Authors
Jansson, Michael
Haldrup, Niels, Prof.

Publication Date
2000-06-01

2000-14

UNIVERSITY OF CALIFORNIA, SAN DIEGO

DEPARTMENT OF ECONOMICS

SPURIOUS REGRESSION, COINTEGRATION, AND NEAR COINTEGRATION: A UNIFYING APPROACH

BY

MICHAEL JANSSON

AND

NIELS HALDRUP

DISCUSSION PAPER 2000-14
JUNE 2000

Spurious Regression, Cointegration, and Near Cointegration: A Unifying Approach

MICHAEL JANSSON AND NIELS HALDRUP*
DEPARTMENT OF ECONOMICS AND CDME, UNIVERSITY OF AARHUS
DK-8000 AARHUS C, DENMARK

First draft: February 1, 1998
This draft: June 2, 2000

ABSTRACT. This paper introduces a representation of an integrated vector time series in which the coefficient of multiple correlation computed from the long-run covariance matrix of the innovation sequences is a primitive parameter of the model. Based on this representation, we propose a notion of near cointegration, which helps bridging the gap between the polar cases of spurious regression and cointegration. Two applications of the model of near cointegration are provided. As a first application, the properties of conventional cointegration methods under near cointegration are characterized, hereby investigating the robustness of cointegration methods. Secondly, we illustrate how to obtain local power functions of cointegration tests that take cointegration as the null hypothesis.

KEYWORDS: Cointegration, spurious regression, near cointegration, cointegration tests, local power function, Brownian motion.

JEL CLASSIFICATION: C12, C13, C22.

1. INTRODUCTION

One of the most important contributions to modern time series econometrics is the development of an asymptotic theory for the analysis of multiple integrated time series. Much of this research has been inspired by the Monte Carlo study conducted by Granger and Newbold (1974). That study considered regressions of independent random walks on each other and found that the usual significance test based on the regression F-statistic tends to overreject the null. To describe this phenomenon, the term spurious regression was coined.[1] The numerical findings of Granger and Newbold were given an analytical explanation by Phillips (1986), while Park, Ouliaris, and

*This paper was completed while the second author was visiting the University of California, San Diego in the spring of 2000. The paper has benefited from the comments of H. Peter Boswijk and seminar participants at Penn State University, Tinbergen Institute, University of California (Riverside, San Diego), the 1998 European Meeting of the Econometric Society and the 1999 NOS-S conference on macroeconomic transmission mechanisms.

[1] Earlier, Yule (1926) had used the term nonsense correlation to describe a similar phenomenon.

Choi (1988) and Park (1990) provided further clarification. These authors considered regressions involving quite general integrated processes and found that the asymptotic properties of the appropriate F-statistic depend crucially on ρ^2, the squared multiple correlation coefficient computed from the long-run covariance matrix of the innovation sequences. If $\rho^2 < 1$, the F-statistic diverges at rate T (where T is the sample size) while $T^{-1} \times F$ has a non-degenerate limiting distribution, which only depends on the dimension of the system. In other words, the regression is spurious whenever the coefficient of correlation is less than unity. In contrast, when $\rho^2 = 1$ the series are cointegrated and $F = O_p(1)$ with a complicated limiting distribution. Conventional asymptotic results therefore depend discontinuously on ρ^2.

On the other hand, it is quite obvious that the finite sample distribution of the F-statistic depends continuously on ρ^2. As a consequence, there is reason to believe that conventional spurious regression asymptotics provide a poor approximation to the finite sample behavior of the F-statistic when the processes are "nearly" cointegrated in the sense that ρ^2 is "close" to unity. More generally, finite sample approximations based on spurious regression theory are likely to be of limited usefulness whenever the limiting behavior of the object of interest (e.g. an estimator or a test statistic) exhibits a discontinuity at $\rho^2 = 1$ and values of ρ^2 close to unity are of particular interest. In contrast, a model of near cointegration in which ρ^2 is a sequence of parameters lying in a shrinking neighborhood of unity as T tends to infinity is much more appealing in such situations.

Motivated by these considerations, the present paper introduces a model in which ρ^2 is a primitive parameter and uses this model to propose a notion of near cointegration.[2] By construction, the limiting behavior of the F-statistic depends continuously on ρ^2 in our setup and the model of near cointegration therefore enables us to bridge the gap between spurious regression and cointegration with respect to the limiting behavior of the F-statistic. The usefulness of our model is by no means limited to the study of the F-statistic. We illustrate this by presenting two further applications of the model. As a first application, the robustness of cointegration methods is investigated. Specifically, we characterize the limiting behavior under near cointegration of the usual Wald statistic devised to test hypotheses on a cointegrating vector. This application complements Elliott's (1998) study, where the implications of near-integration in exactly cointegrated models are examined. Our finding is that under near cointegration the limiting distribution is no longer χ^2. In fact, the results of a simulation study indicate that substantial size distortions are encountered even for moderate values of the noncentrality parameter measuring the deviation from exact cointegration. In our second application, we illustrate how to obtain local power functions of cointegration tests that take cointegration as the null hypothesis. In the literature, several different classes of cointegration tests have been proposed. It is

[2] In the aforementioned papers, ρ^2 is computed from a long-run covariance matrix which is itself defined by taking limits as $T \to \infty$. Therefore, it is not immediately obvious how to model ρ^2 as a sequence of parameters that lie in (say) a $1/T^2$ neighborhood of unity. By working with a representation where ρ^2 is a primitive parameter, we circumvent this potential problem.

therefore desirable to investigate what (if anything) can be said about the relative power properties of these competing test procedures. As a first step in that direction, we characterize the behavior of several regression based cointegration tests under local alternatives and compute the corresponding local power functions. Among the six test statistics under study, four are found to have virtually identical local power properties, while the remaining two are significantly inferior in terms of local power.

The paper proceeds as follows. In Section 2, we present the general model and discuss how the polar cases of spurious regression and cointegration arise as special cases of that model. In addition to these familiar concepts, a notion of near cointegration is introduced. Section 3 discusses the behavior of regression estimators under spurious regression, cointegration, and near cointegration, while Section 4 contains the corresponding results for inference procedures based on these estimators. Specifically, Section 4.1 studies the F-statistic and Section 4.2 investigates the robustness of cointegration methods by characterizing the behavior of a Wald statistic under local alternatives. In Section 5, we report the behavior of several cointegration tests under near cointegration. Finally, Section 6 offers a few concluding remarks. Proofs of all results of the paper are outlined in an Appendix.

Before we begin, a word on notation. The inequality " > 0" signifies positive definiteness when applied to square matrices and $\|A\|$ is the Euclidean norm $(\mathrm{tr}\,(A'A))^{1/2}$. For any symmetric $A > 0$, $A^{-1/2} = \left(A^{1/2}\right)^{-1}$ and $A^{1/2}$ is the upper triangular matrix with positive diagonal elements such that $A^{1/2}A^{1/2\prime} = A$. To simplify the notation, integrals such as $\int_0^1 W(r)\,dr$ and stochastic integrals such as $\int_0^1 W(r)\,dW(r)'$ are typically written as $\int W$ and $\int W dW'$, respectively. We use $\mathcal{L}(X)$ to denote the probability law of X, the symbol " $\stackrel{\mathcal{L}}{=}$ " signifies equality in law, and "$X_T \stackrel{\mathcal{L}_\infty}{=} Y_T$" is shorthand for " $\lim_{T\to\infty} \mathcal{L}(X_T)$ and $\lim_{T\to\infty} \mathcal{L}(Y_T)$ both exist and are equal". Finally, all limits are taken as the sample size $T \to \infty$ unless otherwise stated.

2. Preliminaries

Section 2.1 introduces the general model and Section 2.2 discusses how the polar cases of spurious regression and cointegration arise as special cases of that model. Finally, Section 2.3 introduces a notion of near cointegration.

2.1. The Model and Assumptions.

We assume that $\{z_t : t \geq 0\}$ is an m-vector integrated process generated by

$$\Delta z_t = C(L)\, e_t, \tag{1}$$

where $C(L)$ and $\{e_t : t \in \mathbb{Z}\}$ satisfy the following requirements:

A1. $C(L) = \sum_{i=0}^{\infty} C_i L^i$ is a lag polynomial, $\sum_{i=0}^{\infty} i^2 \|C_i\| < \infty$ and $C(1) = \sum_{i=0}^{\infty} C_i$ is upper triangular with non-negative diagonal elements.

A2. The sequence $\{e_t\}$ is $i.i.d.$ with $E(e_t) = 0$ and $E(e_t e_t') = I_m$.

The memory condition A1 is satisfied whenever $\{\Delta z_t\}$ is a stationary vector ARMA process. Along with the moment condition A2, A1 will enable us to call upon well known results for linear processes (e.g. Phillips and Solo (1992), Phillips (1988b)) when deriving the results of the paper. Assuming that $E(e_t e_t') = I_m$ and $C(1)$ is upper triangular entails essentially no loss of generality. Indeed, suppose $B(L) = \sum_{i=0}^{\infty} B_i L^i$ is a lag polynomial with $\sum_{i=0}^{\infty} i^2 \|B_i\| < \infty$ and suppose the sequence $\{u_t : t \in \mathbb{Z}\}$ is $i.i.d.$ with $E(u_t) = 0$ and $E(u_t u_t') = \Sigma$, a positive definite matrix. Define $\{C_i : i \geq 0\}$ and $\{e_t : t \in \mathbb{Z}\}$ as follows: $C_i = B_i \Sigma^{1/2} \mathcal{O}$ and $e_t = \mathcal{O}' \Sigma^{-1/2} u_t$, where $\mathcal{O}$ is an orthogonal matrix such that $B(1)\Sigma^{1/2}\mathcal{O}$ is upper triangular (with non-negative diagonal elements). Then, for all $i \geq 0, t \in \mathbb{Z}$, $C_i e_t = B_i u_t$ and $E(e_t e_t') = I_m$. Moreover, $C(1)$ is upper triangular (with non-negative diagonal elements) and $\sum_{i=0}^{\infty} i^2 \|C_i\| < \infty$.

Applying the Beveridge-Nelson (1981) decomposition to $C(L)$, we have:

$$z_t = C(1)\xi_t + \tilde{C}(L) e_t + \tilde{z}_0, \tag{2}$$

where $\xi_t = \sum_{s=1}^{t} e_s$, $\tilde{z}_0 = z_0 - \tilde{C}(L) e_0$ and $\tilde{C}(L) = \sum_{i=0}^{\infty} \tilde{C}_i L^i$ is a lag polynomial with coefficients $\tilde{C}_i = -\sum_{j=i+1}^{\infty} C_j$ satisfying $\sum_{i=0}^{\infty} i \left\|\tilde{C}_i\right\| \leq \frac{1}{2}\sum_{i=0}^{\infty} i^2 \|C_i\| < \infty$.

Partition the m-vectors z_t and ξ_t into $m_y = 1$ and $m_x = m - 1$ components as $z_t' = (y_t, x_t')$ and $\xi_t' = \left(\xi_{y,t}, \xi_{x,t}'\right)$. The cointegration rank of $\{z_t\}$ equals the rank deficiency of $C(1)$ and we can therefore parameterize the cointegration rank of $\{z_t\}$ directly by a suitable parameterization of $C(1)$. It turns out to be convenient to parameterize $C(1)$ in terms of the elements of the long-run covariance matrix of Δz_t, viz.

$$\Omega_{zz} = \begin{pmatrix} \omega_{yy} & \omega_{xy}' \\ \omega_{xy} & \Omega_{xx} \end{pmatrix} = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T}\sum_{s=1}^{T} E(\Delta z_t \Delta z_s') = C(1)C(1)', \tag{3}$$

where the partitioning is in conformity with z_t. Specifically, we shall parameterize $C(1)$ as follows:

A3. Let $C(1)$ be partitioned in conformity with ξ_t. Then

$$C(1) = \begin{pmatrix} \omega_{yy}^{1/2}\left(1-\rho^2\right)^{1/2} & \rho\left(\Omega_{xx}^{-1/2}\bar{\omega}_{xy}\right)' \\ 0 & \Omega_{xx}^{1/2} \end{pmatrix},$$

where $\omega_{yy} > 0, \Omega_{xx} > 0, 0 \leq \rho \leq 1$ and $\bar{\omega}_{xy}$ is an m_x-vector satisfying $\bar{\omega}_{xy}' \Omega_{xx}^{-1} \bar{\omega}_{xy} = \omega_{yy}$.

The assumptions $\omega_{yy} > 0$ and $\Omega_{xx} > 0$ in A3 imply that $\{y_t\}$ is an integrated process and $\{x_t\}$ is a non-cointegrated integrated process. Admittedly, the assumption that $\{x_t\}$ is non-cointegrated is somewhat restrictive. On the other hand, the assumption of non-cointegrated regressors is fairly standard in the related literature,[3] so in order to facilitate comparisons with existing results we shall maintain this assumption throughout.

When A3 holds,

$$\Omega_{zz} = C(1)\, C(1)' = \begin{pmatrix} \omega_{yy} & \rho\bar{\omega}_{xy}' \\ \rho\bar{\omega}_{xy} & \Omega_{xx} \end{pmatrix}.$$

The parameters ω_{yy} and Ω_{xx} in A3 therefore coincide with the corresponding long-run variances in (3). The long-run covariance ω_{xy} between Δx_t and Δy_t is given by $\rho\bar{\omega}_{xy}$, where $\bar{\omega}_{xy}$ expresses the direction of the covariance while ρ measures the strength of the covariance. In fact, as the notation suggests,

$$\rho^2 = \frac{\omega_{xy}'\Omega_{xx}^{-1}\omega_{xy}}{\omega_{yy}}$$

is the squared coefficient of multiple correlation computed from Ω_{zz}.

As we shall see shortly, the cointegration properties of $\{z_t\}$ depend solely on the scalar parameter ρ. Indeed, $\{z_t\}$ is cointegrated if and only if $\rho^2 = 1$. For our purposes, this is very convenient since it enables us to introduce a notion of near cointegration by modeling ρ as a sequence of parameters lying in a shrinking neighborhood of unity as T tends to infinity.

To complete the specification of the model, we need to make an assumption concerning the initialization of $\{z_t\}$ at $t = 0$. For convenience, we make the following assumption, which implies that $\tilde{z}_0 = 0$ in (2):

A4. $z_0 = \tilde{C}(L)\, e_0$.

In a well defined sense, A4 is simply a normalization. Indeed, as we shall see in Remark (i) following Lemma 1, it is straightforward to accommodate a non-zero (possibly time-dependent) mean in z_t. Doing so will not alter our results in any interesting way, however, and we therefore retain A4 in order to simplify the exposition.

Together, A3-A4 imply that $(y_t, x_t')'$ can be represented as

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} \omega_{yy}^{1/2}\left(1-\rho^2\right)^{1/2} & \rho\left(\Omega_{xx}^{-1/2}\bar{\omega}_{xy}\right)' \\ 0 & \Omega_{xx}^{1/2} \end{pmatrix} \begin{pmatrix} \xi_{y,t} \\ \xi_{x,t} \end{pmatrix} + \tilde{C}(L)\, e_t, \tag{4}$$

where $\xi_{y,t}$ and $\xi_{x,t}$ are uncorrelated random walks.

[3] Notable exceptions are Park and Phillips (1989, Section 5.2), Choi (1994), and McCabe, Leybourne, and Shin (1997). See also Phillips (1995) and Chang and Phillips (1995).

2.2. Spurious Regression and Cointegration.
Since $\{x_t\}$ is non-cointegrated, $\{z_t\}$ is cointegrated if and only if $\{y_t - \beta_0' x_t\}$ is stationary, where β_0 is the projection coefficient computed from Ω_{zz}, viz.

$$\beta_0 = \Omega_{xx}^{-1} \rho \bar{\omega}_{xy}. \tag{5}$$

Following Park, Ouliaris, and Choi (1988), we shall occasionally refer to β_0 as the fundamental coefficient. From (4) – (5) , we get

$$y_t - \beta_0' x_t = \omega_{yy}^{1/2} \left(1 - \rho^2\right)^{1/2} \xi_{y,t} + \left(\begin{array}{cc} 1 & -\beta_0' \end{array} \right) \tilde{C}(L) e_t.$$

When $\rho < 1$ (and fixed), $\{y_t - \beta' x_t\}$ is an integrated process for any value of β. We shall refer to this as the spurious regression case. In contrast, $\{z_t\}$ is cointegrated when $\rho = 1$. Indeed, when $\rho = 1$, $y_t - \beta_0' x_t = u_t$, where

$$u_t = \left(\begin{array}{cc} 1 & -\bar{\beta}_0' \end{array} \right) \tilde{C}(L) e_t, \qquad \bar{\beta}_0 = \Omega_{xx}^{-1} \bar{\omega}_{xy}.$$

Under spurious regression, our distributional results depend solely on Ω_{zz}. Under cointegration, in contrast, our distributional results depend on the following parameters:

$$\Omega_{ww} = \begin{pmatrix} \omega_{uu} & \omega_{xu}' \\ \omega_{xu} & \Omega_{xx} \end{pmatrix} = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} E\left(w_t w_s'\right),$$

$$\Gamma_{ww} = \left(\begin{array}{cc} \Gamma_{\cdot u} & \Gamma_{\cdot x} \end{array} \right) = \begin{pmatrix} \gamma_{uu} & \gamma_{ux} \\ \gamma_{xu} & \Gamma_{xx} \end{pmatrix} = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} \sum_{i=0}^{t-1} E\left(w_t w_{t-i}'\right),$$

$$\Sigma_{ww} = \begin{pmatrix} \sigma_{uu} & \sigma_{xu}' \\ \sigma_{xu} & \Sigma_{xx} \end{pmatrix} = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} E\left(w_t w_t'\right),$$

where $w_t' = (u_t, \Delta x_t')$ and Ω_{ww}, Γ_{ww} and Σ_{ww} are partitioned in the obvious way.

The case where Ω_{ww} and Γ_{ww} are block lower triangular (i.e. $\omega_{xu}' = \gamma_{ux} = 0_{1 \times m_x}$) is of particular interest, since the asymptotic theory simplifies considerably in this case. In most applications, however, we would not expect the raw data $\{z_t\}$ to satisfy such requirements.[4] On the other hand, we can transform $\{z_t\}$ in such a way that the transformed data, $\left\{z_t^\dagger\right\}$ (say), does meet these requirements. One such transformation, the CCR transformation suggested by Park (1992), is $z_t^{\dagger\prime} = \left(y_t^\dagger, x_t^{\dagger\prime}\right)$, where

$$y_t^\dagger = y_t - \bar{\beta}_0' \Gamma_{\cdot x}' \Sigma_{xx}^{-1} w_t - \omega_{xu}' \Omega_{xx}^{-1} \Delta x_t, \tag{6}$$

$$x_t^\dagger = x_t - \Gamma_{\cdot x}' \Sigma_{xx}^{-1} w_t. \tag{7}$$

Let $w_t^{\dagger\prime} = \left(u_t^\dagger, \Delta x_t^{\dagger\prime}\right)$, where $u_t^\dagger = u_t - \omega_{xu}' \Omega_{xx}^{-1} \Delta x_t$ and define $\Omega_{ww}^\dagger, \Gamma_{ww}^\dagger$ and $\Sigma_{ww}^\dagger$ in analogy with Ω_{ww}, Γ_{ww} and Σ_{ww}. Then $\Omega_{ww}^\dagger$ and $\Gamma_{ww}^\dagger$ are block lower triangular with $\Omega_{xx}^\dagger = \Omega_{xx}$ and $\omega_{uu}^\dagger = \omega_{uu.x} = \omega_{uu} - \omega_{xu}' \Omega_{xx}^{-1} \omega_{xu}$, the conditional variance computed from Ω_{zz}.

Remark. In applications, $\bar{\beta}_0, \Omega_{ww}, \Gamma_{ww}$ and Σ_{ww} are typically unknown and $\{w_t\}$ is unobserved, so the CCR transformations (6) – (7) are infeasible. On the other hand, consistent estimators of $\bar{\beta}_0, \Omega_{ww}, \Gamma_{ww}$ and Σ_{ww} are easily constructed. Indeed, the OLS estimator $\hat{\beta}$ in equation (9) below is consistent for $\bar{\beta}_0$ (Lemma 1 (b)). Likewise, conventional kernel estimators of Ω_{ww}, Γ_{ww} and Σ_{ww} can be shown to be consistent under (near) cointegration (Jansson (1999)). As it turns out, our asymptotic results derived under (near) cointegration are unaffected when $\left\{z_t^\dagger\right\}$ is constructed using a feasible CCR transformation based on consistent estimators of $\bar{\beta}_0, \Omega_{ww}, \Gamma_{ww}$ and Σ_{ww} (e.g. Park (1992)). For convenience, we therefore assume throughout that $\left\{z_t^\dagger\right\}$ is observed and that (nuisance) parameters such as Ω_{ww}, Γ_{ww} and Σ_{ww} are known. ■

2.3. Near Cointegration.

In addition to the familiar concepts of spurious regression and cointegration, we now introduce a notion of near cointegration. We say that $\{z_t\}$ is nearly cointegrated when the following assumption holds:

A5. (i) $1 - \rho^2 = T^{-2} \lambda^2 \omega_{uu.x} / \omega_{yy}$ for some $\lambda \geq 0$, (ii) $\omega_{uu.x} = \omega_{uu} - \omega_{xu}' \Omega_{xx}^{-1} \omega_{xu} > 0$, and (iii) $\begin{pmatrix} 1 & -\bar{\beta}_0' \end{pmatrix} \tilde{C}(1) \begin{pmatrix} 1 & 0_{1 \times m_x} \end{pmatrix}' > 0$.

Of course, near cointegration reduces to cointegration when $\lambda = 0$ in A5 (i). When $\lambda = 0$, A5 (ii) states that the cointegration is regular in the sense of Park (1992, Definition 2.3). On the other hand, when $\lambda \neq 0$, A5(iii) is essentially an identification

[4] A sufficient condition for these block triangularity requirements to hold is that $\{x_t\}$ is strictly exogenous in the sense that $E(\Delta x_t u_s) = 0 \ \forall t \geq 1, s \geq 1$.

assumption.[5] Under A5, ρ (and hence also the fundamental coefficient β_0) is a sequence of parameters. Likewise, $\{z_t\}$ is a triangular array rather than a sequence. As is common in the literature, we follow Phillips (1987, 1988a) and omit an additional subscript T, since it is inessential to the discussion.

The parameter λ introduced in A5 will play a prominent role in the asymptotic theory developed under near cointegration. Under A5,

$$\lambda = \frac{T \cdot \omega_{yy}^{1/2} \left(1 - \rho^2\right)^{1/2}}{\omega_{uu.x}^{1/2}}, \tag{8}$$

and we see that λ can be interpreted as a signal-to-noise ratio. Specifically, the numerator in (8) is proportional to $\omega_{yy}^{1/2} \left(1 - \rho^2\right)^{1/2}$, the long-run standard deviation of Δy_t conditional on Δx_t, while the denominator, $\omega_{uu.x}^{1/2}$, is the long-run standard deviation of u_t conditional on Δx_t. Under cointegration, the former is zero and $\lambda = 0$. Under spurious regression (when $\rho < 1$ is fixed), on the other hand, the right hand side of (8) diverges. Near cointegration corresponds to the intermediate case where the numerator and denominator of (8) are of the same order of magnitude.

In closely related work, Tanaka (1993; 1996, p. 449) has introduced a notion of near cointegration, which might appear to differ slightly from the notion introduced here.[6] Essentially, those works consider the seemingly more general case in which our Assumption A2 is replaced with the following assumption:

A2′. The sequence $\{e_t\}$ is *i.i.d.* with $E(e_t) = 0$ and $E(e_t e_t')$ is positive definite and finite.

[5] For $T \geq 1$, let

$$D_T(L) = \begin{pmatrix} 1 & -\beta_0' \\ 0 & I_{m_x} \end{pmatrix} C(L),$$

where the subscript T on $D(L)$ reflects the fact that β_0 and $C(L)$ depend on T (through ρ). It is not hard to show that the following identification/invertibility condition is sufficient for A5 (iii) to hold:

$$\inf\{|z| : |D_T(z)| = 0\} > 1 \quad \forall T, \lambda > 0.$$

[6] Alternative conditions of near cointegration have appeared in Quintos and Phillips (1993, Section 5) and Phillips (1988a, p. 1025). The (multivariate extension of the) notion of near cointegration introduced by Quintos and Phillips (1993) is more general than the notion suggested here. On the other hand, the notion of near cointegration discussed in Phillips (1988a) is fundamentally different from ours, since the series $\{h'y_t\}$ generated by equation (5) of that paper is nearly integrated.

As discussed previously, working under A2 rather than A2′ entails no loss of generality, so our notion of near cointegration coincides with that of Tanaka. On the other hand, our normalization results in a great simplification of the representation and interpretation of the limiting distributions of interest (cf. the discussion following Theorem 6). For this reason, we prefer the present setup.

3. Behavior of Regression Estimators

Let $\hat{\alpha}$ and $\hat{\beta}$ be the OLS estimators in the multiple regression

$$y_t = \hat{\alpha}' d_t + \hat{\beta}' x_t + \hat{u}_t, \qquad (t = 1, \ldots, T), \tag{9}$$

where $d_t = (1, \ldots, t^{m_d - 1})'$ for some $m_d \geq 1$.[7] In addition to the OLS estimator $\left(\hat{\alpha}', \hat{\beta}'\right)'$, we want to study an estimator that has a compound normal distribution under cointegration. For concreteness, we study the CCR estimator (Park (1992)) $\left(\hat{\alpha}_\dagger', \hat{\beta}_\dagger'\right)'$ obtained from the multiple regression

$$y_t^\dagger = \hat{\alpha}_\dagger' d_t + \hat{\beta}_\dagger' x_t^\dagger + \hat{u}_t^\dagger, \qquad (t = 1, \ldots, T), \tag{10}$$

using the CCR transformed data.[8] Lemma 1 characterizes the limiting behavior of $\left(\hat{\alpha}', \hat{\beta}'\right)$ and $\left(\hat{\alpha}_\dagger', \hat{\beta}_\dagger'\right)$.

Lemma 1. *Suppose* $\{z_t\}$ *is generated by* (1) *and suppose A1-A4 hold.*

(a) When $\rho < 1$ *and fixed (spurious regression),*

$$\begin{aligned} T^{-1}\Psi_T \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} - \beta_0 \end{pmatrix} &\overset{\mathcal{L}_\infty}{=} T^{-1}\Psi_T \begin{pmatrix} \hat{\alpha}_\dagger \\ \hat{\beta}_\dagger - \beta_0 \end{pmatrix} \\ &\overset{\mathcal{L}_\infty}{=} \omega_{yy}^{1/2} \left(1 - \rho^2\right)^{1/2} \left(\int Q_x Q_x'\right)^{-1} \left(\int Q_x U\right), \end{aligned}$$

(b) When A5 holds (near cointegration),

$$\begin{aligned} &\Psi_T \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} - \beta_0 \end{pmatrix} \\ &\overset{\mathcal{L}_\infty}{=} \left(\int Q_x Q_x'\right)^{-1} \left(\omega_{uu.x}^{1/2} \int Q_x dU_\lambda + \int Q_x dX' \Omega_{xx}^{-1} \omega_{xu} + \begin{pmatrix} 0 \\ \gamma_{ux}' \end{pmatrix}\right), \end{aligned}$$

[7] For a justification of the inclusion of d_t in (9), see Remark (i) following Lemma 1.

[8] Alternative estimators with identical asymptotic properties include the estimators proposed by Johansen (1988, 1991), Phillips (1991a, 1991b), Phillips and Hansen (1990), Saikkonen (1991, 1992), and Stock and Watson (1993).

$$\Psi_T \begin{pmatrix} \hat{\alpha}_\dagger \\ \hat{\beta}_\dagger - \beta_0 \end{pmatrix} \overset{\mathcal{L}_\infty}{=} \left(\int Q_x Q_x' \right)^{-1} \left(\left(\omega_{uu}^\dagger\right)^{1/2} \int Q_x dU_\lambda \right),$$

where

$$\Psi_T = \begin{pmatrix} diag\left(T^{1/2}, \ldots, T^{m_d - 1/2}\right) & 0_{m_d \times m_x} \\ 0_{m_x \times m_d} & T \cdot I_{m_x} \end{pmatrix},$$

$Q_x(r)' = \left(D(r)', X(r)'\right)$, $D(r) = (1, r, \ldots, r^{m_d-1})'$, $X(r) = \Omega_{xx}^{1/2} V(r)$,

$$U_\lambda(r) = \lambda \int_0^r U(s)\, ds + U(r),$$

while V and U are independent Wiener processes of dimension m_x and 1, respectively.

Part (a) is well known (e.g. Phillips (1986)), as is part (b) in the case where $\lambda = 0$ (e.g. Phillips and Durlauf (1986)). When $\lambda \neq 0$, the limiting distribution in (b) is a linear combination of the spurious regression distribution reported in (a) and the distribution corresponding to exact cointegration $(\lambda = 0)$. A similar result was obtained by Tanaka (1993, Theorem 6). Under near cointegration, $\hat{\beta}$ and $\hat{\beta}_\dagger$ are super-consistent estimators of β_0. Moreover, the limiting distribution of $T\left(\hat{\beta}_\dagger - \beta_0\right)$ is compound normal (see Remark (iii) below). In important respects, the near cointegration case therefore closely resembles the cointegration case.

Remarks. (i) A non-zero mean of the form $E(z_t) = Ad_t$ (where A is some $m \times m_d$ matrix) is easily accommodated. Suppose we run the regressions

$$Y_t = \hat{\alpha}' d_t + \hat{\beta}' X_t + \hat{u}_t, \qquad (t = 1, \ldots, T),$$

$$Y_t^\dagger = \hat{\alpha}_\dagger' d_t + \hat{\beta}_\dagger' X_t^\dagger + \hat{u}_t^\dagger, \qquad (t = 1, \ldots, T),$$

where $Z_t = (Y_t, X_t')' = Ad_t + z_t$, $Z_t^\dagger = \left(Y_t^\dagger, X_t^{\dagger\prime}\right)' = Ad_t + z_t^\dagger$, while $\{z_t\}$ and $\left\{z_t^\dagger\right\}$ are as before. Then the limiting behavior of $\left(\hat{\alpha}', \hat{\beta}'\right)$ and $\left(\hat{\alpha}_\dagger', \hat{\beta}_\dagger'\right)$ is exactly the same as in Lemma 1 apart from the fact that the limiting distributions of $\hat{\alpha}$ and $\hat{\alpha}_\dagger$ are centered at $\bar{\alpha}_0 = \begin{pmatrix} 1 & -\overline{\beta}_0' \end{pmatrix} A$ rather than zero. In this sense Assumption A4 is merely a normalization whenever the deterministic regressors d_t are included in $(9)-(10)$.

(ii) The results in the case (corresponding to $m_d = 0$) where d_t is omitted from (9) – (10) are completely analogous. Specifically, let $\hat{\beta}$ and $\hat{\beta}_\dagger$ be the OLS estimators in the multiple regressions

$$y_t = \hat{\beta}' x_t + \hat{u}_t, \qquad (t = 1, \ldots, T),$$

$$y_t^\dagger = \hat{\beta}_\dagger' x_t^\dagger + \hat{u}_t^\dagger, \qquad (t = 1, \ldots, T).$$

Then

$$\left(\hat{\beta} - \beta_0\right) \overset{\mathcal{L}_\infty}{=} \left(\hat{\beta}_\dagger - \beta_0\right) \overset{\mathcal{L}_\infty}{=} \omega_{yy}^{1/2} \left(1 - \rho_{xy}^2\right)^{1/2} \left(\int XX'\right)^{-1} \left(\int XU\right),$$

under spurious regression, while

$$T\left(\hat{\beta} - \beta_0\right) \overset{\mathcal{L}_\infty}{=} \left(\int XX'\right)^{-1} \left(\omega_{uu.x}^{1/2} \int X dU_\lambda + \int X dX' \Omega_{xx}^{-1} \omega_{xu} + \gamma_{ux}'\right),$$

$$T\left(\hat{\beta}_\dagger - \beta_0\right) \overset{\mathcal{L}_\infty}{=} \left(\int XX'\right)^{-1} \left(\omega_{uu.x}^{1/2} \int X dU_\lambda\right),$$

under near cointegration.

(iii) Using integration by parts, we obtain

$$\int Q_x dU_\lambda \overset{\mathcal{L}}{=} \int Q_{x,\lambda} dU,$$

where

$$Q_{x,\lambda}(r) = \lambda \underline{Q_x}(r) + Q_x(r), \qquad \underline{Q_x}(r) = \int_r^1 Q_x(s)\, ds.$$

As a consequence, the limiting distribution of

$$\Psi_T \begin{pmatrix} \hat{\alpha}_\dagger \\ \hat{\beta}_\dagger - \beta_0 \end{pmatrix}$$

is compound normal:

$$\left.\left(\int Q_x Q_x'\right)^{-1}\left(\omega_{uu.x}^{1/2}\int Q_x dU_\lambda\right)\right|_{\mathcal{F}_V}$$
$$\overset{\mathcal{L}}{\equiv}\mathcal{N}\left(0,\omega_{uu.x}\left(\int Q_x Q_x'\right)^{-1}\left(\int Q_{x,\lambda}Q_{x,\lambda}'\right)\left(\int Q_x Q_x'\right)^{-1}\right),$$

where " $\cdot|_{\mathcal{F}_V}$ " signifies the conditional distribution relative to $\mathcal{F}_V=\sigma\left(V\left(r\right):0\leq r\leq 1\right)$, the σ-algebra generated by V. ■

4. Inference on Regression Coefficients

This section is concerned with inference on regression coefficients. In Section 4.1, we consider the standard regression F-statistic and demonstrates that it's limiting distribution depends continuously on λ under near cointegration. Section 4.2 considers the behavior of cointegration procedures under near cointegration.

4.1. The *F*-statistic.

Let $F\left(\hat{\beta}\right)$ and $F\left(\hat{\beta}_\dagger\right)$ be the standard F-statistics used to test the null hypothesis $H_0:\beta=\beta_0$ based on the regressions (9) and (10), respectively. As is well known, $F\left(\hat{\beta}\right)$ and $F\left(\hat{\beta}_\dagger\right)$ diverge at rate T under spurious regression (e.g. Phillips (1986)). Indeed, we have:

Lemma 2. *Suppose $\{z_t\}$ is generated by (1) and suppose A1-A4 hold. When $\rho<1$ and fixed,*

$$T^{-1}\left(p\times F\left(\hat{\beta}\right)\right)\overset{\mathcal{L}_\infty}{\equiv}T^{-1}\left(p\times F\left(\hat{\beta}_\dagger\right)\right)\overset{\mathcal{L}_\infty}{\equiv}\frac{\left\|\int\widetilde{V_D}U_D\right\|^2}{\int U_Q^2},$$

where

$$\widetilde{V_D}\left(r\right)=\left(\int_0^1 V_D\left(s\right)V_D\left(s\right)'ds\right)^{-1/2}V_D\left(r\right),$$

$$V_D\left(r\right)=V\left(r\right)-\left(\int_0^1 V\left(s\right)D\left(s\right)'ds\right)\left(\int_0^1 D\left(s\right)D\left(s\right)'ds\right)^{-1}D\left(r\right),$$

$$U_D(r) = U(r) - \left(\int_0^1 U(s) D(s)' ds\right) \left(\int_0^1 D(s) D(s)' ds\right)^{-1} D(r),$$

$$U_Q(r) = U(r) - \left(\int_0^1 U(s) Q(s)' ds\right) \left(\int_0^1 Q(s) Q(s)' ds\right)^{-1} Q(r),$$

$Q(r)' = \left(D(r)', V(r)'\right)$, *while* D, V *and* U *are defined as in Lemma 1.*

Quite remarkably, the limiting distribution of $T^{-1} \cdot F\left(\hat{\beta}\right)$ (and $T^{-1} \cdot F\left(\hat{\beta}_\dagger\right)$) does not depend on any unknown parameters. In particular, it does not depend on ρ. However, as demonstrated by Phillips and Durlauf (1986, Theorem 5.1), the conclusion of the lemma depends crucially on the assumption that $\rho < 1$, since $F\left(\hat{\beta}\right) = O_p(1)$ with a complicated limiting distribution when $\rho = 1$. Theorem 3 generalizes that result to the case of near cointegration.

Theorem 3. *Suppose* $\{z_t\}$ *is generated by* (1) *and suppose A1-A5 hold. Then*

$$p \times F\left(\hat{\beta}\right) \overset{\mathcal{L}_\infty}{=}$$

$$\frac{\omega_{uu.x}}{\sigma_{uu}} \left\| \int \widetilde{V_D} dU_\lambda + \omega_{uu.x}^{-1/2} \left(\int X_D X_D'\right)^{-1/2} \left[\left(\int X_D dX'\right) \Omega_{xx}^{-1} \omega_{xu} + \gamma_{ux}'\right] \right\|^2,$$

$$p \times F\left(\hat{\beta}_\dagger\right) \overset{\mathcal{L}_\infty}{=} \frac{\omega_{uu}^\dagger}{\sigma_{uu}^\dagger} \left\| \int \widetilde{V_D} dU_\lambda \right\|^2,$$

where

$$X_D(r) = X(r) - \left(\int_0^1 X(s) D(s)' ds\right) \left(\int_0^1 D(s) D(s)' ds\right)^{-1} D(r),$$

while X *and* U_λ *are defined as in Lemma 1 and* $\widetilde{V_D}$ *is defined as in Lemma 2.*

Since

$$\int \widetilde{V_D} dU_\lambda \overset{\mathcal{L}}{=} \int V_{D,\lambda} dU,$$

where

$$V_{D,\lambda}(r) = \lambda \underline{V_D}(r) + V_D(r), \qquad \underline{V_D}(r) = \int_r^1 V_D(s)\, ds,$$

we see that the limiting behavior of $F\left(\hat{\beta}\right)$ depends continuously on λ as λ approaches zero. The notion of near cointegration therefore seems to suggest a useful way of bridging the apparent gap between spurious regression and (exact) cointegration.

Remark. Although the motivation underlying the notion of near cointegration is very similar in spirit to the motivation underlying the notion of near integration, the limiting behavior as the noncentrality parameter λ increases without bound is qualitatively different. Under near integration, the asymptotic behavior as the noncentrality parameter approaches it's boundary of definition coincides with the results for the stationary and explosive AR(1)'s (Chan and Wei (1987, Theorem 2), Phillips (1987, Theorem 2)). As emphasized by Phillips (1987, pp. 542-543), these findings do not constitute a rigorous proof of the results for stable and explosive AR(1)'s. None the less, we might expect to discover a close connection between the distributions described in Lemma 2 and Theorem 3. Heuristically, spurious regression corresponds to near cointegration with a "large" λ and in some sense the results in Lemma 2 and Theorem 3 are similar, since both results can be interpreted as suggesting that $F\left(\hat{\boldsymbol{\beta}}\right)$ diverges under spurious regression (letting $T \to \infty$ in Lemma 2 and $\lambda \to \infty$ in the distribution reported in Theorem 3). However, we notice that $1/\lambda^2$ times the limiting distribution in Theorem 3 converges to

$$\frac{\omega_{uu.x}}{\sigma_{uu}} \left\| \int \widetilde{V_D U} \right\|^2,$$

as $\lambda \to \infty$. Therefore, Lemma 2 cannot be deduced from Theorem 3. As such, our results complement Phillips and Moon's (1999, Section 3) recent discussion of multi-index asymptotic theory by providing an illustration of the point that one cannot deduce rigorous asymptotic results that apply for $T \to \infty$ with ρ^2 fixed by telescoping the limits as $T \to \infty$ and $\lambda \to \infty$. ■

4.2. Cointegration Procedures.

Even under cointegration (when $\lambda = 0$), the limiting distribution of $F\left(\hat{\beta}_\dagger\right)$ reported in Theorem 3 is not particularly useful in itself, since it depends on the (unknown) parameter $\omega_{uu}^\dagger/\sigma_{uu}^\dagger$. On the other hand, it is straightforward to modify Wald statistics such as $F\left(\hat{\beta}_\dagger\right)$ in a way that makes the modified test statistic asymptotically pivotal under cointegration. Consider a general linear hypothesis of the form $H_0 : \Phi_\beta \beta = \phi_\beta$, where Φ_β is a $p \times m_x$ matrix of rank p and ϕ_β is a p-vector. Define

$$G\left(\hat{\beta}_{\dagger}\right)=\frac{\left(\Phi_{\beta}\hat{\beta}_{\dagger}-\phi_{\beta}\right)'\left(\Phi_{\beta}\left(\sum_{t=1}^{T}x_{t,d}^{\dagger}x_{t,d}^{\dagger\prime}\right)^{-1}\Phi_{\beta}'\right)^{-1}\left(\Phi_{\beta}\hat{\beta}_{\dagger}-\phi_{\beta}\right)}{\omega_{uu}^{\dagger}}, \tag{11}$$

where

$$x_{t,d}^{\dagger}=x_{t}^{\dagger}-\left(\sum_{s=1}^{T}x_{s}^{\dagger}d_{s}'\right)\left(\sum_{s=1}^{T}d_{s}d_{s}'\right)^{-1}d_{t}.$$

Under cointegration, $G\left(\hat{\beta}_{\dagger}\right)\overset{\mathcal{L}_{\infty}}{=}\chi^{2}\left(p\right)$ when H_0 is true. More generally, under near cointegration, we have:

Theorem 4. *Suppose* $\{z_t\}$ *is generated by* (1) *and suppose A1-A5 hold. When* $H_0:\Phi_{\beta}\beta=\phi_{\beta}$ *is true,*

$$G\left(\hat{\beta}_{\dagger}\right)\overset{\mathcal{L}_{\infty}}{=}\left\|\int\widetilde{V_{D}^{p}}dU_{\lambda}\right\|^{2},$$

where

$$\widetilde{V_{D}^{p}}\left(r\right)=\left(\int_{0}^{1}V_{D}^{p}\left(s\right)V_{D}^{p}\left(s\right)'ds\right)^{-1/2}V_{D}^{p}\left(r\right),$$

$$V_{D}^{p}\left(r\right)=V_{1D}\left(r\right)-\left(\int_{0}^{1}V_{1D}\left(s\right)V_{2D}\left(s\right)'ds\right)\left(\int_{0}^{1}V_{2D}\left(s\right)V_{2D}\left(s\right)'ds\right)^{-1}V_{2D}\left(r\right),$$

$$V_{D}\left(r\right)=\left(\begin{array}{c}V_{1D}\left(r\right)\\V_{2D}\left(r\right)\end{array}\right)\begin{array}{l}\updownarrow p\\\updownarrow m_{x}-p\end{array},$$

while V_D *and* U_{λ} *are defined as in Lemmas 2 and 1, respectively.*

Remarks. (i) Using integration by parts, we obtain

$$\int\widetilde{V_{D}^{p}}dU_{\lambda}\overset{\mathcal{L}}{=}\int V_{D,\lambda}^{p}dU,$$

where

$$V_{D,\lambda}^{p}(r) = \lambda \underline{\widetilde{V_D^p}}(r) + \widetilde{V_D^p}(r), \qquad \underline{\widetilde{V_D^p}}(r) = \int_r^1 \widetilde{V_D^p}(s)\, ds.$$

Now,

$$\int_0^1 \widetilde{V_D^p}(r)\, \widetilde{V_D^p}(r)'\, dr = I_p,$$

$$\int_0^1 \underline{\widetilde{V_D^p}}(r)\, \widetilde{V_D^p}(r)'\, dr + \int_0^1 \widetilde{V_D^p}(r)\, \underline{\widetilde{V_D^p}}(r)'\, dr = \left(\int_0^1 \widetilde{V_D^p}(r)\, dr\right)\left(\int_0^1 \widetilde{V_D^p}(r)\, dr\right)' = 0,$$

so

$$\left.\int \widetilde{V_D^p} dU_\lambda\right|_{\mathcal{F}_V} \overset{\mathcal{L}}{=} \mathcal{N}\left(0, I_p + \lambda^2 \int_0^1 \underline{\widetilde{V_D^p}}(r)\, \underline{\widetilde{V_D^p}}(r)'\, dr\right),$$

where, once more, " $\cdot|_{\mathcal{F}_V}$ " signifies the conditional distribution relative to $\mathcal{F}_V = \sigma\left(V(r) : 0 \leq r \leq 1\right)$. As a consequence,

$$\left.\left\|\int \widetilde{V_D^p} dU_\lambda\right\|^2\right|_{\mathcal{F}_V} \overset{\mathcal{L}}{=} \sum_{i=1}^{p} \left(1 + \lambda^2 \cdot \mu_i\right) \chi_i^2(1),$$

where $\{\chi_i^2(1)\}_{i=1}^p$ are *i.i.d.* $\chi^2(1)$ variables and $0 \leq \mu_1 \leq \ldots \leq \mu_p$ are the eigenvalues of the matrix

$$\int_0^1 \underline{\widetilde{V_D^p}}(r)\, \underline{\widetilde{V_D^p}}(r)'\, dr.$$

The random variable $\left\|\int \widetilde{V_D^p} dU_\lambda\right\|^2$ therefore has a complicated mixture distribution whenever $\lambda \neq 0$, whereas $\left\|\int \widetilde{V_D^p} dU_\lambda\right\|^2 \overset{\mathcal{L}}{=} \chi^2(p)$ under cointegration (when $\lambda = 0$).

(ii) Theorem 4 can be generalized to the case of a nonlinear hypothesis of the form $H_0 : \phi(\theta) = 0$, where $\theta' = (\alpha', \beta')$ and $\phi : \mathbb{R}^{m_d + m_x} \to \mathbb{R}^p$ is assumed to be continuously differentiable with Jacobian $\Phi(\theta) = \partial\phi/\partial\theta'$. For brevity, we merely state the result, whose proof is a bit more tedious than that of Theorem 4 due to the fact that the elements of $\hat{\theta}_{\dagger}'$ converge at different rates. Define

$$H\left(\hat{\theta}_{\dagger}\right) = \frac{\phi\left(\hat{\theta}_{\dagger}\right)'\left(\Phi\left(\hat{\theta}_{\dagger}\right)\left(\sum_{t=1}^{T} q_t^{\dagger} q_t^{\dagger\prime}\right)^{-1}\Phi\left(\hat{\theta}_{\dagger}\right)'\right)^{-1}\phi\left(\hat{\theta}_{\dagger}\right)}{\omega_{uu}^{\dagger}}, \tag{12}$$

where $\hat{\theta}_{\dagger}' = \left(\hat{\alpha}_{\dagger}', \hat{\beta}_{\dagger}'\right)$ and $q_t^{\dagger\prime} = \left(d_t', x_t^{\dagger\prime}\right)$. Suppose $\Phi\left(\bar{\theta}_0\right)$ has rank p, where $\bar{\theta}_0' = \left(0', \bar{\beta}_0'\right)$. Then we can find a sequence $\{\Lambda_T\}_{T\geq 1}$ of invertible $p \times p$ matrices along with a full (row) rank $p \times (m_d + m_x)$ matrix Φ_θ such that $\Lambda_T^{-1}\Phi\left(\theta_0\right)\Psi_T \to \Phi_\theta$. When H_0 is true and A1-A5 hold,

$$H\left(\hat{\theta}_{\dagger}\right) \overset{\mathcal{L}_\infty}{=} \left\|\int \widetilde{Q_x^{\Phi_\theta}} dU_\lambda\right\|^2,$$

where

$$\widetilde{Q_x^{\Phi_\theta}}(r) = \left(\int_0^1 Q_x^{\Phi_\theta}(s)\, Q_x^{\Phi_\theta}(s)'\, ds\right)^{-1/2} Q_x^{\Phi_\theta}(r),$$

$$Q_x^{\Phi_\theta}(r) = \Phi_\theta \left(\int_0^1 Q_x(s)\, Q_x(s)'\, ds\right)^{-1} Q_x(r),$$

while Q_x and U_λ are defined as in Lemma 1. As in remark (i), the limiting distribution is a complicated mixture distribution whenever $\lambda \neq 0$, whereas $H\left(\hat{\theta}_{\dagger}\right) \overset{\mathcal{L}_\infty}{=} \chi^2(p)$ under cointegration. ■

Recently, Elliott (1998) has investigated the robustness of cointegration methods by considering a model in which the regressors are nearly integrated while some linear combination of the regressand and the regressor is exactly stationary. It turns out that the aforementioned χ^2 result can break down when the regressors are not exactly integrated. Theorem 4 enables us to conduct a complimentary experiment: we can investigate the behavior of cointegration methods in a model where the regressors are exactly integrated while some linear combination of the regressand and the regressors is nearly stationary.

It follows from the preceding Remark (i) that the family

$$\left\{\mathcal{L}\left(\left\|\int \widetilde{V_D^p} dU_\lambda\right\|^2\right) : \lambda \geq 0\right\}$$

is stochastically increasing in λ. In other words, $P\left(\left\|\int \widetilde{V_D^p} dU_\lambda\right\|^2 \le t\right)$ is a strictly decreasing function of λ for all $t > 0$ and, in particular,

$$P\left(\left\|\int \widetilde{V_D^p} dU_\lambda\right\|^2 \le t\right) < P\left(\chi^2(p) \le t\right),$$

for all $t > 0$ whenever $\lambda \neq 0$. As a consequence, tests based on the distribution applicable under cointegration (the $\chi^2(p)$ distribution) are over-sized (asymptotically) under near cointegration. For concreteness, consider the case where $m_d = 1$, $\Phi_\beta = I_{m_x}$ and $\phi_\beta = \beta_0$. In other words, consider the null hypothesis $H_0 : \beta = \beta_0$ in a regression of $y_t^\dagger$ on $x_t^\dagger$ and a constant. To illustrate the magnitude of the size distortions encountered under near cointegration, we have simulated the limiting distribution of $G\left(\hat{\beta}_\dagger\right)$ for $m_x = 1, \ldots, 4$ and for various values of λ. Specifically, we have made 20,000 draws from the distribution of the discrete approximations (using 2,000 steps) to the limiting random variables. Figure 1 plots the rejection frequencies corresponding to a test with a nominal size of 5%.

FIGURE 1 ABOUT HERE

The evidence presented in Figure 1 suggests that severe size distortions can occur if conventional cointegration methods are being used when the series are nearly cointegrated rather than exactly cointegrated. In fact, the size increases dramatically as (the absolute value of) λ increases from 0 and substantial size distortions are encountered even for values of λ in the range 5 to 10. Whether or not this is a problem obviously depends on whether or not researchers can be expected to be able to detect such departures from exact cointegration. It is therefore of interest to know whether or not tests for cointegration can be expected to reject the null hypothesis of cointegration when λ is equal to 10, say. A partial answer to this question is provided in the next section, where we illustrate how to obtain the local power functions of several available tests for cointegration.

5. Local Power of Cointegration Tests

During the last decade, numerous cointegration tests taking cointegration as the null hypothesis have been proposed. These test procedures utilize different properties of cointegrated systems and it therefore seems desirable to investigate what, if anything, can be said about the power properties of the different tests. In this section, we characterize the behavior of several regression based cointegration tests[9] under local alternatives and obtain the corresponding local power functions.

[9] Harris (1997) and Snell (1998) have proposed tests for cointegration that utilize principal component methods, while Breitung (1998) has developed a test based on canonical correlation analysis. These tests are not considered here.

All of the cointegration test procedures under study here involve an asymptotically efficient estimation procedure in their original formulations. Different authors have advocated different procedures, but all of the testing procedures can be based on any one of the available estimation procedures. For concreteness, we have decided to present test statistics based on the CCR procedure. This allows us to give a simple, unified treatment that focuses on the question of interest without complicating the discussion unnecessarily. We emphasize, though, that some of the test statistics presented below differ slightly from the test statistics proposed in the original papers.

This section is divided into four parts. Section 5.1 deals with tests based on the variable addition procedure. In Section 5.2, we study tests based on partial score sums, while Section 5.3 is concerned with tests based on residuals from an I(2) regression. Finally, Section 5.4 obtains the local power functions of the different tests and addresses the following important questions:

(i) Does any one of these tests dominate the others in terms of local power?

(ii) Can cointegration tests be expected to detect those departures from cointegration that seriously distort the size of conventional cointegration procedures (cf. Section 4.2)?

5.1. Variable Addition Tests.

The variable addition test procedure proposed by Park (1990) can be motivated using the results from Section 4: under cointegration, appropriately constructed Wald tests (such as $G\left(\hat{\beta}_{\dagger}\right)$ and $H\left(\hat{\theta}_{\dagger}\right)$) on (subsets of) regression coefficients have limiting χ^2 distributions, while they diverge under spurious regression. As a consequence, the null of cointegration can be tested by means of a variable addition test where superfluous regressors are added to (10).

Let k_1 and k_2 be arbitrary non-negative integers such that $k = k_1 + k_2 \geq 1$ and for $t = 1, \ldots, T$, let $r_{1t} = \left(t^{m_d}, \ldots, t^{m_d+k_1-1}\right)'$ (if $k_1 \geq 1$) and (if $k_2 \geq 1$) let $\{r_{2t}\}$ be a k_2-dimensional computer generated random walk such that $\{\Delta r_{2t}\} \sim i.i.d.\ \mathcal{N}\left(0, I_{k_2}\right)$.[10] Finally, let $r_t' = \left(r_{1t}', r_{2t}'\right)$.

Based on the multiple regressions (10) and

$$y_t^{\dagger} = \ddot{\alpha}_{\dagger}' d_t + \ddot{\beta}_{\dagger}' x_t^{\dagger} + \ddot{\gamma}_{\dagger}' r_t + \ddot{u}_t^{\dagger}, \qquad (t = 1, \ldots, T), \tag{13}$$

construct the statistic

$$J_1\left(k_1, k_2\right) = \frac{\sum_{t=1}^{T}\left(\hat{u}_t^{\dagger}\right)^2 - \sum_{t=1}^{T}\left(\ddot{u}_t^{\dagger}\right)^2}{\omega_{uu}^{\dagger}}. \tag{14}$$

[10] This particular choice of superfluous regressors is advocated by Park (1990, Remark b). On the other hand, little guidance on the optimal choice of k_1 and k_2 is provided although Remark c of the paper suggests that $k_1 + k_2 \geq 2$ is preferable.

This is simply the Wald test used to test the significance of the regressor r_t in (13). Under the null hypothesis of cointegration, $J_1(k_1,k_2) \overset{\mathcal{L}_\infty}{\cong} \chi^2(k)$. More generally, under near cointegration, we have:

Theorem 5. *Suppose* $\{z_t\}$ *is generated by* (1) *and suppose A1-A5 hold. Then*

$$J_1(k_1,k_2) \overset{\mathcal{L}_\infty}{\cong} \left\| \int \widetilde{R_Q} dU_\lambda \right\|^2,$$

where

$$\widetilde{R_Q}(r) = \left(\int_0^1 R_Q(s) R_Q(s)' ds \right)^{-1/2} R_Q(r),$$

$$R_Q(r) = R(r) - \left(\int_0^1 R(s) Q(s)' ds \right) \left(\int_0^1 Q(s) Q(s)' ds \right)^{-1} Q(r),$$

$R(r)' = (R_1(r)', R_2(r)')$, $R_1(r) = (r^{m_d}, \ldots, r^{m_d+k_1-1})'$, R_2 *is a* k_2*-dimensional Wiener process independent of* Q *and* U_λ, *while* Q *and* U_λ *are defined as in Lemmas 2 and 1, respectively.*

5.2. Tests Based on Partial Score Sums.

Several cointegration tests based on partial score sums have been proposed. We shall consider the tests due to Shin (1994) and Hansen (1992b). Closely related tests have been proposed by Harris and Inder (1994), Kuo (1998), Leybourne and McCabe (1993), McCabe, Leybourne, and Shin (1997), Quintos and Phillips (1993), and Tanaka (1996, Section 11.6.2).

Shin's (1994) test is based on[11]

$$CI = \frac{T^{-2}\sum_{t=1}^T \left(\hat{S}_t^*\right)^2}{\omega_{uu}^\dagger}, \tag{15}$$

where $\hat{S}_t^* = \sum_{s=1}^t \hat{u}_s^\dagger$. This is simply the stationarity test proposed by Kwiatkowski, Phillips, Schmidt, and Shin (1992) applied to the residuals $\left\{\hat{u}_t^\dagger\right\}$ from (10).

[11] In it's original formulation, Shin's (1994) test uses Saikkonen's (1991) estimator.

Hansen (1992b) notes that a test of cointegration can be based on[12]

$$L_c = T^{-1} \frac{\sum_{t=1}^{T} \hat{S}_t^{**\prime} \hat{S}_t^{**}}{\omega_{uu}^{\dagger}}, \tag{16}$$

where

$$\hat{S}_t^{**} = \left(\sum_{s=1}^{T} q_s^{\dagger} q_s^{\dagger\prime} \right)^{-1/2} \sum_{s=1}^{t} q_s^{\dagger} \hat{u}_s^{\dagger}, \qquad q_t^{\dagger\prime} = \left(d_t^{\prime}, x_t^{\dagger\prime} \right).$$

Theorem 6. *Suppose* $\{z_t\}$ *is generated by* (1) *and suppose A1-A5 hold. Then*

$$CI \overset{\mathcal{L}_\infty}{\cong} \int \left(U_{\lambda,Q}^{*} \right)^2,$$

$$L_c \overset{\mathcal{L}_\infty}{\cong} \int \left\| U_{\lambda,Q}^{**} \right\|^2,$$

where

$$U_{\lambda,Q}^{*}(r) = \int_0^r dU_\lambda(s) - \left(\int_0^r \widetilde{Q}(s)^{\prime} ds \right) \left(\int_0^1 \widetilde{Q}(s) \, dU_\lambda(s) \right),$$

$$U_{\lambda,Q}^{**}(r) = \int_0^r \widetilde{Q}(s) \, dU_\lambda(s) - \left(\int_0^r \widetilde{Q}(s) \widetilde{Q}(s)^{\prime} ds \right) \left(\int_0^1 \widetilde{Q}(s) \, dU_\lambda(s) \right),$$

$$\widetilde{Q}(r) = \left(\int_0^1 Q(s) \, Q(s)^{\prime} ds \right)^{-1/2} Q(r),$$

while Q *and* U_λ *are defined as in Lemmas 2 and 1, respectively.*

Tanaka (1996, Theorem 11.11) reports a result very similar to the result for CI. However, the limiting distribution reported there depends on an m-dimensional parameter. In contrast, both limiting distributions reported here only depend on a scalar parameter, λ. Therefore, our notion of near cointegration yields much simpler representations of the limiting distributions than the notion introduced by Tanaka. In turn, this enables us to visualize our results in an easily interpretable manner.

[12] In Hansen (1992b), the L_c test is based on Phillips and Hansen's (1990) estimator.

5.3. Tests Based on Residuals from an I(2) Regression.

Choi and Ahn (1995) propose three cointegration tests based on the residuals $\{\check{S}_t\}$ from the multiple regression

$$S_t^y = \check{\alpha}' S_t^d + \check{\beta}' S_t^x + \check{S}_t,$$

where $S_t^y = \sum_{s=1}^t y_s^\dagger$, $S_t^d = \sum_{s=1}^t d_s$ and $S_t^x = \sum_{s=1}^t x_s^\dagger$. Consider the test statistics

$$LM_I = \left[\frac{T^{-1} \sum_{t=2}^T \check{S}_{t-1} \Delta \check{S}_t - \frac{1}{2} \left(\omega_{uu}^\dagger - \sigma_{uu}^\dagger \right)}{\omega_{uu}^\dagger} \right]^2, \tag{17}$$

$$LM_{II} = \frac{\left[T^{-1} \sum_{t=2}^T \check{S}_{t-1} \Delta \check{S}_t - \frac{1}{2} \left(\omega_{uu}^\dagger - \sigma_{uu}^\dagger \right) \right]^2}{\omega_{uu}^\dagger \cdot T^{-2} \sum_{t=2}^T \check{S}_{t-1}^2}, \tag{18}$$

$$SBDH_I = \frac{T^{-2} \sum_{t=1}^T \check{S}_t^2}{\omega_{uu}^\dagger}. \tag{19}$$

These tests are intimately related to the stationarity tests proposed by Choi and Ahn (1998).

Theorem 7. *Suppose* $\{z_t\}$ *is generated by* (1) *and suppose A1-A5 hold. Then*

$$LM_I \overset{\mathcal{L}_\infty}{\cong} \left(\int U_{\lambda,\bar{Q}} dU_{\lambda,\bar{Q}} \right)^2,$$

$$LM_{II} \overset{\mathcal{L}_\infty}{\cong} \frac{\left(\int U_{\lambda,\bar{Q}} dU_{\lambda,\bar{Q}} \right)^2}{\int U_{\lambda,\bar{Q}}^2},$$

$$SBDH_I \overset{\mathcal{L}_\infty}{\cong} \int U_{\lambda,\bar{Q}}^2,$$

where

$$U_{\lambda,\bar{Q}}(r) = U_\lambda(r) - \left(\int_0^1 U_\lambda(s) \overline{Q}(s)' ds \right) \left(\int_0^1 \overline{Q}(s) \overline{Q}(s)' ds \right)^{-1} \overline{Q}(r),$$

$$\overline{Q}(r) = \int_0^r Q(s) ds,$$

while Q *and* U_λ *are defined as in Lemmas 2 and 1, respectively.*

5.4. Local Power Functions.

In order to obtain local power functions, we have simulated (the discrete time counterparts of) the limiting distributions of the $J_1(2,2)$,[13] CI, L_c, LM_I, LM_{II} and $SBDH_I$ test statistics in the case where $m_d = 1$. As in Section 4.2, we have used $2,000$ steps and have repeated the procedure 20,000 times. Figures 2-5 show the local power functions for $m_x = 1, \ldots, 4$. The size of the tests is 5%.

FIGURE 2 ABOUT HERE

FIGURE 3 ABOUT HERE

FIGURE 4 ABOUT HERE

FIGURE 5 ABOUT HERE

In short, the figures suggest that the local power properties of $J_1(2,2)$, CI, L_c and $SBDH_I$ are very similar, whereas LM_I and (in particular) LM_{II} are remarkably inferior in terms of local power. Since the local power properties of $J_1(2,2)$, CI, L_c and $SBDH_I$ are almost indistinguishable, our tentative conclusion is that the choice among these tests should be guided by finite sample considerations concerning size distortions.

Remarks. (i) Notice that

$$LM_{II} \overset{\mathcal{L}_\infty}{=} \frac{LM_I}{SBDH_I}.$$

Under fixed alternatives (i.e. under spurious regression), LM_I diverges at a faster rate than $SBDH_I$ and a test based on LM_{II} is therefore consistent (Choi and Ahn (1995, Theorem 2)). In contrast, since both LM_I and $SBDH_I$ are bounded under near cointegration, there seem to be no reasons whatsoever to expect that LM_{II} should be better than LM_I in terms of local power. In fact, if the local power of $SBDH_I$ is higher than the local power of LM_I, LM_{II} might be expected to have rather disastrous local power properties and this is indeed what the figures suggest.

(ii) Remark (i) illustrates an important point. As mentioned by Choi and Ahn (1998, p. 46), the difference between LM_I and LM_{II} lies in how the estimate of the information matrix is chosen. Specifically, LM_I is simply the square of the (scaled) first derivative of the log-likelihood function, whereas LM_{II} involves the (scaled) second derivative of the log-likelihood function. With integrated processes, the (scaled) second-derivative of the log-likelihood function will typically converge weakly to a random variable rather than a non-stochastic limit.[14] Therefore, the asymptotic properties of otherwise identical (Lagrange Multiplier) test statistics will often depend

[13] That is, $r_{1t} = (t, t^2)'$ and r_{2t} is a two-dimensional random walk in (13). Changing the values of k_1 and k_2 does not seem to affect the local power of the J_1 test much.

[14] Here, for instance, the scaled second derivative of the log-likelihood function is $SBDH_I - T^{-2}\check{S}_T^2/\omega_{uu}^{\dagger} \overset{\mathcal{L}_\infty}{=} SBDH_I$.

on whether or not they involve the second derivative of the log-likelihood function and some caution should be exercised whenever a test statistic involves the second derivative of the log-likelihood function.

(iii) Another lesson to be learned from our findings is that the rate of divergence under fixed alternatives might be a poor measure of the (local) power properties of a test. In the present example, for instance, LM_{II} and $SBDH_I$ diverge at the same rate under fixed alternatives and LM_I diverges faster than both of these (Choi and Ahn (1995, Theorem 2)). Evidently, figures 2-5 tell an entirely different story.

(iv) A somewhat related point is that the local power of all the test under study here depends solely on λ, whereas the rate of divergence under fixed alternatives depends on the particular non-parametric estimator used to estimate nuisance parameters such as $\omega_{uu}^{\dagger} = \omega_{uu} - \omega_{xu}' \Omega_{xx}^{-1} \omega_{xu}$. Our results, in contrast with existing results, therefore suggest that trying to improve power by letting the lag truncation number grow slowly (as suggested by e.g. Choi and Ahn (1995, p. 966)) is not worthwhile. Instead, we suggest that the lag truncation number should be chosen so as to minimize finite sample size distortions. ■

In the previous section, we argued that Wald tests based on conventional cointegration methods can encounter severe size distortions when the series are nearly cointegrated and λ exceeds 5. On the other hand, the evidence presented in figures 2-5 indicates that even when $\lambda = 10$ the power of the tests for cointegration can be well below 50%. This suggests that even if the departure from (exact) cointegration is substantial (in the sense that it severely affects the size of the conventional tests), tests for cointegration cannot be expected to detect such departures very frequently. Therefore, whenever a researcher rejects a structural hypothesis (on the coefficient β) using cointegration methods, the result should be interpreted carefully. Indeed, it might be the case that the structural hypothesis is correct, whereas the (possibly auxiliary) assumption of cointegration is not. This of course leaves open the question of how to interpret the coefficient vector in a non-cointegrated system, a question which we shall not attempt to answer here.[15]

6. Concluding Remarks

A notion of near cointegration was proposed and it's usefulness was demonstrated by means of several examples. Throughout, we have deliberately studied the properties of known inference procedures under near cointegration rather than proposed new methods. As a result, several extensions are possible. For instance, a companion paper by one of us (Jansson (2000)) takes the analysis of Section 5 one step further and uses the model of near cointegration to propose a new cointegration test with (essentially) optimal local power properties.

[15] For a recent contribution to this discussion, see Phillips (1998).

7. APPENDIX: PROOFS

This Appendix contains proofs of Lemmas 1-2 and Theorems 3-7. To facilitate the proofs, we start with a preliminary lemma. The lemma follows from standard results (Phillips and Solo (1992), Phillips (1988b), Hansen (1992a)) and is stated without proof.

Lemma 8. *Let* $q_t' = (d_t', x_t')$, $v_t = y_t - \beta_0' x_t$, $q_t^{\dagger\prime} = \left(d_t', x_t^{\dagger\prime}\right)$ *and* $v_t^\dagger = y_t^\dagger - \beta_0' x_t^\dagger$. *Suppose* $\{z_t\}$ *is generated by* (1) *and suppose A1-A4 hold. Then*

(a) $T^{1/2}\Psi_T^{-1} q_{[T\mu]} \overset{\mathcal{L}_\infty}{=} Q_x(\mu)$,
(a†*)* $T^{1/2}\Psi_T^{-1} q_{[T\mu]}^\dagger \overset{\mathcal{L}_\infty}{=} Q_x(\mu)$,

where $[T\mu]$ *denotes the integer part of* $T\mu$. *Moreover, if* $\rho < 1$ *and fixed,*

(b) $T^{-1/2} v_{[T\mu]} \overset{\mathcal{L}_\infty}{=} \omega_{yy}^{1/2}\left(1-\rho^2\right)^{1/2} U(\mu)$,
(b†*)* $T^{-1/2} v_{[T\mu]} \overset{\mathcal{L}_\infty}{=} \omega_{yy}^{1/2}\left(1-\rho^2\right)^{1/2} U(\mu)$.

On the other hand, if A5 holds,

(c†*)* $T^{-1/2}\sum_{t=1}^{[T\mu]} v_t^\dagger \overset{\mathcal{L}_\infty}{=} \left(\omega_{uu}^\dagger\right)^{1/2} U_\lambda(\mu)$,
(d†*)* $T^{-1}\sum_{t=2}^{[T\mu]}\left(\sum_{s=1}^{t-1} v_s^\dagger\right) v_t^\dagger \overset{\mathcal{L}_\infty}{=} \omega_{uu}^\dagger \int_0^\mu U_\lambda(\tau)\, dU_\lambda(\tau) + \frac{1}{2}\left(\omega_{uu}^\dagger - \sigma_{uu}^\dagger\right)\mu$,
(e) $\Psi_T^{-1}\sum_{t=1}^{[T\mu]} q_t v_t \overset{\mathcal{L}_\infty}{=} \omega_{uu.x}^{1/2}\int_0^\mu Q_x(\tau)\, dU_\lambda(\tau) + \int_0^\mu Q_x(\tau)\, dX(\tau)'\, \Omega_{xx}^{-1}\omega_{xu} + \begin{pmatrix} 0 \\ \gamma_{ux}' \end{pmatrix}\mu$,
(e†*)* $\Psi_T^{-1}\sum_{t=1}^{[T\mu]} q_t^\dagger v_t^\dagger \overset{\mathcal{L}_\infty}{=} \left(\omega_{uu}^\dagger\right)^{1/2}\int_0^\mu Q_x(\tau)\, dU_\lambda(\tau)$,
(f) $T^{1/2}\Upsilon_T^{-1} r_{[T\mu]} \overset{\mathcal{L}_\infty}{=} R(\mu)$,
(g†*)* $\Upsilon_T^{-1}\sum_{t=1}^{[T\mu]} r_t v_t^\dagger \overset{\mathcal{L}_\infty}{=} \left(\omega_{uu}^\dagger\right)^{1/2}\int_0^\mu R(\tau)\, dU_\lambda(\tau)$,
(h) $T^{-1}\sum_{t=1}^{T} v_t^2 \overset{\mathcal{L}_\infty}{=} \sigma_{uu}$,
(h†*)* $T^{-1}\sum_{t=1}^{T}\left(v_t^\dagger\right)^2 \overset{\mathcal{L}_\infty}{=} \sigma_{uu}^\dagger$,

where

$$\Upsilon_T = \begin{pmatrix} diag\left(T^{m_d+1/2}, \ldots, t^{m_d+k_1-1/2}\right) & 0_{k_1\times k_2} \\ 0_{k_2\times k_1} & T\cdot I_{k_2} \end{pmatrix},$$

while $\Psi_T, Q_x, U, U_\lambda, X$ *and* R *are defined as in the text.*

In Lemma 8, all random variables are understood to be functionals defined on the unit interval. Equipped with this Lemma, Lemmas 1-2 and Theorems 3-7 can be established using conventional techniques. We merely outline the proofs.

7.1. Proof of Lemma 1. The results for $\left(\hat{\alpha}'_{\dagger}, \hat{\beta}'_{\dagger}\right)$ follow immediately from the results for $\left(\hat{\alpha}', \hat{\beta}'\right)$, so it suffices to consider $\left(\hat{\alpha}', \hat{\beta}'\right)$.

(a) We have:

$$T^{-1}\Psi_T \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} - \beta_0 \end{pmatrix} = \left(\Psi_T^{-1}\left(\sum_{t=T}^{T} q_t q_t'\right)\Psi_T^{-1}\right)^{-1}\left(T^{-1}\Psi_T^{-1}\left(\sum_{t=T}^{T} q_t v_t\right)\right).$$

Now, by Lemma 8 (a)-(b) and the continuous mapping theorem (CMT),

$$\Psi_T^{-1}\left(\sum_{t=T}^{T} q_t q_t'\right)\Psi_T^{-1} \overset{\mathcal{L}_\infty}{\cong} \int Q_x Q_x',$$

$$T^{-1}\Psi_T^{-1}\left(\sum_{t=T}^{T} q_t v_t\right) \overset{\mathcal{L}_\infty}{\cong} \omega_{yy}^{1/2}\left(1-\rho^2\right)^{1/2}\int Q_x U,$$

so

$$T^{-1}\Psi_T \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} - \beta_0 \end{pmatrix} \overset{\mathcal{L}_\infty}{\cong} \omega_{yy}^{1/2}\left(1-\rho^2\right)^{1/2}\left(\int Q_x Q_x'\right)^{-1}\left(\int Q_x U\right),$$

as claimed.

(b) We have:

$$\Psi_T \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} - \beta_0 \end{pmatrix} = \left(\Psi_T^{-1}\left(\sum_{t=T}^{T} q_t q_t'\right)\Psi_T^{-1}\right)^{-1}\left(\Psi_T^{-1}\left(\sum_{t=T}^{T} q_t v_t\right)\right).$$

Now, by Lemma 8 (e),

$$\Psi_T^{-1}\left(\sum_{t=T}^{T} q_t v_t\right) \overset{\mathcal{L}_\infty}{\cong} \left(\omega_{uu.x}^{1/2}\int Q_x dU_\lambda + \int Q_x dX'\Omega_{xx}^{-1}\omega_{xu} + \begin{pmatrix} 0 \\ \gamma_{ux}' \end{pmatrix}\right),$$

and we have

$$\Psi_T \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} - \beta_0 \end{pmatrix} \overset{\mathcal{L}_\infty}{\cong} \left(\int Q_x Q_x'\right)^{-1}\left(\omega_{uu.x}^{1/2}\int Q_x dU_\lambda + \int Q_x dX'\Omega_{xx}^{-1}\omega_{xu} + \begin{pmatrix} 0 \\ \gamma_{ux}' \end{pmatrix}\right),$$

as claimed. ■

7.2. Proof of Lemma 2. We have:

$$p \times F\left(\hat{\beta}\right) = \frac{\left\| \left(\sum_{t=1}^{T} x_{t,d} x_{t,d}'\right)^{-1/2} \left(\sum_{t=1}^{T} x_{t,d} v_{t,d}\right) \right\|^2}{(T - m_x - m_d)^{-1} \sum_{t=1}^{T} v_{t,q}^2},$$

where

$$x_{t,d} = x_t - \left(\sum_{s=1}^{T} x_s d_s'\right) \left(\sum_{s=1}^{T} d_s d_s'\right)^{-1} d_t,$$

$$v_{t,d} = v_t - \left(\sum_{s=1}^{T} v_s d_s'\right) \left(\sum_{s=1}^{T} d_s d_s'\right)^{-1} d_t,$$

$$v_{t,q} = v_t - \left(\sum_{s=1}^{T} v_s q_s'\right) \left(\sum_{s=1}^{T} q_s q_s'\right)^{-1} q_t.$$

Now, by Lemma 8 (a)-(b) and CMT,

$$T^{-2} \sum_{t=1}^{T} x_{t,d} x_{t,d}' \overset{\mathcal{L}_\infty}{=} \int X_D X_D',$$

$$T^{-2} \sum_{t=1}^{T} x_{t,d} v_{t,d} \overset{\mathcal{L}_\infty}{=} \omega_{yy}^{1/2} \left(1 - \rho^2\right)^{1/2} \int X_D U_D,$$

$$T^{-2} \sum_{t=1}^{T} v_{t,q}^2 \overset{\mathcal{L}_\infty}{=} \omega_{yy} \left(1 - \rho^2\right) \int U_Q^2,$$

where $X_D = \Omega_{xx}^{1/2} V_D$ (as in Theorem 3). Clearly,

$$\left(\int X_D X_D'\right)^{-1/2} \int X_D U_D = \int \widetilde{V_D} U_D,$$

and the result for $F\left(\hat{\beta}\right)$ follows immediately. Identical arguments can be used to establish the result for $F\left(\hat{\beta}_\dagger\right)$. ■

7.3. Proof of Theorem 3. We have:

$$p \times F\left(\hat{\beta}\right) = \frac{\left\| \left(\sum_{t=1}^{T} x_{t,d} x_{t,d}'\right)^{-1/2} \left(\sum_{t=1}^{T} x_{t,d} v_t\right) \right\|^2}{\left(T - m_x - m_d\right)^{-1} \sum_{t=1}^{T} v_{t,q}^2},$$

where $\{x_{t,d}\}$ and $\{v_{t,q}\}$ are defined as in the proof of Lemma 2. Now, by Lemma 8 (a), (e), (h) and CMT,

$$T^{-1} \sum_{t=1}^{T} x_{t,d} v_t \overset{\mathcal{L}_\infty}{=} \omega_{uu.x} \int X_D dU_\lambda + \left(\int X_D dX' \right) \Omega_{xx}^{-1} \omega_{xu} + \gamma_{ux}',$$

$$T^{-1} \sum_{t=1}^{T} v_{t,q}^2 \overset{\mathcal{L}_\infty}{=} T^{-1} \sum_{t=1}^{T} v_t^2 \overset{\mathcal{L}_\infty}{=} \sigma_{uu},$$

and the result for $F\left(\hat{\beta}\right)$ follows immediately. Identical arguments can be used to establish the result for $F\left(\hat{\beta}_\dagger\right)$. ■

7.4. Proof of Theorem 4. Since

$$T^{-2} \sum_{t=1}^{T} x_{t,d}^{\dagger} x_{t,d}^{\dagger\prime} \overset{\mathcal{L}_\infty}{=} \int X_D X_D',$$

$$T\left(\hat{\beta}_\dagger - \beta_0\right) \overset{\mathcal{L}_\infty}{=} \left(\int X_D X_D' \right)^{-1} \left(\left(\omega_{uu}^{\dagger}\right)^{1/2} \int X_D dU_\lambda \right),$$

and $X_D = \Omega_{xx}^{1/2} V_D$, we have

$$G\left(\hat{\beta}_\dagger\right) \overset{\mathcal{L}_\infty}{=} \left\| \int \widetilde{X_D^{\Phi_\beta}} dU_\lambda \right\|^2,$$

where

$$\widetilde{X_D^{\Phi_\beta}}(r) = \left(\int_0^1 X_D^{\Phi_\beta}(s) X_D^{\Phi_\beta}(s)' ds \right)^{-1/2} X_D^{\Phi_\beta}(r),$$

$$X_D^{\Phi_\beta}(r) = \left(\Phi_\beta \Omega_{xx}^{-1/2\prime}\right)\left(\int_0^1 V_D(s)\, V_D(s)'\, ds\right) V_D(r).$$

Now, the distribution of $\widetilde{X_D^{\Phi_\beta}}$ depends on Φ_β and Ω_{xx} through $\Phi_\beta \Omega_{xx}^{-1/2\prime}$ and is invariant under transformations of the form

$$\Phi_\beta \Omega_{xx}^{-1/2\prime} \quad \to \quad K \Phi_\beta \Omega_{xx}^{-1/2\prime} \mathcal{O},$$

where K is non-singular $p \times p$ matrix and $\mathcal{O}$ is an orthogonal $m_x \times m_x$ matrix. Take $\mathcal{O}$ such that $\Phi_\beta \Omega_{xx}^{-1/2\prime} \mathcal{O} = \begin{pmatrix} L & 0_{p\times(m_x-p)} \end{pmatrix}$, where L is lower triangular. Setting $K = L^{-1}$, it follows that we can assume that $\Omega_{xx} = I_{m_x}$ and $\Phi_\beta = \begin{pmatrix} I_p & 0_{p\times(m_x-p)} \end{pmatrix}$. The conclusion now follows by applying the partitioned inverse formula. ■

7.5. Proof of Theorem 5. We have:

$$\sum_{t=1}^T \left(\hat{u}_t^\dagger\right)^2 - \sum_{t=1}^T \left(\ddot{u}_t^\dagger\right)^2 = \left\| \left(\sum_{t=1}^T r_{t,q^\dagger} r_{t,q^\dagger}'\right)^{-1/2} \left(\sum_{t=1}^T r_{t,q^\dagger} v_t^\dagger\right) \right\|^2,$$

where

$$r_{t,q^\dagger} = r_t - \left(\sum_{s=1}^T r_s q_s^{\dagger\prime}\right)\left(\sum_{s=1}^T q_s^\dagger q_s^{\dagger\prime}\right)^{-1} q_t^\dagger.$$

Now, by Lemma 8 (a†), (e†), (f), (g†) and CMT,

$$T^{-2}\sum_{t=1}^T r_{t,q^\dagger} r_{t,q^\dagger}' \overset{\mathcal{L}_\infty}{\cong} \int R_Q R_Q',$$

$$T^{-1}\sum_{t=1}^T r_{t,q^\dagger} v_t^\dagger \overset{\mathcal{L}_\infty}{\cong} \left(\omega_{uu}^\dagger\right)^{1/2} \int R_Q dU_\lambda,$$

and the result follows. ■

7.6. Proof of Theorem 6. Using Lemma 8 (a†), (c†), (e†) and CMT, we have:

$$T^{-1/2}\hat{S}^{*}_{[Tr]} \overset{\mathcal{L}_\infty}{\cong} \left(\omega^{\dagger}_{uu}\right)^{1/2} U^{*}_{\lambda}(r),$$

so

$$CI = \frac{T^{-2}\sum_{t=1}^{T}\left(\hat{S}^{*}_{t}\right)^{2}}{\omega^{\dagger}_{uu}} \overset{\mathcal{L}_\infty}{\cong} \int \left(U^{*}_{\lambda,Q}\right)^{2},$$

as claimed. Likewise,

$$\Psi_T^{-1}\left(\sum_{s=1}^{T} q^{\dagger}_{s} q^{\dagger\prime}_{s}\right)\Psi_T^{-1} \overset{\mathcal{L}_\infty}{\cong} \int Q_x Q_x'$$

$$\sum_{s=1}^{[Tr]} q^{\dagger}_{s}\hat{u}^{\dagger}_{s} \overset{\mathcal{L}_\infty}{\cong} \left(\omega^{\dagger}_{uu}\right)^{1/2}\left(\int_0^r Q_x(s)\,dU_\lambda(s) - \left(\int_0^r Q_x(s)\,\widetilde{Q}(s)'\,ds\right)\left(\int_0^1 \widetilde{Q}(s)\,dU_\lambda(s)\right)\right),$$

so

$$T^{-1/2}\hat{S}^{**}_{[Tr]} \overset{\mathcal{L}_\infty}{\cong} \left(\omega^{\dagger}_{uu}\right)^{1/2} U^{**}_{\lambda}(r),$$

and the result follows. ■

7.7. Proof of Theorem 7. Using Lemma 8 (a†), (c†)-(e†) and CMT, it is not hard to show that

$$T^{-1/2}\check{S}_{[Tr]} \overset{\mathcal{L}_\infty}{\cong} \left(\omega^{\dagger}_{uu}\right)^{1/2} U_{\lambda,\bar{Q}}(r),$$

$$T^{-1}\sum_{t=2}^{T}\check{S}_{t-1}\Delta\check{S}_{t} \overset{\mathcal{L}_\infty}{\cong} \omega^{\dagger}_{uu}\int_0^1 U_{\lambda,\bar{Q}}(r)\,dU_{\lambda,\bar{Q}}(r) + \frac{1}{2}\left(\omega^{\dagger}_{uu} - \sigma^{\dagger}_{uu}\right).$$

As a consequence,

$$LM_I = \left[\frac{T^{-1} \sum_{t=2}^{T} \check{S}_{t-1} \Delta \check{S}_t - \frac{1}{2} \left(\omega_{uu}^{\dagger} - \sigma_{uu}^{\dagger} \right)}{\omega_{uu}^{\dagger}} \right]^2 \overset{\mathcal{L}_\infty}{=} \left(\int U_{\lambda,\bar{Q}} dU_{\lambda,\bar{Q}} \right)^2,$$

$$SBDH_I = \frac{T^{-2} \sum_{t=1}^{T} \check{S}_t^2}{\omega_{uu}^{\dagger}} \overset{\mathcal{L}_\infty}{=} \int U_{\lambda,\bar{Q}}^2,$$

$$LM_{II} \overset{\mathcal{L}_\infty}{=} \frac{LM_I}{SBDH_I} \overset{\mathcal{L}_\infty}{=} \frac{\left(\int U_{\lambda,\bar{Q}} dU_{\lambda,\bar{Q}} \right)^2}{\int U_{\lambda,\bar{Q}}^2},$$

as claimed. ■

References

Beveridge, S., and C. R. Nelson (1981): "A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the 'Business Cycle'," *Journal of Monetary Economics*, 7, 151–174.

Breitung, J. (1998): "Canonical Correlation Statistics for Testing the Cointegration Rank in a Reversed Order," Discussion Paper 105, SFB 373, Humboldt-Universität zu Berlin.

Chan, B. H., and C. Z. Wei (1987): "Asymptotic Inference for Nearly Nonstationary AR(1) Processes," *Annals of Statistics*, 15, 1050–1063.

Chang, Y., and P. C. B. Phillips (1995): "Time Series Regression with Mixtures of Integrated Processes," *Econometric Theory*, 11, 1033–1094.

Choi, I. (1994): "Spurious Regressions and Residual-Based Tests for Cointegration when Regressors are Cointegrated," *Journal of Econometrics*, 60, 313–320.

Choi, I., and B. C. Ahn (1995): "Testing for Cointegration in a System of Equations," *Econometric Theory*, 11, 952–983.

——— (1998): "Testing the Null of Stationarity for Multiple Time Series," *Journal of Econometrics*, 88, 41–77.

Elliott, G. (1998): "On the Robustness of Cointegration Methods when Regressors Almost have Unit Roots," *Econometrica*, 66, 149–158.

Granger, C. W. J., and P. Newbold (1974): "Spurious Regressions in Econometrics," *Journal of Econometrics*, 2, 111–120.

Hansen, B. E. (1992a): "Convergence to Stochastic Integrals for Dependent Heterogeneous Processes," *Econometric Theory*, 8, 489–500.

——— (1992b): "Tests for Parameter Instability in Regressions with I(1) Processes," *Journal of Business and Economic Statistics*, 10, 321–335.

Harris, D. (1997): "Principal Components Analysis of Cointegrated Time Series," *Econometric Theory*, 13, 529–557.

Harris, D., and B. Inder (1994): "A Test of the Null Hypothesis of Cointegration," in *Nonstationary Time Series Analysis and Cointegration*, ed. by C. Hargreaves, pp. 133–152. Oxford University Press, Oxford.

Jansson, M. (1999): "Consistent Covariance Matrix Estimation for Linear Processes," Department of Economics Working Paper 1999-29, University of Aarhus.

——— (2000): "Point Optimal Invariant Tests of the Null Hypothesis of Cointegration," Manuscript, University of Aarhus.

Johansen, S. (1988): "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12, 231–254.

——— (1991): "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Econometrica*, 59, 1551–1580.

Kuo, B.-S. (1998): "Test for Partial Parameter Instability in Regressions with I(1) Processes," *Journal of Econometrics*, 86, 337–368.

Kwiatkowski, D., P. C. B. Phillips, P. Schmidt, and Y. Shin (1992): "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure are We that Economic Time Series have a Unit Root," *Journal of Econometrics*, 54, 159–178.

Leybourne, S. J., and B. P. M. McCabe (1993): "A Simple Test for Cointegration," *Oxford Bulletin of Economics and Statistics*, 55, 97–103.

McCabe, B. P. M., S. J. Leybourne, and Y. Shin (1997): "A Parametric Approach to Testing the Null of Cointegration," *Journal of Time Series Analysis*, 18, 395–413.

Park, J. Y. (1990): "Testing for Unit Roots and Cointegration by Variable Addition," in *Advances in Econometrics: Co-integration, Spurious Regression, and Unit Roots*, ed. by T. B. Fomby, and G. F. Rhodes, Jr. Greenwich, CT: JAI Press, 107-133.

——— (1992): "Canonical Cointegrating Regressions," *Econometrica*, 60, 119–143.

Park, J. Y., S. Ouliaris, and B. Choi (1988): "Spurious Regressions and Tests for Cointegration," CAE Working Paper 88-07, Cornell University.

Park, J. Y., and P. C. B. Phillips (1989): "Statistical Inference in Regressions with Integrated Processes: Part 2," *Econometric Theory*, 5, 95–131.

Phillips, P. C. B. (1986): "Understanding Spurious Regressions in Econometrics," *Journal of Econometrics*, 33, 311–340.

——— (1987): "Towards a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74, 535–547.

——— (1988a): "Regression Theory for Near-Integrated Time Series," *Econometrica*, 56, 1021–1043.

——— (1988b): "Weak Convergence of Sample Covariance Matrices to Stochastic Integrals Via Martingale Approximations," *Econometric Theory*, 4, 528–533.

——— (1991a): "Optimal Inference in Cointegrated Systems," *Econometrica*, 59, 283–306.

——— (1991b): "Spectral Regression for Cointegrated Time Series," in *Nonparametric and Semiparametric Methods in Economics and Statistics*, ed. by W. A. Barnett, J. Powell, and G. E. Tauchen. Cambridge, UK: CUP, 413-435.

——— (1995): "Fully Modified Least Squares and Vector Autoregression," *Econometrica*, 63, 1023–1078.

——— (1998): "New Tools for Understanding Spurious Regressions," *Econometrica*, 66, 1299–1325.

PHILLIPS, P. C. B., AND S. N. DURLAUF (1986): "Multiple Time Series Regression with Integrated Processes," *Review of Economic Studies*, 53, 473–495.

PHILLIPS, P. C. B., AND B. E. HANSEN (1990): "Statistical Inference in Instrumental Variables Regression with I(1) Variables," *Review of Economic Studies*, 57, 99–125.

PHILLIPS, P. C. B., AND H. R. MOON (1999): "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica*, 67, 1057–1111.

PHILLIPS, P. C. B., AND V. SOLO (1992): "Asymptotics for Linear Processes," *Annals of Statistics*, 20, 971–1001.

QUINTOS, C. E., AND P. C. B. PHILLIPS (1993): "Parameter Constancy in Cointegrating Regressions," *Empirical Economics*, 18, 675–706.

SAIKKONEN, P. (1991): "Asymptotically Efficient Estimation of Cointegration Regressions," *Econometric Theory*, 7, 1–21.

——— (1992): "Estimation and Testing of Cointegrated Systems by an Autoregressive Approximation," *Econometric Theory*, 8, 1–27.

SHIN, Y. (1994): "A Residual-Based Test of the Null of Cointegration Against the Alternative of No Cointegration," *Econometric Theory*, 10, 91–115.

SNELL, A. (1998): "Testing for r Versus $r-1$ Cointegrating Vectors," *Journal of Econometrics*, 88, 151–191.

STOCK, J. H., AND M. W. WATSON (1993): "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems," *Econometrica*, 61, 783–820.

TANAKA, K. (1993): "An Alternative Approach to the Asymptotic Theory of Spurious Regression, Cointegration, and Near Cointegration," *Econometric Theory*, 9, 36–61.

——— (1996): *Time Series Analysis: Nonstationary and Nonivertible Distribution Theory*. New York: Wiley.

Yule, G. U. (1926): "Why Do We Sometimes Get Nonsense-Correlations Between Time-Series? A Study in Sampling and the Nature of Time-Series," *Journal of the Royal Statistical Society*, 89, 1–63.

8. Figures

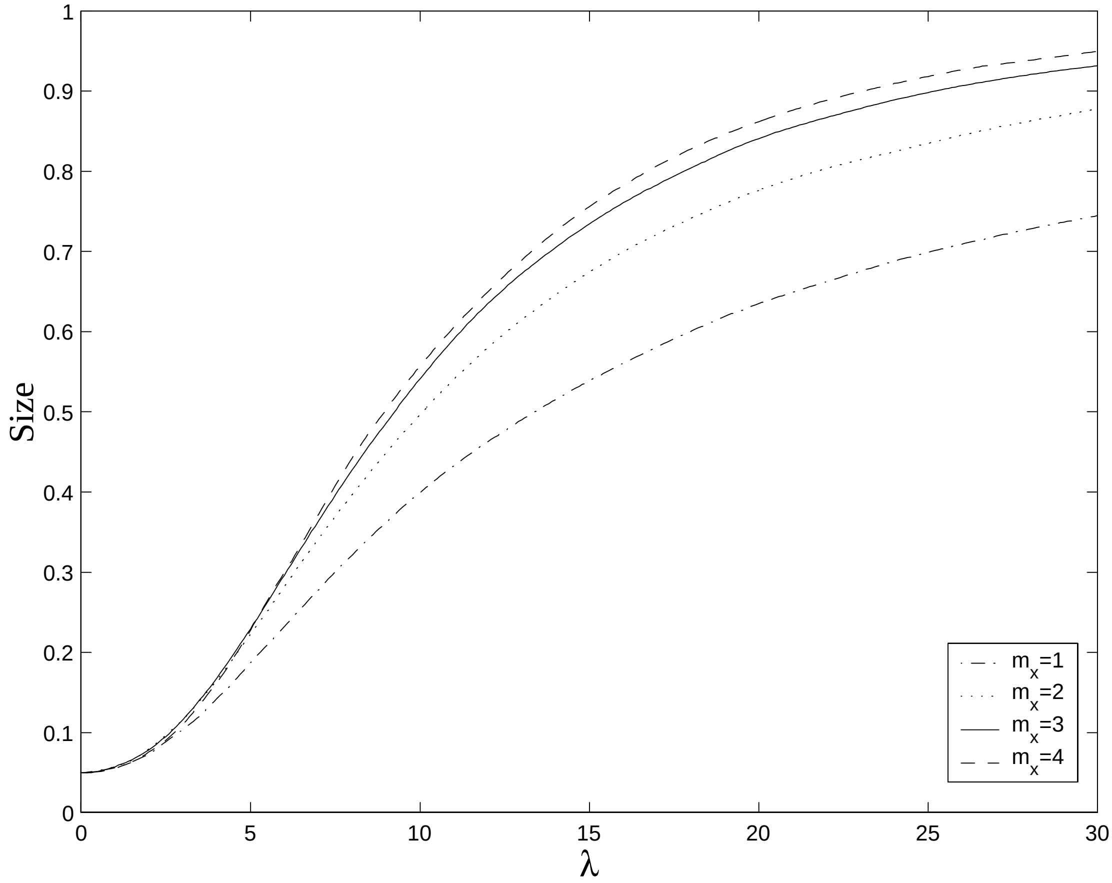

Figure 1: Rejection rates for $G\left(\hat{\beta}_{\dagger}\right)$; Nominal size is 5%.

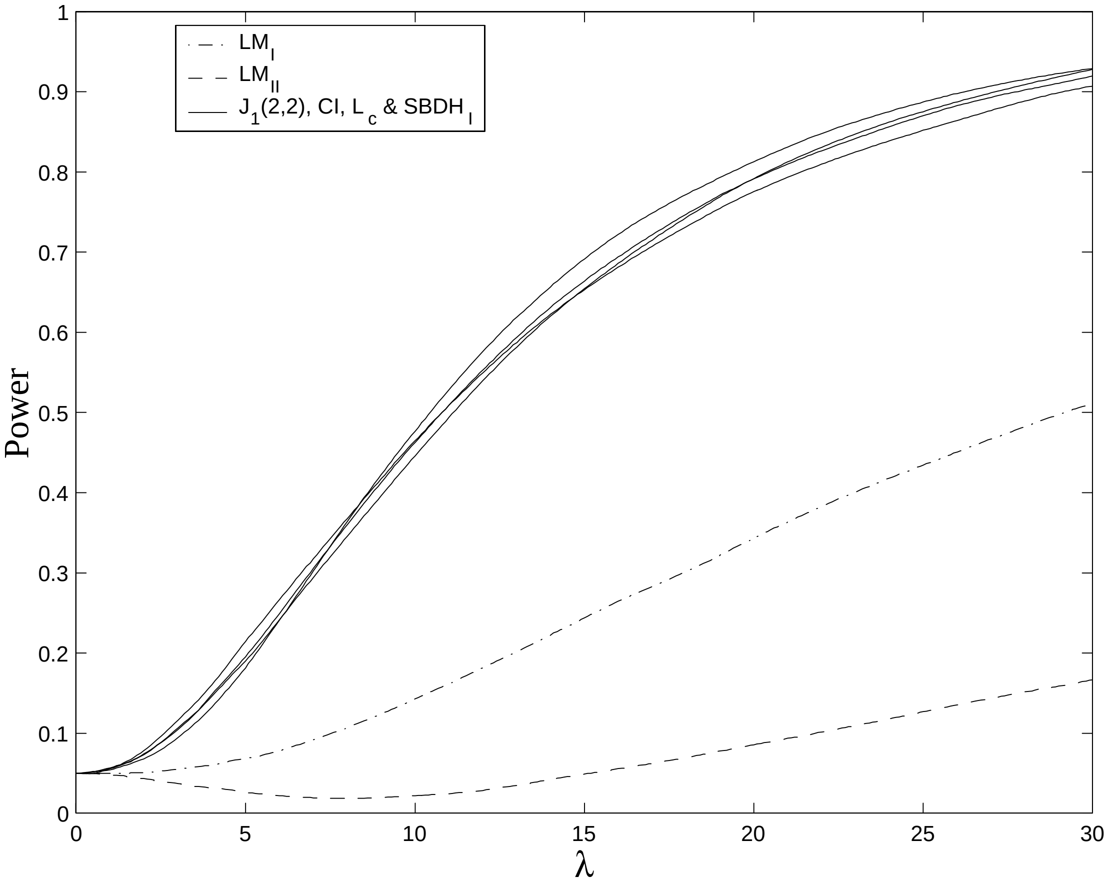

Figure 2: Local Power of Tests for Cointegration; $m_x = 1$.

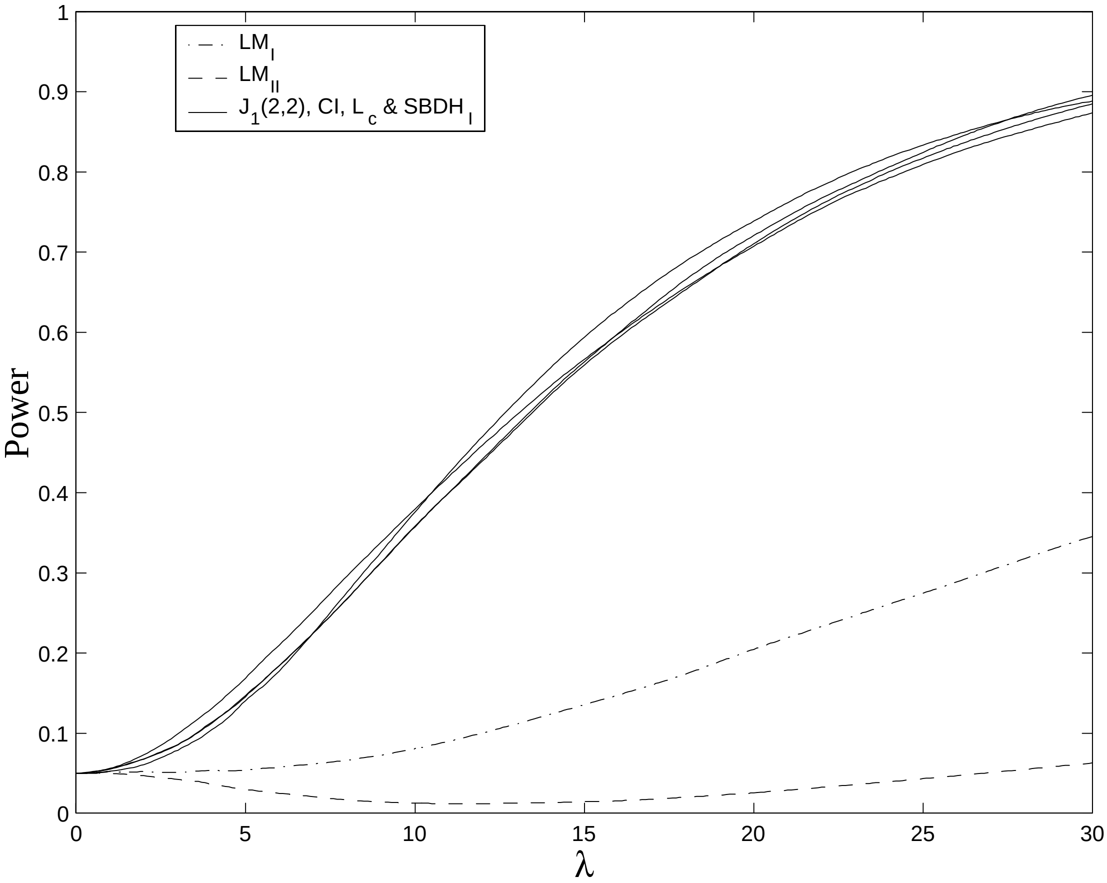

Figure 3: Local Power of Tests for Cointegration; $m_x = 2$.

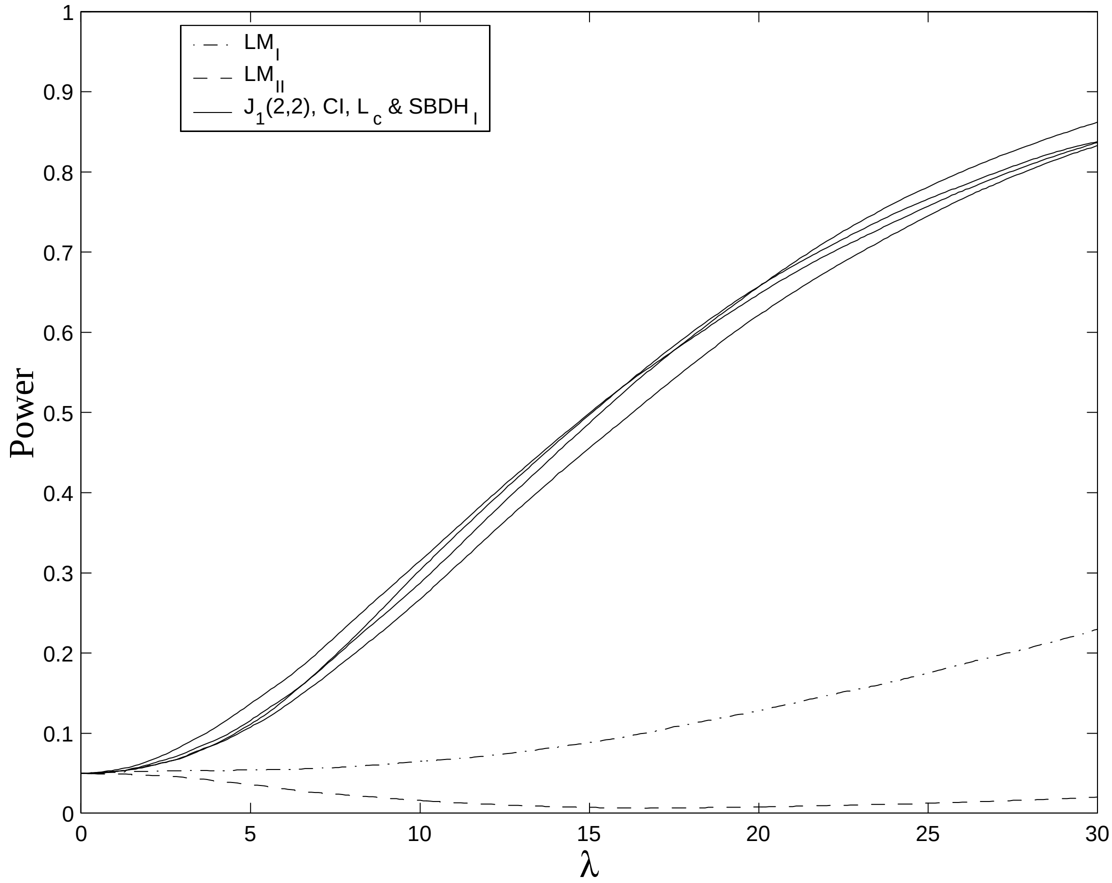

Figure 4: Local Power of Tests for Cointegration; $m_x = 3$.

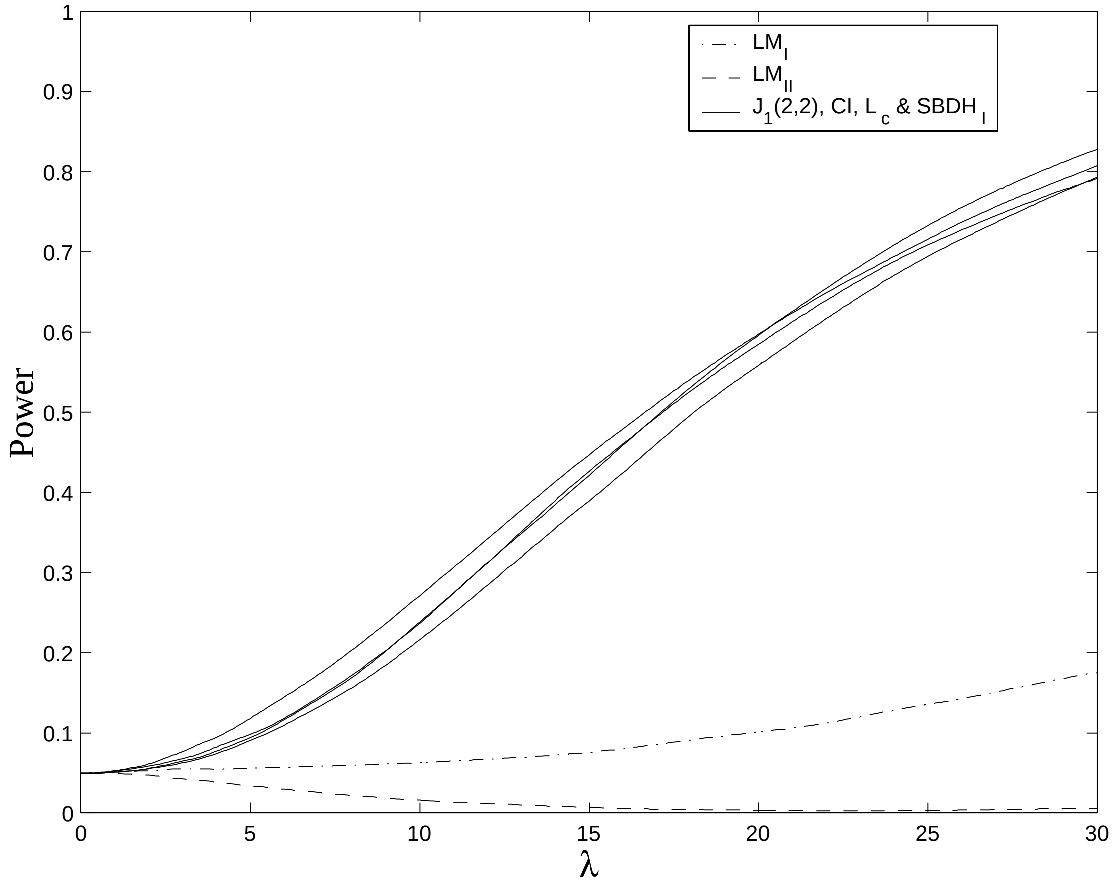

Figure 5: Local Power of Tests for Cointegration; $m_x = 4$.