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

NIELS HALDRUP AND MICHAEL JANSSON*
CENTRE FOR NON-LINEAR MODELLING IN ECONOMICS
DEPARTMENT OF ECONOMICS, UNIVERSITY OF AARHUS
DK-8000 AARHUS C, DENMARK

1 February 1999

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, a notion of near cointegration is proposed and three separate applications of the model of near cointegration are provided. As a first application, we give analytical corroboration of the conjecture that the finite sample behavior of F-statistics based on OLS estimators depends continuously on the aforementioned squared multiple correlation coefficient. Hence, the notion of near cointegration helps to bridge the gap between the polar cases of spurious regression and cointegration. Secondly, we characterize the properties of conventional cointegration methods under near cointegration, hereby investigating the robustness of cointegration methods. Finally, 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. The numerical findings of Granger and Newbold were given an analytical explanation by Phillips (1986), while Park, Ouliaris, and Choi

*This paper has benefited from the comments of H. Peter Boswijk and seminar participants at the Tinbergen Institute, Penn State University, and the 1998 European Meeting of the Econometric Society.

(1988) provided further clarification (see also Park (1990)). 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 (Park, Ouliaris, and Choi (1988, Corollary 2.2)). 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 (Phillips and Durlauf (1986, Theorem 5.1)).

This discontinuity is somewhat discomforting, since intuition suggests 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. In the present paper we provide analytical corroboration of this conjecture. Specifically, we introduce a model in which ρ^2 is a primitive parameter. This model allows us to model ρ^2 as local to unity[1], hereby introducing a notion of near cointegration. Using the model of near cointegration, we obtain a limiting distribution of the F-statistic which depends continuously on the noncentrality parameter measuring the deviation from exact cointegration. Therefore, the model of near cointegration seems to suggest a useful way of bridging the gap between spurious regression and cointegration with respect to the limiting behavior of the F-statistic.

More generally, our model makes it is possible to generalize existing results derived under the assumption of exact cointegration. To demonstrate why such extensions to local alternatives might be useful, we provide two further applications of the model. As a first application, we investigate the robustness of cointegration methods. We do so by characterizing the limiting behavior under near cointegration of the usual Wald statistic devised to test hypotheses on a cointegrating vector. Hence, 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. Secondly, we illustrate how to compute 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 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 cointegration tests under local alternatives

[1] 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.

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, Section 2 introduces a notion of near cointegration. Section 3 discusses the behavior of OLS estimators under spurious regression, cointegration, and near cointegration, while Section 4 contains the corresponding results for the F-statistic based on the OLS estimators. In particular, Section 4 uses the model of near cointegration to provide analytical corroboration of the conjecture that the finite sample behavior of F depends continuously on ρ^2 even when ρ^2 approaches unity. The two further applications of the model of near cointegration alluded to in the previous paragraph are the subjects of Sections 5 and 6. Section 5 investigates the robustness of cointegration methods by characterizing the behavior of a Wald statistic under local alternatives. Similarly, Section 6 reports the behavior of several cointegration tests under near cointegration. Finally, Section 7 offers a few concluding remarks. Proofs of all results of the paper are collected in an Appendix together with a list of notation and definitions of the various stochastic processes appearing throughout the paper.

2. The Model

We assume that $\left\{(y_t, \mathbf{x}_t')'\right\}$ is a $(p+1)$-dimensional zero-mean integrated process generated by

$$\begin{pmatrix} y_t \\ \mathbf{x}_t \end{pmatrix} = \begin{pmatrix} \omega_0\sqrt{1-\rho^2} & \rho\bar{\boldsymbol{\omega}}_{10}'\left(\Omega_1'\right)^{-1} \\ \mathbf{0} & \Omega_1 \end{pmatrix} \begin{pmatrix} \xi_{0t} \\ \boldsymbol{\xi}_{1t} \end{pmatrix} + \mathbf{C}\left(L\right)\mathbf{e}_t, \tag{1}$$

where $\boldsymbol{\xi}_t = (\xi_{0t}, \boldsymbol{\xi}_{1t}')'$ is defined as $\boldsymbol{\xi}_t := \sum_{s=1}^{t}\mathbf{e}_s$, y_t and ξ_{0t} are scalars, $\mathbf{x}_t$ and $\boldsymbol{\xi}_{1t}$ are p-vectors, and

A1. $\{\mathbf{e}_t\}$ is i.i.d. with $\mathrm{E}\left(\mathbf{e}_t\right) = \mathbf{0}$ and $\mathrm{E}\left(\mathbf{e}_t\mathbf{e}_t'\right) = \mathbf{I}_{p+1}$,

A2. $\mathbf{C}\left(L\right) := \sum_{i=0}^{\infty}\mathbf{C}_i L^i$ is a lag polynomial with $\sum_{i=0}^{\infty} i\sqrt{tr\left(\mathbf{C}_i'\mathbf{C}_i\right)} < \infty$,

A3. $\omega_0 > 0$, $0 \le \rho \le 1$, Ω_1 is an upper triangular, nonsingular matrix of dimension p, and $\bar{\boldsymbol{\omega}}_{10}$ is a p-vector satisfying $\bar{\boldsymbol{\omega}}_{10}'\left(\Omega_1\Omega_1'\right)^{-1}\bar{\boldsymbol{\omega}}_{10} = \omega_0^2$.

When $\left\{(y_t, \mathbf{x}_t')'\right\}$ is generated by (1), $\left\{(\Delta y_t, \Delta\mathbf{x}_t')'\right\}$ is a linear process. Conversely, a representation very similar to (1) can be obtained whenever $\left\{(\Delta y_t, \Delta\mathbf{x}_t')'\right\}$ is a linear process by applying the Beveridge-Nelson (1981) decomposition to the original filter. For our purposes, representation (1) is very convenient because it allows us to parameterize the long-run covariance matrix of $\left\{(\Delta y_t, \Delta\mathbf{x}_t')'\right\}$ directly. Assumption A2 is satisfied whenever $\left\{(\Delta y_t, \Delta\mathbf{x}_t')'\right\}$ is a stationary and invertible vector ARMA process. Together, A1 and A2 ensure that we can call upon well known results for linear processes (e.g. Phillips and Solo (1992), Phillips (1988b)) when deriving the results of the paper. At first, the parameterization of the matrix premultiplying $\boldsymbol{\xi}_t$ in (1) might seem slightly peculiar. Notice, however, that the long-run covariance matrix of $(\Delta y_t, \Delta \mathbf{x}_t')'$ is given by[2]

$$\lim_{T\to\infty} T^{-1}\mathrm{E}\begin{pmatrix} y_T^2 & y_T\mathbf{x}_T' \\ \mathbf{x}_T y_T & \mathbf{x}_T\mathbf{x}_T' \end{pmatrix}$$

$$= \begin{pmatrix} \omega_0\sqrt{1-\rho^2} & \rho\bar{\boldsymbol{\varpi}}_{10}'\left(\Omega_1'\right)^{-1} \\ \mathbf{0} & \Omega_1 \end{pmatrix}\begin{pmatrix} \omega_0\sqrt{1-\rho^2} & \rho\bar{\boldsymbol{\varpi}}_{10}'\left(\Omega_1'\right)^{-1} \\ \mathbf{0} & \Omega_1 \end{pmatrix}'$$

$$= \begin{pmatrix} \omega_0^2 & \rho\bar{\boldsymbol{\varpi}}_{10}' \\ \rho\bar{\boldsymbol{\varpi}}_{10} & \Omega_1\Omega_1' \end{pmatrix}. \tag{2}$$

Therefore,

$$\begin{pmatrix} \omega_0\sqrt{1-\rho^2} & \rho\bar{\boldsymbol{\varpi}}_{10}'\left(\Omega_1'\right)^{-1} \\ \mathbf{0} & \Omega_1 \end{pmatrix}$$

is simply an upper triangular matrix "square root" of the long-run covariance matrix of $(\Delta y_t, \Delta \mathbf{x}_t')'$. Moreover, the individual parameters have straightforward interpretations: ω_0^2 is the long-run variance of Δy_t and Ω_1 is an upper triangular matrix "square root" of the long-run covariance matrix of $\Delta \mathbf{x}_t$. The long-run covariance between $\Delta \mathbf{x}_t$ and Δy_t is given by $\rho\bar{\boldsymbol{\varpi}}_{10}$, where $\bar{\boldsymbol{\varpi}}_{10}$ is a p-vector expressing the direction of the covariance. This vector is normalized in such a way that ρ is a unitless quantity measuring the strength of the covariance. In fact, as the notation suggests, ρ is the multiple correlation coefficient computed from the long-run covariance matrix of $(\Delta y_t, \Delta \mathbf{x}_t')'$. Under Assumption A3, $\lim_{T\to\infty} T^{-1}\mathrm{E}\left(y_T^2\right)$ and $\lim_{T\to\infty} T^{-1}\mathrm{E}\left(\mathbf{x}_T\mathbf{x}_T'\right)$ are positive (definite) and finite. This implies that $\{y_t\}$ is an integrated process and that $\{\mathbf{x}_t\}$ is a non-cointegrated integrated process. Admittedly, the assumption that $\{\mathbf{x}_t\}$ is non-cointegrated (i.e. that Ω_1 is nonsingular) 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.

A glance at (1) reveals that the only linear combination of y_t and $\mathbf{x}_t$ that removes the $\boldsymbol{\xi}_{1t}$ component from y_t is $y_t - \boldsymbol{\beta}_0'\mathbf{x}_t$, where $\boldsymbol{\beta}_0 := \left(\Omega_1\Omega_1'\right)^{-1}\rho\bar{\boldsymbol{\varpi}}_{10}$. In other words,

[2]The following discussion assumes that ω_0, ρ, Ω_1 and $\bar{\boldsymbol{\varpi}}_{10}$ are constants. When introducing a notion of near cointegration (Assumption A4 below), we allow ρ to be a function of T.

[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).

$\boldsymbol{\beta}_0$ is the unique value of $\boldsymbol{\beta}$ for which the long-run covariance between the "error" $\Delta y_t - \boldsymbol{\beta}' (\Delta \mathbf{x}_t)$ and the "regressor" $\Delta \mathbf{x}_t$ equals zero. For this reason, we shall follow Park, Ouliaris, and Choi (1988) and refer to $\boldsymbol{\beta}_0$ as the *fundamental coefficient.*

Clearly, $\{(y_t, \mathbf{x}_t')'\}$ is cointegrated if and only if $\{y_t - \boldsymbol{\beta}_0' \mathbf{x}_t\}$ is stationary. Now,

$$y_t - \boldsymbol{\beta}_0' \mathbf{x}_t = \omega_0 \sqrt{1-\rho^2} \xi_{0t} + \begin{pmatrix} 1 & -\boldsymbol{\beta}_0' \end{pmatrix} \mathbf{C}(L) \mathbf{e}_t, \tag{3}$$

and we deduce that the cointegration properties of $\{(y_t, \mathbf{x}_t')'\}$ depend solely on ρ. When $\rho < 1$ (and fixed), $\{y_t - \boldsymbol{\beta}' \mathbf{x}_t\}$ is an integrated process for all values of $\boldsymbol{\beta}$ and we will refer to this as the case of spurious regression. On the other hand, $\{(y_t, \mathbf{x}_t')'\}$ is cointegrated when $\rho = 1$, since $\{y_t - \boldsymbol{\beta}_0' \mathbf{x}_t\}$ is stationary in this case. In addition to these familiar concepts, we introduce a notion of *near cointegration.* We say that $\{(y_t, \mathbf{x}_t')'\}$ is nearly cointegrated if the following assumption holds:

A4. $\sqrt{1-\rho^2} = \min(\kappa, T)/T$ for some $\kappa \geq 0$ and $\sigma^2 > 0$, where

$$\sigma^2 := \begin{pmatrix} 1 & -\bar{\boldsymbol{\beta}}_0' \end{pmatrix} \left(\sum_{i=0}^{\infty} \mathbf{C}_i \mathbf{C}_i' \right) \begin{pmatrix} 1 \\ -\bar{\boldsymbol{\beta}}_0 \end{pmatrix}, \tag{4}$$

and $\bar{\boldsymbol{\beta}}_0 := (\Omega_1 \Omega_1')^{-1} \bar{\boldsymbol{\omega}}_{10}$.[4]

Strictly speaking, ρ (and hence also the fundamental coefficient $\boldsymbol{\beta}_0$) is a sequence of parameters under near cointegration. Similarly, $\{(y_t, \mathbf{x}_t')'\}$ is a triangular array rather than a sequence. Since no ambiguity is likely to arise, we have chosen to simplify the notation slightly by omitting the additional subscript T.

Under near cointegration, equation (2) remains valid, provided ρ is replaced with 1, it's limiting value. So, near cointegration is similar to cointegration in the sense that the long-run covariance matrix of $(\Delta y_t, \Delta \mathbf{x}_t')'$ is independent of κ. In particular, the long-run covariance matrix is singular under both exact cointegration and near cointegration.

[4] The assumption $\sigma^2 > 0$ is needed to rule out the pathological case where $\begin{pmatrix} 1 & -\bar{\boldsymbol{\beta}}_0' \end{pmatrix} \mathbf{C}(L) \mathbf{e}_t = 0$. The crucial implication of Assumption A4 is that $\lim_{T \to \infty} T\sqrt{1-\rho^2}$ exists and is positive. As a consequence, A4 could be replaced with any of the following assumptions:

A4′. $\rho = 1 - \min(\kappa', T)/T^2$ for some $\kappa' \geq 0$ and $\sigma^2 > 0$.

A4″. $\rho^2 = 1 - \min(\kappa'', T)/T^2$ for some $\kappa'' \geq 0$ and $\sigma^2 > 0$.

A4‴. $\rho = \exp(-\kappa'''/T^2)$ for some $\kappa''' \geq 0$ and $\sigma^2 > 0$.

Of course, near cointegration reduces to cointegration when $\kappa = 0$. Under near cointegration both terms on the right hand side of (3) make a non-negligible contribution to the limiting distribution of most statistics that involve $\{y_t - \boldsymbol{\beta}_0'\mathbf{x}_t\}$ and these limiting distributions therefore depend continuously on κ.

In closely related work, Tanaka (1993; 1996, p. 449) has introduced a notion of near cointegration which, although very similar in spirit, differs slightly from the notion introduced here.[5] Essentially, those works consider the seemingly more general case in which Assumption A1 is replaced with the following assumption:

A1′. $\{\mathbf{e}_t\}$ is i.i.d. with $\mathrm{E}(\mathbf{e}_t) = \mathbf{0}$ and $\mathrm{E}(\mathbf{e}_t\mathbf{e}_t')$ is positive definite and finite.

As is readily verified, the only restrictions imposed on the long-run covariance matrix of $(\Delta y_t, \Delta\mathbf{x}_t')'$ in our model are that $\lim_{T\to\infty} T^{-1}\mathrm{E}(y_T^2)$ and $\lim_{T\to\infty} T^{-1}\mathrm{E}(\mathbf{x}_T\mathbf{x}_T')$ are positive (definite) and finite. Therefore, representation (1) is more general than might initially appear to be the case. In particular, working under A1 rather than A1′ entails *no* loss of generality. However, under A1′ one allows ξ_{0t}, the stochastic trend driving $y_t - \boldsymbol{\beta}_0'\mathbf{x}_t$, to be correlated with $\boldsymbol{\xi}_{1t}$, the stochastic trends driving $\mathbf{x}_t$. Therefore, an important difference between A1 and A1′ is that $\boldsymbol{\beta}_0$ always equals the fundamental coefficient under A1, whereas the (pseudo-)true value $\boldsymbol{\beta}_0$ typically differs from the fundamental coefficient under A1′. For this reason, the interpretation of the parameters of the model seems to be more straightforward under A1 than under A1′. Moreover, it turns out that the various distributional results reported in sections 3-6 depend on a scalar parameter under A1, whereas equivalent expressions computed under A1′ involve a $(p+1)$-dimensional parameter (e.g. Tanaka (1996, Theorem 11.11)). To understand this, notice that under A1 the decomposition $y_t = \boldsymbol{\beta}_0'\mathbf{x}_t + (y_t - \boldsymbol{\beta}_0'\mathbf{x}_t)$ is orthogonal in the sense that the long-run covariance between $\Delta(y_t - \boldsymbol{\beta}_0'\mathbf{x}_t)$ and $\Delta\mathbf{x}_t$ is zero. Under A1′, in contrast, the long-run covariance between $\Delta(y_t - \boldsymbol{\beta}_0'\mathbf{x}_t)$ and $\Delta\mathbf{x}_t$ will typically be non-zero. Whenever this is the case, the limiting distributions derived under A1′ depend on a parameter of dimension greater than one and the interpretation of the results seems less straightforward.

3. Behavior of OLS Estimators

Consider the OLS estimators $\hat{\alpha}$ and $\hat{\boldsymbol{\beta}}$ in the multiple regression

$$y_t = \hat{\alpha} + \hat{\boldsymbol{\beta}}'\mathbf{x}_t + \hat{u}_t, \qquad (t = 1, \dots, T) \tag{5}$$

where $\{(y_t, \mathbf{x}_t')'\}$ is generated by (1). It turns out that the limiting behavior of

[5] 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, p. 1025) is fundamentally different from ours, since the series $h'y_t$ generated by equation (5) of that paper is nearly integrated.

$\left(\hat{\alpha}, \hat{\boldsymbol{\beta}}'\right)'$ under near cointegration depends on the following parameters:

$$\boldsymbol{\pi} = \begin{pmatrix} \pi_0 \\ \boldsymbol{\pi}_1 \end{pmatrix} := \left(\sum_{i=0}^{\infty} \mathbf{C}_i\right)' \begin{pmatrix} 1 \\ -\bar{\boldsymbol{\beta}}_0 \end{pmatrix}, \tag{6}$$

$$\boldsymbol{\delta} := \boldsymbol{\pi}_1 + \Omega_1^{-1} \begin{pmatrix} \mathbf{0} & \mathbf{I}_p \end{pmatrix} \left(\sum_{i=0}^{\infty} \mathbf{C}_i \mathbf{C}_i'\right) \begin{pmatrix} 1 \\ -\bar{\boldsymbol{\beta}}_0 \end{pmatrix}, \tag{7}$$

where $\bar{\boldsymbol{\beta}}_0 := (\Omega_1 \Omega_1')^{-1} \bar{\boldsymbol{\omega}}_{10}$. The parameter π_0 is a scalar, and $\boldsymbol{\pi}_1$ and $\boldsymbol{\delta}$ are p-vectors. To see why these parameters are important, notice that under near cointegration it holds that

$$\lim_{T\to\infty} T^{-1} \mathrm{E}\left(\mathbf{x}_T \left(\sum_{t=1}^{T} (y_t - \boldsymbol{\beta}_0' \mathbf{x}_t)\right)\right) = \Omega_1 \boldsymbol{\pi}_1, \tag{8}$$

$$\lim_{T\to\infty} \mathrm{E}\left(\mathbf{x}_T \left(y_T - \boldsymbol{\beta}_0' \mathbf{x}_T\right)\right) = \Omega_1 \boldsymbol{\delta}. \tag{9}$$

As is well known (e.g. Phillips and Durlauf (1986, Theorem 4.1 (a))), the limiting distribution of OLS estimators under cointegration depends on these (nuisance) parameters. A similar situation occurs under near cointegration.

Lemma 1. *Suppose* $\left\{(y_t, \mathbf{x}_t')'\right\}$ *is generated by* (1) *and suppose A1-A3 hold.*
(a) If $\rho < 1$ *and fixed (spurious regression), then as* $T \to \infty$,

$$\begin{pmatrix} T^{-1/2}\hat{\alpha} \\ \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \end{pmatrix} \Rightarrow$$

$$\omega_0 \sqrt{1-\rho^2} \begin{pmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \Omega_1' \end{pmatrix}^{-1} \left(\int_0^1 \mathbf{X}_1(r) \mathbf{X}_1(r)' \, dr\right)^{-1} \left(\int_0^1 \mathbf{X}_1(r) \mathrm{W}_0(r) \, dr\right),$$

(b) If A4 holds (near cointegration), then as $T \to \infty$,

$$\begin{pmatrix} \sqrt{T}\hat{\alpha} \\ T\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \end{pmatrix} \Rightarrow$$

$$\begin{pmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \Omega_1' \end{pmatrix}^{-1} \left(\int_0^1 \mathbf{X}_1(r) \mathbf{X}_1(r)' \, dr \right)^{-1}$$

$$\times \left(\omega_0 \kappa \int_0^1 \mathbf{X}_1(r) \mathrm{W}_0(r) \, dr + \pi_0 \int_0^1 \mathbf{X}_1(r) \, d\mathrm{W}_0(r) + \left(\int_0^1 \mathbf{X}_1(r) \, d\mathbf{W}_1(r)' \right) \boldsymbol{\pi}_1 + \begin{pmatrix} 0 \\ \boldsymbol{\delta} \end{pmatrix} \right),$$

where $(\pi_0, \boldsymbol{\pi}_1')'$ *and* $\boldsymbol{\delta}$ *are defined in* (6) *and* (7), *and* W_0, $\mathbf{W}_1$ *and* $\mathbf{X}_1$ *are defined in the Appendix.*

Under spurious regression, the limiting distribution of $\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)$ has mean zero. This provides us with another interpretation of the fundamental coefficient: Under spurious regression, the fundamental coefficient is simply the limiting expected value of $\hat{\boldsymbol{\beta}}$. Under near cointegration, $\hat{\boldsymbol{\beta}}$ is a super-consistent estimator of $\boldsymbol{\beta}_0$ since $\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) = O_p(T^{-1})$. In this sense, the near cointegration case is qualitatively similar to the cointegration case. Since near cointegration reduces to exact cointegration when $\kappa = 0$, the distributional result in (b) is well known for that special case (e.g. Phillips and Durlauf (1986, Theorem 4.1 (a))). When $\kappa \neq 0$, the limiting distribution in (b) is essentially a mixture of the spurious regression distribution reported in (a) and the distribution corresponding to exact cointegration. A similar result has been obtained by Tanaka (1993, Theorem 6). However, the distribution reported there depends on a $(p+1)$-dimensional noncentrality parameter rather than a scalar.

4. TEST STATISTICS BASED ON OLS ESTIMATORS

Based on the multiple regression (5), standard regression packages can be used to compute the F-statistic

$$F\left(\hat{\boldsymbol{\beta}}\right) := \frac{\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)' \left(\sum_{t=1}^{T} (\mathbf{x}_t - \bar{\mathbf{x}})(\mathbf{x}_t - \bar{\mathbf{x}})' \right) \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)}{p \cdot s^2}, \tag{10}$$

where $\bar{\mathbf{x}} := T^{-1} \sum_{t=1}^{T} \mathbf{x}_t$ and $s^2 := (T - p - 1)^{-1} \sum_{t=1}^{T} \hat{u}_t^2$.

This is simply the F-statistic used to test the null hypothesis $\mathrm{H}_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$. As the following lemma shows, (usual) spurious regression asymptotics predict that $F\left(\hat{\boldsymbol{\beta}}\right)$ diverges at rate T whenever $\rho < 1$.

Lemma 2. *Suppose* $\left\{(y_t, \mathbf{x}_t')'\right\}$ *is generated by* (1) *and suppose A1-A3 hold. If* $\rho < 1$ *and fixed, then as* $T \to \infty$,

$$T^{-1}\left(p \times F\left(\hat{\boldsymbol{\beta}}\right)\right) \Rightarrow$$

$$\frac{\left(\int_0^1 \mathbf{W}_1^\mu(r)\,\mathrm{W}_0^\mu(r)\,dr\right)'\left(\int_0^1 \mathbf{W}_1^\mu(r)\,\mathbf{W}_1^\mu(r)'\,dr\right)^{-1}\left(\int_0^1 \mathbf{W}_1^\mu(r)\,\mathrm{W}_0^\mu(r)\,dr\right)}{\int_0^1 \mathrm{W}_0^\mu(r)^2\,dr - \left(\int_0^1 \mathbf{W}_1^\mu(r)\,\mathrm{W}_0^\mu(r)\,dr\right)'\left(\int_0^1 \mathbf{W}_1^\mu(r)\,\mathbf{W}_1^\mu(r)'\,dr\right)^{-1}\left(\int_0^1 \mathbf{W}_1^\mu(r)\,\mathrm{W}_0^\mu(r)\,dr\right)},$$

where W_0^μ *and* $\mathbf{W}_1^\mu$ *are demeaned Wiener processes defined in the Appendix.*

Quite remarkably, the limiting distribution of $T^{-1} \times F\left(\hat{\boldsymbol{\beta}}\right)$ 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 $F\left(\hat{\boldsymbol{\beta}}\right) = O_p(1)$ with a complicated limiting distribution depending on nuisance parameters when $\rho = 1$. The following proposition generalizes that result to the case of near cointegration.

Proposition 3. *If* $\{(y_t, \mathbf{x}_t')'\}$ *is generated by* (1) *and A1-A4 hold, then as* $T \to \infty$,

$$p \times F\left(\hat{\boldsymbol{\beta}}\right) \Rightarrow$$

$$\frac{1}{\sigma^2}\left(\omega_0\kappa\int_0^1 \mathbf{W}_1^\mu(r)\,\mathrm{W}_0^\mu(r)\,dr + \pi_0\int_0^1 \mathbf{W}_1^\mu(r)\,d\mathrm{W}_0(r) + \left(\int_0^1 \mathbf{W}_1^\mu(r)\,d\mathbf{W}_1(r)'\right)\boldsymbol{\pi}_1 + \boldsymbol{\delta}\right)'$$
$$\times\left(\int_0^1 \mathbf{W}_1^\mu(r)\,\mathbf{W}_1^\mu(r)'\,dr\right)^{-1}$$
$$\times\left(\omega_0\kappa\int_0^1 \mathbf{W}_1^\mu(r)\,\mathrm{W}_0^\mu(r)\,dr + \pi_0\int_0^1 \mathbf{W}_1^\mu(r)\,d\mathrm{W}_0(r) + \left(\int_0^1 \mathbf{W}_1^\mu(r)\,d\mathbf{W}_1(r)'\right)\boldsymbol{\pi}_1 + \boldsymbol{\delta}\right),$$

where $(\pi_0, \boldsymbol{\pi}_1')'$ *and* $\boldsymbol{\delta}$ *are defined in* (6) *and* (7), *and* W_0, W_0^μ, $\mathbf{W}_1$, *and* $\mathbf{W}_1^\mu$ *are defined in the Appendix.*

This proposition provides analytical corroboration of the conjecture that the spurious regression distribution reported in Lemma 2 provides a poor approximation to the finite sample distribution of $F\left(\hat{\boldsymbol{\beta}}\right)$ when $\{(y_t, \mathbf{x}_t')'\}$ is nearly cointegrated in the sense that $\rho^2 \approx 1$. In fact, in accordance with common sense, Proposition 3 suggests that in finite samples the behavior of $F\left(\hat{\boldsymbol{\beta}}\right)$ depends continuously on ρ even when ρ^2 approaches unity. Near cointegration therefore seems to suggest a useful way of bridging the apparent gap between spurious regression and (exact) cointegration.

Although the motivation underlying the notion of near cointegration is very similar in spirit to the motivation underlying the notion of near integration, the associated asymptotic theory is different in a couple of respects. The asymptotic theory under near cointegration can be expressed in terms of simple functionals of Wiener processes. In contrast, the relevant limiting distributions under near integration is typically expressed in terms of functionals of diffusion processes.[6] Moreover, the limiting behavior as the noncentrality parameter increases without bound is qualitatively different, as we now discuss. 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 Proposition 3.[7] In some sense, the results in Lemma 2 and Proposition 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 $\kappa \to \infty$ in the distribution reported in Proposition 3). However, we notice that $1/\kappa^2$ times the limiting distribution in Proposition 3 converges to

$$\frac{\omega_0^2}{\sigma^2}\left(\int_0^1 \mathbf{W}_1^{\mu}(r)\,\mathrm{W}_0^{\mu}(r)\,dr\right)'\left(\int_0^1 \mathbf{W}_1^{\mu}(r)\,\mathbf{W}_1^{\mu}(r)'\,dr\right)^{-1}\left(\int_0^1 \mathbf{W}_1^{\mu}(r)\,\mathrm{W}_0^{\mu}(r)\,dr\right)$$

as $\kappa \to \infty$. Therefore, Lemma 2 cannot be deduced from Proposition 3. As such, our results provide 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 $\kappa \to \infty$ (Phillips (1987, p. 543)).

Even under exact cointegration, when $\kappa = 0$, the distribution reported in proposition 3 is not useful in itself, since it depends on unknown nuisance parameters, notably $\boldsymbol{\pi}_1$ and $\boldsymbol{\delta}$. In the subsequent sections, we will repeatedly make the simplifying assumption that $\boldsymbol{\pi}_1 = \boldsymbol{\delta} = \mathbf{0}$. This is not because we believe it is a realistic assumption in practice[8], but rather because it is known that by using more sophisticated estimation methods than OLS it is possible to obtain parameter estimates that behave "as if" $\boldsymbol{\pi}_1 = \boldsymbol{\delta} = \mathbf{0}$. The question of how to do this is obviously of immense interest and has attracted considerable attention in the literature.[9] However, since the primary purpose of this paper is to investigate the properties of various known

[6] Since these diffusion processes can be represented as functionals of Wiener processes, it is of course possible to express the asymptotic theory in terms of functionals of Wiener processes. Doing so complicates the notation considerably, however.

[7] Heuristically, spurious regression corresponds to near cointegration with a "large" κ.

[8] Since Ω_1 is nonsingular, it follows from equations (8) and (9) that $\boldsymbol{\pi}_1 = \boldsymbol{\delta} = \mathbf{0}$ if and only if $\lim_{T\to\infty} T^{-1}\mathrm{E}\left(\mathbf{x}_T\left(\sum_{t=1}^T\left(y_t - \boldsymbol{\beta}_0'\mathbf{x}_t\right)\right)\right) = \lim_{T\to\infty} T^{-1}\mathrm{E}\left(\mathbf{x}_T\left(y_t - \boldsymbol{\beta}_0'\mathbf{x}_t\right)\right) = \mathbf{0}$. Therefore, a sufficient condition is that $\{\mathbf{x}_t\}$ is strictly exogenous in the sense that it is driven by a process which is independent of the error process $\{y_t - \boldsymbol{\beta}_0'\mathbf{x}_t\}$.

[9] Among the available estimators are those proposed by Johansen (1988, 1991), Johansen and Juselius (1990), Phillips and Hansen (1990), Park (1992), Saikkonen (1991, 1992), and Stock and Watson (1993).

inference procedures under near cointegration rather than to propose new methods, we will focus on the simplest case in the hope that the fundamental messages will be more transparent. Therefore, whenever we impose $\boldsymbol{\pi}_1 = \boldsymbol{\delta} = \mathbf{0}$ (Assumption A5 below) in the sequel, we are implictly considering estimators and test statistics based on more sophisticated procedures than OLS.

From (3) we notice that under exact cointegration, when $\rho = 1$, we have

$$\sum_{t=1}^{T} \left(y_t - \boldsymbol{\beta}_0' \mathbf{x}_t\right) = \pi_0 \xi_{0T} + \boldsymbol{\pi}_1' \boldsymbol{\xi}_{1T} + O_p\left(1\right), \tag{11}$$

where π_0 and $\boldsymbol{\pi}_1$ are the parameters defined in (6). When $\boldsymbol{\pi}_1 = \mathbf{0}$, as will be assumed throughout Sections 5 and 6, $\left\{\sum_{s=1}^{t} \left(y_s - \boldsymbol{\beta}_0' \mathbf{x}_s\right)\right\}$ is integrated if and only if $\pi_0 \neq 0$. More generally, $\left\{\left(\sum_{s=1}^{t} \left(y_s - \boldsymbol{\beta}_0' \mathbf{x}_s\right), \mathbf{x}_t'\right)'\right\}$ is non-cointegrated if and only if $\pi_0 \neq 0$. Therefore, the statistical properties of $\left\{\left(\sum_{s=1}^{t} \left(y_s - \boldsymbol{\beta}_0' \mathbf{x}_s\right), \mathbf{x}_t'\right)'\right\}$ depend crucially on the value of π_0.[10] These considerations, along with the arguments presented in the previous paragraph, lead us to require the following throughout sections 5 and 6:

A5. $\pi_0 \neq 0$ and $\boldsymbol{\pi}_1 = \boldsymbol{\delta} = \mathbf{0}$.

5. BEHAVIOR OF COINTEGRATION METHODS UNDER NEAR COINTEGRATION

A quite remarkable result from the literature on cointegration is the fact that for the purpose of doing inference on cointegration coefficients, test statistics have been developed whose limiting distributions are χ^2.[11] Recently, Elliott (1998) has investigated the robustness of this result by considering a model in which the regressors are *nearly* integrated while some linear combination of the regressand and the regressors is *exactly* stationary. It turns out that the χ^2 result can break down when the regressors are not exactly integrated. The model of near cointegration allows us to conduct a complementary experiment: we can investigate the behavior of test statistics in a model where the regressors are *exactly* integrated while some linear combination of the regressand and the regressors is *nearly* stationary. To that end, we define the following Wald statistic:

$$G\left(\hat{\boldsymbol{\beta}}\right) := \frac{\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)' \left(\sum_{t=1}^{T} \left(\mathbf{x}_t - \bar{\mathbf{x}}\right)\left(\mathbf{x}_t - \bar{\mathbf{x}}\right)'\right) \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)}{\hat{\pi}_0^2}, \tag{12}$$

where $\hat{\pi}_0^2$ is an estimator of π_0^2 based on $\{\hat{u}_t\}$. We shall require that $\hat{\pi}_0^2$ is a consistent estimator of π_0^2 under near cointegration. Since none of our results depend on the

[10] When $\pi_0 = 0$, i.e. when $\left\{\left(\sum_{s=1}^{t} \left(y_s - \boldsymbol{\beta}_0' \mathbf{x}_s\right), \mathbf{x}_t'\right)'\right\}$ is cointegrated, $\left\{\left(y_t, \mathbf{x}_t'\right)'\right\}$ is multicointegrated in the sense of Granger and Lee (1990) and conventional cointegration results no longer hold. Park (1992) refers to this case as singular cointegration.

[11] See e.g. Phillips (1991, Remark (i)) and the references listed in footnote 9.

particular choice of estimator, it suffices to notice that such consistent estimators exist.[12]

Evidently, the test statistics $F\left(\hat{\boldsymbol{\beta}}\right)$ and $G\left(\hat{\boldsymbol{\beta}}\right)$ are very similar. In fact, apart from the constant p the only difference between them is that the denominator of $F\left(\hat{\boldsymbol{\beta}}\right)$ contains a consistent estimate of σ^2, the (asymptotic) variance of $(y_t - \boldsymbol{\beta}_0' \mathbf{x}_t)$, whereas $G\left(\hat{\boldsymbol{\beta}}\right)$ contains a consistent estimate of π_0^2, the conditional long-run variance of $(y_t - \boldsymbol{\beta}_0' \mathbf{x}_t)$.

As the following corollary to Proposition 3 demonstrates, the χ^2 result does not hold under near cointegration.[13]

Corollary 4. *If $\left\{(y_t, \mathbf{x}_t')'\right\}$ is generated by* (1) *and A1-A5 hold, then as $T \to \infty$,*

$$G\left(\hat{\boldsymbol{\beta}}\right) \Rightarrow$$

$$\left(\lambda \int_0^1 \mathbf{W}_1^\mu(r)\, \mathrm{W}_0^\mu(r)\, dr + \int_0^1 \mathbf{W}_1^\mu(r)\, d\mathrm{W}_0(r)\right)' \times \left(\int_0^1 \mathbf{W}_1^\mu(r)\, \mathbf{W}_1^\mu(r)'\, dr\right)^{-1} \times \left(\lambda \int_0^1 \mathbf{W}_1^\mu(r)\, \mathrm{W}_0^\mu(r)\, dr + \int_0^1 \mathbf{W}_1^\mu(r)\, d\mathrm{W}_0(r)\right),$$

where $\lambda := \omega_0 \kappa / \pi_0$, and W_0, W_0^μ, and $\mathbf{W}_1^\mu$ are defined in the Appendix.

Conditional on (the σ-algebra generated by) $\mathbf{W}_1$, $\int_0^1 \mathbf{W}_1^\mu(r)\, d\mathrm{W}_0(r)$ is distributed $\mathcal{N}\left(\mathbf{0}, \left(\int_0^1 \mathbf{W}_1^\mu(r)\, \mathbf{W}_1^\mu(r)'\, dr\right)\right)$ and is independent of $\int_0^1 \mathbf{W}_1^\mu(r)\, \mathrm{W}_0^\mu(r)\, dr$. Therefore, the limiting distribution of $G\left(\hat{\boldsymbol{\beta}}\right)$ is χ_p^2 if and only if $\kappa = 0$. So, under near cointegration (with $\kappa \neq 0$) the limiting distribution is not χ_p^2.

[12] It follows from the proof of Proposition 3 that $T^{-1} \sum_{t=1}^T \hat{u}_t^2 \to_p \sigma^2 = \mathrm{E}(u_t^2)$ (as $T \to \infty$), where $u_t := \begin{pmatrix} 1 & -\boldsymbol{\beta}_0' \end{pmatrix} \mathbf{C}(L)\, \mathbf{e}_t$. Similarly, $T^{-1} \sum_{t=h+1}^T \hat{u}_t \hat{u}_{t-h} \to_p \mathrm{E}(u_t u_{t-h})$ for $h \geq 1$. Therefore, the usual variance estimators that are consistent under (exact) cointegration are also consistent under near cointegration.

[13] The limiting distribution reported in Corrolary 4 could equivalently be written in the following slightly more compact way:

$$\left(\int_0^1 \mathbf{W}_1^\mu(r)\, d\mathrm{V}_\lambda(r)\right)' \left(\int_0^1 \mathbf{W}_1^\mu(r)\, \mathbf{W}_1^\mu(r)'\, dr\right)^{-1} \left(\int_0^1 \mathbf{W}_1^\mu(r)\, d\mathrm{V}_\lambda(r)\right),$$

where $\lambda := \omega_0 \kappa / \pi_0$ and $\mathbf{W}_1^\mu$ and V_λ are defined in the Appendix.

For simplicity, we have only considered the null hypothesis $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$. Extending the result to cover more general (possibly nonlinear) restrictions on $\boldsymbol{\beta}$ is a straightforward exercise, but we shall not do so here.

It is worth noting that the conclusion holds even though $\boldsymbol{\pi}_1 = \boldsymbol{\delta} = \mathbf{0}$. In Elliott (1998), the deviation from the χ^2 distribution arises because the corrections made in order to accomodate endogenous regressors fail to work for those regressors that are nearly (but not exactly) integrated. Therefore, when $\boldsymbol{\pi}_1 = \boldsymbol{\delta} = \mathbf{0}$, i.e. when no correction is needed, the χ^2 result continues to hold even with nearly integrated regressors. Here, the χ^2 result breaks down because of the deviation from exact cointegration and therefore also breaks down when $\boldsymbol{\pi}_1 = \boldsymbol{\delta} = \mathbf{0}$. Similarly, it always breaks down for all subsets of $\boldsymbol{\beta}$, in contrast with the result in Elliott (1998). This is important, since essentially it is this result that justifies the variable addition test for cointegration due to Park (1990), which will by studied in the next section.

As already argued, the limiting distribution of $G\left(\hat{\boldsymbol{\beta}}\right)$ deviates from the χ^2_p distribution whenever $\kappa \neq 0$. The severity of the deviation is seen to depend on $\lambda = \omega_0\kappa/\pi_0$. This scalar parameter reflects the relative size of two conditional variances, as we now explain. It follows from (11) that under exact cointegration, π_0^2 can be interpreted as the long-run variance of $(y_t - \boldsymbol{\beta}_0'\mathbf{x}_t)$ conditional on $\Delta\mathbf{x}_t$. More generally, under near cointegration, π_0^2 can be interpreted as the long-run variance of $\begin{pmatrix} 1 & -\boldsymbol{\beta}_0' \end{pmatrix} \mathbf{C}(L)\,\mathbf{e}_t$, the stationary component of the error $y_t - \boldsymbol{\beta}_0'\mathbf{x}_t$, conditional on $\Delta\mathbf{x}_t$. As for $\omega_0\kappa$, notice that under spurious regression the long-run variance of Δy_t conditional on $\Delta\mathbf{x}_t$ is given by $\omega_0^2(1-\rho^2)$. Under near cointegration, $\omega_0^2(1-\rho^2) = (T\cdot\omega_0\kappa)^2$ and the numerator of λ, $\omega_0\kappa$, represents (the square root of) this conditional variance. So, the coefficient λ on the "spurious regression term" $\int_0^1 \mathbf{W}_1^\mu(r)\,\mathrm{W}_0^\mu(r)\,dr$ in the limiting distribution of $G\left(\hat{\boldsymbol{\beta}}\right)$ reflects the relative magnitude of the two conditional variances corresponding to the random walk part and the stationary part of the error $y_t - \boldsymbol{\beta}_0'\mathbf{x}_t$, respectively. Unlike ω_0 and κ, which are positive scalars by assumption, π_0 can be both positive and negative. Therefore, so can λ.

To illustrate the magnitude of the size distortions encountered under local alternatives, we have simulated (the discrete time counterpart of) the limiting distribution of $G\left(\hat{\boldsymbol{\beta}}\right)$. Specifically, we have generated $\boldsymbol{\xi}_t = (\xi_{0t}, \boldsymbol{\xi}_{1t}')' := \sum_{s=1}^t \mathbf{e}_s$ for $t = 1, \ldots, T = 2000$, where $\mathbf{e}_t \sim$ i.i.d. $\mathcal{N}(\mathbf{0}, \mathbf{I}_{p+1})$. Using these, integrals have been approximated by discrete sums.[14] This procedure was repeated 20.000 times. In Figure 1, we have plotted the rejection frequencies for various values of p and λ. The nominal size of the tests is 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. Therefore, it is 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.

6. Behavior of Tests for Cointegration Under Near Cointegration

During the last decade, numerous cointegration tests taking cointegration as the null hypothesis have been proposed. Since these test procedures utilize different properties of cointegrated systems it seems desirable to know what, if anything, can be said about the relative power properties of the different tests. Also, the evidence presented in the previous section suggests that the absolute power properties of tests for cointegration are worth investigating. In this section we take a first step in that direction by characterizing the behavior of several regression based cointegration tests[15] under local alternatives and obtaining the corresponding local power functions. In their original formulations, all of the test procedures under study here involve (typically non-parametric) corrections in order to accomodate the case where the condition

[14]For instance, $\int_0^1 \mathbf{W}_1^{\mu}(r)\, \mathrm{W}_0^{\mu}(r)\, dr$ is approximated by

$$T^{-2}\sum_{t=1}^{T}\left(\boldsymbol{\xi}_{1t}-T^{-1}\sum_{s=1}^{T}\boldsymbol{\xi}_{1s}\right)\left(\xi_{0t}-T^{-1}\sum_{s=1}^{T}\xi_{0s}\right)$$

$$=T^{-2}\sum_{t=1}^{T}\boldsymbol{\xi}_{1t}\xi_{0t}-T^{-3}\left(\sum_{t=1}^{T}\boldsymbol{\xi}_{1t}\right)\left(\sum_{t=1}^{T}\xi_{0t}\right).$$

[15]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. We note, however, that under cointegration the limiting distribution of the test proposed by Harris (1997) is the same as that of the CI test proposed by Shin (1994) (Harris (1997, Theorem 7)). We therefore conjecture that the local power properties of Harris' test are similar to those of the CI test.

$\boldsymbol{\pi}_1 = \boldsymbol{\delta} = \mathbf{0}$ is violated. Since all tests can be based on simple OLS methods when $\boldsymbol{\pi}_1 = \boldsymbol{\delta} = \mathbf{0}$ holds, we once more simplify the presentation by assuming $\boldsymbol{\pi}_1 = \boldsymbol{\delta} = \mathbf{0}$ throughout. This allows us to focus on the question of interest without complicating the discussion unnecessarily. We emphasize, though, that all of the test statistics studied here are simplified versions of the original tests.

We have divided this section into four parts. Section 6.1 deals with tests based on the variable addition procedure, while Sections 6.2 and 6.3 study tests based on residuals from I(1) and I(2) regressions, respectively. Finally, Section 6.4 compares the local power properties of the different tests.

6.1. Variable Addition Tests.

The variable addition test procedure proposed by Park (1990) is extremely simple to implement and can be motivated as follows. Under cointegration, appropriately constructed Wald tests (such as $G\left(\hat{\boldsymbol{\beta}}\right)$ defined in (12)) 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 regression (5).

Therefore, let k_1 and k_2 be arbitrary non-negative integers such that $k_1 + k_2 \geq 1$ and for $t = 1, \dots, T$, let $\mathbf{d}_t := \left(t, \dots, t^{k_1}\right)'$ (if $k_1 \geq 1$) and let $\{\mathbf{z}_t\}$ be a k_2-dimensional computer generated random walk such that $\{\Delta \mathbf{z}_t\} \sim i.i.d.\ \mathcal{N}\left(\mathbf{0}, \mathbf{I}_{k_2}\right)$.[16]

Based on the multiple regressions (5) and[17]

$$y_t = \tilde{\alpha} + \tilde{\boldsymbol{\beta}}' \mathbf{x}_t + \tilde{\boldsymbol{\gamma}}_1' \mathbf{d}_t + \tilde{\boldsymbol{\gamma}}_2' \mathbf{z}_t + \tilde{u}_t, \qquad (t = 1, \dots, T) \tag{13}$$

we can construct the statistic

$$J_1(k_1, k_2) := \frac{\sum_{t=1}^T \hat{u}_t^2 - \sum_{t=1}^T \tilde{u}_t^2}{\hat{\pi}_0^2}, \tag{14}$$

where $\hat{\pi}_0^2$ is some consistent estimator of π_0^2 based on $\{\hat{u}_t\}$. This is simply the (appropriately standardized) Wald test used to test the null hypothesis $\mathrm{H}_0 : \boldsymbol{\gamma}_1 = \mathbf{0}, \boldsymbol{\gamma}_2 = \mathbf{0}$.

Proposition 5. *If $\{(y_t, \mathbf{x}_t')'\}$ is generated by* (1) *and A1-A5 hold, then as $T \to \infty$,*

$$J_1(k_1, k_2) \Rightarrow \left(\int_0^1 \mathbf{X}_{2.1}(r)\, d\mathrm{V}_\lambda(r)\right)' \left(\int_0^1 \mathbf{X}_{2.1}(r)\, \mathbf{X}_{2.1}(r)'\, dr\right)^{-1} \left(\int_0^1 \mathbf{X}_{2.1}(r)\, d\mathrm{V}_\lambda(r)\right),$$

where $\lambda := \omega_0 \kappa / \pi_0$, and $\mathbf{X}_{2.1}$ and V_λ are defined in the Appendix.

[16] This particular choice of superfluous regressors is advocated by Park (1990, Remark b). However, 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.

[17] In (13), $\mathbf{d}_t$ ($\mathbf{z}_t$) is omitted if $k_1 = 0$ ($k_2 = 0$).

The limiting distribution of J_1 is $\chi^2_{k_1+k_2}$ if and only if $\kappa = 0$. Under fixed alternatives, i.e. under spurious regression, J_1 diverges at a rate depending on the kernel and bandwidth used in constructing $\hat{\pi}_0^2$ (Park (1990, Theorem 4.1 (b))). In contrast, the local power only depends on λ.

6.2. Tests Based on Residuals from an I(1) Regression.

Several cointegration tests based on the residuals $\{\hat{u}_t\}$ from (5) have been proposed in the literature. 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), McCabe, Leybourne, and Shin (1997), Quintos and Phillips (1993), and Tanaka (1996, Section 11.6.2).

The test proposed by Shin (1994) is based on

$$CI := \frac{T^{-2}\sum_{t=1}^{T}\hat{S}_t^2}{\hat{\pi}_0^2}, \tag{15}$$

where $\hat{S}_t := \sum_{s=1}^{t}\hat{u}_s$ for $t = 1, \ldots, T$, and $\hat{\pi}_0^2$ is a consistent estimator of π_0^2. Essentially, this is the stationarity test proposed by Kwiatkowski, Phillips, Schmidt, and Shin (1992) applied to the residuals $\{\hat{u}_t\}$.[18]

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

$$L_c := T^{-1}\frac{\sum_{t=1}^{T}\mathbf{S}_t^{*\prime}\mathbf{M}_T^{-1}\mathbf{S}_t^*}{\hat{\pi}_0^2} = T^{-1}\frac{tr\left[\mathbf{M}_T^{-1}\left(\sum_{t=1}^{T}\mathbf{S}_t^*\mathbf{S}_t^{*\prime}\right)\right]}{\hat{\pi}_0^2}, \tag{16}$$

where

$$\mathbf{S}_t^* := \sum_{s=1}^{t}\begin{pmatrix}\hat{u}_s \\ \mathbf{x}_s\hat{u}_s\end{pmatrix}, \qquad (t = 1, \ldots, T)$$

$$\mathbf{M}_T := \sum_{t=1}^{T}\begin{pmatrix}1 & \mathbf{x}_t' \\ \mathbf{x}_t & \mathbf{x}_t\mathbf{x}_t'\end{pmatrix},$$

and $\hat{\pi}_0^2$ is a consistent estimator of π_0^2.

[18] The simple version of the CI statistic considered here has appeared (at least) twice in the literature. It was proposed by Tanaka (1993), who emphasizes that it is only applicable when the regressors are strictly exogenous. Also, it has been derived by Leybourne and McCabe (1993) (essentially) under the assumption of strictly exogeneous regressors.

Proposition 6. *If $\left\{(y_t, \mathbf{x}_t')'\right\}$ is generated by* (1) *and A1-A5 hold, then as* $T \to \infty$,

$$CI \Rightarrow \int_0^1 \hat{\mathrm{V}}_\lambda (r)^2 \, dr,$$

$$\begin{aligned} L_c \quad &\Rightarrow \quad \int_0^1 \mathbf{V}_\lambda^* (r)' \left(\int_0^1 \mathbf{X}_1 (s) \mathbf{X}_1 (s)' \, ds \right)^{-1} \mathbf{V}_\lambda^* (r) \, dr \\ &= \quad tr \left[\left(\int_0^1 \mathbf{X}_1 (s) \mathbf{X}_1 (s)' \, ds \right)^{-1} \left(\int_0^1 \mathbf{V}_\lambda^* (r) \mathbf{V}_\lambda^* (r)' \, dr \right) \right], \end{aligned}$$

where $\lambda := \omega_0 \kappa / \pi_0$, *and* $\hat{\mathrm{V}}_\lambda$, $\mathbf{X}_1$, *and* $\mathbf{V}_\lambda^*$ *are defined in the Appendix.*

Tanaka (1996, Theorem 11.11) reports a result very similar to the result for CI. However, the limiting distribution reported there depends on a $(p+1)$-dimensional parameter. In contrast, both limiting distributions reported here only depend on a scalar parameter, λ. Under spurious regression, the rate of divergence of CI and L_c depends on the growth rate of the lag truncation number (Shin (1994, Theorem 3), Kuo (1998, Theorem 3)). The local power, on the other hand, only depends on λ.

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

The cointegration tests proposed by Choi and Ahn (1995) are based on the residuals from the multiple regression

$$S_t^y = \breve{\alpha} t + \breve{\boldsymbol{\beta}}' \mathbf{S}_t^{\mathbf{x}} + \breve{S}_t, \qquad (t = 1, \ldots, T) \tag{17}$$

where $S_t^y := \sum_{s=1}^t y_s$ and $\mathbf{S}_t^{\mathbf{x}} := \sum_{s=1}^t \mathbf{x}_s$ for $t = 1, \ldots, T$.
Three different test statistics based on $\left\{\breve{S}_t\right\}$ are proposed:

$$LM_I := \frac{\left(T^{-1} \sum_{t=2}^T \breve{S}_{t-1} \Delta \breve{S}_t - \frac{\breve{\pi}_0^2 - \breve{\sigma}^2}{2} \right)^2}{\breve{\pi}_0^4}, \tag{18}$$

$$LM_{II} := \frac{\left(T^{-1} \sum_{t=2}^T \breve{S}_{t-1} \Delta \breve{S}_t - \frac{\breve{\pi}_0^2 - \breve{\sigma}^2}{2} \right)^2}{\breve{\pi}_0^2 \left(T^{-2} \sum_{t=2}^T \breve{S}_{t-1}^2 \right)}, \tag{19}$$

$$SBDH_I := \frac{T^{-2} \sum_{t=1}^T \breve{S}_t^2}{\breve{\pi}_0^2}, \tag{20}$$

where $\breve{\pi}_0^2$ and $\breve{\sigma}^2$ are consistent estimators of π_0^2 and σ^2 based on $\left\{\Delta \breve{S}_t\right\}$. These tests are intimately related to the stationarity tests proposed by Choi and Ahn (1998).

Proposition 7. *If $\{(y_t, \mathbf{x}_t')'\}$ is generated by* (1) *and A1-A5 hold, then as* $T \to \infty$,

$$LM_I \Rightarrow \left(\int_0^1 \check{\mathrm{V}}_\lambda(r)\, d\check{\mathrm{V}}_\lambda(r)\right)^2,$$

$$LM_{II} \Rightarrow \frac{\left(\int_0^1 \check{\mathrm{V}}_\lambda(r)\, d\check{\mathrm{V}}_\lambda(r)\right)^2}{\int_0^1 \check{\mathrm{V}}_\lambda(r)^2\, dr},$$

$$SBDH_I \Rightarrow \int_0^1 \check{\mathrm{V}}_\lambda(r)^2\, dr,$$

where $\lambda := \omega_0\kappa/\pi_0$ *and* $\check{\mathrm{V}}_\lambda$ *is defined in the Appendix.*

Once more, the behavior under local alternatives depends solely on λ, whereas the behavior under spurious regression depends on the expansion rate of the lag truncation number (Choi and Ahn (1995, Theorem 2)).

6.4. Local Power of Tests for Cointegration.

In order to obtain local power functions, we have simulated[19] (the discrete time counterparts of) the limiting distributions of the $J_1(2,2)$ [20], CI, L_c, LM_I, LM_{II} and $SBDH_I$ test statistics. Figures 2-5 show the local power functions for $p = 1, \ldots, 4$. The size of the tests is 5%.

FIGURE 2 ABOUT HERE
FIGURE 3 ABOUT HERE
FIGURE 4 ABOUT HERE
FIGURE 5 ABOUT HERE

The figures suggest that the local power properties of $J_1(2,2)$, CI, L_c and $SBDH_I$ are very similar. On the other hand, LM_I and (in particular) LM_{II} are remarkably inferior in terms of local power. In the case of LM_{II} this is not surprising. In fact, it follows from the proof of Proposition 7 that

$$LM_{II} = \frac{LM_I}{SBDH_I} + o_p(1).$$

[19] As in Section 5, we set $T = 2.000$ and repeat the procedure 20.000 times.

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

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.

This result 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.[21] Therefore, the asymptotic properties of otherwise identical (Lagrange Multiplier) test statistics will often depend on whether or not they involve the second derivate of the log-likelihood function and some caution should be exercised whenever a test statistic involves the second derivate of the log-likelihood function.

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. 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. 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 number should be chosen so as to minimize finite sample size distortions. Similarly, 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.

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.

[21] Here, for instance, the scaled second derivative of the log-likelihood function is $SBDH_I - T^{-2}\check{S}_T^2/\check{\pi}_0^2 = SBDH_I + o_p(1)$.

Therefore, whenever a reasearcher rejects a structural hypothesis (on the coefficient $\boldsymbol{\beta}$) using cointegration methods, the result should be interpreted carefully, since it might be the case that the 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.[22]

7. Concluding Remarks

A notion of near cointegration was proposed and three applications of the model of near cointegration were provided. As these applications illustrate, the notion of near cointegration is useful in a variety of settings. Throughout the paper, we have deliberately made several simplifying assumptions with respect to the DGP in order to focus specifically on the impact of cointegration failure. As a consequence, several extensions of the analysis are possible. For instance, deterministic terms can be included in the DGP and the condition $\boldsymbol{\pi}_1 = \boldsymbol{\delta} = \mathbf{0}$ can be relaxed with little or no difficulty. Similarly, greater flexibility with respect to the cointegration rank can be achieved by allowing $\{y_t\}$ to be a vector process and/or allowing $\{\mathbf{x}_t\}$ to be cointegrated. These and other extensions are currently being considered by one of the authors and will be reported elsewhere.

8. Appendix

We start by establishing a useful lemma.

Lemma 8. *Let $\boldsymbol{\xi}_t$ be defined as $\boldsymbol{\xi}_t := \sum_{s=1}^{t} \mathbf{e}_s$, where $\{\mathbf{e}_t\}$ is i.i.d. with $\mathrm{E}(\mathbf{e}_t) = \mathbf{0}$ and $\mathrm{E}(\mathbf{e}_t\mathbf{e}_t') = \mathbf{I}_{p+1}$, and let $\mathbf{C}(L) := \sum_{i=0}^{\infty} \mathbf{C}_i L^i$ be a lag polynomial such that $\{\mathbf{C}_i\}$ is a sequence of $(p+1) \times (p+1)$ matrices satifying $\sum_{i=0}^{\infty} i\sqrt{tr(\mathbf{C}_i'\mathbf{C}_i)} < \infty$. Then, as $T \to \infty$, the following hold jointly:*

(a) $T^{-1/2}\boldsymbol{\xi}_{[Tr]} \Rightarrow \mathbf{W}(r)$,

(b) $T^{-1/2}\sum_{t=1}^{[Tr]} \mathbf{C}(L)\mathbf{e}_t \Rightarrow \mathbf{C}(1)\mathbf{W}(r) = \mathbf{C}(1)\int_0^r d\mathbf{W}(s)$,

(c) $T^{-1}\sum_{t=1}^{T} \mathbf{C}(L)\mathbf{e}_t(\mathbf{C}(L)\mathbf{e}_t)' \to_p \sum_{i=0}^{\infty} \mathbf{C}_i\mathbf{C}_i'$,

(d) $T^{-1}\sum_{t=1}^{[Tr]} \boldsymbol{\xi}_t(\mathbf{C}(L)\mathbf{e}_t)' \Rightarrow \left(\int_0^r \mathbf{W}(s)\,d\mathbf{W}(s)' + r\mathbf{I}_{p+1}\right)\mathbf{C}(1)'$,

where $\mathbf{W}$ is a $(p+1)$-dimensional Wiener process.

Proof. Parts (a)-(c) are multivariate versions of Phillips and Solo (1992, Theorems 3.4 and 3.7), while (d) can be established along the lines of Phillips (1988b) and Hansen (1992a, Theorem 4.1). ■

In representation (1), $\boldsymbol{\xi}_t$ is partitioned as $(\xi_{0t}, \boldsymbol{\xi}_{1t}')'$, where ξ_{0t} is a scalar and $\boldsymbol{\xi}_{1t}$ is a p-vector. Similarly, we shall typically partition $\mathbf{W}$ as $(\mathrm{W}_0, \mathbf{W}_1')'$, where W_0 and

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

$\mathbf{W}_1$ are of dimensions 1 and p, respectively. Throughout the appendix, integrals such as $\int_0^1 \mathbf{W}(r)\,dr$ and $\int_0^1 \mathbf{W}(r)\,d\mathbf{W}(r)'$ will often be abbreviated as $\int \mathbf{W}$ and $\int \mathbf{W}d\mathbf{W}'$, respectively, in order to simplify the notation.

8.1. Proof of Lemma 1.

The OLS formula reads:

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \end{pmatrix} = \begin{pmatrix} T & \sum_{t=1}^{T} \mathbf{x}_t' \\ \sum_{t=1}^{T} \mathbf{x}_t & \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^{T} (y_t - \boldsymbol{\beta}_0' \mathbf{x}_t) \\ \sum_{t=1}^{T} \mathbf{x}_t (y_t - \boldsymbol{\beta}_0' \mathbf{x}_t) \end{pmatrix}.$$

Let Υ_T be a diagonal normalization matrix defined as

$$\Upsilon_T := \begin{pmatrix} \sqrt{T} & \mathbf{0}' \\ \mathbf{0} & T \cdot \mathbf{I}_p \end{pmatrix}.$$

From Lemma 8 (a)-(b), we have

$$\sqrt{T}\Upsilon_T^{-1} \begin{pmatrix} 1 \\ \mathbf{x}_{[Tr]} \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \Omega_1 \end{pmatrix} \begin{pmatrix} 1 \\ \mathbf{W}_1(r) \end{pmatrix} = \begin{pmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \Omega_1 \end{pmatrix} \mathbf{X}_1(r), \tag{21}$$

under both spurious regression and near cointegration. By the continuous mapping theorem (CMT),

$$\Upsilon_T^{-1} \begin{pmatrix} T & \sum_{t=1}^{T} \mathbf{x}_t' \\ \sum_{t=1}^{T} \mathbf{x}_t & \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \end{pmatrix} \Upsilon_T^{-1} \Rightarrow \begin{pmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \Omega_1 \end{pmatrix} \left(\int \mathbf{X}_1 \mathbf{X}_1' \right) \begin{pmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \Omega_1 \end{pmatrix}'. \tag{22}$$

Similarly, we can use Lemma 8 (a)-(b) and CMT to show that

$$T^{-1}\Upsilon_T^{-1} \begin{pmatrix} \sum_{t=1}^{T} (y_t - \boldsymbol{\beta}_0' \mathbf{x}_t) \\ \sum_{t=1}^{T} \mathbf{x}_t (y_t - \boldsymbol{\beta}_0' \mathbf{x}_t) \end{pmatrix} \Rightarrow \omega_0 \sqrt{1-\rho^2} \begin{pmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \Omega_1 \end{pmatrix} \int \mathbf{X}_1 \mathrm{W}_0.$$

Therefore,

$$\begin{pmatrix} T^{-1/2}\hat{\alpha} \\ \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \end{pmatrix} = \left(\Upsilon_T^{-1} \begin{pmatrix} T & \sum_{t=1}^{T} \mathbf{x}_t' \\ \sum_{t=1}^{T} \mathbf{x}_t & \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \end{pmatrix} \Upsilon_T^{-1} \right)^{-1} T^{-1}\Upsilon_T^{-1} \begin{pmatrix} \sum_{t=1}^{T} (y_t - \boldsymbol{\beta}_0' \mathbf{x}_t) \\ \sum_{t=1}^{T} \mathbf{x}_t (y_t - \boldsymbol{\beta}_0' \mathbf{x}_t) \end{pmatrix} \Rightarrow$$

$$\omega_0 \sqrt{1-\rho^2} \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \Omega_1' \end{pmatrix}^{-1} \left(\int \mathbf{X}_1 \mathbf{X}_1' \right)^{-1} \left(\int \mathbf{X}_1 \mathrm{W}_0 \right),$$

proving (a). Using Lemma 8 (a)-(b) and (d) and CMT, we get

$$T^{-1/2} \sum_{t=1}^{T} (y_t - \boldsymbol{\beta}_0' \mathbf{x}_t) \Rightarrow \omega_0 \kappa \int \mathrm{W}_0 + \pi_0 \int d\mathrm{W}_0 + \left(\int d\mathbf{W}_1' \right) \boldsymbol{\pi}_1,$$

and

$$T^{-1} \sum_{t=1}^{T} \mathbf{x}_t (y_t - \boldsymbol{\beta}_0' \mathbf{x}_t) \Rightarrow \Omega_1 \left(\omega_0 \kappa \int \mathbf{W}_1 \mathrm{W}_0 + \pi_0 \int \mathbf{W}_1 d\mathrm{W}_0 + \left(\int \mathbf{W}_1 d\mathbf{W}_1' \right) \boldsymbol{\pi}_1 + \boldsymbol{\delta} \right),$$

under near cointegration. In other words,

$$\Upsilon_T^{-1} \begin{pmatrix} \sum_{t=1}^{T} (y_t - \boldsymbol{\beta}_0' \mathbf{x}_t) \\ \sum_{t=1}^{T} \mathbf{x}_t (y_t - \boldsymbol{\beta}_0' \mathbf{x}_t) \end{pmatrix} \Rightarrow$$

$$\begin{pmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \Omega_1 \end{pmatrix} \left(\omega_0 \kappa \int \mathbf{X}_1 \mathrm{W}_0 + \pi_0 \int \mathbf{X}_1 d\mathrm{W}_0 + \left(\int \mathbf{X}_1 d\mathbf{W}_1' \right) \boldsymbol{\pi}_1 + \begin{pmatrix} 0 \\ \boldsymbol{\delta} \end{pmatrix} \right), \quad (23)$$

and we have

$$\begin{pmatrix} \sqrt{T}\hat{\alpha} \\ T\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \end{pmatrix} = \left(\Upsilon_T^{-1} \begin{pmatrix} T & \sum_{t=1}^{T} \mathbf{x}_t' \\ \sum_{t=1}^{T} \mathbf{x}_t & \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \end{pmatrix} \Upsilon_T^{-1} \right)^{-1} \Upsilon_T^{-1} \begin{pmatrix} \sum_{t=1}^{T} (y_t - \boldsymbol{\beta}_0' \mathbf{x}_t) \\ \sum_{t=1}^{T} \mathbf{x}_t (y_t - \boldsymbol{\beta}_0' \mathbf{x}_t) \end{pmatrix} \Rightarrow$$

$$\begin{pmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \Omega_1' \end{pmatrix}^{-1} \left(\int \mathbf{X}_1 \mathbf{X}_1'\right)^{-1} \left(\omega_0 \kappa \int \mathbf{X}_1 \mathrm{W}_0 + \pi_0 \int \mathbf{X}_1 d\mathrm{W}_0 + \left(\int \mathbf{X}_1 d\mathbf{W}_1'\right) \boldsymbol{\pi}_1 + \begin{pmatrix} 0 \\ \boldsymbol{\delta} \end{pmatrix}\right),$$

proving (b).■

Before continuing with the proofs of Lemma 2 and Proposition 3, we note that equation (21) along with CMT implies that

$$T^{-2}\left(\sum_{t=1}^{T}(\mathbf{x}_t - \bar{\mathbf{x}})(\mathbf{x}_t - \bar{\mathbf{x}})'\right) = T^{-2}\sum_{t=1}^{T}\mathbf{x}_t \mathbf{x}_t' - \left(T^{-3/2}\sum_{t=1}^{T}\mathbf{x}_t\right)\left(T^{-3/2}\sum_{t=1}^{T}\mathbf{x}_t\right)'$$

$$\Rightarrow \Omega_1 \left(\int \mathbf{W}_1 \mathbf{W}_1' - \left(\int \mathbf{W}_1\right)\left(\int \mathbf{W}_1\right)'\right)\Omega_1' = \Omega_1 \left(\int \mathbf{W}_1^{\mu} \mathbf{W}_1^{\mu\prime}\right)\Omega_1', \qquad (24)$$

under both spurious regression and near cointegration.

8.2. Proof of Lemma 2.

By application of Lemma 1 (a), we have

$$\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \Rightarrow \omega_0 \sqrt{1-\rho^2}\,(\Omega_1')^{-1}\left(\int \mathbf{W}_1^{\mu} \mathbf{W}_1^{\mu\prime}\right)^{-1}\left(\int \mathbf{W}_1^{\mu} \mathrm{W}_0^{\mu}\right)$$

under spurious regression. Along with (24), this implies that

$$T^{-2}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)'\left(\sum_{t=1}^{T}(\mathbf{x}_t - \bar{\mathbf{x}})(\mathbf{x}_t - \bar{\mathbf{x}})'\right)\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \Rightarrow$$

$$\omega_0^2\left(1-\rho^2\right)\left(\int \mathbf{W}_1^{\mu} \mathrm{W}_0^{\mu}\right)'\left(\int \mathbf{W}_1^{\mu} \mathbf{W}_1^{\mu\prime}\right)^{-1}\left(\int \mathbf{W}_1^{\mu} \mathrm{W}_0^{\mu}\right). \qquad (25)$$

Now,

$$T^{-1}\left(p \times F\left(\hat{\boldsymbol{\beta}}\right)\right) = \frac{T^{-2}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)'\left(\sum_{t=1}^{T}(\mathbf{x}_t - \bar{\mathbf{x}})(\mathbf{x}_t - \bar{\mathbf{x}})'\right)\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)}{T^{-2}\sum_{t=1}^{T}\hat{u}_t^2} \times \frac{T-p-1}{T},$$

while

$$\sum_{t=1}^{T}\hat{u}_t^2 = \sum_{t=1}^{T}\left(y_t-\boldsymbol{\beta}_0'\mathbf{x}_t\right)^2 - T^{-1}\left(\sum_{t=1}^{T}\left(y_t-\boldsymbol{\beta}_0'\mathbf{x}_t\right)\right)^2 - \left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)'\left(\sum_{t=1}^{T}\left(\mathbf{x}_t-\bar{\mathbf{x}}\right)\left(\mathbf{x}_t-\bar{\mathbf{x}}\right)'\right)\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right).$$

So, to complete the proof it suffices to notice that Lemma 8 (a)-(b) implies that

$$T^{-2}\left(\sum_{t=1}^{T}\left(y_t-\boldsymbol{\beta}_0'\mathbf{x}_t\right)^2 - T^{-1}\left(\sum_{t=1}^{T}\left(y_t-\boldsymbol{\beta}_0'\mathbf{x}_t\right)\right)^2\right)$$

$$\Rightarrow \omega_0^2\left(1-\rho^2\right)\int\left(\mathrm{W}_0\right)^2dr - \omega_0^2\left(1-\rho^2\right)\left(\int \mathrm{W}_0\right)^2 = \omega_0^2\left(1-\rho^2\right)\int\left(\mathrm{W}_0^{\mu}\right)^2. \blacksquare$$

8.3. Proof of Proposition 3.

By application of Lemma 1, we have

$$T\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)\Rightarrow$$

$$\left(\Omega_1'\right)^{-1}\left(\int \mathbf{W}_1^{\mu}\mathbf{W}_1^{\mu\prime}\right)^{-1}\left(\omega_0\kappa\int \mathbf{W}_1^{\mu}\mathrm{W}_0^{\mu}+\pi_0\int \mathbf{W}_1^{\mu}d\mathrm{W}_0+\left(\int \mathbf{W}_1^{\mu}d\mathbf{W}_1'\right)\boldsymbol{\pi}_1+\boldsymbol{\delta}\right),$$

under near cointegration. Along with (24) this implies that

$$\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)'\left(\sum_{t=1}^{T}\left(\mathbf{x}_t-\bar{\mathbf{x}}\right)\left(\mathbf{x}_t-\bar{\mathbf{x}}\right)'\right)\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)\Rightarrow$$

$$\left(\omega_0\kappa\int \mathbf{W}_1^{\mu}\mathrm{W}_0^{\mu}+\pi_0\int \mathbf{W}_1^{\mu}d\mathrm{W}_0+\left(\mathbf{W}_1^{\mu}d\mathbf{W}_1'\right)\boldsymbol{\pi}_1+\boldsymbol{\delta}\right)' \times\left(\int \mathbf{W}_1^{\mu}\mathbf{W}_1^{\mu\prime}\right)^{-1} \times\left(\omega_0\kappa\int \mathbf{W}_1^{\mu}\mathrm{W}_0^{\mu}+\pi_0\int \mathbf{W}_1^{\mu}d\mathrm{W}_0+\left(\int \mathbf{W}_1^{\mu}d\mathbf{W}_1'\right)\boldsymbol{\pi}_1+\boldsymbol{\delta}\right).$$

Therefore, the proof is complete once we have established that $s^2 \to_p \sigma^2$. Now, by Lemma 8 (a)-(c),

$$s^2 = \frac{T}{T-p-1} \times T^{-1} \sum_{t=1}^{T} \hat{u}_t^2 = \frac{T}{T-p-1} \times T^{-1} \sum_{t=1}^{T} \left(y_t - \boldsymbol{\beta}_0' \mathbf{x}_t\right)^2 + o_p(1)$$

$$= \frac{T}{T-p-1} \times \left(1 \quad -\bar{\boldsymbol{\beta}}_0' \right) \left(T^{-1} \sum_{t=1}^{T} \mathbf{C}(L)\, \mathbf{e}_t \left(\mathbf{C}(L)\, \mathbf{e}_t\right)' \right) \begin{pmatrix} 1 \\ -\bar{\boldsymbol{\beta}}_0 \end{pmatrix} + o_p(1)$$

$$\to_p \left(1 \quad -\bar{\boldsymbol{\beta}}_0' \right) \left(\sum_{i=0}^{\infty} \mathbf{C}_i \mathbf{C}_i' \right) \begin{pmatrix} 1 \\ -\bar{\boldsymbol{\beta}}_0 \end{pmatrix} = \sigma^2. \blacksquare$$

8.4. Proof of Corollary 4.
Under the assumptions of the corollary, $s^2 \to_p \sigma^2$ and $\hat{\pi}_0^2 \to_p \pi_0^2 \neq 0$. Therefore,

$$G\left(\hat{\boldsymbol{\beta}}\right) = \frac{s^2}{\hat{\pi}_0^2} \left(p \times F\left(\hat{\boldsymbol{\beta}}\right)\right)$$

$$= \left(\frac{\sigma^2}{\pi_0^2} + o_p(1) \right) \cdot \left(p \times F\left(\hat{\boldsymbol{\beta}}\right)\right),$$

and the result follows immediately from Proposition 3 by setting $\boldsymbol{\pi}_1 = \boldsymbol{\delta} = \mathbf{0}. \blacksquare$

8.5. Proof of Proposition 5.
We shall prove Proposition 5 assuming $k_1 \geq 1$ and $k_2 \geq 1$. The case where $k_1 = 0$ or $k_2 = 0$ can be treated in exactly the same fashion. As is readily verified,

$$\sum_{t=1}^{T} \hat{u}_t^2 - \sum_{t=1}^{T} \tilde{u}_t^2 = \left(\sum_{t=1}^{T} \mathbf{X}_{2.1,t} \left(y_t - \boldsymbol{\beta}_0' \mathbf{x}_t\right) \right)' \left(\sum_{t=1}^{T} \mathbf{X}_{2.1,t} \mathbf{X}_{2.1,t}' \right)^{-1} \left(\sum_{t=1}^{T} \mathbf{X}_{2.1,t} \left(y_t - \boldsymbol{\beta}_0' \mathbf{x}_t\right) \right),$$

where

$$\mathbf{X}_{2.1,t} := \begin{pmatrix} \mathbf{d}_t \\ \mathbf{z}_t \end{pmatrix} - \left(\begin{pmatrix} 1 \\ \mathbf{x}_t \end{pmatrix}' \begin{pmatrix} T & \sum_{s=1}^{T} \mathbf{x}_s' \\ \sum_{s=1}^{T} \mathbf{x}_s & \sum_{s=1}^{T} \mathbf{x}_s \mathbf{x}_s' \end{pmatrix}^{-1} \begin{pmatrix} \sum_{s=1}^{T} \mathbf{d}_s' & \sum_{s=1}^{T} \mathbf{z}_s' \\ \sum_{s=1}^{T} \mathbf{x}_s \mathbf{d}_s' & \sum_{s=1}^{T} \mathbf{x}_s \mathbf{z}_s' \end{pmatrix} \right)'.$$

Let Ψ_T be a diagonal normalization matrix defined as

$$\Psi_T := \begin{pmatrix} \sqrt{T} \cdot \operatorname{diag}\left(T, \ldots, T^{k_1}\right) & \mathbf{0} \\ \mathbf{0} & T \cdot \mathbf{I}_{k_2} \end{pmatrix}.$$

Under the assumptions of Proposition 5, the following hold jointly with Lemma 8:

$$\sqrt{T}\Psi_T^{-1} \begin{pmatrix} \mathbf{d}_{[Tr]} \\ \mathbf{z}_{[Tr]} \end{pmatrix} \Rightarrow \begin{pmatrix} r \\ \vdots \\ r^{k_1} \\ \mathbf{W}_2(r) \end{pmatrix} = \mathbf{X}_2(r), \tag{26}$$

$$\Psi_T^{-1} \sum_{t=1}^{T} \begin{pmatrix} \mathbf{d}_t \\ \mathbf{z}_t \end{pmatrix} (y_t - \boldsymbol{\beta}_0' \mathbf{x}_t) \Rightarrow \omega_0 \kappa \int \mathbf{X}_2 \mathrm{W}_0 + \pi_0 \int \mathbf{X}_2 d\mathrm{W}_0, \tag{27}$$

where $\mathbf{W}_2$ is a k_2-dimensional Wiener process independent of $\mathbf{W}$. Recalling that $\boldsymbol{\pi}_1 = \boldsymbol{\delta} = \mathbf{0}$ under the assumptions of Proposition 5, we deduce from (21), (23), (26), and (27) that

$$\Psi_T^{-1} \sum_{t=1}^{T} \mathbf{X}_{2.1,t} (y_t - \boldsymbol{\beta}_0' \mathbf{x}_t) \Rightarrow \omega_0 \kappa \int \mathbf{X}_{2.1} \mathrm{W}_0 + \pi_0 \int \mathbf{X}_{2.1} d\mathrm{W}_0 = \pi_0 \int \mathbf{X}_{2.1} d\mathrm{V}_\lambda,$$

and

$$\Psi_T^{-1} \left(\sum_{t=1}^{T} \mathbf{X}_{2.1,t} \mathbf{X}_{2.1,t}' \right) \Psi_T^{-1} \Rightarrow \int \mathbf{X}_{2.1} \mathbf{X}_{2.1}'.$$

Hence

$$\sum_{t=1}^{T} \hat{u}_t^2 - \sum_{t=1}^{T} \tilde{u}_t^2 \Rightarrow \pi_0^2 \left(\int \mathbf{X}_{2.1} d\mathrm{V}_\lambda \right)' \left(\int \mathbf{X}_{2.1} \mathbf{X}_{2.1}' \right)^{-1} \left(\int \mathbf{X}_{2.1} d\mathrm{V}_\lambda \right),$$

and the conclusion follows since $\hat{\pi}_0^2 \to_p \pi_0^2$. ■

8.6. Proof of Proposition 6.

Using (21), (23), and Lemma 1 (b), we have

$$T^{-1/2}\hat{S}_{[Tr]} = T^{-1/2}\sum_{t=1}^{[Tr]}(y_t - \boldsymbol{\beta}_0'\mathbf{x}_t) - \left(T^{-1/2}\Upsilon_T^{-1}\sum_{t=1}^{[Tr]}\begin{pmatrix}1\\ \mathbf{x}_t\end{pmatrix}\right)'\left(\Upsilon_T\begin{pmatrix}\hat{\alpha}\\ \hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\end{pmatrix}\right)$$

$$\Rightarrow \pi_0 \mathrm{V}_\lambda(r) - \pi_0\left(\int_0^r \mathbf{X}_1(s)\,ds\right)'\left(\int \mathbf{X}_1\mathbf{X}_1\right)^{-1}\left(\int \mathbf{X}_1 d\mathrm{V}_\lambda\right) = \pi_0\hat{\mathrm{V}}_\lambda(r).$$

Therefore, by CMT,

$$CI = \frac{T^{-2}\sum_{t=1}^T \hat{S}_t^2}{\hat{\pi}_0^2} = \frac{T^{-1}\sum_{t=1}^T\left(T^{-1/2}\hat{S}_t\right)^2}{\pi_0^2 + o_p(1)} \Rightarrow \int\left(\hat{\mathrm{V}}_\lambda\right)^2,$$

as claimed. Similar reasoning yields

$$\Upsilon_T^{-1}\mathbf{S}_{[Tr]}^* = \Upsilon_T^{-1}\sum_{t=1}^{[Tr]}\begin{pmatrix}1\\ \mathbf{x}_t\end{pmatrix}\hat{u}_t \Rightarrow$$

$$\pi_0\begin{pmatrix}1 & \mathbf{0}'\\ \mathbf{0} & \Omega_1\end{pmatrix}\left(\int_0^r \mathbf{X}_1(s)\,d\mathrm{V}_\lambda(s) - \left(\int_0^r \mathbf{X}_1(s)\mathbf{X}_1(s)'\,ds\right)'\left(\int \mathbf{X}_1\mathbf{X}_1'\right)^{-1}\left(\int \mathbf{X}_1 d\mathrm{V}_\lambda\right)\right)$$

$$= \pi_0\begin{pmatrix}1 & \mathbf{0}'\\ \mathbf{0} & \Omega_1\end{pmatrix}\int_0^r \mathbf{X}_1(s)\,d\hat{\mathrm{V}}_\lambda(s) = \pi_0\begin{pmatrix}1 & \mathbf{0}'\\ \mathbf{0} & \Omega_1\end{pmatrix}\mathbf{V}_\lambda^*(r). \tag{28}$$

From (22), we have

$$\Upsilon_T^{-1}\mathbf{M}_T\Upsilon_T^{-1} \Rightarrow \begin{pmatrix}1 & \mathbf{0}'\\ \mathbf{0} & \Omega_1\end{pmatrix}\left(\int \mathbf{X}_1\mathbf{X}_1'\right)\begin{pmatrix}1 & \mathbf{0}'\\ \mathbf{0} & \Omega_1'\end{pmatrix}.$$

Along with (28) and CMT, this implies that

$$L_c = T^{-1}\frac{\sum_{t=1}^T \mathbf{S}_t^{*\prime}\mathbf{M}_T^{-1}\mathbf{S}_t^*}{\hat{\pi}_0^2} = \frac{T^{-1}\sum_{t=1}^T\left(\Upsilon_T^{-1}\mathbf{S}_t^*\right)'\left(\Upsilon_T^{-1}\mathbf{M}_T\Upsilon_T^{-1}\right)^{-1}\left(\Upsilon_T^{-1}\mathbf{S}_t^*\right)}{\pi_0^2 + o_p(1)}$$

$$\Rightarrow \int_0^1 \mathbf{V}_\lambda^*(r)' \left(\int_0^1 \mathbf{X}_1(s) \mathbf{X}_1(s)' ds \right)^{-1} \mathbf{V}_\lambda^*(r) dr,$$

as claimed.■

8.7. Proof of Proposition 7.

We have

$$\begin{pmatrix} \check{\alpha} \\ \check{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \end{pmatrix} = \begin{pmatrix} \sum_{t=1}^T t^2 & \sum_{t=1}^T t\mathbf{S}_t^{\mathbf{x}\prime} \\ \sum_{t=1}^T t\mathbf{S}_t^{\mathbf{x}} & \sum_{t=1}^T \mathbf{S}_t^{\mathbf{x}} \mathbf{S}_t^{\mathbf{x}\prime} \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^T tS_t \\ \sum_{t=1}^T \mathbf{S}_t^{\mathbf{x}} S_t \end{pmatrix},$$

where $S_t := \sum_{s=1}^t (y_s - \boldsymbol{\beta}_0' \mathbf{x}_s)$. By (21), (23) and CMT, we have

$$T^{-1/2}\Upsilon_T^{-1} \begin{pmatrix} [Tr] \\ \mathbf{S}_{[Tr]}^{\mathbf{x}} \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \Omega_1 \end{pmatrix} \int_0^r \mathbf{X}_1(s) ds = \begin{pmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \Omega_1 \end{pmatrix} \bar{\mathbf{X}}_1(r), \qquad (29)$$

$$T^{-1/2} S_{[Tr]} \Rightarrow \omega_0 \kappa \int_0^r \mathrm{W}_0(s) ds + \pi_0 \int_0^r d\mathrm{W}_0(s) = \pi_0 \mathrm{V}_\lambda(r). \qquad (30)$$

Consequently, it is easy to show that

$$\Upsilon_T \begin{pmatrix} \check{\alpha} \\ \check{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \end{pmatrix} \Rightarrow \pi_0 \begin{pmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \Omega_1' \end{pmatrix}^{-1} \left(\int \bar{\mathbf{X}}_1 \bar{\mathbf{X}}_1' \right)^{-1} \left(\int \bar{\mathbf{X}}_1 \mathrm{V}_\lambda \right). \qquad (31)$$

Therefore,

$$T^{-1/2} \check{S}_{[Tr]} = T^{-1/2} S_{[Tr]} - \left(T^{-1/2} \Upsilon_T^{-1} \begin{pmatrix} [Tr] \\ \mathbf{S}_{[Tr]}^{\mathbf{x}} \end{pmatrix} \right)' \left(\Upsilon_T \begin{pmatrix} \check{\alpha} \\ \check{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \end{pmatrix} \right) \Rightarrow$$

$$\pi_0 \mathrm{V}_\lambda(r) - \pi_0 \bar{\mathbf{X}}_1(r)' \left(\bar{\mathbf{X}}_1 \bar{\mathbf{X}}_1' \right)^{-1} \left(\int \bar{\mathbf{X}}_1 \mathrm{V}_\lambda \right) = \pi_0 \check{\mathrm{V}}_\lambda(r),$$

and by CMT we have

$$SBDH_I = \frac{T^{-2} \sum_{t=1}^T \check{S}_t^2}{\check{\pi}_0^2} = \frac{T^{-1} \sum_{t=1}^T \left(T^{-1/2} \check{S}_t \right)^2}{\pi_0^2 + o_p(1)} \Rightarrow \int \left(\check{\mathrm{V}}_\lambda \right)^2,$$

as claimed. Now,

$$LM_{II} = \frac{LM_I}{SBDH_I + o_p(1)},$$

while

$$LM_I = \frac{\left(T^{-1}\sum_{t=2}^{T} \check{S}_{t-1}\Delta\check{S}_t - \frac{\check{\pi}_0^2-\check{\sigma}^2}{2}\right)^2}{\check{\pi}_0^4} = \frac{\left(T^{-1}\sum_{t=2}^{T} \check{S}_{t-1}\Delta\check{S}_t - \frac{\pi_0^2-\sigma^2}{2} + o_p(1)\right)^2}{\pi_0^4 + o_p(1)}.$$

Therefore, to complete the proof it suffices to establish that

$$T^{-1}\sum_{t=2}^{T} \check{S}_{t-1}\Delta\check{S}_t \Rightarrow \pi_0^2 \int \check{\mathrm{V}}_\lambda d\check{\mathrm{V}}_\lambda + \frac{\pi_0^2-\sigma^2}{2}. \tag{32}$$

Upon defining

$$\ddot{S}_t := \check{S}_t - S_t = \begin{pmatrix} t \\ \mathbf{S}_t^{\mathbf{x}} \end{pmatrix}' \begin{pmatrix} \check{\alpha} \\ \check{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \end{pmatrix},$$

we have

$$T^{-1}\sum_{t=2}^{T} \check{S}_{t-1}\Delta\check{S}_t =$$

$$T^{-1}\sum_{t=2}^{T} S_{t-1}\Delta S_t - T^{-1}\sum_{t=2}^{T} \ddot{S}_{t-1}\Delta S_t - T^{-1}\sum_{t=2}^{T} S_{t-1}\Delta\ddot{S}_t + T^{-1}\sum_{t=2}^{T} \ddot{S}_{t-1}\Delta\ddot{S}_t. \tag{33}$$

It follows from Phillips (1988b) that the following holds jointly with Lemma 8:

$$\begin{pmatrix} 1 & -\boldsymbol{\beta}_0' \end{pmatrix} \left(T^{-1}\sum_{t=2}^{T}\left(\sum_{s=1}^{t-1} \mathbf{C}(L)\,\mathbf{e}_s\right)(\mathbf{C}(L)\,\mathbf{e}_t)'\right) \begin{pmatrix} 1 \\ -\boldsymbol{\beta}_0 \end{pmatrix} \Rightarrow$$

$$\begin{pmatrix} 1 & -\bar{\boldsymbol{\beta}}_0' \end{pmatrix} \left(\mathbf{C}(1) \left(\int \mathbf{W} d\mathbf{W}' \right) \mathbf{C}(1)' + \Lambda \right) \begin{pmatrix} 1 \\ -\bar{\boldsymbol{\beta}}_0 \end{pmatrix}$$

$$= \pi_0^2 \int \mathrm{W}_0 d\mathrm{W}_0 + \begin{pmatrix} 1 & -\bar{\boldsymbol{\beta}}_0' \end{pmatrix} \Lambda \begin{pmatrix} 1 \\ -\bar{\boldsymbol{\beta}}_0 \end{pmatrix}, \tag{34}$$

where

$$\Lambda := \sum_{k=1}^{\infty} \mathrm{E} \left(\mathbf{C}(L) \mathbf{e}_0 \left(\mathbf{C}(L) \mathbf{e}_k \right)' \right).$$

Now,

$$\Lambda + \Lambda' = \mathbf{C}(1) \mathbf{C}(1)' - \sum_{i=0}^{\infty} \mathbf{C}_i \mathbf{C}_i',$$

and we therefore have

$$\begin{pmatrix} 1 & -\bar{\boldsymbol{\beta}}_0' \end{pmatrix} \Lambda \begin{pmatrix} 1 \\ -\bar{\boldsymbol{\beta}}_0 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & -\bar{\boldsymbol{\beta}}_0' \end{pmatrix} (\Lambda + \Lambda') \begin{pmatrix} 1 \\ -\bar{\boldsymbol{\beta}}_0 \end{pmatrix}$$

$$= \frac{1}{2} \left(\begin{pmatrix} 1 & -\bar{\boldsymbol{\beta}}_0' \end{pmatrix} \mathbf{C}(1) \mathbf{C}(1)' \begin{pmatrix} 1 \\ -\bar{\boldsymbol{\beta}}_0 \end{pmatrix} - \begin{pmatrix} 1 & -\bar{\boldsymbol{\beta}}_0' \end{pmatrix} \left(\sum_{i=0}^{\infty} \mathbf{C}_i \mathbf{C}_i' \right) \begin{pmatrix} 1 \\ -\bar{\boldsymbol{\beta}}_0 \end{pmatrix} \right)$$

$$= \frac{1}{2} \left(\pi_0^2 - \sigma^2 \right). \tag{35}$$

Using (34), (35), and straightforward algebra, we arrive at

$$T^{-1} \sum_{t=2}^{T} S_{t-1} \Delta S_t \Rightarrow \pi_0^2 \int \mathrm{V}_\lambda d\mathrm{V}_\lambda + \frac{\pi_0^2 - \sigma^2}{2}. \tag{36}$$

Moreover, using (29), (30), (31), and integration by parts it is not hard to show that

$$T^{-1} \sum_{t=2}^{T} \ddot{S}_{t-1} \Delta S_t \Rightarrow \pi_0^2 \left(\int \bar{\mathbf{X}}_1 \mathrm{V}_\lambda \right)' \left(\int \bar{\mathbf{X}}_1 \bar{\mathbf{X}}_1' \right)^{-1} \left(\int \bar{\mathbf{X}}_1 d\mathrm{V}_\lambda \right), \tag{37}$$

$$T^{-1}\sum_{t=2}^{T} S_{t-1}\Delta\ddot{S}_t \Rightarrow \pi_0^2\left(\int \mathrm{V}_\lambda d\bar{\mathbf{X}}_1\right)'\left(\int \bar{\mathbf{X}}_1\bar{\mathbf{X}}_1'\right)^{-1}\left(\int \bar{\mathbf{X}}_1\mathrm{V}_\lambda\right), \tag{38}$$

and

$$T^{-1}\sum_{t=2}^{T} \ddot{S}_{t-1}\Delta\ddot{S}_t \Rightarrow$$

$$\pi_0^2\left(\int \bar{\mathbf{X}}_1\mathrm{V}_\lambda\right)'\left(\int \bar{\mathbf{X}}_1\bar{\mathbf{X}}_1'\right)^{-1}\left(\int \bar{\mathbf{X}}_1 d\bar{\mathbf{X}}_1\right)'\left(\int \bar{\mathbf{X}}_1\bar{\mathbf{X}}_1'\right)^{-1}\left(\int \bar{\mathbf{X}}_1\mathrm{V}_\lambda\right). \tag{39}$$

Using (36) – (39), we obtain (32) from (33).■

8.8. Notation.

$:=$	definitional equality.
$[\cdot]$	integer part of.
$\Rightarrow$	weak convergence (of the associated probability measures).
$\rightarrow_p$	convergence in probability.
$o_p(1)$	tends to zero in probability.
$O_p(1)$	bounded in probability.

8.9. Processes Appearing in the Paper.

Let W_0 and $\mathbf{W}_1$ be independent Wiener processes of dimensions 1 and $p \geq 1$, respectively. Moreover, for $k_1 \geq 1$ and $k_2 \geq 1$, let $\mathbf{X}_2 := (\mathbf{f}', \mathbf{W}_2')'$,where $\mathbf{f}(r) := \left(r, \ldots, r^{k_1}\right)'$ and $\mathbf{W}_2$ is a k_2-dimensional Wiener process independent of $(\mathrm{W}_0, \mathbf{W}_1')'$. When $k_1 = 0$ $(k_2 = 0)$, $\mathbf{f}$ $(\mathbf{W}_2)$ is omitted from $\mathbf{X}_2$. Finally, let λ be a scalar parameter. Using W_0, $\mathbf{W}_1$, $\mathbf{X}_2$, and λ, we define the following processes, listed in order of appearance in the paper:

$$\mathbf{X}_1(r) := \begin{pmatrix} 1 \\ \mathbf{W}_1(r) \end{pmatrix},$$

$$\mathrm{W}_0^\mu(r) := \mathrm{W}_0(r) - \int_0^r \mathrm{W}_0(s)\,ds,$$

$$\mathbf{W}_1^\mu(r) := \mathbf{W}_1(r) - \int_0^r \mathbf{W}_1(s)\,ds,$$

$$\mathbf{X}_{2.1}(r) := \mathbf{X}_2(r) - \mathbf{X}_1(r)' \left(\int_0^1 \mathbf{X}_1(s)\mathbf{X}_1(s)'ds\right)^{-1} \left(\int_0^1 \mathbf{X}_1(s)\mathbf{X}_2(s)'ds\right),$$

$$\mathrm{V}_\lambda(r) := \lambda \int_0^r \mathrm{W}_0(s)\,ds + \mathrm{W}_0(r),$$

$$\hat{\mathrm{V}}_\lambda(r) := \mathrm{V}_\lambda(r) - \left(\int_0^r \mathbf{X}_1(s)\,ds\right)' \left(\int_0^1 \mathbf{X}_1(s)\mathbf{X}_1(s)'ds\right)^{-1} \left(\int_0^1 \mathbf{X}_1(s)\,d\mathrm{V}_\lambda(s)\right),$$

$$\mathbf{V}_\lambda^*(r) := \int_0^r \mathbf{X}_1(s)\,d\hat{\mathrm{V}}_\lambda(s),$$

$$\bar{\mathbf{X}}_1(r) := \int_0^r \mathbf{X}_1(s)\,ds,$$

$$\check{\mathrm{V}}_\lambda(r) := \mathrm{V}_\lambda(r) - \bar{\mathbf{X}}_1(r)' \left(\int_0^1 \bar{\mathbf{X}}_1(s)\bar{\mathbf{X}}_1(s)'ds\right)^{-1} \left(\int_0^1 \bar{\mathbf{X}}_1(s)\mathrm{V}_\lambda(s)\,ds\right).$$

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 T. LEE (1990): "Multicointegration," in *Advances in Econometrics: Co-integration, Spurious Regressions, and Unit Roots*, ed. by T. B. Fomby, and G. F. Rhodes, Jr. Greenwich, CT: Jai Press, 71-84.

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.

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.

JOHANSEN, S., AND K. JUSELIUS (1990): "Maximum Likelihood Estimation and Inference on Cointegration - with Applications to the Demand for Money," *Oxford bulletin of Economics and Statistics*, 52, 169–210.

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.

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

——— (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 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.

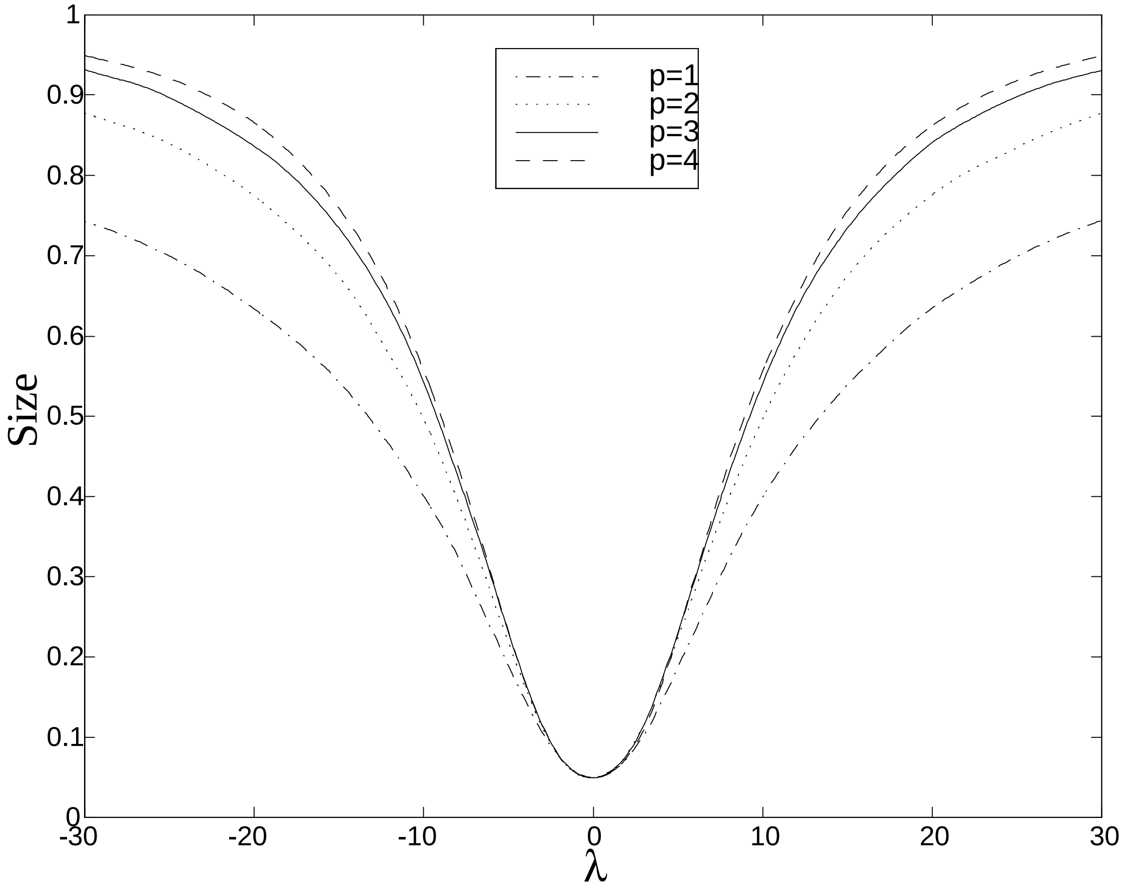

Figure 1: Rejection rates for $G\left(\hat{\boldsymbol{\beta}}\right)$; Nominal Size is 5%.

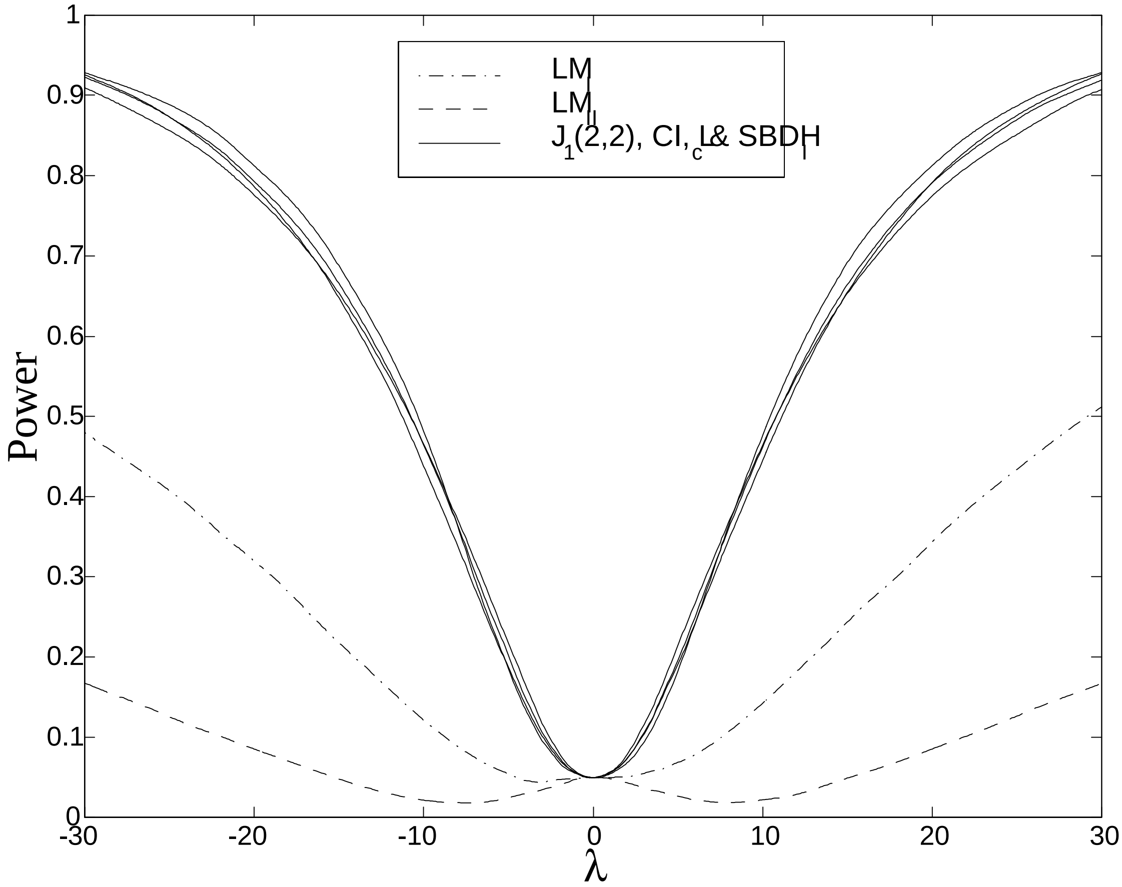

Figure 2: Local Power of Tests for Cointegration; $p = 1$.

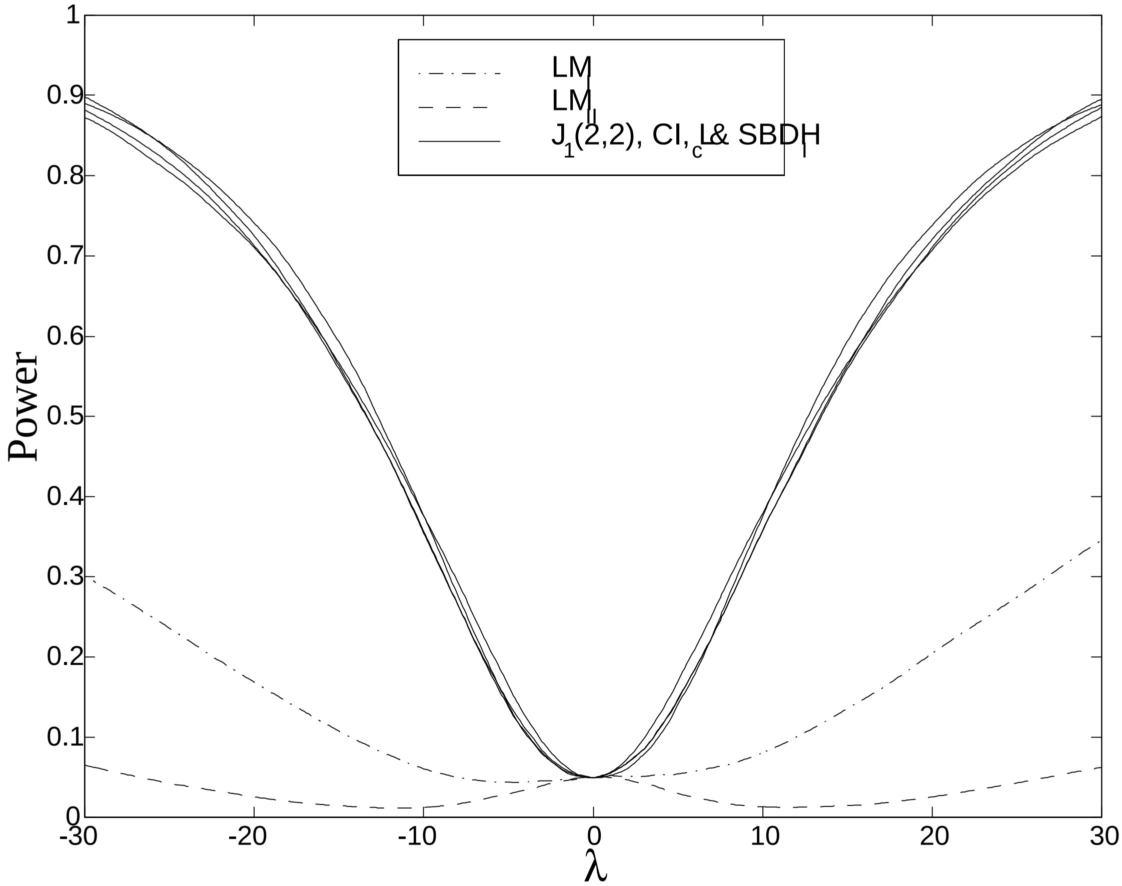

Figure 3: Local Power of Tests for Cointegration; $p = 2$.

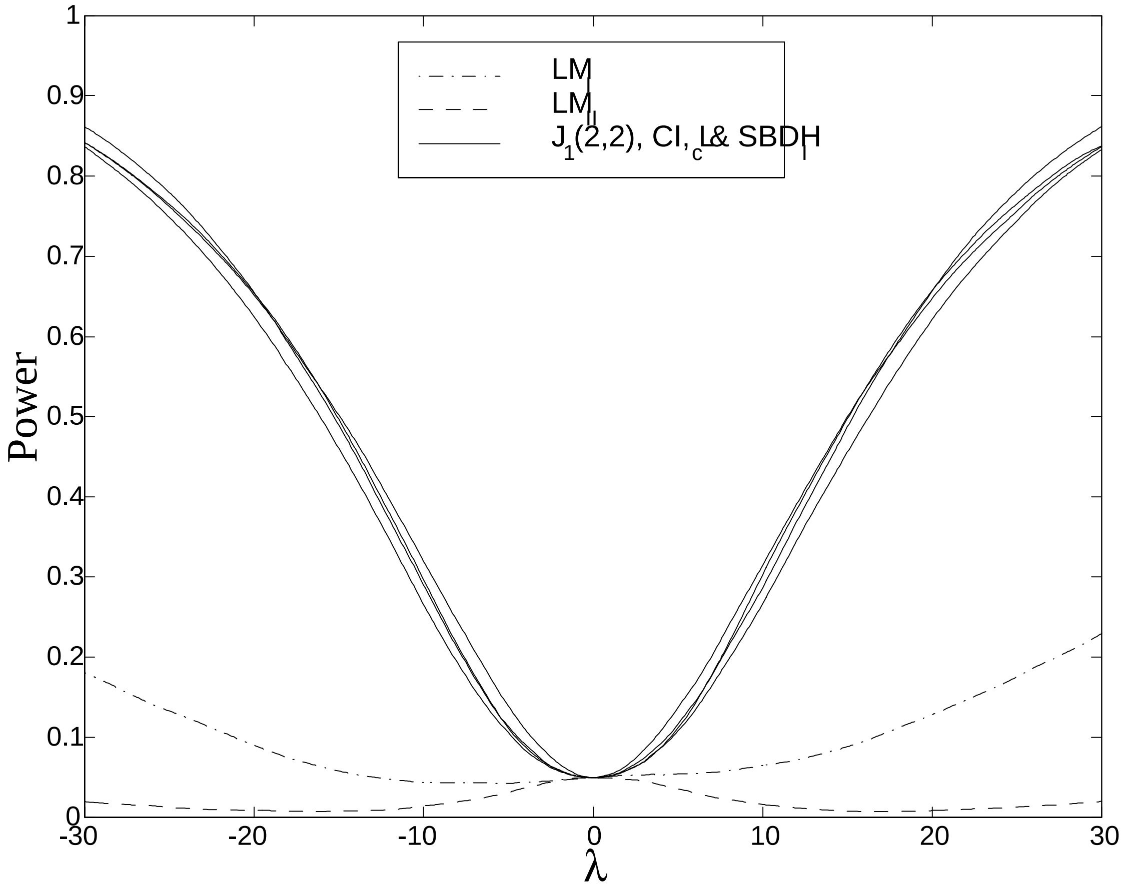

Figure 4: Local Power of Tests for Cointegration; $p = 3$.

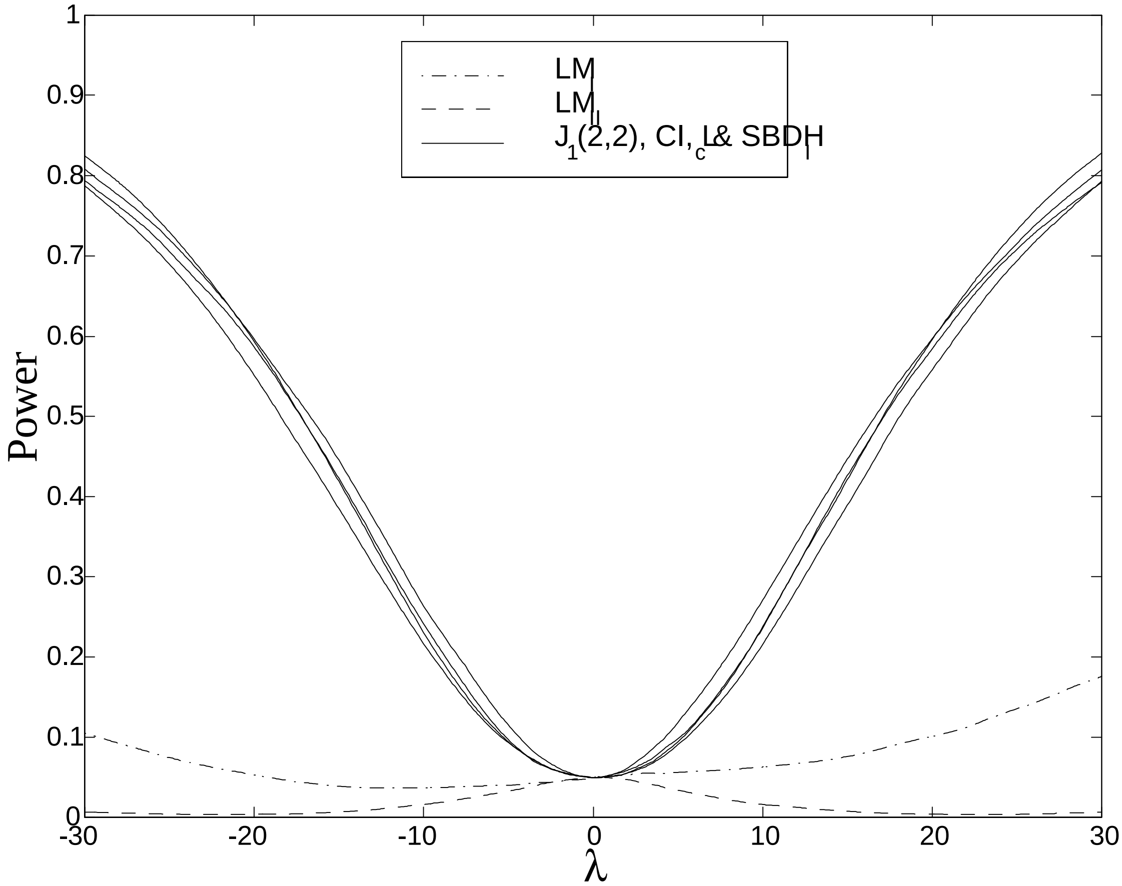

Figure 5: Local Power of Tests for Cointegration; $p = 4$.