Discussion Paper Series

No.190

Test for the null hypothesis of cointegration with reduced size distortion

Eiji Kurozumi
Yoichi Arai

November 2006

Hitotsubashi University Research Unit for Statistical Analysis in Social Sciences
A 21st-Century COE Program

Institute of Economic Research
Hitotsubashi University
Kunitachi, Tokyo, 186-8603 Japan
http://hi-stat.ier.hit-u.ac.jp/

Test for the null hypothesis of cointegration with reduced size distortion

Eiji Kurozumi[1] — Department of Economics, Hitotsubashi University

Yoichi Arai[2] — Faculty of Economics, University of Tokyo

September, 2006

Abstract

This paper considers a single equation cointegrating model and proposes the locally best invariant and unbiased (LBIU) test for the null hypothesis of cointegration. We derive the asymptotic local power functions and compare them with the standard residual-based test, and we show that the LBIU test is more powerful in a wide range of local alternatives. Then, we conduct a Monte Carlo simulation to investigate the finite sample properties of the tests and show that the LBIU test outperforms the residual-based test in terms of both size and power. The advantage of the LBIU test is particularly patent when the error is highly autocorrelated. Further, we point out that finite sample performance of existing tests is largely affected by the initial value condition while our tests are immune to it. We propose a simple transformation of data that resolves the problem in the existing tests.

JEL classification: C12; C22

Key Words: Cointegration; locally best test; point optimal test

---

[1]The earlier version of this paper was written while Eiji Kurozumi was a visiting scholar of Boston University. Kurozumi's research was partially supported by the Ministry of Education, Culture, Sports, Science and Technology under Grants-in-Aid No. 17203016 and 18730142. Correspondence: Eiji Kurozumi, Department of Economics, Hitotsubashi University, Naka 2-1, Kunitachi, Tokyo 186-8601, Japan. E-mail: kurozumi@stat.hit-u.ac.jp

[2]Arai's research was partially supported by Grant-In-Aid for Scientific Research (No. 17730144) from JSPS and MEXT of the Japanese government. We are grateful to Michael Jansson, Yoshihiko Nishiyama, and Katsuto Tanaka for their helpful comments.

1. Introduction

Following the seminal work of Engle and Granger (1987), tests of cointegration have been intensively investigated in the econometric literature. For a single equation model, tests for the null of cointegration are proposed by Hansen (1992a), Quintos and Phillips (1993), Shin (1994), and Jansson (2005), while the null of no cointegration is considered in Engle and Granger (1987) and Phillips and Ouliaris (1990), among others. A system equations approach is also considered in a number of studies.[3]

For the null hypothesis of cointegration, Shin (1994) proposes the residual-based test, while Jansson (2005) develops the point optimal invariant (POI) test. Jansson (2005) shows that the POI test performs better than the residual-based test in a wide range of alternatives based on the local asymptotic power functions. A Monte Carlo experiment conducted to examine the finite sample properties of the test developed by Jansson (2005) demonstrates that the POI test is more powerful than the residual-based test when the error is not persistent; at the same time, it reveals several important drawbacks of the tests. First of all, the POI test suffers from size distortions and power losses when the error is persistent. With respect to the size properties, the test is undersized when the endogeneity of the regressor is low and oversized when it is high. With regard to the power properties, the POI test is outperformed by the residual-based test proposed by Shin (1994). Second, the residual-based test also suffers from the same type of size distortions as the POI test.

In this paper, we consider a single equation cointegrating model and propose the locally best invariant and unbiased (LBIU) test with correct size. In order to do so, we first develop the point optimal test that is invariant to some location-scale transformation of the data under simple assumptions on the error. The transformation deals with directions of transformations that are wider than those in Jansson (2005). Next, we derive the LBIU test based on the POI test. Finally, we generalize the tests to accommodate general assumptions on the error. After we present the test statistics, we study their asymptotic power properties. Comparing the asymptotic local power function of the LBIU test with that of the residual-

[3]See Hubrich, Lütkepohl, and Saikkonen (2001) for an excellent review of system equations methods.

based test, we show that the LBIU test is more powerful in a wide range of local alternatives and that the power properties of the two tests against the hypothesis that is very close to the null are indistinguishable.

To investigate the finite sample properties of our test, we conduct a Monte Carlo experiment. We find that the empirical size of the LBIU test is very close to the nominal size regardless of the degree of persistence in the error and the endogeneity of the regressor. In addition, the LBIU test is generally more powerful than the residual-based test while the POI tests are more powerful than the LBIU and the residual-based test when the error is not persistent. The advantage of the LBIU test over the residual-based test and the POI tests is particularly patent when the error is highly autocorrelated. Based on these facts, the LBIU test becomes a strong candidate for researchers who are perplexed with regard to a size versus power trade-off.

The other important finding in this paper is that Jansson's POI test and the residual-based test are greatly affected by the initial value condition on the stochastic regressors, while our POI and LBIU tests are shown to be free of the initial value condition. We propose a simple transformation of the data that resolves the problem in Jansson's POI and the residual-based tests. Finite sample simulations show that Jansson and Shin's tests suffer from severe size distortions without the transformation.

The remainder of the paper is organized as follows. In Section 2, we derive the POI and LBIU tests for a stylized model and obtain the limiting local power functions. Section 3 generalizes the assumptions by allowing the error term to be weakly dependent; we modify the test statistics such that their limiting distributions are independent of nuisance parameters. We investigate the finite sample properties of our tests through a Monte Carlo simulation in Section 4. Section 5 concludes the paper.

2. LBIU and POI tests

In this section, we first develop a POI test and then derive an LBIU test based on the POI test. Let us consider the following model:

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

$$x_t = \alpha_x' d_t + x_t^0, \quad (1-L)x_t^0 = u_t^x, \tag{2}$$

where $d_t = [1, \cdots, t^p]'$ with $p \geq 0$, y_t and x_t are 1- and k-dimensional observations, L is the lag operator, and $v_0 = u_0^y = 0$. For the error process, we consider the following assumption in this section.

Assumption 1 $u_t = [u_t^y, u_t^{x\prime}]' \sim i.i.d.N(0, \Sigma)$ *with* $\Sigma > 0$.

We divide Σ conformably with u_t as follows:

$$\Sigma = \begin{bmatrix} \sigma_{yy} & \sigma_{yx} \\ \sigma_{xy} & \Sigma_{xx} \end{bmatrix}.$$

We proceed with this restricted assumption in this section; however, we will relax the assumption of normality and consider the dependent case in the next section.

The model is expressed in the vectorized form as

$$y = D\alpha + X\beta + v, \quad L_1 v = L_\theta u^y,$$

$$X = D\alpha_x + \Psi_0^{1/2} U^x,$$

where $y = [y_1, \cdots, y_T]'$, $D = [d_1, \cdots, d_T]'$, and the other vectors and matrices are defined similarly, $\Psi_\theta = \Psi_\theta^{1/2} \Psi_\theta^{1/2\prime}$ with

$$\Psi_\theta^{1/2} = \begin{bmatrix} 1 & & & 0 \\ 1-\theta & 1 & & \\ \vdots & \ddots & \ddots & \\ 1-\theta & \cdots & 1-\theta & 1 \end{bmatrix} \quad \text{and} \quad L_\theta = \begin{bmatrix} 1 & & & 0 \\ -\theta & 1 & & \\ & \ddots & \ddots & \\ 0 & & -\theta & 1 \end{bmatrix}.$$

Since $L_1^{-1} L_\theta = \Psi_0^{1/2} L_\theta = \Psi_\theta^{1/2}$ because $L_1^{-1} = \Psi_0^{1/2}$, the above system can also be expressed as

$$y = D\alpha + X\beta + \Psi_\theta^{1/2} u^y,$$

$$\Psi_0^{-1/2} X = \Psi_0^{-1/2} D\alpha_x + U^x. \tag{3}$$

Note that the first column of $\Psi_0^{-1/2}D$ comprises $e_1 = [1, 0, \cdots, 0]'$, while the other columns are obtained by a nonsingular transformation of the first p columns of D, which corresponds to $[1, \cdots, t^{p-1}]$.

Let us suppose that we are interested in the following testing problem:

$$H_0 : \theta = 1 \quad \text{v.s.} \quad H_1 : \theta < 1.$$

Under the null hypothesis, $v_t = u_t^y$ and subsequently y_t and x_t are cointegrated; however, they are not cointegrated under the alternative because v_t is a unit root process when $\theta \neq 1$.

Based on the observation that x_t is weakly exogenous for θ, it is sufficient for us to consider the distribution of y conditional on X as far as the hypothesis regarding θ is concerned. It is evident that the conditional distribution $y|X$ is given by $N(D\alpha + X\beta + \Psi_\theta^{1/2}U^x\Sigma_{xx}^{-1}\sigma_{xy}, \sigma_{yy\cdot x}\Psi_\theta)$, where $\sigma_{yy\cdot x} = \sigma_{yy} - \sigma_{yx}\Sigma_{xx}^{-1}\sigma_{xy}$. Using (3), the conditional distribution is also expressed as

$$y|X \sim N\left(D\alpha^* + X\beta^* + \Psi_0^{-1/2}X\gamma^* + e_1\delta^*, \sigma_{yy\cdot x}\Psi_\theta\right), \tag{4}$$

where α^*, β^*, γ^*, and δ^* are defined appropriately, and the relation $\Psi_\theta^{1/2}\Psi_0^{-1/2} = L_\theta = \theta\Psi_0^{-1/2} + (1-\theta)I_T$ is employed. It is then observed that the testing problem is invariant under the group of transformations

$$\begin{gathered} y \to sy + Da + Xb + \Psi_0^{-1/2}Xc + e_1 d \\ (\theta, \alpha^*, \beta^*, \gamma^*, \delta^*, \sigma_{yy\cdot x}) \to (\theta, s\alpha^* + a, s\beta^* + b, s\gamma^* + c, s\delta^* + d, s^2\sigma_{yy\cdot x}), \end{gathered} \tag{$\mathcal{G}_y$}$$

where a is a $p+1$-dimensional vector, b and c are k-dimensional vectors, and d and s are scalar with $0 < a < \infty$. Note that in a classical regression context, a location shift in y is considered only in the directions of the regressors D and X, while we additionally consider the directions of $\Psi_0^{-1/2}X$ and e_1. It is noteworthy that in our model, the I(1) regressors X are correlated with the error term, u^y; thus, the conditional mean of y depends on $\Psi_0^{-1/2}X\gamma^*$ and $e_1\delta^*$ in addition to $D\alpha^*$ and $X\beta^*$, as is observed in (4). Since it is natural to consider a location shift in y in the directions of the conditional mean, $[D, X, \Psi_0^{-1/2}X, e_1]$ provides the appropriate directions of the shift in y in our case. We can also see that invariance in the directions of e_1 implies that the tests do not depend on the initial value condition.

Let us define $M = I - Z(Z'Z)^{-1}Z'$, where $Z = [D, X, \Psi_0^{-1/2}X, e_1]$, and select a $T \times (T-q)$ matrix H such that $H'H = I_{T-q}$ and $HH' = M$, where $q = 2k + p + 2$. As $H'Z = 0$, we have

$$H'y|X \sim N(0, \sigma_{yy\cdot x} H'\Psi_\theta H).$$

Then, the distribution of $H'y|X$ is observed to be free of the nuisance parameters α^*, β^*, γ^*, and δ^*. In addition, it is shown that $\eta = H'y/\sqrt{y'HH'y}$ conditional on X is a maximal invariant under the group of transformations $(\mathcal{G}_y)$. In this section, we assume that $\sigma_{yy\cdot x} = 1$ without loss of generality because $\eta|X$ is invariant to scale change in y. The probability density function of $\eta|X$ is given by (see Kariya, 1980 and King, 1980)

$$f(\eta|X;\theta) = \frac{1}{2}\Gamma\left(\frac{T-q}{2}\right)\pi^{-(T-q)/2}|H'\Psi_\theta H|^{-1/2}\left(\eta'(H'\Psi_\theta H)^{-1}\eta\right)^{-(T-q)/2}. \tag{5}$$

Given the density of the maximal invariant under the group of transformations $(\mathcal{G}_y)$, we can now propose the test statistics. First, we develop the POI test. According to the Neyman–Pearson lemma, the POI test against $\theta = \bar{\theta}$ is given by $f(\eta|X;\bar{\theta})/f(\eta|X;1)$, which is normalized as follows in order to have a limiting distribution:

$$\begin{aligned} \mathcal{R}_T(\bar{\theta}) &= T\left\{1 - \left(\frac{f(\eta|X;\bar{\theta})}{f(\eta|X;1)}\right)^{-2/(T-q)}\right\} \\ &= T\left\{1 - \left(\frac{|Z'\Psi_{\bar{\theta}}^{-1}Z|}{|Z'Z|}\right)^{1/(T-q)} \frac{y'(\Psi_{\bar{\theta}}^{-1} - \Psi_{\bar{\theta}}^{-1}Z(Z'\Psi_{\bar{\theta}}^{-1}Z)^{-1}Z'\Psi_{\bar{\theta}}^{-1})y}{y'My}\right\}. \end{aligned}$$

The null hypothesis is rejected when $\mathcal{R}_T(\bar{\theta})$ takes large values. Note that $\mathcal{R}_T(\bar{\theta})$ has an expression that is different from Jansson's POI test statistic, which is constructed by considering only location invariance. One of the reasons for the difference between the two test statistics is the directions of the location shift: Jansson (2005) considers location invariance in the directions of $R = [D, X]$, while we introduced invariance in the directions of $[\Psi_0^{-1/2}X, e_1]$ in addition to R. The other reason for the difference lies in the introduction of scale change, which leads to a distributional difference between the two maximal invariants: the maximal invariant η in our analysis has a nonnormal distribution, as given by (5), while the maximal invariant with only location invariance has a normal density, as shown in Jansson (2005).

To investigate the asymptotic properties of the POI test, we localize the parameters θ and $\bar{\theta}$ such that $\theta = 1 - \lambda/T$ and $\bar{\theta} = 1 - \bar{\lambda}/T$. Then, the limiting distribution of $\mathcal{R}_T(\bar{\theta})$ is given in the following theorem.[4]

Theorem 1 *Under Assumption 1, the limiting distribution of $\mathcal{R}_T(\bar{\theta})$ is given by*

$$\begin{aligned}
\mathcal{R}_T(\bar{\theta}) \;\; \Rightarrow \;\; & 2\bar{\lambda}\int_0^1 V_\lambda^{\bar{\lambda}} dV_\lambda - \bar{\lambda}^2 \int_0^1 (V_\lambda^{\bar{\lambda}})^2 ds \\
& + \left(\int_0^1 Q^{\bar{\lambda}} dV_\lambda^{\bar{\lambda}}\right)' \left(\int_0^1 Q^{\bar{\lambda}} Q^{\bar{\lambda}\prime} ds\right)^{-1} \left(\int_0^1 Q^{\bar{\lambda}} dV_\lambda^{\bar{\lambda}}\right) \\
& - \left(\int_0^1 Q dV_\lambda\right)' \left(\int_0^1 QQ' ds\right)^{-1} \left(\int_0^1 Q dV_\lambda\right) - \log\left|\int_0^1 Q^{\bar{\lambda}} Q^{\bar{\lambda}\prime} ds\right| + \log\left|\int_0^1 QQ' ds\right|,
\end{aligned}$$

where $\Rightarrow$ signifies weak convergence of the associated probability measures, $Q(s) = [1, s, \cdots, s^p, W(s)']'$ with $W(s)$ being a k-dimensional standard Brownian motion, $Q^{\bar{\lambda}}(s) = \int_0^s \exp(-\bar{\lambda}(s-r))dQ(r)$, $V_\lambda(s) = V(s) + \lambda \int_0^s V(r)dr$ with $V(s)$ being a univariate standard Brownian motion that is independent of $W(s)$, and $V_\lambda^{\bar{\lambda}}(s) = \int_0^s \exp(-\bar{\lambda}(s-r))dV_\lambda(r)$.

Remark 1: Although our test statistic $\mathcal{R}_T(\bar{\theta})$ is different from Jansson's $\mathcal{P}_T(\bar{\theta})$, the limiting distribution of $\mathcal{R}_T(\bar{\theta})$ is the same as that of $\mathcal{P}_T(\bar{\theta})$. This is because the additional deterministic and I(0) regressors –e_1 and $\Psi_0^{-1/2}X$, respectively– do not contribute to the asymptotic local distribution, as is shown in the proof of the theorem provided in the Appendix. Our result implies that we can impose scale invariance in addition to location invariance in wider directions without sacrificing local asymptotic power. However, in Section 4, we will see that these additional regressors, particularly e_1, play an important role in finite samples.

In practice, we specify the value of $\bar{\theta}$ or $\bar{\lambda}$ in order to implement our feasible point optimal test. We follow Elliott et al. (1996) and Jansson (2005) for the selection of $\bar{\lambda}$. According to

[4] In Theorem 1, an integral such as $\int_0^1 X(s)dY(s)'$ is simply written as $\int XdY'$ to achieve notational economy.

their approach, $\bar{\lambda}$ should be selected such that the asymptotic local power against the local alternative $\bar{\theta} = 1 - \bar{\lambda}/T$ is approximately 50% when we use the 5% test based on $\mathcal{R}_T(\bar{\theta})$. The recommended values of $\bar{\lambda}$ and the percentiles of $\mathcal{R}_T(\bar{\lambda})$ are given by Table 1 in Jansson (2005).

Next, we consider a locally best test that is also a natural candidate when no uniformly most powerful tests are available as in the present situation. This can be considered as the extreme case of the POI test with $\bar{\theta} \to 1$. According to Ferguson (1967), the locally best invariant (LBI) test is given by $d \log f(\eta|X;\theta)/d\theta|_{\theta=1}$, but in the Appendix, it is shown that $d \log f(\eta|X;\theta)/d\theta|_{\theta=1} = 0$. Then, instead of the LBI test, we consider the LBI and unbiased (i.e., LBIU) test, that rejects the null when the following holds.

$$\left.\frac{d^2 \log f(\eta|X;\theta)}{d\theta^2}\right|_{\theta=1} + \left(\left.\frac{d \log f(\eta|X;\theta)}{d\theta}\right|_{\theta=1}\right)^2 > c_1 + c_2 \left.\frac{d \log f(\eta|X;\theta)}{d\theta}\right|_{\theta=1},$$

where c_1 and c_2 are some constants. See Ferguson (1967) for detailed discussions on the LBIU test. The Appendix shows that the LBIU test statistic is given by

$$\mathcal{L}_T = \frac{y'M\Psi_0 My/T^2}{y'My/(T-q)} + \frac{1}{T^2}\mathrm{tr}\left\{(Z'Z)^{-1}(Z'\Psi_0 Z)\right\}. \tag{6}$$

The null hypothesis is rejected when $\mathcal{L}_T$ takes large values.

Theorem 2 *Under Assumption 1, the limiting distribution of $\mathcal{L}_T$ is given by*

$$\begin{aligned}\mathcal{L}_T \quad \Rightarrow \quad & \int_0^1 \left\{V_\lambda - \int_0^s Q'dr \left(\int_0^1 QQ'dr\right)^{-1} \int_0^1 QdV_\lambda\right\}^2 ds \\ & + tr\left\{\left(\int_0^1 QQ'dr\right)^{-1} \int_0^1 \left(\int_s^1 Qdr\right)\left(\int_s^1 Q'dr\right) ds\right\}.\end{aligned}$$

The percentiles of $\mathcal{L}_T$ are given in Table 1. Figure 1 depicts the Gaussian power envelope of the 5% test based on $\mathcal{R}_T(\theta)$ along with the local asymptotic power functions of four cointegration tests in the constant mean case with $k = 1$.[5] The two tests are the feasible tests proposed in this paper and are denoted by $\mathcal{R}_T$ and $\mathcal{L}_T$. The other two tests are the

[5]The curves are obtained from 20,000 replications from the distribution of the discrete approximation based on 2,000 steps to the limiting distribution given in Theorem 1.

residual-based test proposed by Shin (1994) and the POI test developed by Jansson (2005) and are denoted by $\mathcal{S}_T$ and $\mathcal{P}_T$, respectively. Since the local asymptotic power functions of $\mathcal{P}_T$ and $\mathcal{R}_T$ are found to be the same, only one line is indicated in Figure 1. $\mathcal{S}_T$ is the most commonly used test in applications and is locally optimal under Shin's assumptions. Therefore, it becomes a convenient benchmark for assessing our new tests, $\mathcal{R}_T$ and $\mathcal{L}_T$.

The local asymptotic powers of $\mathcal{P}_T$ and $\mathcal{R}_T$ are close to the envelope for all the values of λ. The local asymptotic powers of $\mathcal{S}_T$ and $\mathcal{L}_T$ are close to the envelope for small values of λ due to their local optimal properties, and they are below the envelope for large values of λ. The asymptotic power of $\mathcal{L}_T$ is closer to the envelope than that of $\mathcal{S}_T$ for large values of λ. Figure 2 shows the case with a linear trend case. Our observations with respect to the constant mean case is also true for this case, although the magnitude of the differences is diminished.

3. Extension to general cases

The POI and LBIU tests derived in the previous section are based on the assumption that the error process is normal and serially independent. However, this assumption is too restrictive in practice, and therefore, we consider more general assumptions where the error term is weakly dependent. The purpose of this section is to construct test statistics having the same local asymptotic properties as those given in Theorems 1 and 2 under general assumptions.

To construct the feasible test statistics, we define the long-run variance of u_t and its one-sided version as

$$\Omega = \Sigma + \Pi + \Pi' \quad \text{and} \quad \Gamma = \Sigma + \Pi,$$

$$\text{where} \quad \Sigma = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} E[u_t u_t'] \quad \text{and} \quad \Pi = \lim_{T\to\infty} T^{-1} \sum_{j=1}^{T-1} \sum_{t=1}^{T-j} E[u_t u_{t+j}'].$$

We divide these matrices conformably with u_t, as in the previous section. We also define the last k rows of Γ as Γ_x; in other words, $\Gamma_x = [0, I_k]\Gamma$.

Assumption 2 *(a)* $\{u_t\}$ *is mean zero and strong mixing with mixing coefficients of size* $-p\alpha/(p-\alpha)$ *and* $E|u_t|^p < \infty$ *for some* $p > \alpha > 5/2$.

(b) The matrix Ω exists with finite elements, $\Omega > 0$, $\omega_{yy} > 0$, and $\Omega_{xx} > 0$.

Assumption 2 ensures that the functional central limit theorem can be applied to the partial sums of u_t.

Let $u_t^* = [u_t^{y \cdot x}, u_t^{x\prime}]$, where $u_t^{y \cdot x} = \kappa' u_t = u_t^y - \omega_{yx}\Omega_{xx}^{-1}u_t^x$ with $\kappa' = [1, -\omega_{yx}\Omega_{xx}^{-1}]$, and let $\hat{u}_t^* = [\hat{u}_t^{y \cdot x}, \hat{u}_t^{x\prime}]'$, where $\hat{u}_t^{y \cdot x}$ and $\hat{u}_t^x$ are the regression residuals of y_t on z_t and x_t on d_t, respectively. We define Ω^*, Σ^*, Π^*, and Γ^* from u_t^* analogously to Ω, Σ, Π, and Γ, respectively, which are defined from u_t, and divide them conformably with u_t^* such that ω_{11}^*, ω_{12}^*, and Ω_{22}^* are $(1,1)$, $(1,2)$, and $(2,2)$ blocks of Ω^*, respectively, and Γ_x^* is the last k rows of Γ^*. Let $\hat{\omega}_{11}^*$, $\hat{\Sigma}^*$, $\hat{\pi}_{11}^*$, and $\hat{\Gamma}_x^*$ be consistent estimators of ω_{11}^*, Σ^*, π_{11}^*, and Γ_x^* based on $\hat{u}_t^*$, which can be obtained by the typical kernel estimators as investigated in Andrews (1991). The proposed test statistics are

$$\begin{aligned}
\mathcal{R}_T^+(\bar{\theta}) &= \hat{\omega}_{11}^{*-1}\left\{y'M^+y - y'(\Psi_{\bar{\theta}}^{-1} - \Psi_{\bar{\theta}}^{-1}Z^+(Z^{+\prime}\Psi_{\bar{\theta}}^{-1}Z^+)^{-1}Z^{+\prime}\Psi_{\bar{\theta}}^{-1})y - 2\bar{\lambda}\hat{\pi}_{11}^*\right\} \\
&\quad - \log\left|Z^{+\prime}\Psi_{\bar{\theta}}^{-1}Z^+\right| + \log\left|Z^{+\prime}Z^+\right|, \\
\mathcal{L}_T^+ &= \frac{1}{T^2}\hat{\omega}_{11}^{*-1}y'M^+\Psi_0M^+y + \frac{1}{T^2}\mathrm{tr}\left\{(Z^{+\prime}Z^+)^{-1}(Z^{+\prime}\Psi_0Z^+)\right\},
\end{aligned}$$

where $M^+ = I_T - Z^+(Z^{+\prime}Z^+)^{-1}Z^{+\prime}$ and $Z^+ = [D, X^+, \Psi_0^{-1/2}X, e_1]$ with the transpose of the t-th row of X^+ being defined by $x_t^+ = x_t - \hat{\Gamma}_x^*\hat{\Sigma}^{*-1}\hat{u}_t^*$. The following theorem yields the limiting distributions of these test statistics.

Theorem 3 *Under Assumption 2, $\mathcal{R}_T^+(\bar{\theta})$ and $\mathcal{L}_T^+$ have the same limiting distributions as $\mathcal{R}_T(\bar{\theta})$ and $\mathcal{L}_T$.*

Although our correction of the test statistics is basically the same as that proposed by Phillips and Hansen (1990), Park (1992), and Jansson (2005), we need not modify y_t to obtain test statistics that are asymptotically independent of nuisance parameters; therefore, our correction of the test statistics is relatively simple. This is because, as explained in the proof of Theorem 3, we can replace y_t in the test statistics with $v_{\theta t}^*$, where $v_{\theta t}^* = \theta u_t^{y \cdot x} + (\lambda/T)\sum_{j=1}^t u_j^{y \cdot x}$. As $u_t^{y \cdot x}$ are (asymptotically) uncorrelated with u_t^x, Brownian motions induced by their partial sums are independent of each other, and hence, a "simultaneous bias correction" is not required for our test statistics.

4. Finite sample evidence

In this section, we investigate the finite sample properties of the tests proposed in Section 3. The data-generating process considered here is the same as that in Jansson (2005). The data are generated according to the system of (1) and (2) with α, β, and α_x normalized to zero. The error term u_t is generated as

$$u_t = \psi(L)\Theta(\rho)\varepsilon_t, \tag{7}$$

where $\varepsilon_t = (\varepsilon_t^y, \varepsilon_t^x)' \sim i.i.d.N(0, I_2)$, $\psi(L) = (1-a)\sum_{i=0}^{\infty} a^i L^i$, and

$$\Theta(\rho) = \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix}.$$

The parameters a and ρ control the strength of autocorrelation for the error and the endogeneity of the regressor, respectively. We set $a = 0, 0.5, 0.8$, $\rho = 0, 0.5, 0.8$, $\theta = 1, 0.975, 0.95, 0.925, 0.90$, and sample size $T = 200$. The initial value, u_0, is drawn from its stationary distribution, and y_0 is set to be equal to zero. We experiment with two initial values for x_0, 0 and 10.

The estimation method used for Σ, Ω, and Γ is the same as that in Jansson (2004).[6] We estimate Σ using $\hat{\Sigma} = T^{-1}\sum_{t=1}^{T} \hat{u}_t^* \hat{u}_t^{*\prime}$ and Ω and Γ using the VAR(1) prewhitened kernel estimator. The rejection frequencies for the 5% level tests with $x_0 = 0$ are reported in Tables 2 and 3 for the cases of the constant mean and linear trend, respectively (we suppress the superscript $^+$ and the argument $\bar{\theta}$ from the test statistics). Cases 1 and 2 describe the results for the cases of $x_0 = 0$ and $x_0 = 10$, respectively. For the sake of comparison, we also show the results for the feasible versions of $\mathcal{P}_T$ and $\mathcal{S}_T$. The test statistic $\mathcal{S}_T$ is based not on the parametric approach by Shin (1994) but on the nonparametric one by Choi and Ahn (1995).

[6] The Matlab code provided by Michael Jansson was very useful in conducting our simulation experiments.

For Case 1, the results are consistent with the analysis of the local asymptotic powers shown in Figures 1 and 2 when the error is not persistent and the endogeneity is low, i.e., when $a \leq 0.5$ and $\rho \leq 0.5$. The empirical sizes of all the tests are satisfactorily close to the nominal one. When the error is serially uncorrelated, i.e., $a = 0$, the robustness of $\mathcal{R}_T$ and $\mathcal{L}_T$ to the endogeneity is pronounced. For $\rho = 0.8$, the results show nontrivial power gain by $\mathcal{R}_T$ and $\mathcal{L}_T$. This is because $\mathcal{R}_T$ and $\mathcal{L}_T$ are invariant under $(\mathcal{G}_y)$, which takes into account the location shift in the direction of $\Psi_0^{-1/2}X$. The most distinctive results can be observed when the error is persistent, i.e., when $a = 0.8$. $\mathcal{S}_T$ and $\mathcal{P}_T$ are undersized when the endogeneity of the regressor is low and oversized when it is high. This is obviously undesirable in practice. On the other hand, the performance of $\mathcal{L}_T$ is highly stable regardless of the degree of persistence. Based on these facts, the LBIU test becomes a strong candidate for researchers who are perplexed with regard to a size versus power trade-off.

Case 2 shows the results with the nonzero initial value of the regressor, $x_0 = 10$. The results on $\mathcal{L}_T$ and $\mathcal{R}_T$ are not presented because they are robust to the initial value, producing exactly the same results as those for Case 1. Tables 2 and 3 reveal that all the appropriate properties of $\mathcal{P}_T$ and $\mathcal{S}_T$ with respect to Case 1 are lost unless the endogeneity is absent, i.e., unless ρ=0. This is an important observation since the initial value is not equal to zero in almost all economic applications. Fortunately, applying the simple transformation of the data that involves subtracting the initial value x_0 from all the observations of x solves the problem. In other words, if we transform the data such that $\tilde{x}_t = x_t - x_0$ for $t = 0, 1, \ldots, T$ and construct $\mathcal{P}_T$ and $\mathcal{S}_T$ using $\tilde{x}_t$ for x_t, the test statistics become invariant to x_0 and perform the same way as Case 1 in finite samples. Researchers who use $\mathcal{P}_T$ and $\mathcal{S}_T$ should always apply this transformation.

5. Conclusions

In this paper, we investigate the LBIU test for the null hypothesis of cointegration. We develop the POI test and then derive the LBIU test among a class of tests that are invariant to some location-scale transformation in the dependent variable. We calculate the asymptotic local power functions and compare them with the standard residual-based test, and we show that the LBIU test is more powerful in a wide range of local alternatives. Our finite sample evidence shows that the LBIU test outperforms the residual-based test in terms of both size and power. The advantage of the LBIU test is particularly patent when the error is persistent. The performance of the LBIU test is highly stable regardless of the degree of persistence and the endogeneity whereas that of the other formerly proposed tests depend considerably on whether the error is persistent or not. Further, we also point out that the finite sample performance of the existing tests is largely affected by the initial value condition, while our tests are immune to it. We propose a simple transformation of data that resolves the problem in the existing tests.

Appendix

Proof of Theorem 1

The POI test statistic can be written as

$$\begin{aligned}\mathcal{R}_T(\bar\theta) &= T\left(1-\mathcal{R}_{1T}^{1/T}(\bar\theta)\mathcal{R}_{2T}(\bar\theta)\right)\\ &= \mathcal{R}_{1T}^{1/T}(\bar\theta)\times T\left(1-\mathcal{R}_{2T}(\bar\theta)\right)+T\left(1-\mathcal{R}_{1T}^{1/T}(\bar\theta)\right)\end{aligned}$$

where

$$\mathcal{R}_{1T}(\bar\theta)=\frac{|Z'\Psi_{\bar\theta}^{-1}Z|}{|Z'Z|},\qquad \mathcal{R}_{2T}(\bar\theta)=\frac{y'(\Psi_{\bar\theta}^{-1}-\Psi_{\bar\theta}^{-1}Z(Z'\Psi_{\bar\theta}^{-1}Z)^{-1}Z'\Psi_{\bar\theta}^{-1})y}{y'My},$$

and we replaced $T-q$ with T for simplicity without loss of generality. We first show that

$$\mathcal{R}_{1T}(\bar\theta)\Rightarrow\frac{|\int_0^1 Q^{\bar\lambda}Q^{\bar\lambda\prime}ds|}{|\int_0^1 QQ'ds|}. \tag{A.1}$$

To show (A.1), notice that in (3) there exist a $k\times(p+1)$ matrix G_{31} and a $k\times 1$ vector g_{34} such that $u_t^x = G_{31}d_t+(1-(1-1_t)L)x_t+g_{34}1_t$, where $1_t=1$ for $t=1$ and $1_t=0$ otherwise. Then, we can transform z_t using a $q\times q$ nonsingular matrix G such that

$$z_t^*=Gz_t,\quad\text{where}\quad G=\begin{bmatrix} I_p & 0 & 0 & 0\\ -\Sigma_{xx}^{-1/2}\alpha_x' & \Sigma_{xx}^{-1/2} & 0 & 0\\ \Sigma_{xx}^{-1/2}G_{31} & 0 & \Sigma_{xx}^{-1/2} & \Sigma_{xx}^{-1/2}g_{34}\\ 0 & 0 & 0 & 1\end{bmatrix},$$

and $z_t^*=[z_{1t}^{*\prime},z_{2t}^{*\prime}]'$ with $z_{1t}^*=[d_t',(\Sigma_{xx}^{-1/2}x_t^0)']'$ and $z_{2t}^*=[(\Sigma_{xx}^{-1/2}u_t^x)',1_t]'$. This is also expressed as $ZG'=Z^*=[Z_1^*,Z_2^*]$ in the matrix form. Then, we have

$$\begin{aligned}\mathcal{R}_{1T}(\bar\theta) &= \left|\frac{1}{T}\Upsilon_T^{-1}GZ'\Psi_{\bar\theta}^{-1}ZG'\Upsilon_T^{-1}\right|\Big/\left|\frac{1}{T}\Upsilon_T^{-1}GZ'ZG'\Upsilon_T^{-1}\right|\\ &= \left|\frac{1}{T}\Upsilon_T^{-1}Z^{\bar\theta\prime}Z^{\bar\theta}\Upsilon_T^{-1}\right|\Big/\left|\frac{1}{T}\Upsilon_T^{-1}Z^{*\prime}Z^*\Upsilon_T^{-1}\right|\end{aligned}$$

where $\Upsilon_T=\text{diag}\{\Upsilon_{1T},\Upsilon_{2T}\}$ with $\Upsilon_{1T}=\text{diag}\{1,T,\cdots,T^p,T^{1/2}I_k\}$ and $\Upsilon_{2T}=\text{diag}\{I_k,T^{-1/2}\}$ and $Z^{\bar\theta}=\Psi_{\bar\theta}^{-1/2}Z^*$. Note that the transpose of the t-th row of $Z^{\bar\theta}$ is expressed as

$$z_t^{\bar\theta}=\bar\theta z_{t-1}^{\bar\theta}+(1-L)z_t^*\quad\text{with}\quad z_1^{\bar\theta}=z_1^*.$$

We partition $z_t^{\bar\theta}$ into $z_{1t}^{\bar\theta}$ and $z_{2t}^{\bar\theta}$ conformably with z_{1t}^* and z_{2t}^*.

Lemma A.1 *For $0 \leq s \leq 1$, the following convergences hold jointly.*

$$\begin{aligned} (i) \quad \Upsilon_{1T}^{-1} z_{1[Ts]}^{*} &\Rightarrow Q(s), \\ (ii) \quad \Upsilon_{1T}^{-1} z_{1[Ts]}^{\bar{\theta}} &\Rightarrow Q^{\bar{\lambda}}(s). \end{aligned}$$

Proof of Lemma A.1: (i) is obtained by using the functional central limit theorem (FCLT). With regard to (ii), from the definition of $z_{1t}^{\bar{\theta}}$, we can express $z_{1t}^{\bar{\theta}}$ as

$$z_{1t}^{\bar{\theta}} = z_{1t}^{*} - \frac{\bar{\lambda}}{T} \sum_{j=1}^{t-1} \left(1 - \frac{\bar{\lambda}}{T}\right)^{t-j-1} z_{1j}^{*}. \tag{A.2}$$

See also the proof of Lemma 7 in Jansson (2004). Then, according to (i) and the continuous mapping theorem (CMT), we have

$$\begin{aligned} \Upsilon_{1T}^{-1} z_{1[Ts]}^{\bar{\theta}} &\Rightarrow Q(s) - \bar{\lambda} \int_0^s e^{-\bar{\lambda}(s-r)} Q dr \\ &= \int_0^s e^{-\bar{\lambda}(s-r)} dQ(r), \end{aligned}$$

where the last equality holds by the partial integration formula.□

From Lemma A.1 (ii) and the CMT we have

$$\frac{1}{T} \Upsilon_{1T}^{-1} Z_1^{\bar{\theta}\prime} Z_1^{\bar{\theta}} \Upsilon_{1T}^{-1} \Rightarrow \int_0^1 Q^{\bar{\lambda}} Q^{\bar{\lambda}\prime} ds. \tag{A.3}$$

$z_{2t}^{\bar{\theta}}$ is expressed in exactly the same way as (A.2) as

$$\begin{aligned} z_{2t}^{\bar{\theta}} &= z_{2t}^{*} - \frac{\bar{\lambda}}{T} \sum_{j=1}^{t-1} \left(1 - \frac{\bar{\lambda}}{T}\right)^{t-j-1} z_{2j}^{*} \\ &= \begin{bmatrix} \Sigma_{xx}^{-1/2} \left(u_t^x - \frac{\bar{\lambda}}{T} \sum_{j=1}^{t-1} \left(1 - \frac{\bar{\lambda}}{T}\right)^{t-j-1} u_j^x \right) \\ 1_t - (1 - 1_t) \frac{\bar{\lambda}}{T} \left(1 - \frac{\bar{\lambda}}{T}\right)^{t-2} \end{bmatrix}. \end{aligned} \tag{A.4}$$

Then, according to the weak law of large numbers (WLLN) and Theorem 4.1 in Hansen (1992b), we have

$$\frac{1}{T} \Upsilon_{2T}^{-1} Z_2^{\bar{\theta}\prime} Z_2^{\bar{\theta}} \Upsilon_{2T}^{-1} \xrightarrow{p} I_{k+1}, \tag{A.5}$$

where $\xrightarrow{p}$ signifies convergence in probability and

$$\frac{1}{T}\Upsilon_{1T}^{-1}Z_1^{\bar{\theta}\prime}Z_2^{\bar{\theta}}\Upsilon_{2T}^{-1} \xrightarrow{p} 0. \tag{A.6}$$

Combining (A.3), (A.5), and (A.6) we obtain

$$\frac{1}{T}\Upsilon_T^{-1}Z^{\bar{\theta}\prime}Z^{\bar{\theta}}\Upsilon_T^{-1} \Rightarrow \begin{bmatrix} \int_0^1 Q^{\bar{\lambda}}Q^{\bar{\lambda}\prime}ds & 0 \\ 0 & I_{k+1} \end{bmatrix}. \tag{A.7}$$

Similarly, we have $T^{-1}\Upsilon_T^{-1}Z^{*\prime}Z^*\Upsilon_T^{-1} \Rightarrow \mathrm{diag}\{\int_0^1 Q(s)Q(s)'ds, I_{k+1}\}$. We then obtain (A.1).

Using (A.1), we can show that

$$\mathcal{R}_{1T}^{1/T} \xrightarrow{p} 1 \qquad \text{and} \qquad T\left(1-\mathcal{R}_{1T}^{1/T}(\bar{\theta})\right) \Rightarrow -\log\left|\int_0^1 Q^{\bar{\lambda}}Q^{\bar{\lambda}\prime}ds\right| + \log\left|\int_0^1 QQ'ds\right| \tag{A.8}$$

because $a^{1/T} \to 1$ and $T(1-a^{1/T}) \to -\log a$ for a given $a > 0$ as $T \to \infty$.

Next, we investigate the asymptotic behavior of $T(1-\mathcal{R}_{2T}(\bar{\theta}))$. To do this, we decompose v_t as

$$\begin{aligned} v_t &= u_t^y + (1-\theta)\sum_{j=1}^{t-1} u_j^y \\ &= \theta u_t^y + \frac{\lambda}{T}\sum_{j=1}^{t} u_j^y \\ &= v_{\theta t}^* + r_{\theta t}, \end{aligned}$$

where $v_{\theta t}^* = \theta u_t^{y\cdot x} + (\lambda/T)\sum_{j=1}^t u_j^{y\cdot x}$ with $u_t^{y\cdot x} = u_t^y - \sigma_{yx}\Sigma_{xx}^{-1}u_t^x$ and $r_{\theta t} = \theta\sigma_{yx}\Sigma_{xx}^{-1}u_t^x + (\lambda/T)\sigma_{yx}\Sigma_{xx}^{-1}\sum_{j=1}^t u_j^x$. Let v_θ^* and r_θ be the vectorized forms of $v_{\theta t}^*$ and $r_{\theta t}$. Since

$$\begin{aligned} r_\theta &= \left\{\theta U^x + (\lambda/T)\Psi_0^{1/2}U^x\right\}\Sigma_{xx}^{-1}\sigma_{xy} \\ &= \left\{\theta\left(\Psi_0^{-1/2}X - \Psi_0^{-1/2}D\alpha_x\right) + (\lambda/T)\left(X - D\alpha_x\right)\right\}\Sigma_{xx}^{-1}\sigma_{xy}, \end{aligned}$$

the conditional likelihood is independent of the change in the direction of r_θ, so that we can replace y in the test statistic with v_θ^*. Then, we can observe that

$$\begin{aligned} T\left(1-R_{2T}(\bar{\theta})\right) &= T\left(1 - \frac{v_\theta^{*\prime}(\Psi_{\bar{\theta}}^{-1} - \Psi_{\bar{\theta}}^{-1}Z(Z'\Psi_{\bar{\theta}}^{-1}Z)^{-1}Z'\Psi_{\bar{\theta}}^{-1})v_\theta^*}{v_\theta^{*\prime}Mv_\theta^*}\right) \\ &= \frac{\mathcal{R}_{21T}(\bar{\theta}) + \mathcal{R}_{22T}(\bar{\theta})}{v_\theta^{*\prime}Mv_\theta^*/T}, \end{aligned} \tag{A.9}$$

where

$$
\begin{aligned}
\mathcal{R}_{21T}(\bar{\theta}) &= 2\left(v_\theta^* - \Psi_{\bar{\theta}}^{-1/2} v_\theta^*\right)' v_\theta^* - \left(v_\theta^* - \Psi_{\bar{\theta}}^{-1/2} v_\theta^*\right)' \left(v_\theta^* - \Psi_{\bar{\theta}}^{-1/2} v_\theta^*\right) \\
&= 2\left(v_\theta^* - v_\theta^{\bar{\theta}}\right)' v_\theta^* - \left(v_\theta^* - v_\theta^{\bar{\theta}}\right)' \left(v_\theta^* - v_\theta^{\bar{\theta}}\right)
\end{aligned}
$$

and

$$
\begin{aligned}
\mathcal{R}_{22T}(\bar{\theta}) &= v_\theta^{*\prime} \Psi_{\bar{\theta}}^{-1} Z (Z' \Psi_{\bar{\theta}}^{-1} Z)^{-1} Z' \Psi_{\bar{\theta}}^{-1} v_\theta^* - v_\theta^{*\prime} Z (Z'Z)^{-1} Z' v_\theta^* \\
&= \left(\frac{1}{\sqrt{T}} v_\theta^{\bar{\theta}\prime} Z^{\bar{\theta}} \Upsilon_T^{-1}\right) \left(\frac{1}{T} \Upsilon_T^{-1} Z^{\bar{\theta}\prime} Z^{\bar{\theta}} \Upsilon_T^{-1}\right)^{-1} \left(\frac{1}{\sqrt{T}} \Upsilon_T^{-1} Z^{\bar{\theta}\prime} v_\theta^{\bar{\theta}}\right) \\
&\quad - \left(\frac{1}{\sqrt{T}} v_\theta^{*\prime} Z^* \Upsilon_T^{-1}\right) \left(\frac{1}{T} \Upsilon_T^{-1} Z^{*\prime} Z^* \Upsilon_T^{-1}\right)^{-1} \left(\frac{1}{\sqrt{T}} \Upsilon_T^{-1} Z^{*\prime} v_\theta^*\right),
\end{aligned}
$$

with $v_\theta^{\bar{\theta}} = \Psi_{\bar{\theta}}^{-1/2} v_\theta^*$. As the denominator in (A.9) is shown to converge to $\sigma_{yy\cdot x} = 1$ in probability by the WLLN under the local alternative, we focus on the derivation of the limiting distributions of $\mathcal{R}_{21T}(\bar{\theta})$ and $\mathcal{R}_{22T}(\bar{\theta})$ in the following.

Lemma A.2 *For $0 \le s \le 1$, the following convergences hold jointly.*

$$
\begin{aligned}
(i) \quad \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} v_{\theta t}^* &\Rightarrow V_\lambda(s), \\
(ii) \quad \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} v_{\theta t}^{\bar{\theta}} &\Rightarrow V_\lambda^{\bar{\lambda}}(s), \ \textit{and} \\
(iii) \quad \sqrt{T}\left(v_{\theta[Ts]}^* - v_{\theta[Ts]}^{\bar{\theta}}\right) &\Rightarrow \bar{\lambda} V_\lambda^{\bar{\lambda}}(s).
\end{aligned}
$$

Proof of Lemma A.2: (i) is obtained from the definition of $v_{\theta t}^*$, the FCLT, and the CMT. With regard to (ii), from the definition of $v_{\theta t}^{\bar{\theta}}$ we have

$$
\sum_{j=1}^{t} v_{\theta j}^{\bar{\theta}} - \bar{\theta} \sum_{j=1}^{t-1} v_{\theta j}^{\bar{\theta}} = v_{\theta t}^*.
$$

Then, in exactly the same way as (A.2), it is seen that

$$
\sum_{j=1}^{t} v_{\theta j}^{\bar{\theta}} = \sum_{j=1}^{t} v_{\theta j}^* - \frac{\bar{\lambda}}{T} \sum_{j=1}^{t-1} \left(1 - \frac{\bar{\lambda}}{T}\right)^{t-j-1} \left(\sum_{i=1}^{j} v_{\theta i}^*\right).
$$

Using (i), the CMT, and the partial integration formula, we obtain (ii).

With regard to (iii), from the definition of $v_{\theta t}^{\bar{\theta}}$, we have

$$v_{\theta t}^{*} - v_{\theta t}^{\bar{\theta}} = (1-\bar{\theta}) \sum_{j=1}^{t-1} v_{\theta j}^{\bar{\theta}}.$$

Then, (iii) is obtained using (ii).□

Using Lemma A.2, the CMT, and Theorem 4.1 in Hansen (1992b), we have

$$\mathcal{R}_{21T}(\bar{\theta}) \Rightarrow 2\bar{\lambda} \int_0^1 V_{\lambda}^{\bar{\lambda}} dV_{\lambda} - \bar{\lambda}^2 \int_0^1 (V_{\lambda}^{\bar{\lambda}})^2 ds. \tag{A.10}$$

For $\mathcal{R}_{22T}(\bar{\theta})$, we can see that

$$\frac{1}{\sqrt{T}} \Upsilon_{1T}^{-1} Z_1^{\bar{\theta}\prime} v_{\theta}^{\bar{\theta}} \Rightarrow \int_0^1 Q^{\bar{\lambda}} dV_{\lambda}^{\bar{\lambda}}, \tag{A.11}$$

using Lemmas A.1, A.2, and Theorem 4.1 in Hansen (1992b), while

$$\frac{1}{\sqrt{T}} \Upsilon_{2T}^{-1} Z_2^{\bar{\theta}\prime} v_{\theta}^{\bar{\theta}} = \begin{bmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\Sigma_{xx}^{-1/2} u_t^{x\bar{\theta}}) v_{\theta t}^{\bar{\theta}} \\ \sum_{t=1}^{T} 1_t^{\bar{\theta}} v_{\theta t}^{\bar{\theta}} \end{bmatrix},$$

where $u_t^{x\bar{\theta}}$ is the transpose of the t-th row of $\Psi_{\bar{\theta}}^{-1/2} U^x$. In exactly the same way as (A.2) we have

$$v_{\theta t}^{\bar{\theta}} = v_{\theta t}^{*} - \frac{\bar{\lambda}}{T} \sum_{j=1}^{t-1} \left(1 - \frac{\bar{\lambda}}{T}\right)^{t-j-1} v_{\theta j}^{*}. \tag{A.12}$$

Then, from (A.4) and (A.12), we can see that

$$\begin{aligned} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\Sigma_{xx}^{-1/2} u_t^{x\bar{\theta}}) v_{\theta t}^{\bar{\theta}} &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\Sigma_{xx}^{-1/2}) u_t^{x} u_t^{y \cdot x} + O_p(T^{-1/2}) \\ &\Rightarrow N(1), \end{aligned} \tag{A.13}$$

by the FCLT, where $N(s)$ is a k dimensional standard Brownian motion that is independent of $W(s)$ and $V(s)$.

On the other hand, using (A.4) we have

$$\begin{aligned} \sum_{t=1}^{T} 1_t^{\bar{\theta}} v_{\theta t}^{\bar{\theta}} &= v_{\theta 1}^{\bar{\theta}} - \frac{\bar{\lambda}}{T} \sum_{t=2}^{T} \left(1 - \frac{\bar{\lambda}}{T}\right)^{t-2} v_{\theta t}^{\bar{\theta}} \\ &\xrightarrow{p} v_{\theta 1}^{\bar{\theta}} = v_{\theta 1}^{*}. \end{aligned} \tag{A.14}$$

Then, combining (A.7), (A.11), (A.13), and (A.14) we have

$$\begin{aligned}
&\left(\frac{1}{\sqrt{T}} v_\theta^{\bar{\theta}\prime} Z^{\bar{\theta}} \Upsilon_T^{-1}\right) \left(\frac{1}{T} \Upsilon_T^{-1} Z^{\bar{\theta}\prime} Z^{\bar{\theta}} \Upsilon_T^{-1}\right)^{-1} \left(\frac{1}{\sqrt{T}} \Upsilon_T^{-1} Z^{\bar{\theta}\prime} v_\theta^{\bar{\theta}}\right) \\
&\quad \Rightarrow \left(\int_0^1 Q^{\bar{\lambda}} dV_\lambda^{\bar{\lambda}}\right)' \left(\int_0^1 Q^{\bar{\lambda}} Q^{\bar{\lambda}\prime} ds\right)^{-1} \left(\int_0^1 Q^{\bar{\lambda}} dV_\lambda^{\bar{\lambda}}\right) + N(1)^2 + v_{\theta 1}^{*2}.
\end{aligned}$$

In exactly the same way we have

$$\begin{aligned}
&\left(\frac{1}{\sqrt{T}} v_\theta^{*\prime} Z^* \Upsilon_T^{-1}\right) \left(\frac{1}{T} \Upsilon_T^{-1} Z^{*\prime} Z^* \Upsilon_T^{-1}\right)^{-1} \left(\frac{1}{\sqrt{T}} \Upsilon_T^{-1} Z^{*\prime} v_\theta^*\right) \\
&\quad \Rightarrow \left(\int_0^1 Q dV_\lambda\right)' \left(\int_0^1 Q Q' ds\right)^{-1} \left(\int_0^1 Q dV_\lambda\right) + N(1)^2 + v_{\theta 1}^{*2}.
\end{aligned}$$

Then, we can see that

$$\begin{aligned}
\mathcal{R}_{22T}(\bar{\theta}) \;\Rightarrow\; & \left(\int_0^1 Q^{\bar{\lambda}} dV_\lambda^{\bar{\lambda}}\right)' \left(\int_0^1 Q^{\bar{\lambda}} Q^{\bar{\lambda}\prime} ds\right)^{-1} \left(\int_0^1 Q^{\bar{\lambda}} dV_\lambda^{\bar{\lambda}}\right) \\
& - \left(\int_0^1 Q dV_\lambda\right)' \left(\int_0^1 Q Q' ds\right)^{-1} \left(\int_0^1 Q dV_\lambda\right). \qquad \text{(A.15)}
\end{aligned}$$

By combining (A.10) and (A.15), we have

$$\begin{aligned}
T\left(1 - \mathcal{R}_{2T}(\bar{\theta})\right) \;\Rightarrow\; & 2\bar{\lambda} \int_0^1 V_\lambda^{\bar{\lambda}} dV_\lambda - \bar{\lambda}^2 \int_0^1 (V_\lambda^{\bar{\lambda}})^2 ds \\
& + \left(\int_0^1 Q^{\bar{\lambda}} dV_\lambda^{\bar{\lambda}}\right)' \left(\int_0^1 Q^{\bar{\lambda}} Q^{\bar{\lambda}\prime} ds\right)^{-1} \left(\int_0^1 Q^{\bar{\lambda}} dV_\lambda^{\bar{\lambda}}\right) \\
& - \left(\int_0^1 Q dV_\lambda\right)' \left(\int_0^1 Q Q' ds\right)^{-1} \left(\int_0^1 Q dV_\lambda\right). \qquad \text{(A.16)}
\end{aligned}$$

The required distribution is obtained from (A.8) and (A.16).□

Proof of Theorem 2

We first derive the LBIU test statistic (6). Note that

$$\begin{aligned}
\frac{d\Psi_\theta}{d\theta} &= (I_T - \Psi_0^{1/2}) \Psi_\theta^{1/2\prime} + \Psi_\theta^{1/2} (I_T - \Psi_0^{1/2\prime}) \\
\frac{d^2\Psi_\theta}{d\theta^2} &= 2(I_T - \Psi_0^{1/2})(I_T - \Psi_0^{1/2\prime}) = 2(\Psi_0 - i_T i_T'),
\end{aligned}$$

where $i_T = [1, \cdots, 1]$ is a $T \times 1$ vector and the relation $I_T - \Psi_0^{1/2} - \Psi_0^{1/2\prime} = -i_T i_T'$ is used. Then, as $\Psi_1^{1/2} = I_T$, $H'H = I_{T-q}$, and $H' i_T = 0$, we have

$$\left.\frac{d(H'\Psi_\theta H)}{d\theta}\right|_{\theta=1} = H'(I_T - i_T i_T')H = I_{T-q} \qquad \text{(A.17)}$$

$$\left.\frac{d^2(H'\Psi_\theta H)}{d\theta^2}\right|_{\theta=1} = 2H'\Psi_0 H. \tag{A.18}$$

From (A.17) and the standard matrix differential calculus, we can show that

$$\begin{aligned}\left.\frac{d\log f(\eta|X;\theta)}{d\theta}\right|_{\theta=1} &= -\frac{1}{2}\mathrm{tr}\left\{(H'\Psi_1 H)^{-1}\left.\frac{d(H'\Psi_\theta H)}{d\theta}\right|_{\theta=1}\right\} \\ &\quad+\frac{T-q}{2}\frac{\eta'(H'\Psi_1 H)^{-1}\left.\frac{d(H'\Psi_\theta H)}{d\theta}\right|_{\theta=1}(H'\Psi_1 H)^{-1}\eta}{\eta'(H'\Psi_1 H)^{-1}\eta} = 0.\end{aligned} \tag{A.19}$$

For the second derivative, it should be noted that

$$\begin{aligned}\left.\frac{d^2\log|H'\Psi_\theta H|}{d\theta^2}\right|_{\theta=1} &= \mathrm{tr}\left\{-I_{T-q}+2H'\Psi_0 H\right\} \\ &= -(T-q)+2\mathrm{tr}\left\{M\Psi_0\right\} \\ &= (T^2+q)-2\mathrm{tr}\left\{(Z'Z)^{-1}Z'\Psi_0 Z\right\},\end{aligned} \tag{A.20}$$

which is obtained using (A.17), (A.18), and $HH'=M$, and

$$\begin{aligned}\left.\frac{d^2\log\{\eta'(H'\Psi_\theta H)^{-1}\eta\}}{d\theta^2}\right|_{\theta=1} &= 1-2\eta' H\Psi_0 H\eta \\ &= 1-2\frac{y'M\Psi_0 My}{y'My}.\end{aligned} \tag{A.21}$$

Then, from (A.20) and (A.21) we have

$$\left.\frac{d^2\log f(\eta|X;\theta)}{d\theta^2}\right|_{\theta=1} = const+\frac{y'M\Psi_0 My}{y'My/(T-q)}+\mathrm{tr}\left\{(Z'Z)^{-1}Z'\Psi_0 Z\right\},$$

so that we obtain (6).

Next, we derive the limiting distribution of the LBIU test statistic. For the same reason given in the proof of Theorem 1, we can replace y in the test statistic with v_θ^* and then we have $\Psi_0^{1/2\prime}My=\Psi_0^{1/2\prime}Mv_\theta^*$. Noting that $\Psi_0^{1/2\prime}=i_T i_T'-\bar{\Psi}_0^{1/2}$, where $\bar{\Psi}_0^{1/2}$ is a $T\times T$ lower triangular matrix with diagonal elements 0 and the other lower elements 1, we have

$$\begin{aligned}\frac{1}{\sqrt{T}}\Psi_0^{1/2\prime}My &= \frac{1}{\sqrt{T}}(i_T i_T'-\bar{\Psi}_0^{1/2})Mv_\theta^* \\ &= -\frac{1}{\sqrt{T}}\left\{\bar{\Psi}_0^{1/2}v_\theta^*-\bar{\Psi}_0^{1/2}Z\left(Z'Z\right)^{-1}Z'v_\theta^*\right\} \\ &= -\left\{\frac{1}{\sqrt{T}}\bar{\Psi}_0^{1/2}v_\theta^*-\frac{1}{T}\bar{\Psi}_0^{1/2}Z^*\Upsilon_T^{-1}\left(\frac{1}{T}\Upsilon_T^{-1}Z^{*\prime}Z^*\Upsilon_T^{-1}\right)^{-1}\frac{1}{\sqrt{T}}\Upsilon_T^{-1}Z^{*\prime}v_\theta^*\right\},\end{aligned}$$

where the second equality holds because $i_T' M = 0$. As the t-th rows of $\bar{\Psi}_0^{1/2} v_\theta^*$ and $\bar{\Psi}_0^{1/2} Z^*$ are $\sum_{j=1}^{t-1} v_{\theta j}^*$ and $\sum_{j=1}^{t-1} z_t^{*\prime}$, we have, using Lemmas A.1, A.2, the FCLT, and the CMT,

$$\frac{v_\theta^{*\prime} M \Psi_0 M v_\theta^* / T^2}{v_\theta^{*\prime} M v_\theta^* /(T-q)} \Rightarrow \int_0^1 \left\{ V_\lambda(s) - \int_0^s Q' dr \left(\int_0^1 QQ' dr \right)^{-1} \int_0^1 Q dV_\lambda \right\}^2 ds. \qquad \text{(A.22)}$$

Similarly, we can also see that

$$\frac{1}{T^2} \mathrm{tr} \left\{ (Z'Z)^{-1} (Z' \Psi_0 Z) \right\} = \mathrm{tr} \left\{ \left(\frac{1}{T} \Upsilon_T^{-1} Z^{*\prime} Z^* \Upsilon_T^{-1} \right) \left(\frac{1}{T^3} \Upsilon_T^{-1} Z^{*\prime} \Psi_0 Z^* \Upsilon_T^{-1} \right) \right\}.$$

Noting that the transpose of the t-th row of $\Psi_0^{1/2\prime} Z^*$ is given by $\sum_{j=t}^T z_j^*$, we have, from Lemma A.1 and the CMT,

$$\frac{1}{T^2} \mathrm{tr} \left\{ (Z'Z)^{-1} (Z' \Psi_0 Z) \right\} \Rightarrow \mathrm{tr} \left\{ \left(\int_0^1 QQ' dr \right)^{-1} \int_0^1 \left(\int_s^1 Q dr \right) \left(\int_s^1 Q' dr \right) ds \right\}. \qquad \text{(A.23)}$$

From (A.22) and (A.23), we obtain the result.□

Proof of Theorem 3

The proof proceeds in the same way as the proofs of Theorem 1 in the last section and Theorem 2 in Jansson (2005). We provide only an outline. First, note that we can obtain the same results in Lemma A.1 by replacing $\Sigma_{xx}^{-1/2}$ in G with $\Omega_{xx}^{-1/2}$. We can also see that, as in the proof of Theorem 1, y_t in the test statistics can be replaced by $v_{\theta t}^*$, where under general assumptions $u_t^{y \cdot x}$ is defined as $u_t^{y \cdot x} = u_t^y - \omega_{yx} \Omega_{xx}^{-1} u_t^x$, so that the limiting distributions in Lemma A.2 should be multiplied by $\omega_{11}^{*1/2}$. Then, applying Lemma 1 in Sims, Stock, and Watson (1990) and Lemma 7 in Jansson (2004), we can see that

$$2(v_\theta^* - v_\theta^{\bar{\theta}})' v_\theta^* \Rightarrow 2\bar{\lambda} \omega_{11}^* \int_0^1 V_\lambda^{\bar{\lambda}} dV_\lambda + 2\bar{\lambda} \pi_{11}^*,$$

and then

$$\mathcal{R}_{2T}(\bar{\theta}) \Rightarrow \omega_{11}^* \left(2\bar{\lambda} \int_0^1 V_\lambda^{\bar{\lambda}} dV_\lambda - \bar{\lambda}^2 \int_0^1 (V_\lambda^{\bar{\lambda}})^2 ds \right) + 2\bar{\lambda} \pi_{11}^*.$$

Similarly, we can see that

$$\begin{aligned} \frac{1}{\sqrt{T}} \Upsilon_{1T}^{-1} Z_1^{+\bar{\theta}\prime} v_\theta^{\bar{\theta}} &\Rightarrow \omega_{11}^{*1/2} \int_0^1 Q^{\bar{\lambda}} dV_\lambda^{\bar{\lambda}} \\ \frac{1}{\sqrt{T}} \Upsilon_{1T}^{-1} Z_1^{+*\prime} v_\theta^* &\Rightarrow \omega_{11}^{*1/2} \int_0^1 Q dV_\lambda. \end{aligned}$$

By combining these results, we obtain the theorem.□

References

[1] Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.

[2] Choi, I. and B. C. Ahn (1995). Testing for cointegration in a system of equations. *Econometric Theory* 11, 952–983.

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

[4] Ferguson, T. S. (1967). *Mathematical statistics: A decision theoretic approach.* Academic Press, New York.

[5] Hansen, B. E. (1992a). Tests for parameter instability in regressions with I(1) processes. *Journal of Business and Economic Statistics* 10, 321–335.

[6] Hansen, B. E. (1992b). Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory* 8, 489–500.

[7] Hubrich, K., H. Lütkepohl, and P. Saikkonen (2001). A review of systems cointegrating tests. *Econometric Reviews* 20, 247–318.

[8] Jansson, M. (2004). Stationarity testing with covariates. *Econometric Theory* 20, 56–94.

[9] Jansson, M. (2005). Point optimal tests of the null hypothesis of cointegration. *Journal of Econometrics* 124, 187–201.

[10] Kariya, T. (1980). Locally robust tests for serial correlation in least squares regression. *Annals of Statistics* 8, 1065–1070.

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

[12] Park, J. Y. (1992). Canonical cointegrating regressions. *Econometrica* 60, 119–143.

[13] 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.

[14] Phillips, P. C. B. and S. Ouliaris (1990). Asymptotic properties of residual based tests for cointegration. *Econometrica* 58, 165–193.

[15] Quintos, C. E. and P. C. B. Phillips (1993). Parameter constancy in cointegrating regressions. *Empirical Economics* 18, 675–706.

[16] Rao, C. R. (1973). *Linear statistical inference and its applications, 2nd ed.* Wiley, New York.

[17] Shin, Y. (1994). A residual-based test of the null of cointegration against the alternative of no cointegration. *Econometric Theory* 10, 91–115.

[18] 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.

Table 1. Percentiles of the Limiting Distribution of $\mathcal{L}_T$

(a) constant mean	$k = 1$	2	3	4	5	6
90%	0.6095	0.5739	0.5512	0.5376	0.5303	0.5246
95%	0.6803	0.6235	0.5823	0.5609	0.5483	0.5387
97.5%	0.7632	0.6795	0.6182	0.5874	0.5706	0.5538
99%	0.8940	0.7667	0.6825	0.6320	0.6037	0.5750
(b) linear trend	$k = 1$	2	3	4	5	6
90%	0.5419	0.5348	0.5277	0.5228	0.5196	0.5165
95%	0.5651	0.5527	0.5425	0.5352	0.5297	0.5255
97.5%	0.5894	0.5716	0.5594	0.5490	0.5410	0.5352
99%	0.6223	0.5997	0.5831	0.5674	0.5570	0.5475

Table 2. Size and Power (conatant mean, $T = 200$)

		(Case 1)												(Case 2)					
		$\mathcal{L}_\mathcal{T}$			$\mathcal{S}_\mathcal{T}$			$\mathcal{R}_\mathcal{T}$			$\mathcal{P}_\mathcal{T}$			$\mathcal{S}_\mathcal{T}$			$\mathcal{P}_\mathcal{T}$		
		ρ			ρ			ρ			ρ			ρ			ρ		
a	θ	0	0.5	0.8	0	0.5	0.8	0	0.5	0.8	0	0.5	0.8	0	0.5	0.8	0	0.5	0.8
	1	5.8	5.5	5.5	4.9	4.5	4.0	5.3	5.4	5.2	5.2	5.5	6.3	5.0	9.6	33.1	6.1	33.5	90.8
	0.975	23.6	22.2	22.7	20.8	19.5	19.2	22.4	21.3	22.1	22.3	19.9	17.7	21.0	24.1	40.7	23.1	47.8	91.4
0	0.95	47.6	46.7	47.1	43.5	42.3	41.0	52.9	53.1	52.5	52.8	51.0	43.4	43.4	45.3	55.7	53.5	68.9	94.0
	0.925	64.8	64.1	63.8	59.6	59.1	57.9	73.9	73.7	73.9	73.6	71.6	64.0	59.9	61.0	67.3	74.6	82.2	95.8
	0.9	75.2	75.3	74.6	71.2	70.3	69.3	85.4	85.7	85.3	85.8	84.9	78.7	71.2	71.9	75.4	86.3	90.4	97.0
	1	5.3	5.5	5.3	4.3	4.9	6.5	4.6	4.8	4.5	3.9	4.0	5.7	5.0	12.8	49.4	6.2	46.4	98.3
	0.975	21.8	20.5	21.3	18.2	18.4	22.5	19.7	18.7	19.0	17.8	16.1	18.3	18.8	24.5	52.8	20.9	54.9	98.0
0.5	0.95	42.4	41.9	41.6	37.1	38.0	42.0	45.7	45.6	45.4	43.6	42.3	43.2	37.6	41.8	61.2	47.5	69.4	97.7
	0.925	56.6	56.7	55.9	49.5	51.7	55.3	64.3	64.5	64.7	62.8	61.4	61.8	50.2	53.7	67.5	65.5	78.7	97.2
	0.9	64.8	64.9	64.2	56.6	58.7	63.7	75.1	75.1	73.8	73.5	72.9	72.9	56.9	60.8	71.3	76.0	84.0	97.0
	1	4.6	5.2	5.5	3.8	6.1	11.5	2.7	2.8	3.1	1.9	3.8	10.4	6.0	22.4	63.3	13.0	64.6	97.9
	0.975	17.6	16.3	17.6	12.8	15.8	24.4	9.8	9.5	10.5	6.3	9.4	19.9	15.4	28.7	61.7	21.0	64.1	97.0
0.8	0.95	28.7	28.5	28.4	20.1	24.6	36.4	18.1	17.7	18.0	10.6	16.4	32.2	22.7	34.7	60.2	30.9	61.8	94.1
	0.925	30.8	29.9	29.6	18.8	25.1	39.4	18.1	18.3	18.5	10.3	15.9	34.9	23.8	33.5	56.3	35.0	56.8	88.4
	0.9	27.3	25.3	25.1	14.5	20.2	36.6	15.4	15.7	16.3	9.0	13.2	30.7	22.0	29.4	51.0	37.3	52.9	81.1

Table 3. Size and Power (linear trend, $T = 200$)

		(Case 1)												(Case 2)					
		$\mathcal{L}_T$			$\mathcal{S}_T$			$\mathcal{R}_T$			$\mathcal{P}_T$			$\mathcal{S}_T$			$\mathcal{P}_T$		
		ρ			ρ			ρ			ρ			ρ			ρ		
a	θ	0	0.5	0.8	0	0.5	0.8	0	0.5	0.8	0	0.5	0.8	0	0.5	0.8	0	0.5	0.8
	1	6.1	6.5	6.2	5.1	5.1	3.2	5.4	5.4	5.1	5.2	5.8	5.5	5.6	12.0	47.4	6.9	39.2	90.9
	0.975	13.4	13.2	13.6	11.9	10.4	8.4	11.1	11.1	10.7	11.3	10.8	8.0	12.4	18.6	49.2	13.0	42.8	90.9
0	0.95	31.4	31.3	31.5	28.6	27.2	23.5	30.7	30.4	30.6	30.5	26.9	17.9	28.9	33.5	56.2	32.5	55.7	91.4
	0.925	50.7	50.7	50.5	46.7	46.0	41.8	53.0	52.6	53.2	52.6	47.3	34.4	47.3	49.8	64.9	53.9	70.0	92.7
	0.9	66.3	65.5	65.0	62.1	61.1	57.5	70.6	70.9	71.5	70.9	66.5	51.8	62.6	64.4	72.7	72.0	81.3	94.2
	1	5.8	5.9	5.6	4.7	5.2	6.5	4.8	5.0	5.0	4.0	3.9	4.3	5.8	16.8	70.2	8.2	55.1	98.9
	0.975	12.1	11.5	12.3	9.9	10.2	12.4	9.6	9.7	9.8	8.6	7.6	7.3	11.4	21.9	68.5	13.1	57.0	98.7
0.5	0.95	27.1	26.6	27.2	23.4	23.6	27.7	25.2	24.7	24.9	23.4	20.9	19.5	24.8	33.5	70.0	28.5	62.8	98.5
	0.925	42.5	42.9	42.7	36.9	38.7	42.5	42.4	42.6	42.0	40.2	37.6	34.2	38.2	45.9	71.5	45.7	69.7	97.8
	0.9	55.2	54.5	54.3	48.3	49.8	54.0	56.4	56.2	56.0	54.2	51.8	47.7	49.4	54.6	74.0	58.8	75.0	96.7
	1	4.8	4.9	5.1	3.2	5.8	12.2	2.2	2.6	2.6	1.4	2.8	6.7	7.4	33.8	81.8	16.4	71.7	98.5
	0.975	8.8	8.9	9.7	6.3	9.5	17.7	3.6	3.7	4.3	2.2	3.7	8.7	11.3	34.2	79.7	20.2	69.7	98.1
0.8	0.95	16.6	16.0	16.5	11.1	16.2	27.6	7.3	7.4	7.6	4.1	6.4	14.4	17.2	37.3	76.6	26.1	66.3	96.5
	0.925	20.8	20.5	20.1	12.5	18.9	34.0	8.6	8.8	9.2	4.5	7.1	17.8	20.7	36.9	72.6	29.8	60.7	93.3
	0.9	20.4	19.5	18.4	11.0	17.7	35.6	7.5	7.8	9.0	4.1	6.0	17.2	20.6	34.3	67.0	33.0	54.6	87.6

Figure 1. 5% Level tests, m=1 constant mean

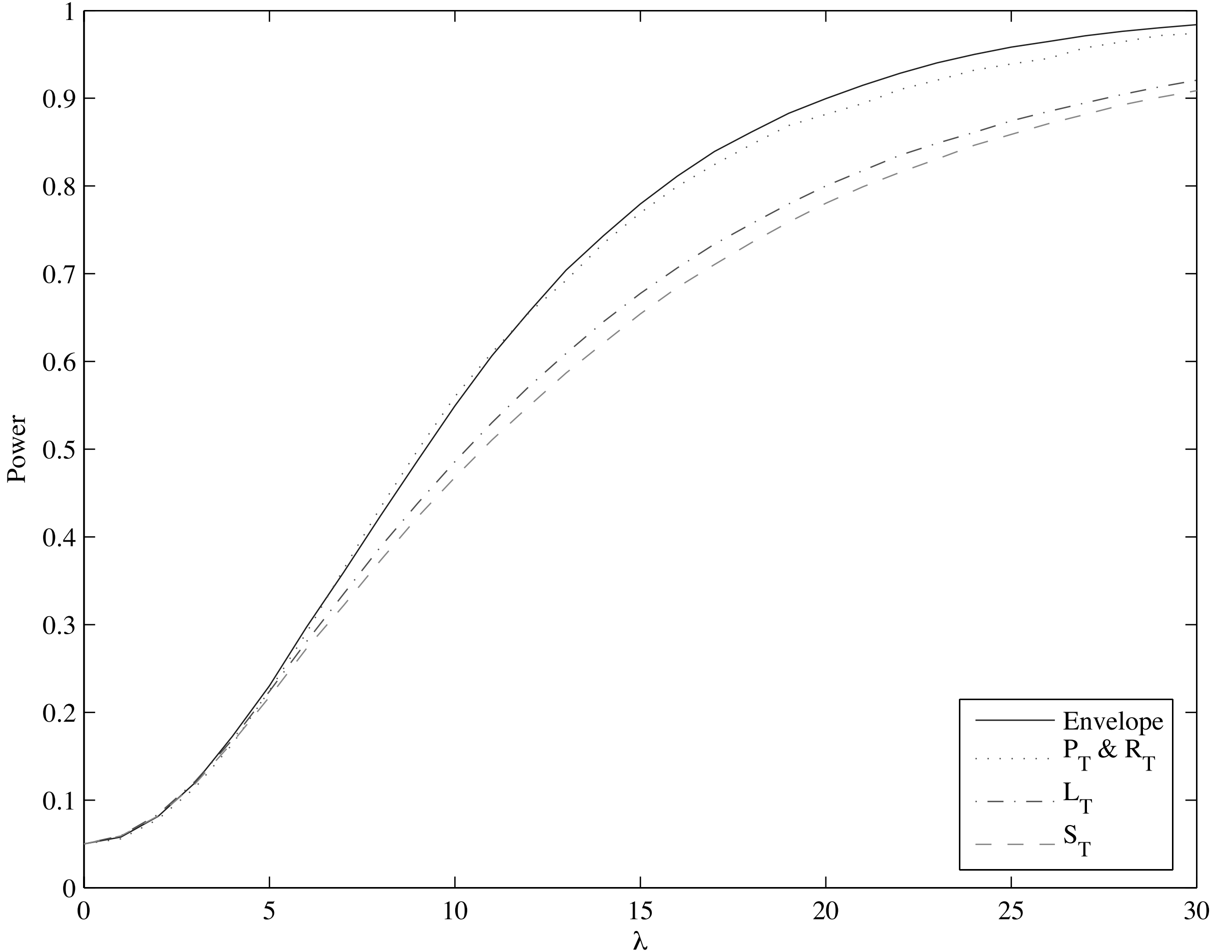

Figure 2. 5% Level tests, m=1 linear trend