \left(S' \left(\int_0^1 J_R J_R' \right)^{-1} S \right)^{1/2} \cdot \eta / \omega_{1\cdot2}$$

$$\equiv \eta' \eta / \omega_{1\cdot2} \equiv \chi_q^2. \quad \square$$

Proof of Theorem 6

Notice that

$$\sqrt{T} \hat{W}_T^{-1} R = \begin{pmatrix} -\Omega_{22}^{1/2} \Pi \\ \sqrt{T} \delta_T \end{pmatrix} \to \begin{pmatrix} -\Omega_{22}^{1/2} \Pi \\ 0 \end{pmatrix} = \underline{R},$$

say, as $T \to \infty$, where $\mathrm{rank}(\underline{R}) = m$. Therefore,

$$T\hat{\beta}^+ = TR'(\hat{\gamma}^+ - \gamma) = \sqrt{T} R' \hat{W}_T'^{-1} \sqrt{T} \hat{W}_T' (\hat{\gamma}^+ - \gamma)$$

$$\Rightarrow \underline{R}' \left(\int_0^1 J_U J_U' \right)^{-1} \int_0^1 J_U \, dB_{1\cdot2}$$

and

$$
T^2 R' \left(\sum_1^T x_t x_t' \right)^{-1} R \Rightarrow \underline{R'} \left(\int_0^1 J_U J_U' \right)^{-1} \underline{R}.
$$

Standard manipulations and the CMT yield the result. □

Proof of Theorem 7

The proofs of (a) and (c) are similar. We show (c). Notice that

$$
\tilde{v}_t = v_t - (\tilde{\alpha} - \alpha)' y_{2t} = \tilde{b}' \xi_{Tt}, \tag{A.13}
$$

where

$$
\tilde{b} = \tilde{W}_{1T} (1, -(\tilde{\alpha} - \alpha)')',
$$

$$
\tilde{W}_{1T} = \mathrm{diag}\{1, T^{-1/2} \tilde{W}_T C^{-1}\},
$$

$$
\xi_{Tt} = W_{1T}'^{-1} \begin{pmatrix} v_t \\ y_{2t} \end{pmatrix}.
$$

Now

$$
\tilde{b} = \left(\begin{matrix} 1 \\ -\left(\dfrac{1}{T} \tilde{W}_T^{-1} C' \sum_1^T y_{2t} y_{2t}' C \tilde{W}_T^{-1} \right)^{-1} \left(\dfrac{1}{T} \tilde{W}_T^{-1} \sum_1^T C' y_{2t} v_t T^{-1/2} \right) \end{matrix} \right)
$$

$$
\Rightarrow \left(\begin{matrix} 1 \\ -\left(\int_0^1 J_R J_R' \right)^{-1} \int_0^1 J_R B_1 \end{matrix} \right) = \eta, \quad \text{say.} \tag{A.14}
$$

Defining

$$
S_{et} = \sum_1^t e_j, \qquad e_j = (u_{1j}, u_{2j}' C_1)',
$$

we have

$$
T^{-1/2}\xi_{Tt} = T^{-1/2}\begin{pmatrix} v_t \\ C_1'S_{2t} \\ \sqrt{T}\,\delta_T^{-1}(k_t + C_2'S_{2t}) \end{pmatrix} = \begin{pmatrix} T^{-1/2}S_{et} \\ \delta_T^{-1}(k_t + C_2'S_{2t}) \end{pmatrix}
$$

$$
\Rightarrow \begin{pmatrix} B_1(r) \\ J_R(r) \end{pmatrix} = \begin{pmatrix} B_1(r) \\ W_{n-m}(r) \\ k(r) \end{pmatrix}
$$

$$
= \begin{pmatrix} B_e(r) \\ k(r) \end{pmatrix} = \xi(r), \quad \text{say.} \tag{A.15}
$$

Now

$$
B_1(r) = \bar{\omega}W_1(r) + \omega_{12}C_1W_{n-m}(r), \qquad \bar{\omega} = \omega_{11} - \omega_{12}C_1C_1'\omega_{21}. \tag{A.16}
$$

Thus

$$
\eta'\xi(r) = B_1(r) - \int_0^1 B_1 J_R'\left(\int_0^1 J_R J_R'\right)^{-1} J_R(r)
$$

$$
= \bar{\omega}^{1/2}\left(W_1(r) - \int_0^1 W_1 J_R'\left(\int_0^1 J_R J_R'\right)^{-1} J_R(r)\right)
$$

$$
= \bar{\omega}^{1/2}W_{J_R}(r). \tag{A.17}
$$

We see from (A.13), (A.14), (A.15), and (A.17) that

$$
T^{-2}\sum_1^T \tilde{v}_t^2 = \bar{b}'T^{-2}\sum_1^T \xi_{Tt}\xi_{Tt}'\bar{b}
$$

$$
\Rightarrow \eta'\int_0^1 \xi\xi'\eta = \int_0^1 (\eta'\xi)^2 = \bar{\omega}\int_0^1 W_{J_R}^2. \tag{A.18}
$$

Now

$$\tilde{\lambda} = T^{-1} \sum_{j=1}^{T} w(j/M) \sum_t \tilde{\varepsilon}_{t-j} \tilde{\varepsilon}_t$$

$$= T^{-1} \sum_j w(j/M) \sum_t \Delta \tilde{v}_{t-j} \Delta \tilde{v}_t + o_p(1)$$

$$= \tilde{b}' \sum_j w(j/M) T^{-1} \sum_t \Delta \xi_{Tt-j} \Delta \xi'_{Tt} \tilde{b} + o_p(1)$$

$$= \tilde{b}' \begin{pmatrix} A_T & C_T \\ B_T & D_T \end{pmatrix} \tilde{b} + o_p(1),$$

where

$$A_T = \sum_j w(j/M) T^{-1} \sum_t e_{t-j} e'_t \to_p \Lambda_e = \sum_{j=0}^{\infty} \mathrm{E}(e_0 e'_j),$$

$$B_T = \sum_j w(j/M) T^{-1} \sum_t \left(\delta_T^{-1} \Delta k_{t-j} + C'_2 u_{2t-j} \right) e'_t,$$

$$C_T = \sum_j w(j/M) T^{-1} \sum_t e_{t-j} \left(\Delta k'_t \delta_T^{-1} + u'_{2t} C_2 \right),$$

$$D_T = \sum_j w(j/M) T^{-1} \sum_t \left(\delta_T^{-1} \Delta k_{t-j} + C'_2 u_{2t-j} \right) \left(\Delta k'_t \delta_T^{-1} + u'_{2t} C_2 \right).$$

(U), (W), and (M) are sufficient for

$$A_T \to_p \Lambda_e = \sum_{j=0}^{\infty} \mathrm{E}(e_0 e'_j)$$

by Andrews (1991, proposition 1). By arguments similar to those used in (A.5), (A.6), and (A.7), we can show that

$$B_T \to_p 0, \qquad C_T \to_p 0, \qquad D_T \to_p 0.$$

Thus

$$\tilde{\lambda} = \tilde{b}' \begin{pmatrix} \Lambda_e & 0 \\ 0 & 0 \end{pmatrix} \tilde{b} + o_p(1). \tag{A.19}$$

Now by Hansen (1990, theorem 4.1)

$$T^{-1} \sum_1^T S_{et} e_{t+1} \Rightarrow \int_0^1 B_e \, \mathrm{d}B_e' + \Lambda_e,$$

and by (A.1) and the CMT

$$T^{-1/2} \sum_1^T S_{et} \, \Delta k_{t+1}' \delta_T^{-1} \Rightarrow \int_0^1 B_e \, \mathrm{d}k',$$

$$T^{-1/2} \sum_1^T \delta_T^{-1} k_t e_{t+1} \Rightarrow \int_0^1 k \, \mathrm{d}B_e,$$

$$\sum_1^T \delta_T^{-1} k_t \, \Delta k_t' \delta_T^{-1} \Rightarrow \int_0^1 k \, \mathrm{d}k'.$$

Combined with (A.13), (A.15), and (A.19),

$$T^{-1} \sum_1^T \tilde{v}_t \, \Delta \tilde{v}_{t+1}' - \tilde{\lambda}$$

$$= \tilde{b}' T^{-2} \sum_1^T \xi_{Tt} \, \Delta \xi_{Tt}' \tilde{b} - \tilde{b}' \begin{pmatrix} \Lambda_e & 0 \\ 0 & 0 \end{pmatrix} \tilde{b} + \mathrm{o}_p(1)$$

$$= \tilde{b}' T^{-1} \sum_1^T \begin{pmatrix} (S_{et} e_{t+1}' - \Lambda_e) & S_{et} \, \Delta k_{t+1}' \delta_T^{-1} T^{1/2} \\ T^{1/2} \delta_T^{-1} k_t e_{t+1}' & T \delta_T^{-1} k_t \, \Delta k_{t+1}' \delta_T^{-1} \end{pmatrix} \tilde{b} + \mathrm{o}_p(1)$$

$$\Rightarrow \eta' \begin{pmatrix} \int_0^1 B_e \, \mathrm{d}B_e' & \int_0^1 B_e \, \mathrm{d}k' \\ \int_0^1 k \, \mathrm{d}B_e' & \int_0^1 k \, \mathrm{d}k' \end{pmatrix} \eta = \eta' \int_0^1 \xi \, \mathrm{d}\xi' \eta$$

$$= \int_0^1 \eta' \xi \, \mathrm{d}(\eta' \xi) = \bar{\omega} \int_0^1 W_{J_R} \, \mathrm{d}W_{J_R}, \tag{A.20}$$

by (A.17). Putting (A.18) and (A.20) together,

$$T(\tilde{\rho}^* - 1) = \frac{T^{-1} \sum_1^T \tilde{v}_t \, \Delta\tilde{v}_{t+1} - \tilde{\lambda}}{T^{-2} \sum_1^T \tilde{v}_t^2} \Rightarrow \frac{\overline{\omega} \int_0^1 W_{J_R} \, \mathrm{d}W_{J_R}}{\overline{\omega} \int_0^1 W_{J_R}^2} = \frac{\int_0^1 W_{J_R} \, \mathrm{d}W_{J_R}}{\int_0^1 W_{J_R}^2},$$

as required.

The proofs of (b) and (d) are similar. We show (d). By the previous analysis,

$$\frac{T^{-1} \sum_1^T \tilde{v}_t \, \Delta\tilde{v}_{t+1} - \tilde{\lambda}}{\left(T^{-2} \sum_1^T \tilde{v}_t^2\right)^{1/2}} \Rightarrow \overline{\omega}^{1/2} \frac{\int_0^1 W_{J_R} \, \mathrm{d}W_{J_R}}{\left(\int_0^1 W_{J_R}^2\right)^{1/2}} = \overline{\omega}^{1/2} \int_0^1 Q(W_1, J_R) \, \mathrm{d}W_{J_R}.$$

Defining $\Omega_e = \sum_{-\infty}^\infty \mathrm{E}(e_0 e_j')$,

$$\tilde{\omega}_{11} = T^{-1} \sum_j w(j/M) \sum_t \tilde{\varepsilon}_{t-j} \tilde{\varepsilon}_t$$

$$= T^{-1} \sum_j w(j/M) \sum_t \Delta\tilde{v}_{t-j} \Delta\tilde{v}_t + \mathrm{o}_p(1)$$

$$= \tilde{b}' T^{-1} \sum_j w(j/M) \sum_t \Delta\xi_{Tt-j} \Delta\xi_{Tt}' \tilde{b} + \mathrm{o}_p(1)$$

$$\Rightarrow \eta' \begin{pmatrix} \Omega_e & 0 \\ 0 & 0 \end{pmatrix} \eta$$

$$= \left(\Omega_e^{1/2}\eta\right)' \begin{pmatrix} I_{n-m+1} \\ 0 \end{pmatrix} (I_{n-m+1} \quad 0)\left(\Omega_e^{1/2}\eta\right)$$

$$= \overline{\omega} K_R' K_R.$$

Indeed,

$$\Omega_e = \begin{pmatrix} \omega_{11} & 0 \\ C_1' \omega_{21} & I \end{pmatrix}, \qquad \Omega_e^{1/2} = \begin{pmatrix} \overline{\omega}^{1/2} & 0 \\ C_1' \omega_{21} & I \end{pmatrix},$$

so

$$\Omega_e^{1/2}\eta = \begin{pmatrix} \overline{\omega}^{1/2} \\ -\left(\left(\int_0^1 J_R J_R'\right)^{-1}\int_0^1 J_R B_1 - C_1'\omega_{21}\right) \end{pmatrix}.$$

Thus,

$$(I_{n-m+1} \quad 0)\Omega_e^{1/2}\eta$$

$$= \begin{pmatrix} \overline{\omega}^{1/2} \\ -(I_{n-m} \quad 0)\left(\left(\int_0^1 J_R J_R'\right)^{-1}\int_0^1 J_R B_1 - C_1'\omega_{21}\right) \end{pmatrix}$$

$$= \begin{pmatrix} \overline{\omega}^{1/2} \\ -(I_{n-m} \quad 0)\left(\left(\int_0^1 J_R J_R'\right)^{-1} \\ \times\left(\int_0^1 J_R W_1 \overline{\omega}^{1/2} + \int_0^1 J_R W_{n-m}' C_1'\omega_{21}\right) - C_1'\omega_{21}\right) \end{pmatrix} \qquad (A.21)$$

[using (A.16)]. Since

$$(I_{n-m} \quad 0)\left(\int_0^1 J_R J_R'\right)^{-1}\int_0^1 J_R W_{n-m}'$$

$$= (I_{n-m} \quad 0)\left(\int_0^1 J_R J_R'\right)^{-1}\int_0^1 J_R J_R'\begin{pmatrix} I_{n-m} \\ 0 \end{pmatrix} = I_{n-m},$$

(A.21) equals

$$\begin{pmatrix} 1 \\ -(I_{n-m} \quad 0)\left(\left(\int_0^1 J_R J_R'\right)^{-1}\int_0^1 J_R W_1\right) \end{pmatrix}\overline{\omega}^{1/2} = K_R\overline{\omega}^{1/2}.$$

Therefore

$$
Z(\tilde{t}^*) = \cfrac{T^{-1}\sum_{1}^{T}\tilde{v}_t\,\Delta\tilde{v}_{t+1} - \tilde{\lambda}}{\left(T^{-2}\sum_{1}^{T}\tilde{v}_t^2\right)^{1/2}\;\bar{\omega}_{11}^{1/2}}
$$

$$
\Rightarrow \cfrac{\bar{\omega}^{1/2}\int_0^1 Q(W_1, J_R)\,\mathrm{d}W_{J_R}}{\left(\bar{\omega}K'_{J_R}K_{J_R}\right)^{1/2}}
$$

$$
= \int_0^1 Q(W_1, J_R)\,\mathrm{d}S(W_1, J_R). \quad \square
$$

References

Andrews, D.W.K., 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, Econometrica 59, 817–858.

Engle, R.F. and C.W.J. Granger, 1987, Co-integration and error-correction: Representation, estimation and testing, Econometrica 55, 251–276.

Engle, R.F. and B.S. Yoo, 1987, Forecasting and testing in co-integrated systems, Journal of Econometrics 35, 143–159.

Fuller, W.A., 1976, Introduction to statistical time series (Wiley, New York, NY).

Granger, C.W.J. and T.-H. Lee, 1990, Multicointegration, Advances in Econometrics 8, 71–84.

Hansen, B.E., 1990, Convergence to stochastic integrals for dependent heterogeneous processes, Manuscript (University of Rochester, Rochester, NY).

Herrndorf, N., 1984, A functional central limit theorem for weakly dependent sequences of random variables, The Annals of Probability 12, 141–153.

Johansen, S., 1988a, Statistical analysis of cointegration vectors, Journal of Economic Dynamics and Control 12, 231–254.

Johansen, S., 1988b, Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, Manuscript.

Newey, W. and K.D. West, 1987, A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix, Econometrica 55, 703–708.

Park, J.Y. and P.C.B. Phillips, 1988, Statistical inference in regressions with integrated processes: Part 1, Econometric Theory 4, 468–498.

Park, J.Y. and P.C.B. Phillips, 1989, Statistical inference in regressions with integrated processes: Part 2, Econometric Theory 5, 95–132.

Parzen, E., 1957, On consistent estimates of the spectrum of a stationary time series, Annals of Mathematical Statistics 28, 329–348.

Phillips, P.C.B., 1987, Time series regression with a unit root, Econometrica 55, 277–301.

Phillips, P.C.B., 1988, Reflections on econometric methodology, Economic Record 64, 344–359.

Phillips, P.C.B., 1991, Optimal inference in cointegrated systems, Econometrica 59, 283–306.

Phillips, P.C.B. and S.N. Durlauf, 1986, Multiple time series with integrated variables, Review of Economic Studies 53, 473–496.

Phillips, P.C.B. and B.E. Hansen, 1990, Statistical inference in instrumental variables regression with $I(1)$ processes, Review of Economic Studies 57, 99–125.

Phillips, P.C.B. and S. Ouliaris, 1990, Asymptotic properties of residual based tests for cointegration, Econometrica 58, 165–193.

Said, S.E. and D.A. Dickey, 1984, Testing for unit roots in autoregressive-moving average models of unknown order, Biometrika 71, 599–607.

Sims, C.A., J.H. Stock, and M.W. Watson, 1990, Inference in linear time series models with some unit roots, Econometrica 58, 113–144.

Stock, J.H., 1987, Asymptotic properties of least squares estimators of cointegrating vectors, Econometrica 55, 1035–1056.

Stock, J.H. and M.W. Watson, 1988, Testing for common trends, Journal of the American Statistical Association 83, 1097–1107.

West, K.D., 1988, Asymptotic normality, when regressors have a unit root, Econometrica 6, 1397–1418.