Available online at www.sciencedirect.com
SCIENCE DIRECT®

JOURNAL OF Econometrics

Journal of Econometrics 124 (2005) 187–201
www.elsevier.com/locate/econbase

Point optimal tests of the null hypothesis of cointegration

Michael Jansson*

Department of Economics, University of California, Berkeley, 549 Evans Hall #3880, Berkeley, CA 94720-3380, USA

Accepted 9 February 2004

Abstract

This paper obtains an asymptotic Gaussian power envelope for tests of the null hypothesis of cointegration. In addition, the paper proposes a feasible point optimal cointegration test whose local asymptotic power function is found to be close to the asymptotic Gaussian power envelope.

JEL classification: C22

Keywords: Cointegration; Local asymptotic power; Point optimal test; Power envelope

1. Introduction

The concept of cointegration (Engle and Granger, 1987) has attracted considerable attention in the literature and answers to a variety of questions concerning inference in cointegrated systems have been provided. Asymptotic optimality theory for the problem of estimating cointegrating vectors under normality has been developed by Phillips (1991) and several asymptotically efficient estimation procedures have been proposed (for a review, see Watson, 1994). Moreover, an asymptotic Gaussian power envelope for tests of the unit root assumption underlying these cointegration methods has been obtained by Elliott et al. (1996). In contrast, although numerous papers have considered the problem of testing the null hypothesis of cointegration against the alternative of no cointegration (examples include Choi and Ahn, 1995; Park, 1990; Shin, 1994;

* Tel.: +1-510-642-4639; fax: +1-510-642-6615.
E-mail address: mjansson@econ.berkeley.edu (M. Jansson).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2004.02.011

Xiao and Phillips, 2002), no asymptotic optimality theory for that testing problem has been developed.

This paper obtains an asymptotic Gaussian power envelope for tests of the null hypothesis of cointegration and proposes a feasible point optimal cointegration test. By construction, the point optimal test attains the asymptotic Gaussian power envelope at a prespecified alternative. Against other alternatives, the local asymptotic power of the point optimal test is close to the asymptotic Gaussian power envelope. Moreover, the point optimal test is found to perform well in a Monte Carlo experiment.

In terms of the methodology employed, the present paper is related to Elliott et al.'s (1996) and Saikkonen and Luukkonen's (1993) studies of unit root testing in autoregressive and moving average models, respectively. As do Elliott et al. (1996) and Saikkonen and Luukkonen (1993), this paper develops asymptotic optimality results by obtaining limiting distributions of optimal test statistics derived under a normality assumption. By accommodating stochastically trending and (possibly) endogenous regressors, the results obtained here generalize the fixed-regressor results of Saikkonen and Luukkonen (1993).

Section 2 presents the model. In Section 3, the asymptotic Gaussian power envelope is derived. Section 4 constructs a feasible point optimal test. Finally, Section 5 reports local asymptotic power results and investigates the finite sample performance of the test proposed in this paper, while all mathematical derivations are collected in an Appendix.

2. The model and assumptions

Let z_t be an observed m-vector time series generated by

$$z_t = \mu_t^z + z_t^0, \quad 1 \leqslant t \leqslant T, \tag{1}$$

where μ_t^z is a deterministic component and z_t^0 is a zero-mean stochastic component. For concreteness, the deterministic component is assumed to be a pth-order polynomial time trend:

$$\mu_t^z = \alpha_z' d_t, \tag{2}$$

where $d_t = (1, \ldots, t^p)'$ and α_z is a $(p+1) \times m$ matrix of parameters. The cases of particular interest are the constant mean and linear trend cases corresponding to $d_t = 1$ and $d_t = (1, t)'$, respectively.

Partition z_t^0 into a scalar y_t^0 and a k-vector x_t^0 $(k = m - 1)$ as $z_t^0 = (y_t^0, x_t^{0\prime})'$ and suppose z_t^0 is generated by the potentially cointegrated system

$$y_t^0 = \beta' x_t^0 + v_t, \tag{3}$$

$$\Delta x_t^0 = u_t^x,$$

where v_t is an error process with generating mechanism

$$\Delta v_t = u_t^y - \theta u_{t-1}^y, \quad 1 \leqslant t \leqslant T. \tag{4}$$

In (3) and (4) $\beta \in \mathbb{R}^k$ and $\theta \in (-1, 1]$ are unknown parameters, and $u_t = (u_t^y, u_t^{x'})$ satisfies the following assumption.

A1. $u_t = \sum_{i=0}^{\infty} C_i \varepsilon_{t-i}$, where $\{\varepsilon_t : t \in \mathbb{Z}\}$ is *i.i.d.* $(0, I_m)$, $\sum_{i=0}^{\infty} C_i$ has full rank and $\sum_{i=0}^{\infty} i\|C_i\| < \infty$, where $\|\cdot\|$ denotes the Euclidean norm.

The system is initialized at $t=0$ with $v_0 = u_0^y = 0$ and $x_0^0 = 0$. The moving average specification (4) implies that y_t^0 and x_t^0 are cointegrated if and only if $\theta = 1$. Assumption A1 imposes the standard, but important, regularity condition that x_t^0 is a non-cointegrated integrated process. In addition, it is assumed that the cointegration between y_t^0 and x_t^0 is regular (in the sense of Park, 1992) when $\theta = 1$. It would be of interest to relax one or both of these assumptions, but doing so is beyond the scope of the present paper.

Conformably with z_t^0, partition z_t and α_z as $z_t = (y_t, x_t')'$ and $\alpha_z = (\alpha_y, \alpha_x)$. Defining $\alpha = \alpha_y - \alpha_x \beta$, the model can be written in triangular form as

$$y_t = \alpha' d_t + \beta' x_t + v_t, \quad \Delta v_t = u_t^y - \theta u_{t-1}^y,$$

$$\Delta x_t = \alpha_x' \Delta d_t + u_t^x.$$

This paper considers the problem of testing

$$\mathrm{H}_0 : \theta = 1 \quad vs. \quad \mathrm{H}_1 : \theta < 1, \qquad (\mathscr{P})$$

treating α and β as unknown nuisance parameters. Section 3 derives an asymptotic power envelope under the following strengthening of A1.

A1*. $u_t \sim$ i.i.d. $\mathcal{N}(0, \Sigma)$, where Σ is positive definite.

As a by-product of the analysis, a point optimal test statistic is obtained. Section 4 relaxes A1* and constructs a feasible point optimal test which is applicable whenever A1 holds.

Remarks. (i) Some papers in the existing literature on cointegration testing (e.g., Shin, 1994) employ an alternative model for v_t, the so-called "local-level" unobserved components model. As does the moving average model (4), the unobserved components model parameterizes cointegration as a point. Using the fact that $\Delta u_t^y + \Delta \xi_t$ has an MA (1) representation (e.g., Stock, 1994), it can be shown that the power envelope developed in the next section coincides with power envelope for the testing problem

$$\mathrm{H}_0 : \sigma_\xi^2 = 0 \quad \text{vs.} \quad \mathrm{H}_1 : \sigma_\xi^2 > 0$$

in the model

$$y_t = \alpha' d_t + \beta' x_t + u_t^y + \xi_t,$$

where $\Delta x_t = \alpha_x' \Delta d_t + u_t^x$, $u_t = (u_t^y, u_t^{x'})'$ satisfies A1*, $\Delta \xi_t \sim$ i.i.d. $\mathcal{N}(0, \sigma_\xi^2)$ and $\{\Delta \xi_t\}$ is independent of $\{u_t\}$. For concreteness, and in order to conserve space, the present paper only studies the moving average model.

(ii) The triangular model formulation (3) and (4) does not treat y_t and x_t symmetrically. In particular, it is assumed that a linear combination of y_t and x_t is stationary only if it puts non-zero weight on y_t. In contrast, the fact that $\beta = 0$ is permitted implies that it is possible for a linear combination of y_t and x_t to be stationary even if it puts zero weight on x_t. As a consequence, the testing problem ($\mathscr{P}$) is not invariant under orthogonal transformations of $(y_t, x_t')'$ (such as a relabelling of y_t and x_t in the case where x_t is a scalar). The assumption that one variable, y_t, is known to appear with a non-zero coefficient in the (potentially) cointegrating relation would appear to be satisfied in most empirical applications.

3. An asymptotic Gaussian power envelope

Under A1*, the model is fully parametric and classical statistical theory can be employed to develop optimality theory for the testing problem ($\mathscr{P}$). In particular, it is possible to construct an attainable upper bound on the local asymptotic power of a class of cointegration tests. The upper bound is constructed by eliminating the nuisance parameters α, β, α_x and Σ from the problem and then applying the Neyman–Pearson lemma. Two different elimination methods are used: α_x and Σ are treated as known, while the principle of invariance (e.g., Lehmann, 1994) is used to handle α and β.

The power bound derived under the assumption that α_x and Σ are known is attainable even when these parameters are unknown. As a consequence, it can (and will) be assumed that α_x and Σ are known. The remaining nuisance parameters, α and β, are eliminated by proceeding along the lines of numerous other papers on tests concerning the covariance structure of the error term in a linear regression model (e.g., King, 1980; King and Hillier, 1985; Dufour and King, 1991).

Define the matrices $Y = (y_1, \ldots, y_T)'$, $X = (x_1, \ldots, x_T)'$ and $D = (d_1, \ldots, d_T)'$. Consider the group of transformations of the form

$$g_{a,b}(Y, X) = (Y + D \cdot a + X \cdot b, X), \qquad (\mathscr{G})$$

where $a \in \mathbb{R}^{p+1}$ and $b \in \mathbb{R}^k$. Each $g_{a,b}$ induces a transformation

$$\bar{g}_{a,b}(\theta, \alpha, \beta) = (\theta, \alpha + a, \beta + b) \qquad (\bar{\mathscr{G}})$$

in the parameter space. Because θ is invariant under ($\bar{\mathscr{G}}$), the testing problem ($\mathscr{P}$) is invariant under ($\mathscr{G}$). It therefore seems natural to restrict attention to tests that are invariant under ($\mathscr{G}$). All previously proposed tests (of which the author is aware) are invariant under this group of transformations, so the class of tests that are invariant under ($\mathscr{G}$) is quite large. Because θ is a maximal invariant under ($\bar{\mathscr{G}}$), it follows from Theorem 6.3 of Lehmann (1994) that the distribution of any invariant statistic (i.e. any statistic invariant under ($\mathscr{G}$)) depends on (θ, α, β) only through θ, implying that the nuisance parameters α and β can be eliminated by restricting attention to statistics that are invariant under ($\mathscr{G}$). A natural maximal invariant under ($\mathscr{G}$) is $((R_\perp' Y)', vec(X)')'$, where $R_\perp$ is a matrix whose columns form an orthonormal basis for the orthogonal complement of the column space of $R=(X,D)$. The distributional properties of this maximal invariant are developed next.

Partition Σ in conformity with u_t as

$$\Sigma=\begin{pmatrix}\sigma_{yy} & \sigma'_{xy}\\ \sigma_{xy} & \Sigma_{xx}\end{pmatrix}$$

and for any θ^*, define $\Psi_{\theta^*}=\Psi^{1/2}_{\theta^*}\Psi^{1/2\prime}_{\theta^*}$, where $\Psi^{1/2}_{\theta^*}$ a lower triangular $T\times T$ matrix with ones on the diagonal and $1-\theta^*$ below the diagonal:

$$\Psi^{1/2}_{\theta^*}=\begin{pmatrix}1 & 0 & \cdots & 0\\ 1-\theta^* & 1 & \ddots & \vdots\\ \vdots & \ddots & \ddots & 0\\ 1-\theta^* & \cdots & 1-\theta^* & 1\end{pmatrix}.$$

Under A1*, $vec(X)\sim\mathcal{N}((\alpha'_x\otimes I_T)vec(D),\Sigma_{xx}\otimes\Psi_0)$ and

$$R'_\perp Y|_X\sim\mathcal{N}(\theta(R'_\perp\Psi_0^{-1/2}X)\Sigma_{xx}^{-1}\sigma_{xy},\sigma_{yy\cdot x}(R'_\perp\Psi_\theta R_\perp)),$$

where $\sigma_{yy\cdot x}=\sigma_{yy}-\sigma'_{xy}\Sigma_{xx}^{-1}\sigma_{xy}$ and "$\cdot|_X$" signifies the conditional distribution relative to the σ-algebra generated by X. Apart from an additive term that does not depend on θ, minus 2 times the log density of the maximal invariant can be written as

$$\mathcal{L}_T(\theta)=\log|R'\Psi_\theta^{-1}R|+\sigma_{yy\cdot x}^{-1}Y'_\theta(\Psi_\theta^{-1}-\Psi_\theta^{-1}R(R'\Psi_\theta^{-1}R)^{-1}R'\Psi_\theta^{-1})Y_\theta, \tag{6}$$

where $Y_\theta=Y-\theta\cdot\Psi_0^{-1/2}X\Sigma_{xx}^{-1}\sigma_{xy}$ and the derivation of (6) makes use of the relations

$$R_\perp(R'_\perp\Psi_\theta R_\perp)^{-1}R'_\perp=\Psi_\theta^{-1}-\Psi_\theta^{-1}R(R'\Psi_\theta^{-1}R)^{-1}R'\Psi_\theta^{-1}$$

and $|R'_\perp\Psi_\theta R_\perp|=|R'\Psi_\theta^{-1}R|\cdot|R'R|^{-1}$.

Expression (6) differs from its fixed-regressor counterpart (e.g., King, 1980) in two respects. In fixed-regressor settings, the term corresponding to $\log|R'\Psi_\theta^{-1}R|$ is non-random and can be omitted. Moreover, the definition of Y_θ reflects the fact that correlations between u_t^y and u_t^x must be taken into account when X is random. As in the fixed-regressor case, the term

$$Y'_\theta(\Psi_\theta^{-1}-\Psi_\theta^{-1}R(R'\Psi_\theta^{-1}R)^{-1}R'\Psi_\theta^{-1})Y_\theta$$

in (6) can be interpreted as the weighted sum of squared residuals from a GLS regression (of Y_θ on R using the covariance matrix Ψ_θ).

It follows from the Neyman–Pearson Lemma that the test which rejects when $\mathcal{L}_T(1)-\mathcal{L}_T(\bar\theta)$ is large is the most powerful invariant test of $\theta=1$ vs. $\theta=\bar\theta<1$. An asymptotic analogue of that optimality result can be obtained by deriving the limiting distribution of $\mathcal{L}_T(1)-\mathcal{L}_T(\bar\theta)$ under a local-to-unity reparameterization of θ and $\bar\theta$ in which $\lambda=T(1-\theta)\geqslant 0$ and $\bar\lambda=T(1-\bar\theta)>0$ are held constant as T increases without bound. A formal statement is provided in Theorem 1, the proof of which represents the limiting distribution of $\mathscr{L}_T(1) - \mathscr{L}_T(\bar{\theta})$ in terms of the random functional

$$\varphi(\lambda;\bar{\lambda}) = 2\bar{\lambda}\int_0^1 V_{\bar{\lambda}}^{\lambda}\,\mathrm{d}V^{\lambda} - \bar{\lambda}^2\int_0^1 (V_{\bar{\lambda}}^{\lambda})^2 + \left(\int_0^1 Q_{\bar{\lambda}}\,\mathrm{d}V_{\bar{\lambda}}^{\lambda}\right)'\left(\int_0^1 Q_{\bar{\lambda}}Q_{\bar{\lambda}}'\right)^{-1}\left(\int_0^1 Q_{\bar{\lambda}}\,\mathrm{d}V_{\bar{\lambda}}^{\lambda}\right) - \log\left|\int_0^1 Q_{\bar{\lambda}}Q_{\bar{\lambda}}'\right| - \left(\int_0^1 Q\,\mathrm{d}V^{\lambda}\right)'\left(\int_0^1 QQ'\right)^{-1}\left(\int_0^1 Q\,\mathrm{d}V^{\lambda}\right) + \log\left|\int_0^1 QQ'\right|,$$

where $V_{\bar{\lambda}}^{\lambda}(s) = \int_0^s \mathrm{e}^{-\bar{\lambda}(s-t)}\,\mathrm{d}V^{\lambda}(t)$, $V^{\lambda}(s) = V(s) + \lambda\int_0^s V(t)\,\mathrm{d}t$, $Q_{\bar{\lambda}}(s) = \int_0^s \mathrm{e}^{-\bar{\lambda}(s-t)}\,\mathrm{d}Q(t)$, $Q(s) = (1,\ldots,s^p, W(s)')'$, and V and W are independent Wiener processes of dimensions 1 and k, respectively.

Theorem 1. *Let z_t be generated by (1)–(4) and suppose* A1* *holds. An upper bound on the local asymptotic power of any invariant test of $\theta = 1$ against $\theta = \theta_T = 1 - T^{-1}\lambda$ is given by $\pi^{\delta}(\lambda) = [\varphi(\lambda;\lambda) > c^{\delta}(\lambda)]$, where δ is the asymptotic level of the test, $c^{\delta}(\lambda)$ satisfies $[\varphi(0;\lambda) > c^{\delta}(\lambda)] = \delta$ and invariance is with respect to transformations of the form ($\mathscr{G}$).*

Under the reparameterization employed in Theorem 1, the null and alternative hypotheses are $\lambda = 0$ and $\lambda > 0$, respectively. The upper bound provided by the Gaussian power envelope is sharp in the sense that $\pi^{\delta}(\lambda)$ can be attained for any given λ by the test which rejects for large values of the corresponding point optimal test statistic, viz. $\mathscr{L}_T(1) - \mathscr{L}_T(1 - T^{-1}\lambda)$. Indeed, previous research on special cases of the testing problem considered here (e.g., Saikkonen and Luukkonen, 1993) suggests that a test based on $\mathscr{L}_T(1) - \mathscr{L}_T(\bar{\theta})$ should have a local asymptotic power function close to $\pi^{\delta}(\cdot)$ if $\bar{\theta}$ is chosen appropriately. The function $\pi^{\delta}(\cdot)$ therefore constitutes a useful benchmark against which the local power function of any invariant test of the null hypothesis of cointegration can be compared.

The power envelope depends on p, the order of the deterministic component, and k, the dimension of x_t. On the other hand, although the form of the point optimal test statistic depends on Σ, the power envelope does not depend on the covariance matrix of the underlying errors u_t. In particular, the power envelope does not depend on the extent to which the regressors are endogenous in the sense that u_t^x is correlated with the latent error u_t^y. Jansson (2004a) derives the asymptotic Gaussian power envelope in a model isomorphic to the present model under the assumption that β is known. That power envelope depends on $\sigma_{yy}^{-1}\sigma_{xy}'\Sigma_{xx}^{-1}\sigma_{xy}$, the squared coefficient of multiple correlation between u_t^y and u_t^x, and substantial power gains over conventional tests are available when the correlation is non-zero. In contrast, it follows from Theorem 1 that it is impossible to exploit any correlations between u_t^x and u_t^y when testing the null hypothesis that y_t and x_t are cointegrated with an unknown cointegrating vector β.

4. Feasible point optimal tests

This section constructs a test statistic which can be computed without knowledge of any nuisance parameters and has a limiting distribution of the form $\varphi(\lambda;\bar{\lambda})$ under A1. Let $\hat{\Sigma},\hat{\Gamma}$ and $\hat{\Omega}$ denote estimators of $\Sigma = E(u_1 u_1')$, $\Gamma = \sum_{t=2}^{\infty} E(u_t u_1')$ and $\Omega = \Sigma + \Gamma + \Gamma'$, respectively. Partition the matrices $\hat{\Gamma}$ and $\hat{\Omega}$ in conformity with u_t and let $\hat{\Gamma}_{x\cdot} = (\hat{\gamma}_{xy}, \hat{\Gamma}_{xx})$, $\hat{\omega}_{yy\cdot x} = \hat{\kappa}'\hat{\Omega}\hat{\kappa}$ and $\hat{\gamma}_{yy\cdot x} = \hat{\kappa}'\hat{\Gamma}\hat{\kappa}$, where $\hat{\kappa} = (1, -\hat{\omega}'_{xy}\hat{\Omega}_{xx}^{-1})'$. Let $Y_D = M_D Y$ and $X_D = M_D X$, where $M_D = I - D(D'D)^{-1}D'$, and define $R^+ = (D, X + \hat{U}\hat{\Sigma}^{-1}\hat{\Gamma}'_{x\cdot})$, where $\hat{U} = (Y_D - X_D\hat{\beta}, \Psi_0^{-1/2} X_D)$ and $\hat{\beta} = (X_D' X_D)^{-1} X_D' Y_D$. Finally, let $Y_{\theta^*}^+ = Y - \theta^* \cdot \Psi_0^{-1/2} X \hat{\Omega}_{xx}^{-1}\hat{\omega}_{xy} + \hat{U}\hat{\Sigma}^{-1}\hat{\Gamma}'_{x\cdot}\hat{\beta}$ for any θ^*.

The proposed test statistic is

$$P_T(\bar{\lambda}) = \mathscr{L}_T^+(1) - \mathscr{L}_T^+(1 - T^{-1}\bar{\lambda}) - 2\bar{\lambda}\hat{\omega}_{yy\cdot x}^{-1}\hat{\gamma}_{yy\cdot x}, \tag{7}$$

where $\bar{\lambda} > 0$ is a prespecified constant and

$$\mathscr{L}_T^+(\theta^*) = \log|R^{+'}\Psi_{\theta^*}^{-1}R^+| + \hat{\omega}_{yy\cdot x}^{-1} Y_{\theta^*}^{+'}(\Psi_{\theta^*}^{-1} - \Psi_{\theta^*}^{-1}R^+(R^{+'}\Psi_{\theta^*}^{-1}R^+)^{-1}R^{+'}\Psi_{\theta^*}^{-1})Y_{\theta^*}^+.$$

Under the assumptions of Theorem 1, $P_T(\bar{\lambda})$ is asymptotically equivalent to $\mathscr{L}_T(1) - \mathscr{L}_T(\bar{\theta})$. As Theorem 2 shows, $P_T(\bar{\lambda})$ has a limiting distribution of the form $\varphi(\lambda;\bar{\lambda})$ even when u_t exhibits serial correlation of the form permitted under A1. This robustness property, not shared by $\mathscr{L}_T(1) - \mathscr{L}_T(\bar{\theta})$, is achieved by employing two serial correlation corrections. The first correction, employed in the construction of $Y_{\theta^*}^+$ and R^+, is similar to the correction proposed by Park (1992) in the context of estimation of a cointegrating regression. Indeed, the purpose of this correction is to remove "serial correlation bias" from the limiting distribution of the estimator of β appearing in $\mathscr{L}_T^+(\theta^*)$. The second correction term, $-2\bar{\lambda}\hat{\omega}_{yy\cdot x}^{-1}\hat{\gamma}_{yy\cdot x}$, in (7) resembles the correction term in Phillips' (1987) Z_α test for an autoregressive unit root and accommodates serial correlation in $u_t^y - \omega_{xy}'\Omega_{xx}^{-1}u_t^x$.

Theorem 2. *Let z_t be generated by (1)–(4), suppose A1 holds and suppose $\theta = \theta_T = 1 - T^{-1}\lambda$ for some $\lambda \geqslant 0$. If $(\hat{\Sigma},\hat{\Gamma},\hat{\Omega}) \to_p (\Sigma,\Gamma,\Omega)$, then $P_T(\bar{\lambda}) \to_d \varphi(\lambda;\bar{\lambda})$.*

A consistent estimator of Σ is $\hat{\Sigma} = T^{-1}\hat{U}'\hat{U}$, while Γ and Ω can be estimated consistently by means of conventional kernel estimators (for reviews, see den Haan and Levin (1997) and Robinson and Velasco (1997)). Primitive sufficient conditions for consistency can be found in previous work by the author (Jansson, 2002, 2004a; Jansson and Haldrup, 2002). Suffice it to say that the consistency requirement of Theorem 2 is met by a variety of estimators $\hat{\Sigma},\hat{\Gamma}$ and $\hat{\Omega}$.

To implement the feasible point optimal test, a value of $\bar{\lambda}$ must be specified. Following Elliott et al. (1996), the approach advocated here is to choose $\bar{\lambda}$ in such a way that the asymptotic local power against the alternative $\theta = 1 - T^{-1}\bar{\lambda}$ is approximately equal to 50% when the 5% test based on $P_T(\bar{\lambda})$ is used. That is, the recommendation is to use the test which is (asymptotically) 0.50-optimal, level 0.05 in the sense of Davies (1969). Table 1a (1b) tabulates the recommended values of $\bar{\lambda}$ for $k = 1,\ldots,6$

Table 1
Percentiles of $P_T(\bar{\lambda})$

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$
(a) Constant mean: $d_t = 1$						
$\bar{\lambda}$	9	10.5	12.5	14	16	17.5
90%	0.71	0.81	0.80	0.83	0.87	0.89
95%	1.70	1.82	1.82	1.87	1.88	1.91
97.5%	2.71	2.77	2.81	2.91	2.87	2.97
99%	3.93	4.20	4.03	4.27	4.34	4.40
(b) Linear trend: $d_t = (1, t)'$						
$\bar{\lambda}$	13.5	15.5	16.5	18	20	21.5
90%	0.84	0.82	0.94	0.98	1.01	1.09
95%	1.88	1.95	2.01	2.03	2.14	2.27
97.5%	2.87	3.04	3.12	3.03	3.28	3.29
99%	4.09	4.52	4.39	4.45	4.76	4.67

Note: The percentiles were computed by generating 20,000 draws from the discrete time approximation (based on 2000 steps) to the limiting random variables.

regressors in the constant mean (linear trend) case and reports selected percentiles of the asymptotic null distributions of the corresponding $P_T(\bar{\lambda})$ statistics.

5. Power properties

5.1. Local asymptotic power

Fig. 1 plots the 5% asymptotic Gaussian power envelope $\pi^{0.05}(\cdot)$ along with the local asymptotic power functions of two feasible cointegration tests in the constant mean case with scalar x_t.[1] The two feasible tests, denoted P_T (9) and S_T, are the tests proposed in Section 4 and the test due to Shin (1994), respectively. The latter test appears to be the most widely used cointegration test in applications. Moreover, S_T is known to enjoy local optimality properties under the assumptions of Theorem 1 (Harris and Inder, 1994). In addition, previous research (Jansson and Haldrup, 2002; Jansson, 2004b) indicates that none of the tests due to Choi and Ahn (1995), Park (1990) and Xiao and Phillips (2002) dominate Shin's (1994) test in terms of local asymptotic power. For these reasons, it seems natural to use the performance of S_T as a benchmark when evaluating new tests such as P_T (9).

As might be expected, the local asymptotic power of S_T is close to the envelope for small values of λ (smaller than 5, say). For larger values of λ, on the other hand, the local asymptotic power of the locally optimal test is well below the envelope. In contrast, the local asymptotic power of P_T (9) is close to the envelope for all values of λ. In particular, the local asymptotic power properties of P_T (9) are similar to those

[1] The curves were generated by taking 20,000 draws from the distribution of the discrete approximation (based on 2000 steps) to the limiting random variables.

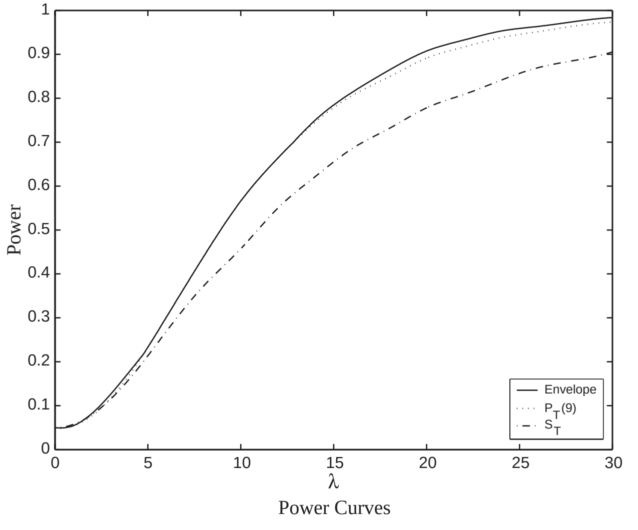

Fig. 1. 5% Level tests, constant mean, scalar x.

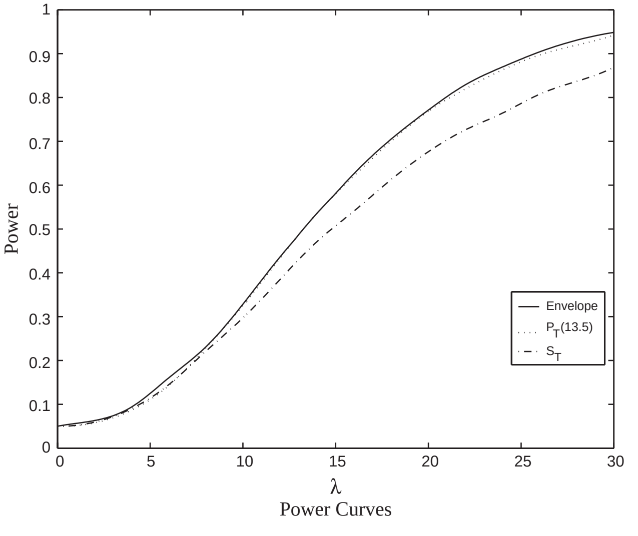

Fig. 2. 5% Level tests, linear trend, scalar x. 5% Level tests; (a) Constant mean, and (b) linear trend scalar x.

of S_T for small values of λ (when the latter is optimal) and P_T (9) dominates S_T in terms of local asymptotic power for larger values of λ. As is apparent from Fig. 2, the situation is similar in the linear trend case although the magnitude of the differences is smaller.

5.2. Finite sample evidence

To gauge the extent to which the predictions from the asymptotic power results of Section 5.1 can be expected to be borne out in sample sizes encountered in practice, a small Monte Carlo experiment is conducted. Samples of size $T = 200$ are generated according to the bivariate version of (1)–(4) with α, β and α_x normalized to zero. The errors u_t are generated by the bivariate model

$$\begin{pmatrix} u_t^y \\ u_t^x \end{pmatrix} = \psi(L) \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix} \begin{pmatrix} \varepsilon_t^y \\ \varepsilon_t^x \end{pmatrix}, \tag{8}$$

where $(\varepsilon_t^y, \varepsilon_t^x)' \sim$ i.i.d. $\mathcal{N}(0, I_2)$ and $\psi(L) = (1-a)\sum_{i=0}^{\infty} a^i L^i$, corresponding to an AR (1) model for u_t^y and u_t^x. When the errors are generated according to (8), the long-run covariance matrix of u_t is given by

$$\Omega = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

As a consequence, the parameter ρ in (8) is the correlation coefficient computed from Ω and therefore controls the endogeneity of the regressors. In the simulations, the endogeneity parameter ρ and the persistence parameter a both take on values in the set $\{0, 0.5, 0.8\}$, while the parameter of interest, θ, takes on values in the set $\{1, 0.975, 0.95, 0.925, 0.90\}$.

The matrix Σ is estimated by $\hat{\Sigma} = T^{-1}\hat{U}'\hat{U}$, while Ω and Γ are estimated using VAR(1) prewhitened kernel estimators.[2] Tables 2a (2b) reports observed rejection rates of 5% level tests using the constant mean (linear trend) versions of the test statistics $P_T(\bar{\lambda})$ and S_T. To facilitate comparisons, the $P_T(\bar{\lambda})$ and S_T test are implemented using the same estimation strategy. That is, both tests use the same estimators $\hat{\Sigma}$ and $\hat{\Gamma}$ and both tests are based on a correction in the spirit of Park (1992). The version of S_T constructed in this fashion is described in Choi and Ahn (1995), where it is denoted $SBDH_{\mathrm{II}}$.

For moderate values of the persistence parameter, the Monte Carlo evidence is consistent with the asymptotic results of Section 5.1. Specifically, when $a \in \{0, 0.5\}$ rejection rates are reasonably close to 5% under the null in and the point optimal test dominates the locally optimal test in terms of power in most cases. Indeed, the simulation results suggest that non-trivial power gains can be achieved by employing the test proposed in this paper. For $a = 0.8$, on the other hand, both tests fail to control the null rejection probability and power is disappointingly low. The size distortions of the point optimal test appear to be slightly smaller than those of the locally optimal test, but the locally optimal test is superior to the point optimal test in terms of (size-adjusted) power.

[2] These prewhitened kernel estimators, developed in Jansson (2004a), are modified versions of Andrews and Monahan's (1992) estimators.

Table 2
Monte carlo rejection rates

		P_T (9), ρ=			S_T, ρ=		
a	θ	0	0.5	0.8	0	0.5	0.8
(a) 5% Level tests, constant mean, $T=200$							
	1.000	5.1	5.7	5.8	5.0	4.6	3.8
	0.975	21.8	21.8	17.0	20.9	21.0	20.0
0	0.950	52.9	52.0	43.0	43.4	43.8	42.8
	0.925	74.0	73.2	65.5	60.0	60.5	59.5
	0.900	85.7	84.8	79.0	71.2	71.8	71.1
	1.000	3.9	4.0	5.2	4.6	5.3	6.8
	0.975	17.3	17.2	17.6	18.4	20.0	21.7
0.5	0.950	43.8	43.1	42.1	37.9	38.5	41.9
	0.925	63.0	61.8	61.2	49.8	51.2	63.0
	0.900	73.8	72.6	71.9	57.3	59.2	63.0
	1.000	1.8	3.8	8.6	4.0	6.2	11.0
	0.975	6.5	9.2	18.9	12.4	16.3	24.5
0.8	0.950	10.3	16.4	31.5	19.8	25.2	36.1
	0.925	9.5	15.5	34.6	18.2	25.0	39.5
	0.900	8.2	13.0	30.6	14.0	20.8	37.0
		P_T (13.5), ρ=			S_T, ρ=		
a	θ	0	0.5	0.8	0	0.5	0.8
(b) 5% Level tests, linear trend, $T=200$							
	1.000	5.0	6.4	5.4	5.0	5.2	2.9
	0.975	10.9	11.1	7.3	11.0	10.5	8.2
0	0.950	29.9	26.9	17.0	28.3	27.5	23.4
	0.925	52.6	48.7	33.6	47.2	45.8	42.1
	0.900	70.4	67.1	51.4	62.0	60.9	57.8
	1.000	4.0	4.3	4.1	4.3	5.0	6.7
	0.975	8.4	8.3	7.5	9.9	11.2	13.1
0.5	0.950	22.3	21.9	18.6	22.9	25.8	28.2
	0.925	39.6	38.3	33.5	36.5	39.4	42.6
	0.900	53.5	51.9	47.4	48.2	49.5	54.1
	1.000	1.6	2.3	5.1	3.5	5.3	11.7
	0.975	2.6	3.3	7.5	5.9	8.8	16.5
0.8	0.950	4.1	6.1	12.7	11.1	15.8	26.6
	0.925	4.5	6.3	16.0	12.8	18.4	33.6
	0.900	3.7	5.6	15.9	10.7	17.1	25.2

Note: Based on 5000 Monte Carlo replications.

Acknowledgements

This paper draws on material in Chapter 2 of the author's Ph.D. dissertation at University of Aarhus, Denmark. Helpful suggestions from Peter Boswijk, Bent J. Christensen, Graham Elliott, Niels Haldrup, Svend Hylleberg, Nick Kiefer, Ulrich Müller, Jim Powell, Peter Robinson, Michael Svarer, and the three referees are gratefully acknowledged. The paper has benefited from the comments of seminar participants at University of Aarhus, University of California, Berkeley, University of British Columbia, Cambridge University, Tilburg University, the 1999 Econometric Society European Meeting, the 1999 CNMLE Conference, and the 2002 NBER Summer Institute. A MATLAB program that implements the tests proposed in this paper is available at http://elsa.Berkeley.EDU/users/mjansson.

Appendix

Throughout the Appendix, $\lfloor\cdot\rfloor$ denotes the integer part of the argument and all functions are understood to be CADLAG functions defined on the unit interval (equipped with the Skorohod topology). The following Lemma, adapted from Jansson (2004a), will be used in the proof of Theorems 1 and 2 to derive limiting representations of sample moments of GLS transformed data from limiting representations of the original data. Indeed, the following relation, in which $\{F_{Tt}(\bar{\lambda})\}$ is expressed in terms of $\{F_{Tt}\}$, motivates the transformations considered in Lemma 3:

$$(F_{T1}(\bar{\lambda}),F_{T2}(\bar{\lambda}),\ldots,F_{TT}(\bar{\lambda}))' = \Psi^{-1/2}_{1-T^{-1}\bar{\lambda}}(F_{T1},F_{T2},\ldots,F_{TT})'.$$

Lemma 3. *Let* $\{F_{Tt} : 0 \leqslant t \leqslant T, T \geqslant 1\}$ *and* $\{g_{Tt} : 1 \leqslant t \leqslant T, T \geqslant 1\}$ *be triangular arrays of (vector) random variables with* $F_{T0} = 0$ *for all* T. *Let* $\bar{\lambda} > 0$ *be given and define* $F_{Tt}(\bar{\lambda}) = \Delta F_{Tt} + (1 - T^{-1}\bar{\lambda})F_{T,t-1}(\bar{\lambda})$ *and* $g_{Tt}(\bar{\lambda}) = \Delta g_{Tt} + (1 - T^{-1}\bar{\lambda})g_{T,t-1}(\bar{\lambda})$ *with initial conditions* $F_{T0}(\bar{\lambda}) = F_{T0}$ *and* $g_{T1}(\bar{\lambda}) = g_{T1}$. *If*

$$\begin{pmatrix} F_{T,\lfloor T\cdot\rfloor} \\ T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor} g_{Tt} \\ T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor} F_{Tt}g_{Tt}' \\ T^{-2}\sum_{t=2}^{\lfloor T\cdot\rfloor}\left(\sum_{i=1}^{t-1} g_{Ti}\right)g_{Tt}' \end{pmatrix} \to_{\mathrm{d}} \begin{pmatrix} F(\cdot) \\ G(\cdot) \\ \int_0^{\cdot} F(s)\,\mathrm{d}G(s)' + \Gamma_{FG}(\cdot) \\ \int_0^{\cdot} G(s)\,\mathrm{d}G(s)' + \Gamma_{GG}(\cdot) \end{pmatrix}, \qquad (9)$$

where F and G are continuous semimartingales and Γ_{FG} and Γ_{GG} are continuous, then

$$\begin{pmatrix} F_{T,\lfloor T\cdot\rfloor}(\bar{\lambda}) \\ g_{T,\lfloor T\cdot\rfloor} - g_{T,\lfloor T\cdot\rfloor}(\bar{\lambda}) \\ T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor} F_{Tt}(\bar{\lambda})g_{Tt}(\bar{\lambda})' \\ T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor}(g_{Tt} - g_{Tt}(\bar{\lambda}))g_{Tt}' \end{pmatrix} \to_{\mathrm{d}} \begin{pmatrix} F_{\bar{\lambda}}(\cdot) \\ \bar{\lambda}G_{\bar{\lambda}}(\cdot) \\ \int_0^{\cdot} F_{\bar{\lambda}}(s)\,\mathrm{d}G_{\bar{\lambda}}(s)' + \Gamma_{FG}(\cdot) \\ \bar{\lambda}\left(\int_0^{\cdot} G_{\bar{\lambda}}(s)\,\mathrm{d}G(s)' + \Gamma_{GG}(\cdot)\right) \end{pmatrix}, \quad (10)$$

where $F_{\bar{\lambda}}(s) = \bar{\lambda}\int_0^s \exp(-\bar{\lambda}(s-t))\,\mathrm{d}F(t)$ and $G_{\bar{\lambda}}(s) = \bar{\lambda}\int_0^s \exp(-\bar{\lambda}(s-t))\,\mathrm{d}G(t)$. Joint convergence in (9) and (10) also applies.

Proof of Theorems 1 and 2. Under the assumptions of Theorem 1, $\Gamma = 0$ and $\Sigma = \Omega$. Theorem 1 therefore follows from the Neyman–Pearson Lemma, Theorem 2 and the fact that the distribution of $\varphi(\lambda, \lambda)$ is continuous.

Next, to prove Theorem 2, let $\bar{\theta} = \bar{\theta}_T = 1 - T^{-1}\bar{\lambda}$ and for $\theta^* \in \{1, \bar{\theta}\}$, let $(y_1^+(\theta^*), \ldots, y_T^+(\theta^*))' = Y_{\theta^*}^+$. Let $x_t^+ = x_t - \hat{\Gamma}_{x\cdot}\hat{\Sigma}^{-1}\hat{u}_t$, where $(\hat{u}_1, \ldots, \hat{u}_T)' = \hat{U}$. Now, $y_t^+(\theta^*)$ can be written as

$$y_t^+(\theta^*) = \alpha(\theta^*)'d_t + \beta' x_t^+ + v_t^+(\theta^*),$$

where $\alpha(\theta^*)$ satisfies $\alpha(\theta^*)'d_t = (\alpha + \alpha_x'\Omega_{xx}^{-1}\omega_{xy}(1-\theta))'d_t - \theta^*\hat{\omega}'_{xy}\hat{\Omega}_{xx}^{-1}\alpha_x'\Delta d_t$ and $v_t^+(\theta^*) = v_t^{++} + \hat{v}_t^{++}(\theta^*)$, where $v_t^{++} = (1-\theta)\sum_{s=1}^t u_s^{y\cdot x} + \theta u_t^{y\cdot x}$, $u_t^{y\cdot x} = u_t^y - \omega_{xy}'\Omega_{xx}^{-1}u_t^x$ and $\hat{v}_t^{++}(\theta^*) = -(\theta^*\hat{\Omega}_{xx}^{-1}\hat{\omega}_{xy} - \theta\Omega_{xx}^{-1}\omega_{xy})'u_t^x - (\hat{\beta} - \beta - (1-\theta)\Omega_{xx}^{-1}\omega_{xy})\hat{u}_t$. Moreover, $x_t^+ = x_t^{++} + \hat{x}_t^{++}$, where $x_t^{++} = x_t - \Gamma_{x\cdot}\Sigma^{-1}u_t$ and $\hat{x}_t^{++} = (\Gamma_{x\cdot}\Sigma^{-1} - \hat{\Gamma}_{x\cdot}\hat{\Sigma}^{-1})\,u_t + \hat{\Gamma}_{x\cdot}\hat{\Sigma}^{-1}(u_t - \hat{u}_t)$. Similarly, $r_t^+ = (d_t', x_t^{+'})' = r_t^{++} + \hat{r}_t^{++}$, where $r_t^{++} = (d_t', x_t^{++'})$ and $\hat{r}_t^{++} = (0, \hat{x}_t^{++'})$.

By proceeding as in the proof of Jansson and Haldrup (2002,Lemma 6), standard weak convergence results for linear processes (e.g., Phillips and Solo, 1992; Phillips, 1988; Hansen, 1992) can be used to show that the following hold jointly:

$$T^{1/2}\Upsilon_T r_{\lfloor T\cdot\rfloor}^+ = T^{1/2}\Upsilon_T r_{\lfloor T\cdot\rfloor}^{++} + o_p(1) \to_{\mathrm{d}} Q(\cdot) \quad (11)$$

$$T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor} v_t^+(\theta^*) = T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor} v_t^{++} + o_p(1) \to_{\mathrm{d}} \omega_{yy\cdot x}^{1/2}V^\lambda(\cdot), \quad (12)$$

$$\Upsilon_T\sum_{t=1}^{\lfloor T\cdot\rfloor} r_t^+ v_t^+(\theta^*) = \Upsilon_T\sum_{t=1}^{\lfloor T\cdot\rfloor} r_t^{++}v_t^{++} + o_p(1) \to_{\mathrm{d}} \omega_{yy\cdot x}^{1/2}\int_0^{\cdot} Q(s)\,\mathrm{d}V^\lambda(s), \quad (13)$$

$$T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor}\left(\sum_{s=1}^{t-1} v_s^+(\theta^*)\right)v_t^+(\theta^*) = T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor}\left(\sum_{s=1}^{t-1} v_s^{++}\right)v_t^{++} + o_p(1)$$
$$\to_{\mathrm{d}} \omega_{yy\cdot x}\int_0^{\cdot} V^\lambda(s)\,\mathrm{d}V^\lambda(s) + \gamma_{yy\cdot x}\int_0^{\cdot}\mathrm{d}s, \quad (14)$$

for $\theta^* \in \{1, \bar{\theta}\}$, where

$$\Upsilon_T = \begin{pmatrix} diag(T^{-1/2}, \ldots, T^{-(p+1/2)}) & 0' \\ -T^{-1}\Omega_{xx}^{-1/2}\alpha_x' & T^{-1}\Omega_{xx}^{-1/2} \end{pmatrix}.$$

Since $\mathscr{L}_T^+(\theta^*)$ is invariant under transformations of the form $Y_{\theta^*}^+ \to Y_{\theta^*}^+ + D \cdot a + X^+ \cdot b$, $P_T(\bar{\lambda})$ can be written as $P_T^1(\bar{\lambda}) + P_T^2(\bar{\lambda}) + P_T^3(\bar{\lambda})$, where $P_T^1(\bar{\lambda}) = \log|R^{+'}R^+| - \log|R^{+'}\Psi_{\bar{\theta}}^{-1}R^+|$, $P_T^2(\bar{\lambda}) = \hat{\omega}_{yy\cdot x}^{-1}(V_1^{+'}V_1^+ - V_{\bar{\theta}}^{+'}\Psi_{\bar{\theta}}^{-1}V_{\bar{\theta}}^+ - 2\bar{\lambda}\hat{\gamma}_{yy\cdot x})$,

$$P_T^3(\bar{\lambda}) = \hat{\omega}_{yy\cdot x}^{-1}V_{\bar{\theta}}^{+'}\Psi_{\bar{\theta}}^{-1}R^+(R^{+'}\Psi_{\bar{\theta}}^{-1}R^+)^{-1}R^{+'}\Psi_{\bar{\theta}}^{-1}V_{\bar{\theta}}^+ - \hat{\omega}_{yy\cdot x}^{-1}V_1^{+'}R^+(R^{+'}R^+)^{-1}R^{+'}V_1^+,$$

and $V_{\theta^*}^+ = (v_1^+(\theta^*), \ldots, v_T^+(\theta^*))'$ for $\theta^* \in \{1, \bar{\theta}\}$.

Using (11), Lemma 3, the continuous mapping theorem (CMT) and standard manipulations,

$$P_T^1(\bar{\lambda}) = \log|\Upsilon_T R^{+'}R^+\Upsilon_T'| - \log|\Upsilon_T R^{+'}\Psi_{\bar{\theta}}^{-1}R^+\Upsilon_T'| \to_{\mathrm{d}} \log\left|\int_0^1 QQ'\right| - \log\left|\int_0^1 Q_{\bar{\lambda}}Q_{\bar{\lambda}}'\right|.$$

Similarly, using (12) and (14), $(\hat{\omega}_{yy\cdot x}, \hat{\gamma}_{yy\cdot x}) \to_{\mathrm{p}} (\omega_{yy\cdot x}, \gamma_{yy\cdot x})$, Lemma 3 and CMT,

$$\begin{aligned} P_T^2(\bar{\lambda}) &= 2\hat{\omega}_{yy\cdot x}^{-1}((V_1^+ - \Psi_{\bar{\theta}}^{-1/2}V_{\bar{\theta}}^+)'V_1^+ - \bar{\lambda}\hat{\gamma}_{yy\cdot x}) \\ &\quad -\hat{\omega}_{yy\cdot x}^{-1}(V_1^+ - \Psi_{\bar{\theta}}^{-1/2}V_{\bar{\theta}}^+)'(V_1^+ - \Psi_{\bar{\theta}}^{-1/2}V_{\bar{\theta}}^+) \\ &= 2\omega_{yy\cdot x}^{-1}((V^{++} - \Psi_{\bar{\theta}}^{-1/2}V^{++})'V^{++} - \bar{\lambda}\gamma_{yy\cdot x}) \\ &\quad -\omega_{yy\cdot x}^{-1}(V^{++} - \Psi_{\bar{\theta}}^{-1/2}V^{++})'(V^{++} - \Psi_{\bar{\theta}}^{-1/2}V^{++}) + o_p(1) \\ &\to_{\mathrm{d}} 2\bar{\lambda}\int_0^1 V_{\bar{\lambda}}^{\lambda}\,\mathrm{d}V^{\lambda} - \bar{\lambda}^2\int_0^1 (V_{\bar{\lambda}}^{\lambda})^2, \end{aligned}$$

where $V^{++} = (v_1^{++}, \ldots, v_T^{++})'$.

Finally, using (11) and (13), $\hat{\omega}_{yy\cdot x} \to_{\mathrm{p}} \omega_{yy\cdot x}$, Lemma 3 and CMT,

$$\begin{aligned} P_T^3(\bar{\lambda}) \to_{\mathrm{d}} &\left(\int_0^1 Q_{\bar{\lambda}}\,\mathrm{d}V_{\bar{\lambda}}^{\lambda}\right)'\left(\int_0^1 Q_{\bar{\lambda}}Q_{\bar{\lambda}}'\right)^{-1}\left(\int_0^1 Q_{\bar{\lambda}}\,\mathrm{d}V_{\bar{\lambda}}^{\lambda}\right) \\ &-\left(\int_0^1 Q\,\mathrm{d}V^{\lambda}\right)'\left(\int_0^1 QQ'\right)^{-1}\left(\int_0^1 Q\,\mathrm{d}V^{\lambda}\right). \end{aligned}$$

The convergence results in the preceding displays hold jointly. Combining these results, Theorem 2 follows. □

References

Andrews, D.W.K., Monahan, J.C., 1992. An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. Econometrica 60, 953–966.

Choi, I., Ahn, B.C., 1995. Testing for cointegration in a system of equations. Econometric Theory 11, 952–983.

Davies, R.B., 1969. Beta-optimal tests and an application to the summary evaluation of experiments. Journal of the Royal Statistical Society B 31, 524–538.

den Haan, W.J., Levin, A.T., 1997. A practitioner's guide to robust covariance matrix estimation. In: Maddala, G.S., Rao, C.R. (Eds.), Handbook of Statistics, Vol. 15. North-Holland, New York, pp. 299–342.

Dufour, J.-M., King, M.L., 1991. Optimal invariant tests for the autocorrelation coefficient in linear regressions with stationary or nonstationary AR(1) errors. Journal of Econometrics 47, 115–143.

Elliott, G., Rothenberg, T.J., Stock, J.H., 1996. Efficient tests for an autoregressive unit root. Econometrica 64, 813–836.

Engle, R.F., Granger, C.W.J., 1987. Cointegration and error correction: representation, estimation, and testing. Econometrica 55, 251–276.

Hansen, B.E., 1992. Convergence to stochastic integrals for dependent heterogeneous processes. Econometric Theory 8, 489–500.

Harris, D., Inder, B., 1994. A test of the null hypothesis of cointegration. In: Hargreaves, C. (Ed.), Nonstationary Time Series Analysis and Cointegration. Oxford University Press, Oxford, pp. 133–152.

Jansson, M., 2002. Consistent covariance matrix estimation for linear processes. Econometric Theory 18, 1449–1459.

Jansson, M., 2004a. Stationarity testing with covariates. Econometric Theory 20, 56–94.

Jansson, M., 2004b. Tests of the null hypothesis of cointegration based on efficient tests for a unit MA root. In: Andrews, D.W.K., Stock, J.H. (Eds.), Identification and Inference in Econometric Models: Essays in Honor of Thomas J. Rothenberg. Cambridge University Press, Cambridge, forthcoming.

Jansson, M., Haldrup, N., 2002. Regression theory for nearly cointegrated time series. Econometric Theory 18, 1309–1335.

King, M.L., 1980. Robust tests for spherical symmetry and their application to least squares regression. Annals of Statistics 8, 1265–1271.

King, M.L., Hillier, G.H., 1985. Locally best invariant tests of the error covariance matrix of the linear regression model. Journal of the Royal Statistical Society B 47, 98–102.

Lehmann, E.L., 1994. Testing Statistical Hypotheses, 2nd Edition. Chapman & Hall, New York.

Park, J.Y., 1990. Testing for unit roots and cointegration by variable addition. In: Fomby, T.B., Rhodes, G.F. (Eds.), Advances in Econometrics, Vol. 8. JAI Press, Greenwich, CT, pp. 107–133.

Park, J.Y., 1992. Canonical cointegrating regressions. Econometrica 60, 119–143.

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

Phillips, P.C.B., 1988. Weak convergence of sample covariance matrices to stochastic integrals via martingale approximations. Econometric Theory 4, 528–533.

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

Phillips, P.C.B., Solo, V., 1992. Asymptotics for linear processes. Annals of Statistics 20, 971–1001.

Robinson, P.M., Velasco, C., 1997. Autocorrelation-robust inference. In: Maddala, G.S., Rao, C.R. (Eds.), Handbook of Statistics, Vol. 15. North-Holland, New York, pp. 267–298.

Saikkonen, P., Luukkonen, R., 1993. Point optimal tests for testing the order of differencing in ARIMA models. Econometric Theory 9, 343–362.

Shin, Y., 1994. A residual-based test of the null of cointegration against the alternative of no cointegration. Econometric Theory 10, 91–115.

Stock, J.H., 1994. Unit roots, structural breaks and trends. In: Engle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics, Vol. 4. North-Holland, New York, pp. 2739–2841.

Watson, M.W., 1994. Vector autoregressions and cointegration. In: Engle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics, Vol. 4. North-Holland, New York, pp. 2843–2915.

Xiao, Z., Phillips, P.C.B., 2002. A CUSUM test for cointegration using regression residuals. Journal of Econometrics 108, 43–61.