Improved Likelihood Ratio Tests for Cointegration Rank in the VAR Model*

H. Peter Boswijk
University of Amsterdam

Michael Jansson†
UC Berkeley and CREATES

Morten Ørregaard Nielsen
Queen's University and CREATES

March 5, 2012

Abstract. We suggest improved tests for cointegration rank in the vector autoregressive (VAR) model and develop asymptotic distribution theory and local power results. The tests are (quasi-)likelihood ratio tests based on a Gaussian likelihood, but of course the asymptotic results apply more generally. The power gains relative to existing tests are due to two factors. First, instead of basing our tests on the conditional (with respect to the initial observations) likelihood, we follow the recent unit root literature and base our tests on the full likelihood as in, e.g., Elliott, Rothenberg, and Stock (1996). Secondly, our tests incorporate a "sign" restriction which generalizes the one-sided unit root test. We show that the asymptotic local power of the proposed tests dominates that of existing cointegration rank tests.

Keywords: Cointegration rank, efficiency, likelihood ratio test, vector autoregression
JEL Codes: C12, C32

1. Introduction

The cointegrated vector autoregressive (VAR) model has been and continues to be of great importance in time series econometrics. Driven equally by theoretical interest and the needs of applied work, the seminal work of Johansen (1988) and Johansen (1991) developed cointegration rank tests within the VAR model.[1] Subsequent contributions have generalized and refined this work in a variety of ways, notably by proposing tests with (local asymptotic) power properties superior to those of Johansen (e.g., Xiao and Phillips (1999), Hubrich, Lütkepohl, and Saikkonen (2001), and the references therein). The purpose of this paper is propose cointegration rank tests that share with the Johansen tests the feature that they are of (quasi-)likelihood ratio type, yet enjoy the additional attraction that they dominate existing tests (including those of Johansen) in terms of local asymptotic power.

In the related unit root testing literature, it has long been recognized that in models with an unknown mean and/or a linear trend, the class of nearly efficient unit root tests does not

*We are grateful to seminar participants at the University of Exeter and the Tinbergen Institute for comments, and to the Social Sciences and Humanities Research Council of Canada (SSHRC grant no. 410-2009-0183) and the Center for Research in Econometric Analysis of Time Series (CREATES, funded by the Danish National Research Foundation) for financial support.

†Corresponding author. Address: University of California, Berkeley, Department of Economics, 508-1 Evans Hall #3880, Berkeley, CA 94720. Email: mjansson@econ.berkeley.edu

[1] For a synthesis of the work by Johansen, see Johansen (1995).

contain the Dickey and Fuller (1979, 1981, henceforth Dickey-Fuller) tests[2], which can be derived from a conditional (with respect to the initial observation) likelihood similar to the Johansen cointegration rank tests. It was pointed out by Elliott, Rothenberg, and Stock (1996) that the initial observation is very informative about the parameters governing the deterministic component, and, indeed, Jansson and Nielsen (2011) showed that a likelihood ratio test derived from the full likelihood implied by an Elliott-Rothenberg-Stock-type model has superior power properties to those of the Dickey-Fuller tests in models with deterministic components.

Like the Dickey-Fuller tests for unit roots, the cointegration rank tests due to Johansen (1991) are derived from a conditional likelihood. In this paper we suggest improved tests for cointegration rank in the VAR model, which are based on the full likelihood similar to the unit root tests of Elliott, Rothenberg, and Stock (1996) and Jansson and Nielsen (2011). We show that their qualitative finding about the relative merits of likelihood ratio tests derived from conditional and full likelihoods extends to tests of cointegration rank. In addition, our tests incorporate a "sign" restriction which generalizes the one-sided unit root test. We develop the asymptotic distribution theory and show that the asymptotic local power of the proposed tests dominates that of existing cointegration rank tests.

The remainder of the paper is laid out as follows. Section 2 contains our results on the likelihood ratio tests for cointegration rank, which are derived in several steps with each subsection adding another layer of complexity. Section 3 evaluates the asymptotic null distributions and local power functions of the newly proposed tests by Monte Carlo simulation. Some additional discussion is given in Section 4. The proofs of our theorems are provided in Section 5.

2. Likelihood Ratio Statistics

Our development of test statistics proceeds in four steps, each step involving accommodation of nuisance parameters not present in the previous step.

2.1. Unit Root Testing in the Zero-mean VAR(1) Model. We initially consider the simplest special case, namely likelihood ratio tests of the multivariate unit root hypothesis $\Pi = 0$ in the p-dimensional zero-mean Gaussian VAR(1) model

$$\Delta y_t = \Pi y_{t-1} + \varepsilon_t, \tag{1}$$

where $y_0 = 0$, $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0, I_p)$, and $\Pi \in \mathbb{R}^{p \times p}$ is an unknown parameter of interest.

In our investigation of the large-sample properties of test statistics, we will follow much of the recent literature on unit root and cointegration testing and use "local-to-unity" asymptotics in order to obtain local asymptotic power results. When testing the multivariate unit root hypothesis $\Pi = 0$ in the model (1), this amounts to employing the reparameterization

$$\Pi = \Pi_T(C) = T^{-1}C \tag{2}$$

and holding $C \in \mathbb{R}^{p \times p}$ fixed as $T \to \infty$.

The statistics we consider are of the form

$$LR_T(\mathcal{C}) = \sup_{\bar{C} \in \mathcal{C}} L_T(\bar{C}) - L_T(0), \tag{3}$$

[2] For a review focusing on power, see Haldrup and Jansson (2006).

where

$$L_T(C) = -\frac{1}{2}\sum_{t=1}^{T} \|\Delta y_t - \Pi_T(C)\, y_{t-1}\|^2$$

is the log likelihood function (modulo an unimportant constant), $\|\cdot\|$ is the Euclidean norm, and $\mathcal{C}$ is some subset of $\mathbb{R}^{p\times p}$. As the notation suggests, the statistic $LR_T(\mathcal{C})$ is a likelihood ratio statistic. Specifically, $LR_T(\mathcal{C})$ is a likelihood ratio statistic associated with the problem of testing the null hypothesis $C = 0$ against the alternative $C \in \mathcal{C}\backslash\{0\}$.[3] Equivalently, $LR_T(\mathcal{C})$ is a likelihood ratio statistic associated with the problem of testing the null hypothesis $\Pi = 0$ against the alternative $\Pi \in \Pi_T(\mathcal{C}) \setminus \{0\}$, where $\Pi_T(\mathcal{C}) = \{\Pi_T(C) : C \in \mathcal{C}\}$.

To give examples of statistics that can be represented as in (3), let $\mathcal{M}_p(r)$ denote the set of elements of $\mathbb{R}^{p\times p}$ with rank no greater than r. For $r = 1, \ldots, p$, it can be shown that

$$LR_T(\mathcal{M}_p(r)) = \frac{1}{2}\sum_{j=1}^{r} \lambda_j,$$

where $\lambda_1 \geq \ldots \geq \lambda_p \geq 0$ are the eigenvalues of the matrix

$$\left(\sum_{t=2}^{T} y_{t-1}\Delta y_t'\right)' \left(\sum_{t=2}^{T} y_{t-1}y_{t-1}'\right)^{-1} \left(\sum_{t=2}^{T} y_{t-1}\Delta y_t'\right).$$

The choices $\mathcal{C} = \mathcal{M}_p(1)$ and $\mathcal{C} = \mathcal{M}_p(p)$ are therefore seen to give rise to "known variance" versions of the so-called maximum eigenvalue and trace statistics, respectively, e.g., Johansen (1995).[4]

Setting $\mathcal{C}$ equal to a set of the form $\mathcal{M}_p(r)$ is computationally and analytically convenient insofar as it gives rise to a statistic $LR_T(\mathcal{C})$ admitting a closed form solution. However, the fact that $\mathcal{C}$ implicitly characterizes the maintained hypothesis of the testing problem suggests that improvements in power against cointegrating alternatives might be achieved by choosing $\mathcal{C}$ in a manner that reflects restrictions implied by cointegration. To be specific, consider the univariate case; that is, suppose $p = 1$. In this case, the (maximal eigenvalue and trace) statistic $LR_T(\mathbb{R})$ corresponds to a squared Dickey-Fuller-type t-statistic (i.e., an F-statistic), while the more conventional, and more powerful, one-sided Dickey-Fuller t-test can be interpreted as being based on the statistic $LR_T(\mathbb{R}_-)$, where $\mathbb{R}_- = (-\infty, 0]$ is the non-positive half-line. In other words, incorporation of the natural restriction $C \leq 0$, or $\Pi \leq 0$, is well known to be advantageous from the point of view of power in the univariate case. On the other hand, we are not aware of any multivariate unit root tests incorporating such "sign" restrictions, so it seems worthwhile to develop (possibly) multivariate tests which incorporate "sign" restrictions and explore whether power gains can be achieved by employing these tests. Doing so is one of the purposes of this paper.

---

[3] The statistic is defined here as the log-likelihood ratio, without the usual multiplication factor 2.

[4] The maximum eigenvalue and trace statistics have been derived by Johansen (1995) for the model with unknown error covariance matrix, but they would reduce to statistics mentioned here if the covariance matrix is treated as known. Under the assumptions of Theorem 1, the maximum eigenvalue and trace statistics of Johansen (1995) are asymptotically equivalent to their "known variance" counterparts $LR_T(\mathcal{M}_p(1))$ and $LR_T(\mathcal{M}_p(p))$.

To describe our proposed "sign" restriction, let $\mathcal{M}_p^-(r)$ denote the subset of $\mathcal{M}_p(r)$ whose members have eigenvalues with non-positive real parts. When $p=1$, $\mathcal{M}_p^-(p)$ is simply the non-positive half-line and the test based on $LR_T\left(\mathcal{M}_p^-(p)\right)$ therefore reduces to the one-sided Dickey-Fuller t-test. For any p, imposing the restriction $C \in \mathcal{M}_p^-(p)$ is equivalent to imposing a nonpositivity restriction on the real parts of the eigenvalues of Π. Doing so also when $p>1$ can be motivated as follows. On the one hand, if the characteristic polynomial $A(z)=I_p-(I_p+\Pi)z$ satisfies the well known condition that $|z|>1$ or $z=1$ whenever $|A(z)|=0$ (e.g., Johansen (1995, Assumption 1)), then the non-zero eigenvalues of Π have non-positive real part. On the other hand, and partially conversely, the set of matrices Π satisfying Johansen (1995, Assumption 1) is approximated (in the sense of Chernoff (1954, Definition 2)) by the closed cone $\mathcal{M}_p^-(p)$ consisting of those elements of $\mathbb{R}^{p\times p}$ whose eigenvalues have non-positive real parts.[5] The latter approximation property implies that under (2), imposing Johansen (1995, Assumption 1) is (asymptotically) equivalent to imposing $C \in \mathcal{M}_p^-(p)$. In particular, we can obtain "sign-restricted" versions of the maximum eigenvalue and trace statistics by setting $\mathcal{C}$ equal to $\mathcal{M}_p^-(1)$ and $\mathcal{C}=\mathcal{M}_p^-(p)$, respectively.

The following result characterizes the large sample properties of $LR_T(\mathcal{C})$ under the assumption that $\mathcal{C}$ is a closed cone. As demonstrated by the examples just given, the assumption that $\mathcal{C}$ is a (closed) cone is without loss of relevance in the sense that the cases of main interest satisfy this restriction. Moreover, the assumption that $\mathcal{C}$ is a cone seems natural insofar as it ensures that the implied maintained hypothesis $\Pi \in \Pi_T(\mathcal{C})$ on Π is T-invariant in the sense that $\Pi_T(\mathcal{C})$ does not depend on T.[6]

Theorem 1. *Suppose $\{y_t\}$ is generated by (1) and (2), with C held fixed as $T\to\infty$. If $\mathcal{C}\subseteq\mathbb{R}^{p\times p}$ is a closed cone, then $LR_T(\mathcal{C}) \to_d \max_{\bar{C}\in\mathcal{C}} \Lambda_{p,C}\left(\bar{C}\right)$, where*

$$\Lambda_{p,C}\left(\bar{C}\right)=\operatorname{tr}\left[\bar{C}\int_0^1 W_C(u)\,dW_C(u)'-\frac{1}{2}\bar{C}'\bar{C}\int_0^1 W_C(u)\,W_C(u)'\,du\right],$$

$W_C(u)=\int_0^u \exp(C(u-s))\,dW(s)$, and $W(\cdot)$ is a p-dimensional Wiener process.

2.2. Deterministics. As an initial generalization of the model (1), suppose

$$y_t=\mu' d_t+v_t, \qquad \Delta v_t=\Pi v_{t-1}+\varepsilon_t, \tag{4}$$

where $d_t=1$ or $d_t=(1,t)'$, μ is an unknown parameter (of conformable dimension), $v_0=0$, and $\varepsilon_t \sim i.i.d.\ \mathcal{N}(0,I_p)$. This model differs from (1) only by accommodating deterministics. Under (2), the model gives rise to a log likelihood function that can be expressed in terms C and μ as

$$L_T^d(C,\mu)=-\frac{1}{2}\sum_{t=1}^{T}\left\|Y_{Tt}(C)-D_{Tt}(C)\operatorname{vec}(\mu)\right\|^2,$$

[5] In other words, $\mathcal{M}_p^-(p)$ is the tangent cone (e.g., Drton (2009, Definition 2.3)) at the point $\Pi=0$ of the set of matrices Π satisfying Johansen (1995, Assumption 1).

[6] Proceeding as in the proof of Theorem 1 it can be shown that if $\mathcal{C}$ is a set whose closure, $cl(\mathcal{C})$, contains zero, then $LR_T(\mathcal{C})$ equals $\max_{\bar{C}\in cl(\mathcal{C})} L_T\left(\bar{C}\right)-L_T(0)$ and has an asymptotic representation of the form $\max_{\bar{C}\in cl(\mathcal{C})}\Lambda_{p,C}\left(\bar{C}\right)$. Therefore, the properties of $LR_T(\mathcal{C})$ depend on $\mathcal{C}$ only through its closure and no generality is lost by assuming that $\mathcal{C}$ is closed.

where, setting $y_0 = 0$ and $d_0 = 0$, $Y_{Tt}(C) = \Delta y_t - \Pi_T(C)\, y_{t-1}$ and $D_{Tt}(C) = I_p \otimes \Delta d_t' - \Pi_T(C) \otimes d_{t-1}'$.[7]

In the presence of the nuisance parameter μ, a likelihood ratio statistic for testing the null hypothesis $C = 0$ against the alternative $C \in \mathcal{C} \backslash \{0\}$ is given by

$$LR_T^d(\mathcal{C}) = \sup_{\bar{C} \in \mathcal{C}, \mu} L_T^d\left(\bar{C}, \mu\right) - \max_\mu L_T^d(0, \mu).$$

This statistic can be expressed in semi-closed form as

$$LR_T^d(\mathcal{C}) = \sup_{\bar{C} \in \mathcal{C}} \mathcal{L}_T^d\left(\bar{C}\right) - \mathcal{L}_T^d(0),$$

where the profile log likelihood $\mathcal{L}_T^d(C) = \max_\mu L_T^d(C, \mu)$ is given by

$$\mathcal{L}_T^d(C) = -\frac{1}{2} Q_{YY,T}(C) + \frac{1}{2} Q_{YD,T}(C)\, Q_{DD,T}(C)^{-1}\, Q_{DY,T}(C),$$

with

$$\begin{aligned} Q_{YY,T}(C) &= \sum_{t=1}^T Y_{Tt}(C)'\, Y_{Tt}(C), \\ Q_{YD,T}(C) &= \sum_{t=1}^T Y_{Tt}(C)'\, D_{Tt}(C) = Q_{DY,T}(C)', \\ Q_{DD,T}(C) &= \sum_{t=1}^T D_{Tt}(C)'\, D_{Tt}(C). \end{aligned}$$

Unlike the zero-mean case considered in Section 2.1, the statistic $LR_T^d(\mathcal{C})$ does not admit a closed form expression even when $\mathcal{C}$ is of the form $\mathcal{M}_p(r)$. Because this computational nuisance can be avoided by dropping the "$t = 1$" contribution from the sum defining $L_T^d(C, \mu)$, it is perhaps tempting to do so. On the other hand, it is by now well understood that likelihood ratio ratio tests constructed from the resulting conditional (on y_1) likelihood function have unnecessarily low power in models with deterministics (e.g., Xiao and Phillips (1999), Hubrich, Lütkepohl, and Saikkonen (2001), and the references therein). The formulation adopted here, which retains the "$t = 1$" contribution in the sum defining $L_T^d(C, \mu)$, is inspired by Jansson and Nielsen (2011), where an analogous formulation was shown to provide an "automatic" way of avoiding the aforementioned power loss in the scalar case (i.e., when $p = 1$).

In the scalar case studied by Jansson and Nielsen (2011), the local-to-unity asymptotic distribution of the likelihood ratio statistic accommodating deterministics was found to be identical that of its no deterministic counterparts in the constant mean case (i.e., when $d_t = 1$), but not in the linear trend case (i.e., when $d_t = (1, t)'$). The following multivariate result shares these qualitative features.

[7] The observed data are $(y_1, \ldots, y_T)$; setting $y_0 = 0$ and $d_0 = 0$ is a notational convention that allows the first likelihood contribution $-\frac{1}{2}\left\|y_1 - \mu' d_1\right\|^2$ to be expressed in the same way as the other terms in the summation.

Theorem 2. *Suppose* $\{y_t\}$ *is generated by* (4) *and* (2), *with* C *held fixed as* $T \to \infty$. *Moreover, suppose* $\mathcal{C} \subseteq \mathbb{R}^{p\times p}$ *is a closed cone.*

(a) If $d_t = 1$, *then* $LR_T^d(\mathcal{C}) \to_d \max_{\bar{C}\in\mathcal{C}} \Lambda_{p,C}(\bar{C})$, *where* $\Lambda_{p,C}$ *is defined in Theorem 1.*

(b) If $d_t = (1,t)'$, *then* $LR_T^d(\mathcal{C}) \to_d \max_{\bar{C}\in\mathcal{C}} \Lambda_{p,C}^{\tau}(\bar{C})$, *where, with* $\bar{C}_s = \frac{1}{2}(\bar{C}+\bar{C}')$ *and* $\bar{C}_a = \frac{1}{2}(\bar{C}-\bar{C}')$ *denoting the symmetric and antisymmetric parts of* $\bar{C}$,

$$\Lambda_{p,C}^{\tau}(\bar{C}) = \Lambda_{p,C}(\bar{C}) + \frac{1}{2}\lambda_{p,C}(\bar{C})'\left(I_p - \bar{C}_s + \frac{1}{3}\bar{C}'\bar{C}\right)^{-1}\lambda_{p,C}(\bar{C}) - \frac{1}{2}\lambda_{p,C}(0)'\lambda_{p,C}(0),$$

$$\lambda_{p,C}(\bar{C}) = (I_p - \bar{C}_s)W_C(1) - \bar{C}_a\left(\int_0^1 W_C(u)\,du - \int_0^1 u\,dW_C(u)\right) + \bar{C}'\bar{C}\int_0^1 uW_C(u)\,du.$$

Theorem 2(a) implies in particular that in the constant mean case, the local asymptotic power of the test based on $LR_T^d(\mathcal{M}_p(p))$ coincides with that of the no-deterministics trace test. This property is shared by the (trace) test proposed by Saikkonen and Luukkonen (1997), which was found by Hubrich, Lütkepohl, and Saikkonen (2001) to be superior to its main rivals, notably the tests proposed by Johansen (1991). A further implication of Theorem 2(a) is that the relative merits of $LR_T^d(\mathcal{M}_p(p))$ and $LR_T^d(\mathcal{M}_p^-(p))$ are the same as those of their no deterministics counterparts analyzed in Section 2.1, so also in the constant mean case positive (albeit slight) power gains can be achieved by imposing "sign" restrictions. In Section 3 we analyze the asymptotic local power functions of our newly proposed tests and compare with those of the Johansen (1991) and Saikkonen and Luukkonen (1997) tests.

Our interpretation of the comprehensive simulation evidence reported in Hubrich, Lütkepohl, and Saikkonen (2001) is that in the linear trend case, the most powerful currently available tests are those of Lütkepohl and Saikkonen (2000) and Saikkonen and Lütkepohl (2000). Under the assumptions of Theorem 2(b), the so-called GLS (trace) statistics proposed in these papers all have asymptotic representations of the form

$$\operatorname{tr}\left[\left(\int_0^1 \tilde{W}_C(u)\,d\tilde{W}_C(u)'\right)'\left(\int_0^1 \tilde{W}_C(u)\tilde{W}_C(u)'\,du\right)^{-1}\left(\int_0^1 \tilde{W}_C(u)\,d\tilde{W}_C(u)'\right)'\right],$$

where $\tilde{W}_C(u) = W_C(u) - uW_C(1)$.

For the purposes of comparing this representation (as well as certain representations that have arisen in the univariate case) with that obtained in Theorem 2(b), it turns out to be convenient to define

$$\Lambda_{p,C}^{GLS}(\bar{C};\bar{C}^*) = \operatorname{tr}\left[\bar{C}\int_0^1 \tilde{W}_{C,\bar{C}^*}(u)\,d\tilde{W}_{C,\bar{C}^*}(u)' - \frac{1}{2}\bar{C}'\bar{C}\int_0^1 \tilde{W}_{C,\bar{C}^*}(u)\tilde{W}_{C,\bar{C}^*}(u)'\,du - \frac{1}{2}\tilde{W}_{C,\bar{C}^*}(1)\tilde{W}_{C,\bar{C}^*}(1)'\right],$$

where, letting $D_{\bar{C}^*}(u) = I_p - \bar{C}^*u$, the process

$$\tilde{W}_{C,\bar{C}^*}(u) = W_C(u) - u\left[\int_0^1 D_{\bar{C}^*}(s)'D_{\bar{C}^*}(s)ds\right]^{-1}\int_0^1 D_{\bar{C}^*}(s)'\left[dW_C(s) - \bar{C}^*W_C(s)ds\right],$$

can be interpreted as a GLS-detrended Ornstein-Uhlenbeck process (the multivariate version of the process $V_c(t, \bar{c})$ defined by Elliott, Rothenberg, and Stock (1996, Section 2.3)).

Using this notation, the asymptotic representation of one half times the GLS trace statistics of Lütkepohl and Saikkonen (2000) and Saikkonen and Lütkepohl (2000) can be written as $\mathcal{LR}_{p,C}^{GLS}\left(\mathcal{M}_p\left(p\right);0\right)$, where $\mathcal{LR}_{p,C}^{GLS}\left(\mathcal{C};\bar{C}_{GLS}\right)=\max_{\bar{C}\in\mathcal{C}}\Lambda_{p,C}^{GLS}\left(\bar{C};\bar{C}_{GLS}\right)$.[8] In the univariate case, a test with the same asymptotic properties was proposed by Schmidt and Lee (1991). Another class of (univariate) tests whose large sample properties can be characterized using representations of the same form are the DF-GLS statistics of Elliott, Rothenberg, and Stock (1996), which can be shown to correspond to $\mathcal{LR}_{1,C}^{GLS}\left(\mathbb{R}_-;\bar{C}_{ERS}\right)$, where $\bar{C}_{ERS}$ is a user-chosen constant set equal to -13.5 by Elliott, Rothenberg, and Stock (1996). Calculations outlined in the proof of Theorem 2(b) show that our test statistics admit asymptotic representations of the form $\max_{\bar{C}\in\mathcal{C}}\Lambda_{p,C}^{GLS}\left(\bar{C};\bar{C}\right)$. As a consequence, our test statistics cannot be interpreted as multivariate generalizations of the DF-GLS statistics of Elliott, Rothenberg, and Stock (1996).

2.3. Reduced Rank Hypotheses. Next, we consider the problem of testing more general reduced rank hypotheses on the matrix Π in the model (4). For the purposes of developing tests of the hypothesis that Π is of rank r_0 (for some $r_0 < p$), it turns out to be useful to consider the case where Π is parameterized as

$$\Pi=\Pi_T\left(C;r_0,\alpha,\alpha_\perp,\beta\right)=\alpha\beta'+T^{-1}\alpha_\perp C\alpha_\perp', \tag{5}$$

where $\alpha\in\mathbb{R}^{p\times r_0}$, $\alpha_\perp\in\mathbb{R}^{p\times q}$, and $\beta\in\mathbb{R}^{p\times r_0}$ with $(\alpha,\alpha_\perp)$ orthogonal, the eigenvalues of $I_{r_0}+\alpha'\beta$ are less than one in absolute value, and $C\in\mathbb{R}^{q\times q}$ is an unknown parameter of interest. Here and throughout $q=p-r_0$. The eigenvalue assumption implies that the matrix $(\beta,\alpha_\perp)$ is non-singular, so that the matrix Π is unrestricted by this reparametrization.

In (5), Π has rank r_0 if and only if $C=0$. Conversely, any $\Pi\in\mathbb{R}^{p\times p}$ of rank r_0 can be expressed as $\alpha\beta'$ for some (semi-orthogonal) $\alpha\in\mathbb{R}^{p\times r_0}$ and some $\beta\in\mathbb{R}^{p\times r_0}$ of full column rank. Moreover, it turns out that likelihood ratio statistics corresponding to hypotheses concerning C in (5) depend on $(\alpha,\alpha_\perp,\beta)$ in a sufficiently nice way that it is of relevance to proceed "as if" these parameters were known. For our purposes, a further attraction of the specification (5) is that restrictions on Π implied by cointegration are "sign" restrictions on C of the exact same form as those discussed earlier.

Assuming (counterfactually) that $(\alpha,\alpha_\perp,\beta)$ is known, a likelihood ratio statistic for test-

[8]It can be shown that if the assumptions of Theorem 2(b) hold, then

$$LR_T^{GLS}\left(\mathcal{C};\bar{C}_{GLS}\right)\to_d\mathcal{LR}_{p,C}^{GLS}\left(\mathcal{C};\bar{C}_{GLS}\right),$$

where letting $\hat{\mu}_T\left(\bar{C}_{GLS}\right)=\arg\max_\mu L_T^d\left(\bar{C}_{GLS},\mu\right)$,

$$LR_T^{GLS}\left(\mathcal{C};\bar{C}_{GLS}\right)=\sup_{\bar{C}\in\mathcal{C}}L_T^d\left(\bar{C},\hat{\mu}_T\left(\bar{C}_{GLS}\right)\right)-L_T^d\left(0,\hat{\mu}_T\left(\bar{C}_{GLS}\right)\right).$$

As a consequence, every limiting representation (indexed by $\mathcal{C}$ and $\bar{C}_{GLS}$) of the form $\mathcal{LR}_{p,C}^{GLS}\left(\mathcal{C};\bar{C}_{GLS}\right)$ is achievable. It is beyond the scope of this paper to attempt to isolate "optimal" choices of $\mathcal{C}$ and $\bar{C}_{GLS}$. Instead, our aim is to clarify the relationship between our tests and certain tests already in the literature.

ing the null hypothesis $C = 0$ against the alternative $C \in \mathcal{C} \backslash \{0\}$ is given by

$$LR_T^d(\mathcal{C}; r_0, \alpha_\perp) = \sup_{\bar{C} \in \mathcal{C}, \mu} L_T^d\left(\bar{C}, \mu; r_0, \alpha, \alpha_\perp, \beta\right) - \max_\mu L_T^d(0, \mu; r_0, \alpha, \alpha_\perp, \beta),$$

where

$$L_T^d(C, \mu; r_0, \alpha, \alpha_\perp, \beta) = -\frac{1}{2} \sum_{t=1}^{T} \| Y_{Tt}(C; r_0, \alpha, \alpha_\perp, \beta) - D_{Tt}(C; r_0, \alpha, \alpha_\perp, \beta) \operatorname{vec}(\mu) \|^2,$$

with $y_0 = 0$, $d_0 = 0$, and

$$Y_{Tt}(C; r_0, \alpha, \alpha_\perp, \beta) = \Delta y_t - \Pi_T(C; r_0, \alpha, \alpha_\perp, \beta) y_{t-1},$$
$$D_{Tt}(C; r_0, \alpha, \alpha_\perp, \beta) = I_p \otimes \Delta d_t' - \Pi_T(C; r_0, \alpha, \alpha_\perp, \beta) \otimes d_{t-1}'.$$

As the notation suggests, the likelihood ratio statistic depends on $(\alpha, \alpha_\perp, \beta)$ only through $\alpha_\perp$. Indeed, as shown in the proof of Theorem 3 the statistic $LR_T^d(\mathcal{C}; r_0, \alpha_\perp)$ is simply the statistic $LR_T^d(\mathcal{C})$ of the previous subsection applied to $\{\alpha_\perp' y_t\}$ rather than $\{y_t\}$. As a consequence, one would expect the large sample distributions of $LR_T^d(\mathcal{C}; r_0, \alpha_\perp)$ to be of the same form as those obtained in Theorem 2. That conjecture is confirmed by the following result, which furthermore gives a simple condition (on the estimator $\hat{\alpha}_{\perp,T}$) under which a "plug-in" statistic of form $LR_T^d(\mathcal{C}; r_0, \hat{\alpha}_{\perp,T})$ is asymptotically equivalent to $LR_T^d(\mathcal{C}; r_0, \alpha_\perp)$.

Theorem 3. *Suppose $\{y_t\}$ is generated by (4) and (5), with (($\alpha, \alpha_\perp, \beta$) and) C held fixed as $T \to \infty$. Moreover, suppose $\mathcal{C} \subseteq \mathbb{R}^{p \times p}$ is a closed cone and suppose $\hat{\alpha}_{\perp,T} \to_p \alpha_\perp$.*

(a) If $d_t = 1$, then $LR_T^d(\mathcal{C}; r_0, \hat{\alpha}_{\perp,T}) \to_d \max_{\bar{C} \in \mathcal{C}} \Lambda_{q,C}\left(\bar{C}\right)$, where $\Lambda_{q,C}$ is defined in Theorem 1.

(b) If $d_t = (1, t)'$, then $LR_T^d(\mathcal{C}; r_0, \hat{\alpha}_{\perp,T}) \to_d \max_{\bar{C} \in \mathcal{C}} \Lambda_{q,C}^\tau\left(\bar{C}\right)$, where $\Lambda_{q,C}^\tau$ is defined in Theorem 2.

The consistency requirement on $\hat{\alpha}_{\perp,T}$ is mild. Let $N_L : \mathcal{M}_p(r_0) \backslash \mathcal{M}_p(r_0 - 1) \to \mathbb{R}^{p \times q}$ be a function which returns a semi-orthogonal matrix spanning the left null space of its argument, i.e. satisfies $N_L(M)'N_L(M) = I_q$ and $N_L(M)'M = 0$ for every $M \in \mathcal{M}_p(r_0) \backslash \mathcal{M}_p(r_0 - 1)$, the set of $p \times p$ matrices of rank r_0. Under the other assumptions of Theorem 3, the matrix $\Pi_0 = \alpha\beta'$ is of rank r_0 and is consistently estimable, as is its left null space. The latter is spanned by the columns of $\alpha_\perp$, so an estimator $\hat{\alpha}_{\perp,T}$ of the form $\hat{\alpha}_{\perp,T} = N_L(\hat{\Pi}_{0,T})$ will be consistent provided $\hat{\Pi}_{0,T} \to_p \Pi_0$ and provided the function $N_L(\cdot)$ is chosen to be continuous in its argument.

2.4. Serial Correlation and Unknown Error Distribution. As a final generalization of the model (4) we assume that the stochastic part of the model is a VAR of order $k + 1$, which we write in error correction form. Thus, suppose

$$y_t = \mu' d_t + v_t, \qquad [\Gamma(L)(1 - L) - \Pi L] v_t = \varepsilon_t, \tag{6}$$

where $d_t = 1$ or $d_t = (1, t)'$, μ is an unknown parameter, $\Gamma(L) = I_p - \Gamma_1 L - \ldots - \Gamma_k L^k$ is a matrix lag polynomial with $|\Gamma(z)(1 - z) - \Pi z| \neq 0$ for $|z| < 1$, the initial condition is

$\max(\|v_0\|, \ldots, \|v_{-k}\|) = o_p(\sqrt{T})$, and the ε_t form a conditionally homoskedastic martingale difference sequence with unknown (full rank) covariance matrix Σ and $\sup_t E\|\varepsilon_t\|^{2+\delta} < \infty$ for some $\delta > 0$.

To develop tests of the hypothesis that Π is of rank r_0, it once again proves convenient to employ a very particular parameterization of Π. Specifically, it turns out to be useful to consider the case where Π is parameterized as

$$\Pi = \Pi_T(C; r_0, \theta) = \alpha\beta' + T^{-1}\Sigma\alpha_\perp C\alpha_\perp'\Gamma(1), \tag{7}$$

where $\alpha \in \mathbb{R}^{p\times r_0}$, $\alpha_\perp \in \mathbb{R}^{p\times q}$, and $\beta \in \mathbb{R}^{p\times r_0}$ with $\left(\Sigma^{-1/2}\alpha, \Sigma^{1/2}\alpha_\perp\right)$ orthogonal, and where $|\Gamma(z)(1-z) - \alpha\beta' z| = 0$ has exactly $q = p - r_0$ roots equal to one and all other roots outside the unit circle, $C \in \mathbb{R}^{q\times q}$ is an unknown parameter of interest, and $\theta = (\alpha, \alpha_\perp, \beta, \Sigma, \Gamma_1, \ldots, \Gamma_k)$ contains all nuisance parameters other than μ.

The Gaussian quasi-log likelihood function corresponding to the model with $u_0 = \ldots = u_{-k} = 0$ and with θ known can be expressed, up to a constant, as

$$L_T^d(C, \mu; r_0, \theta) = -\frac{1}{2}\sum_{t=1}^{T}\left\|\Sigma^{-1/2}\left[Y_{Tt}(C; r_0, \theta) - D_{Tt}(C; r_0, \theta)\operatorname{vec}(\mu)\right]\right\|^2,$$

where, setting $y_0 = \ldots = y_{-k} = 0$ and $d_0 = \ldots = d_{-k} = 0$,

$$Y_{Tt}(C; r_0, \theta) = \Gamma(L)\Delta y_t - \Pi_T(C; r_0, \theta) y_{t-1}$$

and

$$D_{Tt}(C; r_0, \theta) = \Gamma(L) \otimes \Delta d_t' - \Pi_T(C; r_0, \theta) \otimes d_{t-1}'.$$

Replacing θ by an estimator $\hat{\theta}_T$ we are led to consider quasi-likelihood ratio type statistics of the form

$$\widehat{LR}_T^d(\mathcal{C}; r_0) = \sup_{\bar{C}\in\mathcal{C},\mu} L_T^d\left(\bar{C}, \mu; r_0, \hat{\theta}_T\right) - \max_\mu L_T^d\left(0, \mu; r_0, \hat{\theta}_T\right).$$

Theorem 4. *Suppose $\{y_t\}$ is generated by (6) and (7), with (θ and) C held fixed as $T \to \infty$. Moreover, suppose $\mathcal{C} \subseteq \mathbb{R}^{p\times p}$ is a closed cone and suppose $\hat{\theta}_T \to_p \theta$.*

(a) If $d_t = 1$, then $\widehat{LR}_T^d(\mathcal{C}; r_0) \to_d \max_{\bar{C}\in\mathcal{C}} \Lambda_{q,C}\left(\bar{C}\right)$, where $\Lambda_{q,C}$ is defined in Theorem 1.

(b) If $d_t = (1, t)'$, then $\widehat{LR}_T^d(\mathcal{C}; r_0) \to_d \max_{\bar{C}\in\mathcal{C}} \Lambda_{q,C}^\tau\left(\bar{C}\right)$, where $\Lambda_{q,C}^\tau$ is defined in Theorem 2.

An obvious choice for the consistent estimator $\hat{\theta}_T$ would be the maximizer of the conditional quasi-likelihood, obtained as the density of $(y_{k+2}, \ldots, y_T)$ conditional on starting values $(y_1, \ldots, y_{k+1})$. The corresponding model under the null hypothesis may be expressed as

$$\Delta y_t = \alpha\beta' y_{t-1} + \Gamma_1\Delta y_{t-1} + \ldots + \Gamma_k\Delta y_{t-k} + \Phi d_t + \varepsilon_t, \qquad t = k+2, \ldots, T,$$

where $\Phi d_t = \Gamma(L)\mu'\Delta d_t - \alpha\beta'\mu' d_{t-1}$. As analyzed in Johansen (1995), conditional likelihood estimation of the parameters of the model in case (a) leads to reduced rank regression applied to the system

$$\Delta y_t = \alpha(\beta', \rho_0)(y'_{t-1}, 1)' + \Gamma_1 \Delta y_{t-1} + \ldots + \Gamma_k \Delta y_{t-k} + \varepsilon_t,$$

where $\rho_0 = -\beta'\mu'$; in case (b), reduced rank regression is applied to

$$\Delta y_t = \alpha(\beta', \rho_1)(y'_{t-1}, t)' + \Gamma_1 \Delta y_{t-1} + \ldots + \Gamma_k \Delta y_{t-k} + \Phi_1 + \varepsilon_t,$$

where $\rho_1 = -\beta'\mu'(0,1)'$ and Φ_1 is unrestricted. Johansen (1995) shows that the resulting estimator of θ is consistent under the null hypothesis, and this result can be extended to local alternatives of the type (7).

3. Critical Values and Local Power

To enable application of the newly proposed tests in practice, and to assess the magnitude of the power gains achievable by using the full likelihood and imposing the "sign" restriction discussed above, we used the results in Theorems 1 and 2 to compute asymptotic critical values and local power functions of the tests for $\mathcal{C} = \mathcal{M}_q(q)$ and $\mathcal{C} = \mathcal{M}_q^-(q)$.

The results in this section are based on simulations conducted in Ox, see Doornik (2007). The "sign" restriction was imposed using the `MaxSQP` sequential quadratic programming optimization routine, while the results without the "sign" restriction were obtained using the `MaxBFGS` routine. Replications where the `maxSQP` routine did not converge have not been discarded, in order to avoid the possibility that the power of the "sign-restricted" tests might be biased upward due to selectivity of convergent replications.

Next we study the power of the tests for the univariate ($q = 1$) and bivariate ($q = 2$) cases. In the univariate case, the local power is simply plotted against $\ell = -c$, where ℓ ranges from 0 to 25 in the case of a constant mean, and from 0 to 50 in the case of a linear trend. In the bivariate case, we consider only cases with $\text{rank}(C) = 1$, and adopt the following variation of the parametrization proposed by Hubrich, Lütkepohl, and Saikkonen (2001), see also Johansen (1995, Chapter 14),

$$C = \ell \begin{bmatrix} -\sqrt{1-\rho^2} & 0 \\ \rho & 0 \end{bmatrix}, \qquad \ell \geq 0, \qquad \rho \in [0,1].$$

Here $\ell = \|C\|$ and ρ determines angle between a and $b_\perp$, where $C = ab'$. The parametrization has been chosen such that local power increases monotonically in both ℓ and ρ. Note that the value $\rho = 1$ corresponds to the process

$$W_C(u) = W(u) + \begin{bmatrix} 0 & 0 \\ \ell & 0 \end{bmatrix} \int_0^u W(s)ds,$$

which is an $I(2)$ process in continuous time. Because the test is proposed to detect stationary linear combinations in y_t, local power against alternatives with $\rho = 1$ is not our main interest, but these cases are included in the results below. In particular, we consider $\rho \in \{0, 0.5, 0.75, 1\}$ and $\ell \in [0, 50]$.

Table 1: Simulated quantiles of the distributions of $\max_{\bar{C}\in\mathcal{C}} \Lambda_{q,0}(\bar{C})$ and $\max_{\bar{C}\in\mathcal{C}} \Lambda_{q,0}^{\tau}(\bar{C})$

	$\mathcal{C} = \mathcal{M}_q(q)$				$\mathcal{C} = \mathcal{M}_q^-(q)$				
q	90%	95%	99%	99.9%	90%	95%	99%	99.9%	NC
Panel A: $\max_{\bar{C}\in\mathcal{C}} \Lambda_{q,0}(\bar{C})$									
1	1.477	2.054	3.486	5.513	1.294	1.861	3.271	5.262	0.0%
2	5.228	6.135	8.104	10.55	5.032	5.925	7.825	10.31	0.6%
3	10.86	12.11	14.73	17.99	10.67	11.90	14.50	17.85	1.6%
4	18.45	20.01	23.16	27.28	18.27	19.82	22.94	27.17	2.9%
5	27.99	29.88	33.73	38.18	27.81	29.68	33.51	37.94	3.9%
6	39.49	41.66	45.97	51.34	39.32	41.49	45.81	51.12	4.4%
Panel B: $\max_{\bar{C}\in\mathcal{C}} \Lambda_{q,0}^{\tau}(\bar{C})$									
1	3.203	3.974	5.665	7.999	3.203	3.974	5.665	7.999	0.0%
2	7.809	8.861	11.05	13.74	7.802	8.848	11.03	13.74	0.1%
3	14.33	15.68	18.54	22.27	14.31	15.65	18.49	22.20	0.2%
4	22.71	24.35	27.70	31.67	22.67	24.32	27.65	31.54	0.4%
5	33.05	35.01	38.82	43.50	33.01	34.96	38.79	43.44	0.7%
6	45.24	47.47	51.87	57.60	45.19	47.40	51.82	57.51	1.0%

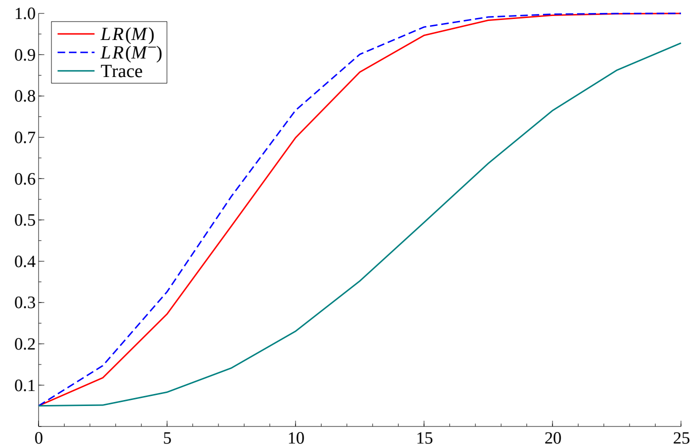
Note: The table presents simulated quantiles, where Wiener processes are approximated by 1000 discrete steps with standard Gaussian innovations. The column labeled NC contains the percentage of the replications where the numerical optimization procedure did not converge when $\mathcal{C} = \mathcal{M}_q^-(q)$. No replications had convergence problems for the case with $\mathcal{C} = \mathcal{M}_q(q)$. All entries are based on $100,000$ Monte Carlo replications.

For the case of a constant mean, we compare the two likelihood ratio tests, indicated by $LR(M)$ and $LR(M^-)$, with the standard Johansen trace test for an unknown mean (i.e., with a restricted constant), indicated by Trace. We use the power function of the trace test as the (only) benchmark because the trace test seems to be the most popular test in applications and because the local power of the trace test was found by Lütkepohl, Saikkonen, and Trenkler (2001) to be very similar to that of its closest rival, the maximum eigenvalue test (i.e., the test corresponding to $\mathcal{C} = \mathcal{M}_p(1)$). Note that the power of the likelihood ratio test with $\mathcal{C} = \mathcal{M}_q(q)$ is in fact identical to the power of Johansen's trace test for a known mean (equal to zero).

In Figures 1 and 2 we display the asymptotic local power functions for the constant mean case. It is clear that imposing the sign restriction does lead to a local power gain in the univariate case, but appears to make very little difference with $q = 2$. More importantly, both versions of the LR test have much higher asymptotic local power than the trace test, both in the univariate and in the bivariate case, although the power difference decreases as ρ approaches the $I(2)$ boundary $\rho = 1$.

Figures 3 and 4 display the asymptotic local power functions for the linear trend case. In this case we have also included the asymptotic local power function of the tests proposed by Lütkepohl and Saikkonen (2000) and Saikkonen and Lütkepohl (2000), indicated by SL. Now the gains from imposing the "sign" restriction vanish entirely. The power difference between the likelihood ratio tests and the trace test are comparable to the constant mean case. The likelihood ratio tests also dominate the SL test in local power, especially for local

Figure 1: Asymptotic local power functions of cointegration tests, constant mean, $q = 1$.



Note: The asymptotic local power functions (5% level) against ℓ are generated using 100,000 Monte Carlo replications, where Wiener processes are approximated by 1000 discrete steps with standard Gaussian innovations.

alternatives relatively far from the null hypothesis (i.e., for large ℓ), where the local power of SL appears to approach one only very slowly.

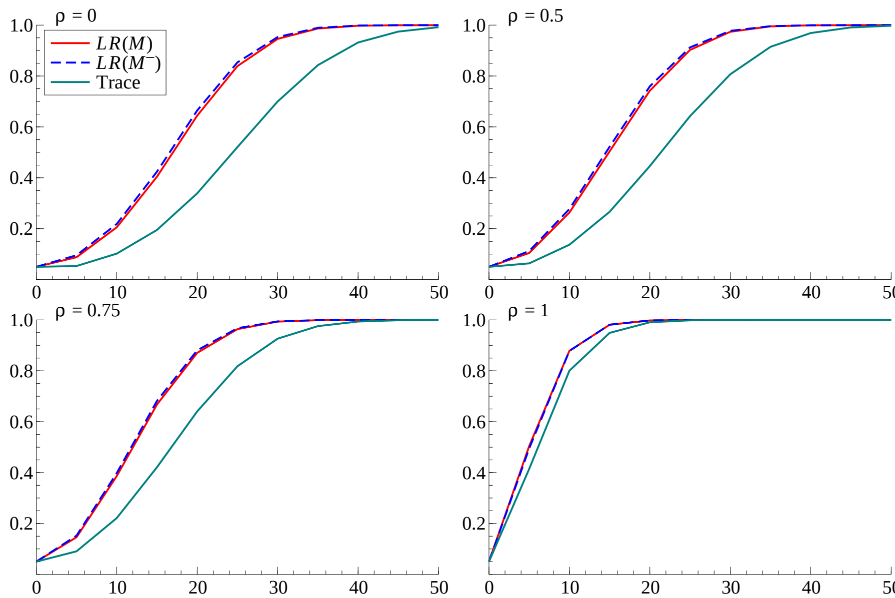
4. Discussion and Conclusions

In this paper, we have suggested improved tests for cointegration rank in the vector autoregressive (VAR) model and developed relevant asymptotic distribution theory and local power results. The tests are (quasi-)likelihood ratio tests based on a Gaussian likelihood, but of course the asymptotic results apply more generally. The power gains relative to existing tests are due to two factors. First, instead of basing our tests on the conditional (with respect to the initial observations) likelihood, we follow the recent unit root literature and base our tests on the full likelihood as in, e.g., Elliott, Rothenberg, and Stock (1996). Secondly, our tests incorporate a "sign" restriction which generalizes the one-sided unit root test. We show that the asymptotic local power of the proposed tests dominates that of existing cointegration rank tests.

Computationally, the new tests require numerical optimization; for the tests that do not impose the sign restriction, this numerical optimization is fast and does not have any convergence problems. In fact, it is possible to devise a convenient switching algorithm for optimizing the likelihood function in such cases.

To deal with the nuisance parameters, we use a plug-in approach for those parameters that are irrelevant in the asymptotic distributions (and asymptotic local power). On the other hand, the likelihood is maximized with respect to those parameters that are important for the asymptotic distributions and power. Existing tests based on GLS detrending, e.g. Xiao and Phillips (1999), do the opposite and use a plug-in approach for the asymptotically relevant parameters and maximize the likelihood with respect to the asymptotically irrelevant parameters.

Figure 2: Asymptotic local power functions of cointegration tests, constant mean, $q = 2$.



Note: The asymptotic local power functions (5% level) against ℓ are generated using 100,000 Monte Carlo replications, where Wiener processes are approximated by 1000 discrete steps with standard Gaussian innovations.

By proposing cointegration rank tests with power superior to those of existing tests, this paper has demonstrated by example that these existing tests are suboptimal in terms of local asymptotic power. In the univariate case, our tests reduce to those of Jansson and Nielsen (2011) and were shown there to be "nearly efficient" (in the sense of Elliott, Rothenberg, and Stock (1996)). Generalizing the optimality theory of Elliott, Rothenberg, and Stock (1996) to multivariate settings is beyond the scope of this paper, however, so it remains an open question whether the tests developed herein themselves enjoy any optimality properties.

5. PROOFS

5.1. Proof of Theorem 1. We use a method of proof similar to that of Jansson and Nielsen (2011). Expanding $L_T(C)$ around $C = 0$, we have

$$L_T(C) - L_T(0) = F(C, S_T, H_T) = \mathrm{tr}\left(CS_T - \frac{1}{2}C'CH_T\right),$$

where

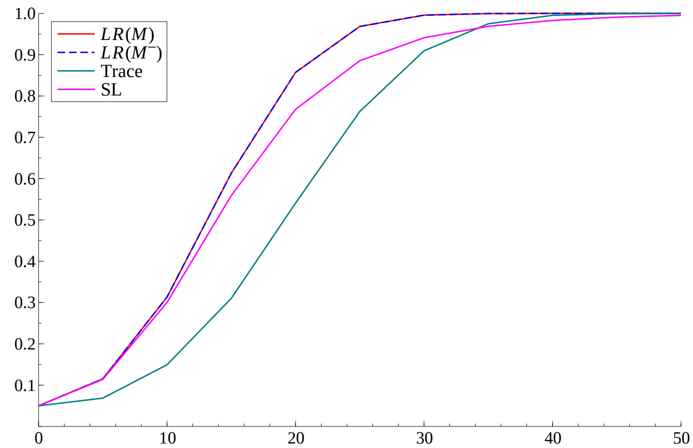
$$(S_T, H_T) = \left(\frac{1}{T}\sum_{t=2}^{T} y_{t-1}\Delta y_t', \frac{1}{T^2}\sum_{t=2}^{T} y_{t-1}y_{t-1}'\right).$$

Therefore, $LR_T(\mathcal{C})$ can be represented as $LR_T(\mathcal{C}) = \max_{\bar{C}\in\mathcal{C}} F\left(\bar{C}, S_T, H_T\right)$.

Under the assumptions of Theorem 1 it follows from Phillips (1988) that

$$(S_T, H_T) \to_d (\mathcal{S}_C, \mathcal{H}_C) = \left(\int_0^1 W_C(u)\, dW_C(u)', \int_0^1 W_C(u)\, W_C(u)'\, du\right),$$

Figure 3: Asymptotic local power functions of cointegration tests, linear trend, $q = 1$.



Note: The asymptotic local power functions (5% level) against ℓ are generated using 100,000 Monte Carlo replications, where Wiener processes are approximated by 1000 discrete steps with standard Gaussian innovations.

implying in particular that $F\left(\bar{C}, S_T, H_T\right) \to_d F\left(\bar{C}, \mathcal{S}_C, \mathcal{H}_C\right) = \Lambda_C\left(\bar{C}\right)$ for every $\bar{C} \in \mathcal{C}$. Using this convergence result and the fact that the set $\mathbb{X}$ of pairs (S, H) of $p \times p$ matrices for which H is symmetric and positive definite satisfies $\Pr\left[(\mathcal{S}_C, \mathcal{H}_C) \in \mathbb{X}\right] = 1$, Theorem 1 will follow from the continuous mapping theorem if it can be shown that the functional $\max_{\bar{C} \in \mathcal{C}} F\left(\bar{C}, \cdot\right)$ is continuous on $\mathbb{X}$.

Using simple bounds (and the fact that H_0 is positive definite whenever $(S_0, H_0) \in \mathbb{X}$), it can be shown that any $(S_0, H_0) \in \mathbb{X}$ admits a finite constant K and an open set $\mathbb{X}_0 \subseteq \mathbb{X}$ containing (S_0, H_0) such that

$$\sup_{(S,H) \in \mathbb{X}_0, \left\|\bar{C}\right\| > K} F\left(\bar{C}, S, H\right) \leq 0.$$

Specifically, the asserted property of $F(\cdot)$ follows from the fact that

$$\lim_{K \to \infty} \sup_{\left\|\bar{C}\right\| > K} \left\|\bar{C}\right\|^{-2} \left|F\left(\bar{C}, S, H\right) - F^*\left(\bar{C}, H\right)\right| \to 0,$$

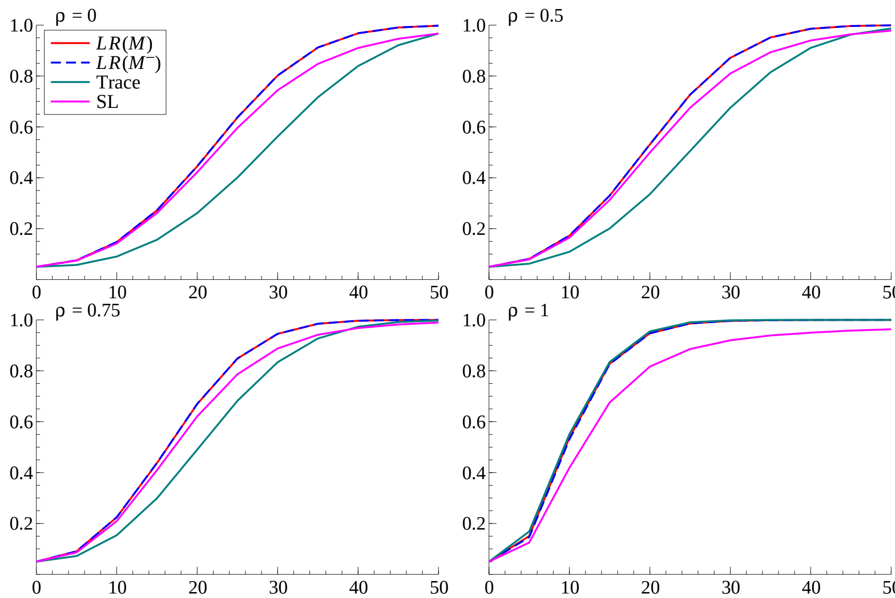
where $F^*(C, H) = -\frac{1}{2}\operatorname{tr}\left(C'CH\right)$, the convergence is uniform (in (S, H)) on compacta, and $\overline{\lim}_{K \to \infty} \sup_{\left\|\bar{C}\right\| > K} \left\|\bar{C}\right\|^{-2} F^*\left(\bar{C}, \cdot\right)$ is negative and continuous on the set of positive definite matrices.

Therefore, because $F(0, S, H) = 0$ and because $\mathcal{C}$ is closed and contains the zero matrix, it holds for any $(S, H) \in \mathbb{X}_0$ that

$$\max_{\bar{C} \in \mathcal{C}} F\left(\bar{C}, S, H\right) = \max_{\bar{C} \in \mathcal{C}, \left\|\bar{C}\right\| \leq K} F\left(\bar{C}, S, H\right).$$

Because $\left\{\bar{C} \in \mathcal{C} : \left\|\bar{C}\right\| \leq K\right\}$ is compact, the theorem of the maximum (e.g., Stokey and Lucas (1989, Theorem 3.6)) can be used to show that $\max_{\bar{C} \in \mathcal{C}} F\left(\bar{C}, \cdot\right)$ is continuous at (S_0, H_0).

Figure 4: Asymptotic local power functions of cointegration tests, linear trend, $q = 2$.



Note: The asymptotic local power functions (5% level) against ℓ are generated using 100,000 Monte Carlo replications, where Wiener processes are approximated by 1000 discrete steps with standard Gaussian innovations.

5.2. Proof of Theorem 2. Because the profile log likelihood function $\mathcal{L}_T^d(\cdot)$ is invariant under transformations of the form $y_t \to y_t + m'd_t$ we can assume without loss of generality that $\mu = 0$, so that $v_t = y_t$ in the proof. Moreover, the proofs of parts (a) and (b) are very similar, so to conserve space we omit the details for part (a).

Proceeding as in the proof of Theorem 1, it can be shown that $\mathcal{L}_T^d(\bar{C}) - \mathcal{L}_T^d(0)$ can be written as $F^d(\bar{C}, X_T^d)$ for some X_T^d satisfying a convergence property of the form $X_T^d \to_d \mathcal{X}_C^d$ and some function $F^d(\cdot)$ enjoying the property that the functional $\max_{\bar{C}\in\mathcal{C}} F^d(\bar{C}, \cdot)$ is continuous on a set $\mathbb{X}^d$ satisfying $\Pr\left[\mathcal{X}_C^d \in \mathbb{X}^d\right] = 1$. By implication, $\max_{\bar{C}\in\mathcal{C}} \mathcal{L}_T^d(\bar{C}) - \mathcal{L}_T^d(0) \to_d \max_{\bar{C}\in\mathcal{C}} F^d(\bar{C}, \mathcal{X}_C)$, so it suffices to show that $\Lambda_{p,C}^\tau(\bar{C})$ is the pointwise (in $\bar{C}$) weak limit of $\mathcal{L}_T^d(\bar{C}) - \mathcal{L}_T^d(0)$.

To do so, note that

$$\Lambda_{p,C}^\tau(\bar{C}) = \Lambda_{p,C}(\bar{C}) + \frac{1}{2} Q_{DY,T}(\bar{C})' Q_{DD,T}(\bar{C})^{-1} Q_{DY,T}(\bar{C}) - \frac{1}{2} Q_{DY,T}(0)' Q_{DD,T}(0)^{-1} Q_{DY,T}(0)$$

and let $d_0 = 0$ and $y_0 = 0$ and define $\Psi_T = I_p \otimes \mathrm{diag}\left(1, 1/\sqrt{T}\right)$ and $\tilde{d}_{Tt} = \mathrm{diag}\left(1, 1/\sqrt{T}\right) d_t$.

For any $\bar{C} \in \mathcal{C}$, we have

$$\begin{aligned}
\Psi_T Q_{DD,T}\left(\bar{C}\right)\Psi_T &= I_p \otimes \left(\sum_{t=1}^{T} \Delta\tilde{d}_{Tt}\Delta\tilde{d}_{Tt}'\right) + \left(\bar{C}'\bar{C}\right) \otimes \left(\frac{1}{T^2}\sum_{t=1}^{T} \tilde{d}_{T,t-1}\tilde{d}_{T,t-1}'\right) \\
&\quad - \bar{C}' \otimes \left(\frac{1}{T}\sum_{t=1}^{T} \tilde{d}_{T,t-1}\Delta\tilde{d}_{Tt}'\right) - \bar{C} \otimes \left(\frac{1}{T}\sum_{t=1}^{T} \Delta\tilde{d}_{Tt}\tilde{d}_{T,t-1}'\right) \\
&\to I_p \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \bar{C}_s \otimes \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} + \left(\bar{C}'\bar{C}\right) \otimes \begin{pmatrix} 0 & 0 \\ 0 & 1/3 \end{pmatrix} \\
&= I_p \otimes \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \left(I_p - \bar{C}_s + \frac{1}{3}\bar{C}'\bar{C}\right) \otimes \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \\
&= \left[I_p \otimes \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \left(I_p - \bar{C}_s + \frac{1}{3}\bar{C}'\bar{C}\right)^{-1} \otimes \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\right]^{-1},
\end{aligned}$$

where the last equality can be verified directly by using the so-called mixed-product property of the Kronecker product.

Next, using Phillips (1988) and the identity $\int_0^1 W_C(u)\,du = W_C(1) - \int_0^1 u dW_C(u)$,

$$\begin{aligned}
\Psi_T Q_{DY,T}\left(\bar{C}\right) &= \mathrm{vec}\left(\sum_{t=1}^{T} \Delta\tilde{d}_{Tt}\Delta v_t'\right) + \mathrm{vec}\left[\left(\frac{1}{T^2}\sum_{t=1}^{T} \tilde{d}_{T,t-1}v_{t-1}'\right)\bar{C}'\bar{C}\right] \\
&\quad - \mathrm{vec}\left[\left(\frac{1}{T}\sum_{t=1}^{T} \Delta\tilde{d}_{Tt}v_{t-1}'\right)\bar{C}'\right] - \mathrm{vec}\left[\left(\frac{1}{T}\sum_{t=1}^{T} \tilde{d}_{T,t-1}\Delta v_t'\right)\bar{C}\right] \\
&\to_d \mathrm{vec}\begin{pmatrix} \mathcal{Y}' \\ W_C(1)' \end{pmatrix} + \mathrm{vec}\left[\begin{pmatrix} 0 \\ \int_0^1 uW_C(u)'\,du \end{pmatrix}\bar{C}'\bar{C}\right] \\
&\quad - \mathrm{vec}\left[\begin{pmatrix} 0 \\ W_C(1)' - \int_0^1 udW_C(u)' \end{pmatrix}\bar{C}'\right] - \mathrm{vec}\left[\begin{pmatrix} 0 \\ \int_0^1 udW_C(u)' \end{pmatrix}\bar{C}\right] \\
&= \mathrm{vec}\begin{pmatrix} \mathcal{Y}' \\ \lambda_C\left(\bar{C}\right)' \end{pmatrix},
\end{aligned}$$

where $\mathcal{Y}$ is a random variable independent of $W_C(\cdot)$. Combining these results with that obtained in the proof of Theorem 1, the desired conclusion follows.

The definition $D_{\bar{C}}(u) = I_p - \bar{C}u$ immediately implies

$$\int_0^1 D_{\bar{C}}(u)'D_{\bar{C}}(u)du = I_p - \bar{C}_s + \frac{1}{3}\bar{C}'\bar{C}.$$

Next, using $\bar{C}_s = \bar{C} - \bar{C}_a$ and the identity $\int_0^1 W_C(u)\,du = W_C(1) - \int_0^1 udW_C(u)$, straightforward algebra shows that $\lambda_C(\bar{C})$ may be expressed as

$$\lambda_{p,C}(\bar{C}) = \int_0^1 D_{\bar{C}}(u)'\left[dW_C(u) - \bar{C}W_C(u)du\right].$$

This leads to

$$\Lambda^{\tau}_{p,C}(\bar{C}) - \Lambda_{p,C}(\bar{C}) = \frac{1}{2}\lambda_{p,C}(\bar{C})'\left(\int_0^1 D_{\bar{C}}(u)'D_{\bar{C}}(u)du\right)^{-1}\lambda_{p,C}(\bar{C}) - \frac{1}{2}W_C(1)'W_C(1)$$
$$= \frac{1}{2}b_C(\bar{C})'\left(\int_0^1 D_{\bar{C}}(u)'D_{\bar{C}}(u)du\right)b_C(\bar{C}) - \frac{1}{2}W_C(1)'W_C(1),$$

where

$$b_C(\bar{C}) = \left(\int_0^1 D_{\bar{C}}(u)'D_{\bar{C}}(u)du\right)^{-1}\int_0^1 D_{\bar{C}}(u)'\left[dW_C(u) - \bar{C}W_C(u)du\right],$$

is the GLS estimated slope parameter in $\tilde{W}_{C,\bar{C}}(r) = W_C(u) - ub_C(\bar{C})$, i.e. the estimated coefficient from continuous-time GLS regression of $W_C(u)$ on u.

From this expression, it can be shown (after substantial rearrangement of terms) that $\Lambda^{\tau}_{p,C}(\bar{C}) = \Lambda^{GLS}_{p,C}\left(\bar{C};\bar{C}\right)$, implying in particular that $LR^d_T(\mathcal{C}) \to_d \Lambda^{GLS}_{p,C}\left(\bar{C};\bar{C}\right)$, as claimed in the main text.

5.3. Proof of Theorem 3. Because $(\alpha,\alpha_\perp)$ is orthogonal and replacing y_t by $y^*_t = \left(y^{*\prime}_{1,t}, y^{*\prime}_{2,t}\right)' = (y'_t\alpha, y'_t\alpha_\perp)'$ if necessary, we can assume without loss of generality that $(\alpha,\alpha_\perp) = I_p$. In that special case, the implied model for $y_{2t} = \alpha'_\perp y_t$ is of the form (4) with $\Pi = T^{-1}C \in \mathbb{R}^{q\times q}$ (as in (2)). Moreover, it follows from simple algebra that, for any $\bar{C}$,

$$\max_\mu L^d_T\left(\bar{C},\mu;r_0,\alpha,\alpha_\perp,\beta\right) - \max_\mu L^d_T\left(0,\mu;r_0,\alpha,\alpha_\perp,\beta\right) = \mathcal{L}^d_T\left(\bar{C};r_0\right) - \mathcal{L}^d_T\left(0;r_0\right),$$

where $\mathcal{L}^d_T(C;r_0)$ is the statistic $\mathcal{L}^d_T(C)$ of Section 2.2 computed using y_{2t} rather y_t; that is,

$$\mathcal{L}^d_T(C;r_0) = -\frac{1}{2}Q_{YY,T}(C;r_0) + \frac{1}{2}Q_{YD,T}(C;r_0)\,Q_{DD,T}(C;r_0)^{-1}\,Q_{DY,T}(C;r_0), \qquad (8)$$

where, setting $y_{2,0} = 0$ and $d_0 = 0$ and defining $Y_{Tt}(C;r_0) = \Delta y_{2t} - T^{-1}Cy_{2,t-1}$ and $D_{Tt}(C) = I_{p-r_0}\otimes\Delta d'_t - T^{-1}C\otimes d'_{t-1}$,

$$Q_{YY,T}(C;r_0) = \sum_{t=1}^T Y_{Tt}(C;r_0)'\,Y_{Tt}(C;r_0),$$
$$Q_{YD,T}(C;r_0) = \sum_{t=1}^T Y_{Tt}(C;r_0)'\,D_{Tt}(C;r_0) = Q_{DY,T}(C;r_0)',$$
$$Q_{DD,T}(C;r_0) = \sum_{t=1}^T D_{Tt}(C;r_0)'\,D_{Tt}(C;r_0).$$

Theorem 3 therefore follows from Theorem 2 in the special case where $\hat{\alpha}_{\perp,T} = \alpha_\perp$. Since the statistics of interest are smooth functionals of the process $T^{-1/2}\hat{\alpha}'_{\perp,T}y_{\lfloor T\cdot\rfloor}$, the more general result, with $\hat{\alpha}_{\perp,T}$ a consistent estimator of $\alpha_\perp$, follows from the result for $\hat{\alpha}_{\perp,T} = \alpha_\perp$ combined with the fact that

$$\sup_{0\le u\le 1} T^{-1/2}\left|\hat{\alpha}'_{\perp,T}y_{\lfloor Tu\rfloor} - \alpha'_\perp y_{\lfloor Tu\rfloor}\right| = \sup_{0\le u\le 1}\left|(\hat{\alpha}_{\perp,T}-\alpha_\perp)'\,T^{-1/2}y_{\lfloor Tu\rfloor}\right| \to_p 0,$$

which holds because $\hat{\alpha}_{\perp,T} \to_p \alpha_\perp$ and $T^{-1/2}y_{\lfloor T\cdot\rfloor}$ is tight.

5.4. Proof of Theorem 4. First consider the special case where $\hat{\theta}_T = \theta$. Because $(\Sigma^{-1/2}\alpha, \Sigma^{1/2}\alpha_\perp)$ is orthogonal, the matrix $(\Sigma^{-1}\alpha, \alpha_\perp)$ is non-singular. Transforming $v_t = y_t - \mu' d_t$ by this matrix leads to transformed errors $\varepsilon_t^* = (\varepsilon_{1t}^{*\prime}, \varepsilon_{2t}^{*\prime})' = (\varepsilon_t'\Sigma^{-1}\alpha, \varepsilon_t'\alpha_\perp)'$ with covariance matrix I_p and the transformed system

$$\begin{aligned}\alpha'\Sigma^{-1}\Gamma(L)\Delta v_t &= \alpha'\Sigma^{-1}\alpha\beta' v_{t-1} + \varepsilon_{1t}^*,\\ \alpha_\perp'\Gamma(L)\Delta v_t &= T^{-1}C\alpha_\perp'\Gamma(1)v_{t-1} + \varepsilon_{2t}^*.\end{aligned}$$

Because the first equation does not involve the parameter C, and the two disturbances ε_{1t}^* and ε_{2t}^* are independent, the profile likelihood function is defined only from the second equation. In other words, analogously to the proof of Theorem 3, we find that for any $\bar{C}$,

$$\max_\mu L_T^d\left(\bar{C}, \mu; r_0, \theta\right) - \max_\mu L_T^d\left(0, \mu; r_0, \theta\right) = \mathcal{L}_T^d\left(\bar{C}; r_0\right) - \mathcal{L}_T^d\left(0; r_0\right),$$

where $\mathcal{L}_T^d\left(C; r_0\right)$ is defined as in (8), but with $Y_{Tt}(C; r_0)$ now defined as

$$Y_{Tt}(C; r_0) = \alpha_\perp'\Gamma(L)\Delta y_t - T^{-1}C\alpha_\perp'\Gamma(1)y_{t-1}.$$

Define $w_t = \alpha_\perp'\Gamma(1)v_t$ and $w_t^* = \alpha_\perp'\Gamma(L)v_t$. The solution to Exercise 14.1 in Hansen and Johansen (1998) implies that

$$T^{-1/2}w_{\lfloor Tu\rfloor} = T^{-1/2}w_{\lfloor Tu\rfloor}^* + o_p(1) \to_d W_C(u) = \int_0^u \exp(C(u-s))dW(s),$$

where $W(\cdot)$ is a $(p-r_0)$-dimensional Wiener process, obtained as the limit in distribution of $T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor}\alpha_\perp'\varepsilon_t$. This result is obtained by replacing $\alpha_1\beta_1'$ in the notation of Hansen and Johansen (1998) by $\Sigma\alpha_\perp C\alpha_\perp'\Gamma(1)$, so that their "standardized" mean-reversion parameter ab' becomes

$$ab' = (\alpha_\perp'\Sigma\alpha_\perp)^{-1/2}\alpha_\perp'\alpha_1\beta_1'\beta_\perp(\alpha_\perp'\Gamma(1)\beta_\perp)^{-1}(\alpha_\perp'\Sigma\alpha_\perp)^{1/2} = C.$$

With Ψ_T and $\tilde{d}_{Tt}$ defined as in the proof of Theorem 2 we then find, analogously to the proof of that theorem (and again assuming $\mu = 0$ without loss of generality), that

$$\begin{aligned}\Psi_T Q_{DY,T}\left(\bar{C}; r_0\right) &= \mathrm{vec}\left(\sum_{t=1}^T \Delta\tilde{d}_{Tt}\Delta w_t^{*\prime}\right) + \mathrm{vec}\left[\left(\frac{1}{T^2}\sum_{t=1}^T \tilde{d}_{T,t-1}w_{t-1}'\right)\bar{C}'\bar{C}\right]\\ &\quad - \mathrm{vec}\left[\left(\frac{1}{T}\sum_{t=1}^T \Delta\tilde{d}_{Tt}w_{t-1}'\right)\bar{C}'\right] - \mathrm{vec}\left[\left(\frac{1}{T}\sum_{t=1}^T \tilde{d}_{T,t-1}\Delta w_t^{*\prime}\right)\bar{C}\right]\\ &\quad \to_d \mathrm{vec}\begin{pmatrix}\mathcal{Y}'\\ \lambda_C\left(\bar{C}\right)'\end{pmatrix},\end{aligned}$$

whereas $\Psi_T Q_{DD,T}(\bar{C}, r_0)\Psi_T$ has the same limit as before. This leads to the required result for the case where $\hat{\theta}_T = \theta$.

If $\hat{\theta}_T$ is a consistent estimator, then w_t and w_t^* in the equation above need to be replaced by $\hat{w}_t = \hat{\alpha}_{\perp,T}'\hat{\Gamma}_T(1)y_t$ and $\hat{w}_t^* = \hat{\alpha}_{\perp,T}'\hat{\Gamma}_T(L)y_t$, respectively. As in the proof of Theorem 3, consistency of $\hat{\theta}_T$ implies

$$\sup_{0\le u\le 1} T^{-1/2}\left|\hat{w}_{\lfloor Tu\rfloor} - w_{\lfloor Tu\rfloor}\right| = \sup_{0\le u\le 1}\left|\left[\hat{\alpha}_{\perp,T}'\hat{\Gamma}_T(1) - \alpha_\perp'\Gamma(1)\right]T^{-1/2}y_{\lfloor Tu\rfloor}\right| \to_p 0.$$

Furthermore, because

$$w_t = \alpha_\perp' \Gamma(L) y_t = \alpha_\perp' \Gamma(1) y_t + \alpha_\perp \Gamma^*(L) \Delta y_t = w_t^* + \alpha_\perp \Gamma^*(L) \Delta y_t,$$

where $\Gamma^*(z) = [\Gamma(z) - \Gamma(1)]/(1-z)$, it follows that

$$\sup_{0 \le u \le 1} T^{-1/2} \left| w_{\lfloor Tu \rfloor}^* - w_{\lfloor Tu \rfloor} \right| \to_p 0,$$

and analogously we have $\sup_{0 \le u \le 1} T^{-1/2} \left| \hat{w}_{\lfloor Tu \rfloor}^* - \hat{w}_{\lfloor Tu \rfloor} \right| \to_p 0$.

References

Chernoff, H. (1954): "On the Distribution of the Likelihood Ratio," *Annals of Mathematical Statistics*, 25, 573–578.

Dickey, D. A., and W. A. Fuller (1979): "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74, 427–431.

——— (1981): "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," *Econometrica*, 49, 1057–1072.

Doornik, J. A. (2007): *Object-Oriented Matrix Programming Using Ox*. Third Edition. London: Timberlake Consultants Press.

Drton, M. (2009): "Likelihood Ratio Tests and Singularities," *Annals of Statistics*, 37, 979–1012.

Elliott, G., T. J. Rothenberg, and J. H. Stock (1996): "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64, 813–836.

Haldrup, N., and M. Jansson (2006): "Improving Size and Power in Unit Root Testing," in *Palgrave Handbook of Econometrics, Volume 1: Econometric Theory*, ed. by T. C. Mills, and K. Patterson. New York: Palgrave Macmillan, 252-277.

Hansen, P. R., and S. Johansen (1998): *Workbook on Cointegration*. Oxford: Oxford University Press.

Hubrich, K., H. Lütkepohl, and P. Saikkonen (2001): "A Review of Systems Cointegration Tests," *Econometric Reviews*, 20, 247–318.

Jansson, M., and M. Ø. Nielsen (2011): "Nearly Efficient Likelihood Ratio Tests of the Unit Root Hypothesis," *Econometrica*, forthcoming.

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.

——— (1995): *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. New York: Oxford University Press.

Lütkepohl, H., and P. Saikkonen (2000): "Testing for the Cointegrating Rank of a VAR Process with a Time Trend," *Journal of Econometrics*, 95, 177–198.

Lütkepohl, H., P. Saikkonen, and C. Trenkler (2001): "Maximum Eigenvalue Versus Trace Tests for the Cointegrating Rank of a VAR Process," *Econometrics Journal*, 4, 287–310.

PHILLIPS, P. C. B. (1988): "Regression Theory for Near-Integrated Time Series," *Econometrica*, 56, 1021–1043.

SAIKKONEN, P., AND H. LÜTKEPOHL (2000): "Trend Adjustment Prior to Testing for the Cointegrating Rank of a Vector Autoregressive Process," *Journal of Time Series Analysis*, 21, 435–456.

SAIKKONEN, P., AND R. LUUKKONEN (1997): "Testing Cointegration in Infinite Order Vector Autoregressive Processes," *Journal of Econometrics*, 81, 93–126.

SCHMIDT, P., AND J. LEE (1991): "A Modification of the Schmidt-Phillips Unit Root Test," *Economics Letters*, 36, 285–289.

STOKEY, N. L., AND R. E. LUCAS (1989): *Recursive Methods in Economic Dynamics*. Cambridge: Harvard University Press.

XIAO, Z., AND P. C. B. PHILLIPS (1999): "Efficient Detrending in Cointegrating Regression," *Econometric Theory*, 15, 519–548.