Queen's Economics Department Working Paper No. 1297

Improved Likelihood Ratio Tests for Cointegration Rank in the VAR Model

H. Peter Boswijk
University of Amsterdam

Michael Jansson
UC Berkeley and CREATES

Morten ÃŸrregaard Nielsen
Queen's University and CREATES

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

8-2012

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

January 15, 2014

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 as usual the asymptotic results do not require normally distributed innovations. Our tests differ from existing tests in two respects. 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, 1991) developed cointegration rank tests within the VAR model.[1] Related methods have been proposed by, among others, Phillips and Durlauf (1986), Stock and Watson (1988), Fountis and Dickey (1989), and Ahn and Reinsel (1990).

Subsequent contributions have generalized and refined this work in a variety of ways, notably by proposing tests with (asymptotic local) power properties superior to those of Johansen (e.g., Xiao and Phillips (1999), Hubrich, Lütkepohl, and Saikkonen (2001), Perron and Rodriguez (2012), and the references therein). The purpose of this paper is to propose

*We thank an anonymous associate editor, two anonymous referees, Joon Park, Benoit Perron, seminar participants at New York University, Brown University, Rutgers, Columbia University, Pennsylvania State University, Michigan State University, Indiana University, Rice University, Aarhus University, Freie Universität Berlin, the Tinbergen Institute, the University of Exeter, and Yale University, and participants at the 2013 HEC Montréal-CIRPÉE Applied Financial Time Series Workshop, the 2013 CIREQ Econometrics Conference and the 2013 New York Camp Econometrics for comments and discussion. Financial support from the Canada Research Chairs program, the Social Sciences and Humanities Research Council of Canada (SSHRC), and the Center for Research in Econometric Analysis of Time Series (CREATES, funded by the Danish National Research Foundation) is gratefully acknowledged.

†Corresponding author. Address: University of California, Berkeley, Department of Economics, 530 Evans Hall #3880, Berkeley, CA 94720. Email: mjansson@econ.berkeley.edu

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

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 asymptotic local power.

In the related unit root testing literature, it has long been recognized that in models with an unknown mean and/or linear trend, the class of nearly efficient unit root tests does not contain the Dickey and Fuller (1979, 1981, henceforth Dickey-Fuller) tests.[2] The Dickey-Fuller tests 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 (2012) 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 (2012). We show that their qualitative findings about the relative merits of likelihood ratio tests derived from conditional and full likelihoods extend to tests of cointegration rank. In addition, our tests are capable of incorporating a "sign" restriction which makes the tests interpretable as generalizations of one-sided unit root tests. 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 an additional layer of complexity. Section 3 evaluates the asymptotic null distributions and local power functions of the newly proposed tests, and Section 4 presents a Monte Carlo simulation study of the finite sample properties of the tests. Some additional discussion is given in Section 5. The proofs of our theorems are provided in the Appendix.

2. Likelihood Ratio Statistics

Our development proceeds in four steps, culminating in the derivation of statistics designed to test reduced rank hypotheses about the matrix $\Pi \in \mathbb{R}^{p\times p}$ in a $VAR(k+1)$ model of the form

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

where $d_t = 1$ or $d_t = (1,t)'$, μ is an unknown parameter, ε_t is an innovation sequence, and $\Gamma(L) = I_p - \Gamma_1 L - \ldots - \Gamma_k L^k$ is an unknown matrix lag polynomial satisfying the condition (e.g, Johansen (1995, Assumption 1)) that if the determinant of $\Gamma(z)(1-z) - \Pi z$ is zero, then either $|z| > 1$ or $z = 1$.

As mentioned in the introduction, the test statistics proposed in this paper differ from existing tests in two respects, namely by (possibly) imposing "sign" restrictions on Π and by handling deterministic components (i.e., eliminating the nuisance parameter μ) in a way that turns out to be superior from the point of view of asymptotic local power. The main goal in Section 2.1 is to present the "sign" restriction. Accordingly, that section considers a

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

very special case in which Π is the only unknown parameter of the model and where the null hypothesis is $\Pi = 0$ (i.e., that Π has rank zero). Section 2.2 then introduces deterministics and describes our approach to elimination of μ when testing $\Pi = 0$.

Although very simple, the testing problems considered in Section 2.2 turn out to be "canonical" in the sense that from an asymptotic perspective the problem of testing reduced rank hypotheses about Π in the general model (1) can be reduced to a problem of the form considered in Section 2.2. The reduction is achieved by combining two distinct insights and it seems natural to proceed in a manner which reflect this. Accordingly, Section 2.3 explains how to test general reduced rank hypotheses about Π within the modeling framework of Section 2.2, while Section 2.4 considers the general model (1) and demonstrates that (the variance of ε_t and) the nuisance parameters $\Gamma_1, \ldots, \Gamma_k$ governing short-run dynamics can be treated "as if" they are known when basing inference about Π on a Gaussian quasi-likelihood.

2.1. Multivariate 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,

$$\Delta y_t = \Pi y_{t-1} + \varepsilon_t, \tag{2}$$

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 asymptotic local power results. When testing the multivariate unit root hypothesis $\Pi = 0$ in the model (2), this amounts to employing the reparameterization

$$\Pi = \Pi_T(C) = T^{-1}C \tag{3}$$

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

The statistics we consider are of the form

$$\mathrm{LR}_T(\mathcal{C}) = \sup_{\bar{C}\in\mathcal{C}} L_T(\bar{C}) - L_T(0), \tag{4}$$

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 $\mathrm{LR}_T(\mathcal{C})$ is a likelihood ratio statistic. Specifically, $\mathrm{LR}_T(\mathcal{C})$ is a likelihood ratio statistic associated with the problem of testing the null hypothesis $C = 0$ against the alternative $C \in \mathcal{C}\setminus\{0\}$.[3] Equivalently, $\mathrm{LR}_T(\mathcal{C})$ 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 (4), 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

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

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

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 $\mathrm{LR}_T(\mathcal{C})$ 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 $\mathrm{LR}_T(\mathbb{R})$ 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 $\mathrm{LR}_T(\mathbb{R}_-)$, 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 such tests. Doing so is one of the purposes of this paper.

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 $\mathrm{LR}_T\left(\mathcal{M}_p^-(p)\right)$ 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 $A(z) = I_p(1-z) - \Pi z$ satisfies the well known condition (e.g., Johansen (1995, Assumption 1)) that $|z| > 1$ or $z = 1$ whenever the determinant of $A(z)$ is zero, 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 (3), 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.

---

[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 the 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 $\mathrm{LR}_T(\mathcal{M}_p(1))$ and $\mathrm{LR}_T(\mathcal{M}_p(p))$.

[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).

As in Horvath and Watson (1995) another restriction that could be imposed on $\mathcal{C}$ is that some cointegration vectors are prespecified. For specificity, suppose it is known that the vector $\beta \in \mathbb{R}^p$ is in the cointegration space (under the alternative). When combined with rank restrictions, this knowledge is useful as it imposes a restriction on the coimage of the members of $\mathcal{C}$. If the members of $\mathcal{C}$ have rank no greater than $r < p$, then this rank restriction can be combined with the knowledge that β is in the cointegration space by setting $\mathcal{C}$ equal to either $\mathcal{M}_p(r;\beta) = \left\{a\beta' + a_1\beta_1' : a \in \mathbb{R}^p, a_1 \in \mathbb{R}^{p\times(r-1)}, \beta_1 \in \mathbb{R}^{p\times(r-1)}\right\}$ or $\mathcal{M}_p^-(r;\beta) = \mathcal{M}_p(r;\beta) \cap \mathcal{M}_p^-(r)$ depending on whether the "sign" restriction discussed above is also imposed.

The following result, which can be thought of as multivariate unit root analogue of Chernoff's theorem (e.g., Theorem 2.6 of Drton (2009)), characterizes the large sample properties of $\mathrm{LR}_T(\mathcal{C})$ 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* (2) *and* (3), *with* C *held fixed as* $T \to \infty$. *If* $\mathcal{C} \subseteq \mathbb{R}^{p\times p}$ *is a closed cone, then* $\mathrm{LR}_T(\mathcal{C}) \to_d \max_{\bar{C}\in\mathcal{C}} \Lambda_{p,C}(\bar{C})$, *where*

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

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

An explicit characterization of the limiting distribution of $\mathrm{LR}_T(\mathcal{C})$ is available in some special cases. In particular, for $r = 1,\ldots,p$, we have

$$\max_{\bar{C}\in\mathcal{M}_p(r)} \Lambda_{p,C}(\bar{C}) = \frac{1}{2}\sum_{j=1}^{r}\xi_j,$$

where $\xi_1 \geq \ldots \geq \xi_p \geq 0$ are the eigenvalues of the matrix

$$\left(\int_0^1 W_C(u)\,dW_C(u)'\right)'\left(\int_0^1 W_C(u)\,W_C(u)'\,du\right)^{-1}\left(\int_0^1 W_C(u)\,dW_C(u)'\right).$$

On the other hand, unlike the univariate (i.e., $p = 1$) situation investigated by Jansson and Nielsen (2012) the more general multivariate (i.e., $p > 1$) situation covered here is one where the "sign-restricted" versions of the maximum eigenvalue and trace statistics do not seem to have asymptotic representations expressible in closed form.

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

2.2. Deterministic Terms. As an initial generalization of the model (2), suppose

$$y_t = \mu' d_t + v_t, \qquad \Delta v_t = \Pi v_{t-1} + \varepsilon_t, \tag{5}$$

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 (2) only by accommodating deterministic terms. Under (3), the model gives rise to a log-likelihood function that can be expressed in terms C and μ as

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

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

$$\mathrm{LR}_T^d(\mathcal{C}) = \sup_{\bar{C} \in \mathcal{C}, \mu} L_T^d(\bar{C}, \mu) - \max_\mu L_T^d(0, \mu).$$

This statistic can be expressed in semi-closed form as

$$\mathrm{LR}_T^d(\mathcal{C}) = \sup_{\bar{C} \in \mathcal{C}} \mathcal{L}_T^d(\bar{C}) - \mathcal{L}_T^d(0),$$

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

$$\mathcal{L}_T^d(C) = -\frac{1}{2} Q_{YY,T}(C) + \frac{1}{2} Q_{DY,T}(C)' Q_{DD,T}(C)^{-1} Q_{DY,T}(C),$$

with

$$Q_{YY,T}(C) = \sum_{t=1}^{T} Y_{Tt}(C)' Y_{Tt}(C),$$

$$Q_{DY,T}(C) = \sum_{t=1}^{T} D_{Tt}(C)' Y_{Tt}(C),$$

$$Q_{DD,T}(C) = \sum_{t=1}^{T} D_{Tt}(C)' D_{Tt}(C).$$

Unlike the zero-mean case considered in Section 2.1, the statistic $\mathrm{LR}_T^d(\mathcal{C})$ 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. However, 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,

[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 $-\frac{1}{2}\|y_1 - \mu' d_1\|^2$ to be expressed in the same way as the other terms in the summation.

Lütkepohl, and Saikkonen (2001), and the references therein).[8] 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 (2012), where an analogous formulation was shown to provide an "automatic" way of avoiding the aforementioned power loss in the scalar case (i.e., when $p = 1$).

As discussed in more detail in the simulation study in Section 4.1 below, numerical maximization of $\mathcal{L}_T^d(C)$ with respect to $C \in \mathcal{C}$ is computationally straightforward in the cases of main interest. Nevertheless, we mention here for completeness that if $\mathcal{C}$ is of the form $\mathcal{M}_p(r)$, so that no "sign" restrictions are imposed, then $L_T^d(C, \mu)$ can be maximized over (C, μ) by a switching algorithm. Specifically, for a fixed value of C the maximum likelihood estimator $\arg\max_\mu L_T^d(C, \mu)$ of μ has a closed-form expression of the GLS type, a fact that was also exploited in the derivation of $\mathcal{L}_T^d(C)$ above. Similarly, for a fixed value of μ the maximum likelihood estimator $\arg\max_{C \in \mathcal{M}_p(r)} L_T^d(C, \mu)$ of C can be obtained by reduced rank regression applied to the error correction model for $v_t = y_t - \mu' d_t$. Therefore, $L_T^d(C, \mu)$ can be maximized over (C, μ) by alternating between maximization over μ given C and maximization over C given μ.

In the scalar case studied by Jansson and Nielsen (2012), 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.

Theorem 2. *Suppose $\{y_t\}$ is generated by (5) and (3), 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 $\mathrm{LR}_T^d(\mathcal{C}) \to_d \max_{\bar{C} \in \mathcal{C}} \Lambda_{p,C}(\bar{C})$, where $\Lambda_{p,C}$ is defined in Theorem 1.

(b) If $d_t = (1, t)'$, then $\mathrm{LR}_T^d(\mathcal{C}) \to_d \max_{\bar{C} \in \mathcal{C}} \Lambda_{p,C}^\tau(\bar{C})$, 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 $\bar{C}$,

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

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

Theorem 2(a) implies in particular that in the constant mean case, the asymptotic local power of the test based on $\mathrm{LR}_T^d(\mathcal{M}_p(p))$ 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 $\mathrm{LR}_T^d(\mathcal{M}_p(p))$ and $\mathrm{LR}_T^d(\mathcal{M}_p^-(p))$ 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.

[8] The low power appears to be attributable to inefficiency of OLS (relative to GLS) as an estimator of deterministic components in models with highly persistent data. For details, see e.g. Phillips and Lee (1996) and Canjels and Watson (1997).

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 (2000b). Under the assumptions of Theorem 2(b), the so-called GLS (trace) statistics proposed in those papers all have asymptotic representations of the form

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

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

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

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.[9]

Using this notation, the asymptotic representation of one-half times the GLS trace statistics of Lütkepohl and Saikkonen (2000) and Saikkonen and Lütkepohl (2000b) can be written as $\mathcal{LR}_{p,C}^{GLS}\left(\mathcal{M}_p(p);0\right)$, where $\mathcal{LR}_{p,C}^{GLS}\left(\mathcal{C};\bar{C}^*\right) = \max_{\bar{C}\in\mathcal{C}} \Lambda_{p,C}^{GLS}\left(\bar{C};\bar{C}^*\right)$.[10] 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

[9] In other words, $\tilde{W}_{C,\bar{C}^*}(u)$ is a multivariate version of the process $V_c(u,\bar{c}^*)$ defined by Elliott, Rothenberg, and Stock (1996, Section 2.3).

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

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

where, letting $\hat{\mu}_T^* = \arg\max_\mu L_T^d\left(\bar{C}^*,\mu\right)$, $\mathrm{LR}_T^{GLS}\left(\mathcal{C};\bar{C}^*\right) = \sup_{\bar{C}\in\mathcal{C}} L_T^d\left(\bar{C},\hat{\mu}_T^*\right) - L_T^d\left(0,\hat{\mu}_T^*\right)$. As a consequence, every limiting representation (indexed by $\mathcal{C}$ and $\bar{C}^*$) of the form $\mathcal{LR}_{p,C}^{GLS}\left(\mathcal{C};\bar{C}^*\right)$ is achievable. It is beyond the scope of this paper to attempt to isolate "optimal" choices of $\mathcal{C}$ and $\bar{C}^*$. Instead, our aim is to clarify the relationship between our tests and certain tests already in the literature.

interpreted as multivariate generalizations of the DF-GLS statistics of Elliott, Rothenberg, and Stock (1996).

The results of Theorem 2 could be extended to more general deterministic specifications, allowing for a structural change in the mean or trend slope at a known break date. Analogous derivations as in the proof of Theorem 2 would lead to an asymptotic representation of the form $\max_{\bar{C}\in\mathcal{C}} \Lambda_{p,C}^{GLS}\left(\bar{C};\bar{C}\right)$ with $\tilde{W}_{C,\bar{C}}\left(u\right)$ as defined above, but with an appropriately adapted definition of $D_{\bar{C}}(u)$.

2.3. Reduced Rank Hypotheses. Next, we consider the problem of testing more general reduced rank hypotheses on the matrix Π in the model (5). 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 define $q = p - r_0$ and consider the case where Π is parameterized as

$$\Pi = \Pi_T\left(C; r_0, \alpha, \alpha_\perp, \beta\right) = \alpha\beta' + T^{-1}\alpha_\perp C\alpha_\perp', \tag{6}$$

where $C \in \mathbb{R}^{q\times q}$ is an unknown parameter of interest while $\alpha \in \mathbb{R}^{p\times r_0}$, $\alpha_\perp \in \mathbb{R}^{p\times q}$, and $\beta \in \mathbb{R}^{p\times r_0}$ are nuisance parameters satisfying the following: $(\alpha, \alpha_\perp)$ is orthogonal and the eigenvalues of $I_{r_0} + \alpha'\beta$ are less than one in absolute value.

In (6), 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 (6) 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 (6) 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 testing the null hypothesis $C = 0$ against the alternative $C \in \mathcal{C}\backslash\{0\}$ is given by

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

where

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

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

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

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 $\mathrm{LR}_T^d\left(\mathcal{C}; r_0, \alpha_\perp\right)$ is simply the statistic $\mathrm{LR}_T^d\left(\mathcal{C}\right)$ 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 $\mathrm{LR}_T^d\left(\mathcal{C}; r_0, \alpha_\perp\right)$ 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 $\mathrm{LR}_T^d\left(\mathcal{C}; r_0, \hat{\alpha}_{\perp,T}\right)$ is asymptotically equivalent to $\mathrm{LR}_T^d\left(\mathcal{C}; r_0, \alpha_\perp\right)$.

Theorem 3. *Suppose $\{y_t\}$ is generated by* (5) *and* (6), *with* (($\alpha, \alpha_\perp, \beta$) *and*) C *held fixed as* $T \to \infty$. *Moreover, suppose* $\mathcal{C} \subseteq \mathbb{R}^{q \times q}$ *is a closed cone and suppose* $\hat{\alpha}_{\perp,T} \to_p \alpha_\perp$.

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

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

The consistency requirement on $\hat{\alpha}_{\perp,T}$ is mild because the matrix $\Pi_0 = \alpha\beta'$ is of rank r_0 and is consistently estimable under the other assumptions of Theorem 3. To be specific, let $N(\cdot; r_0)$ be an $\mathbb{R}^{p \times q}$-valued function which satisfies $N(M; r_0)' N(M; r_0) = I_q$ and $N(M; r_0)' M = 0$ for every $p \times p$ matrix M of rank r_0. Then $\hat{\alpha}_{\perp,T} = N(\hat{\Pi}_{0,T}; r_0)$ will be consistent for $\alpha_\perp = N(\Pi_0; r_0)$ provided $\hat{\Pi}_{0,T} \to_p \Pi_0$ and provided the function $N(\cdot; r_0)$ is chosen to be continuous on $\mathcal{M}_p(r_0) \backslash \mathcal{M}_p(r_0 - 1)$, the set of $p \times p$ matrices of rank r_0.

2.4. Serial Correlation and Unknown Error Distribution. As a final generalization, we consider the model (1) under the following assumption on the initial condition and the errors.

Assumption 1. *(a) The initial condition satisfies* $\max(\|v_0\|, \ldots, \|v_{-k}\|) = o_p(\sqrt{T})$.
(b) 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* $\delta > 0$.

As argued by Müller and Elliott (2003) in a univariate context, relaxing Assumption 1(a) may be of interest and will affect the large sample power properties (but not the large sample size properties) of cointegration tests. To conserve space we develop asymptotic theory only under Assumption 1(a), but in the Monte Carlo experiments below we investigate the consequences of replacing Assumption 1(a) by assumptions of Müller and Elliott (2003) type, and find that the tests proposed herein remain competitive with (and often superior to) their rivals in that case.

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, collecting all nuisance parameters other than μ in the matrix $\theta = (\alpha, \alpha_\perp, \beta, \Sigma, \Gamma_1, \ldots, \Gamma_k)$, 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 $C \in \mathbb{R}^{q \times q}$ is an unknown parameter of interest while $\alpha \in \mathbb{R}^{p \times r_0}$, $\alpha_\perp \in \mathbb{R}^{p \times q}$, $\beta \in \mathbb{R}^{p \times r_0}$, and $\Gamma_1, \ldots, \Gamma_k \in \mathbb{R}^{p \times p}$ are nuisance parameters satisfying the following: $(\Sigma^{-1/2}\alpha, \Sigma^{1/2}\alpha_\perp)$ is orthogonal, the solutions to the determinantal equation $\det[\Gamma(z)(1 - z) - \alpha\beta' z] = 0$ satisfy $z = 1$ or $|z| > 1$, and the matrix $\beta'\Gamma(1)^{-1}\alpha$ is nonsingular.

The Gaussian quasi-log-likelihood function corresponding to the model with $v_0 = \ldots = v_{-k} = 0$ and with θ known can be expressed, up to a constant, as

$$L_T^d(C, \mu; r_0, \theta) = -\frac{1}{2}\sum_{t=1}^{T} \left\|\Sigma^{-1/2}\left[Y_{Tt}(C; r_0, \theta) - D_{Tt}(C; r_0, \theta)\operatorname{vec}(\mu)\right]\right\|^2, \tag{8}$$

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

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

Theorem 4. *Suppose* $\{y_t\}$ *is generated by* (1) *under Assumption 1, and with* Π *satisfying* (7) *with (*θ *and)* C *held fixed as* $T \to \infty$. *Moreover, suppose* $\mathcal{C} \subseteq \mathbb{R}^{q\times q}$ *is a closed cone and suppose* $\hat{\theta}_T \to_p \theta$.

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

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

A possible choice for the consistent estimator $\hat{\theta}_T$ is 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

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

where $\rho_1 = -\beta'\mu'$; in case (b), reduced rank regression is applied to

$$\Delta y_t = \alpha(\beta', \rho_2)(y_{t-1}', t)' + \Gamma_1 \Delta y_{t-1} + \ldots + \Gamma_k \Delta y_{t-k} + \Phi_2 + \varepsilon_t,$$

where $\rho_2 = -\beta'\mu'(0, 1)'$ and Φ_2 are 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 $\mathcal{C} = \mathcal{M}_q^-(q)$.

Critical values of the tests are given in Table 1. This table, as well as the local power functions in this section, are based on simulations conducted in Ox, see Doornik (2007).

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

	$\mathcal{C} = \mathcal{M}_q(q)$				$\mathcal{C} = \mathcal{M}^-_q(q)$				
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: $\max_{\bar{C}\in\mathcal{C}} \Lambda^{\tau}_{q,0}(\bar{C})$									
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 each of the 100,000 replications, we simulated the q-variate Brownian motion process $W(u)$, approximated by a Gaussian vector random walk with 1000 steps. To simulate local power for some value of $C \neq 0$, we used the simulated $W(u)$ to generate, using an Euler discretization, the q-variate Ornstein-Uhlenbeck process $W_C(u)$. These were then used to calculate the limiting log-likelihood ratio functions $\Lambda_{p,C}(\bar{C})$ and $\Lambda^{\tau}_{p,C}(\bar{C})$ given in Theorems 1 and 2, where $C = 0$ for the simulations under the null hypothesis (to obtain critical values) and specific values of $C \neq 0$, discussed below, were used for local power calculations. For each replication, the functions $\Lambda_{p,C}(\bar{C})$ and $\Lambda^{\tau}_{p,C}(\bar{C})$ were maximized over $\bar{C}$. The "sign" restriction was imposed by maximization 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 the angle between a and $b_{\perp}$, where $C = ab'$.[11] The para-

[11] An alternative interpretation of ρ is as the long-run correlation between the errors v_{1t} and v_{2t} in a

metrization 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 $\ell \in [0, 50]$.

For the case of a constant mean, we compare the two likelihood ratio tests, indicated by LR($\mathcal{M}$) and LR($\mathcal{M}^-$), 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 $\mathcal{C} = \mathcal{M}_q(1)$). 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). As mentioned in Section 2.2, the test proposed by Saikkonen and Luukkonen (1997), which also allows for an unknown mean, has the same asymptotic local power function as that of LR($\mathcal{M}$).

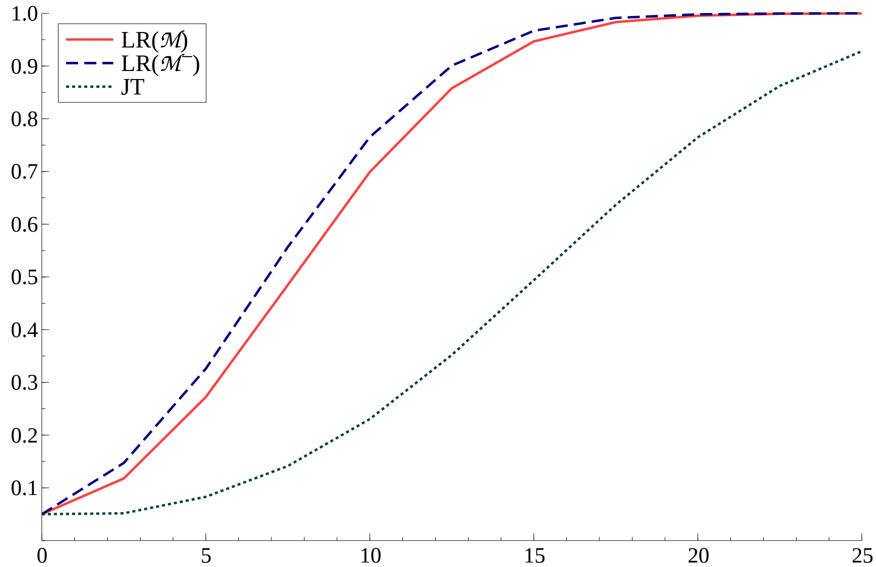
In Figures 1 and 2 we display the asymptotic local power functions for the constant mean case. It is clear that imposing the sign restriction does lead to a local power gain in the univariate case, but appears to make very little difference with $q = 2$. More importantly, both versions of the LR test 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 $\rho = 1$. This highlights the power gains that can be obtained from using the full likelihood instead of the conditional likelihood.

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 functions of the tests proposed by Lütkepohl and Saikkonen (2000) and Saikkonen and Lütkepohl (2000b), indicated by SL. Now the gains from imposing the "sign" restriction vanish entirely. In the univariate case, it is well known that allowing for a linear trend leads to a shift to the left in the distribution of the autoregressive coefficient estimator, relative to the case of a constant mean, which leads to a reduction of the probability of the estimator ending up in the explosive region (both under the null and under local alternatives), to such an extent that truncating the distribution at one does not affect asymptotic local power. Figures 3 and 4 suggest that apparently the same phenomenon occurs in higher-dimensional cases as well. The power difference between the likelihood ratio tests and the trace test in the linear trend case in Figures 3 and 4 are comparable to that in the constant mean case in Figures 1 and 2. The likelihood ratio tests also dominate the SL tests in terms of asymptotic local power, especially for local alternatives relatively far from the null hypothesis (i.e., for large ℓ), where the local power of the SL tests appear to approach one only very slowly.

As remarked by Hubrich, Lütkepohl, and Saikkonen (2001), further power gains are possible in case the process y_t has a linear trend, but it is known that the linear trend in

cointegrating regression framework $y_{1t} = \beta y_{2t} + u_{1t}$, where $\Delta u_{1t} = (-\ell/T)u_{1,t-1} + v_{1t}$ and $\Delta y_{2t} = v_{2t}$; see, e.g., Perron and Rodriguez (2012).

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. $q = p - r_0$ refers to the number of integrated linear combinations under the null hypothesis. LR($\mathcal{M}^-$) and LR($\mathcal{M}$) refer to the likelihood ratio tests allowing for a constant mean, with or without the "sign" restriction imposed, and JT refers to Johansen's trace test with a restricted constant.

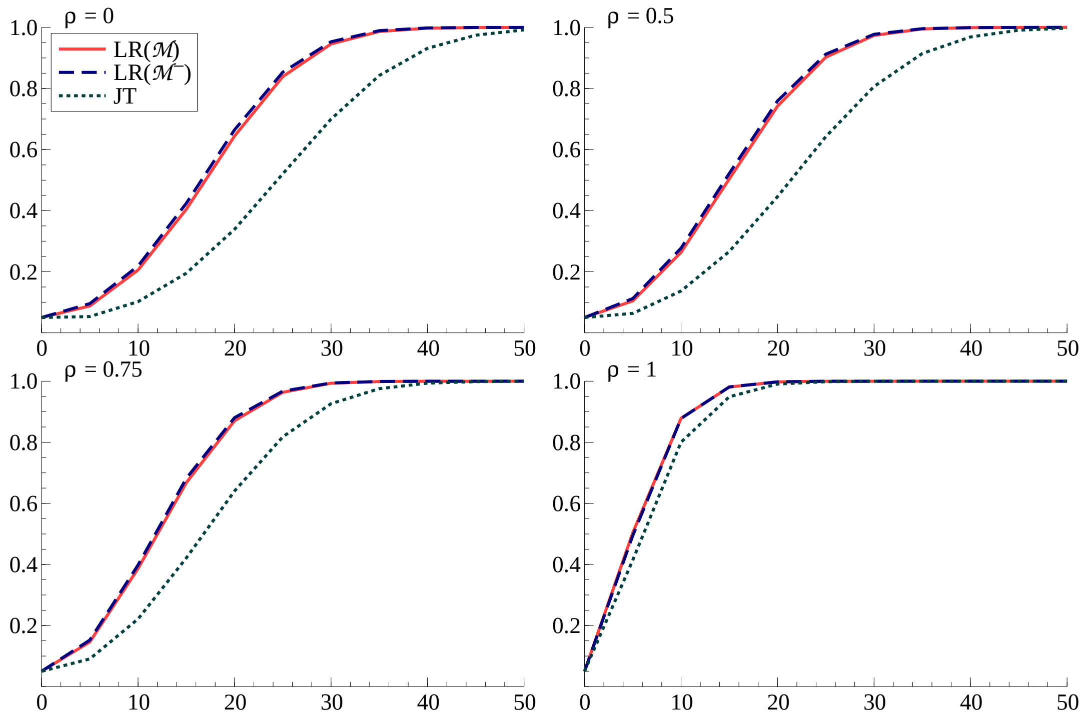
the cointegrating linear combinations $\beta' y_t$ is zero. This corresponds to the model with an unrestricted constant as analyzed by Johansen (1991, Theorem 2.1), and by Saikkonen and Lütkepohl (2000a). A drawback of this class of tests is that they are not asymptotically similar: their asymptotic null distribution depends on whether the trend in the integrated linear combinations is zero or not. For this reason, and because such tests are based on a different set of assumptions than our tests, we have not included them explicitly in the local power comparison.

4. Monte Carlo Simulations

In this section we present Monte Carlo simulation results to illustrate finite sample properties of the proposed tests and to compare with existing tests. The first subsection outlines the practical details on the implementation of the our tests.

4.1. Implementation of the LR Tests. Suppose we are calculating the LR test for $H_0 : r = r_0$, for some $0 \leq r_0 \leq p - 1$. The first thing that is needed to implement the LR tests in practice, is a preliminary estimate of $\theta = (\alpha, \alpha_\perp, \beta, \Sigma, \Gamma_1, \ldots, \Gamma_k)$. Here we note that $\alpha_\perp$ should be normalized such that $\alpha_\perp' \Sigma \alpha_\perp = I_p$, and also that if $r_0 = 0$ then $\alpha = \beta = 0$ and $\alpha_\perp = \Sigma^{-1/2}$. This preliminary estimate could be obtained in many different ways, but we apply the suggestion following Theorem 4 and obtain the estimate from the Johansen procedure under the appropriate rank restriction and with the appropriate deterministic components present. The preliminary estimate $\hat{\theta} = (\hat{\alpha}, \hat{\alpha}_\perp, \hat{\beta}, \hat{\Sigma}, \hat{\Gamma}_1, \ldots, \hat{\Gamma}_k)$,

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. $q = p - r_0$ refers to the number of integrated linear combinations under the null hypothesis. LR($\mathcal{M}^-$) and LR($\mathcal{M}$) refer to the likelihood ratio tests allowing for a constant mean, with or without the "sign" restriction imposed, and JT refers to Johansen's trace test with a restricted constant.

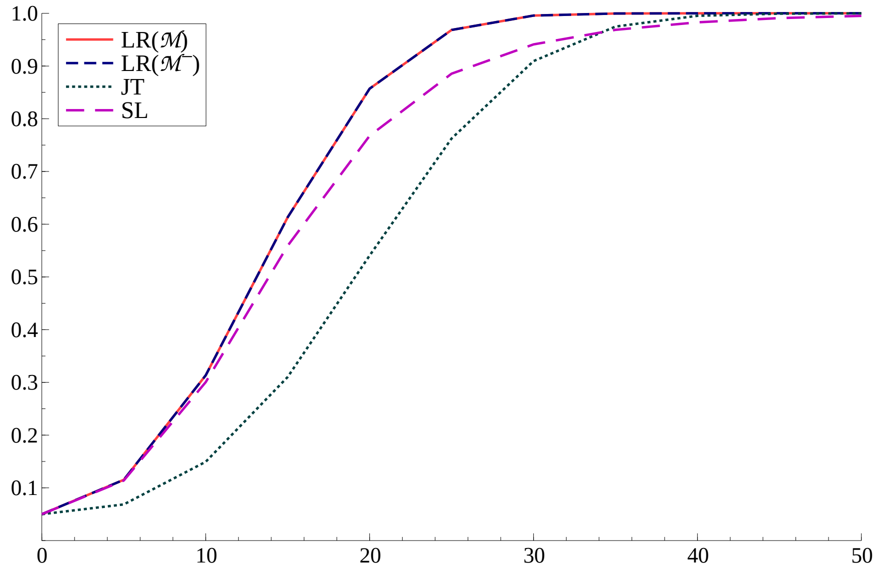
with $\hat{\alpha}_\perp' \hat{\Sigma} \hat{\alpha}_\perp = I_p$, and if $r_0 = 0$ then with $\hat{\alpha} = \hat{\beta} = 0$ and $\hat{\alpha}_\perp = \hat{\Sigma}^{-1/2}$, is taken as given and is fixed throughout the remainder of the procedure.

We then proceed to calculate the likelihood for a given value of the parameter C. First, we calculate $\hat{Y}_{Tt}(C) = \hat{\alpha}_\perp' \hat{\Gamma}(L)\Delta y_t - T^{-1} C \hat{\alpha}_\perp' \hat{\Gamma}(1) y_{t-1}$ and $\hat{D}_{Tt}(C) = \hat{\alpha}_\perp' \hat{\Gamma}(L) \otimes \Delta d_t' - T^{-1} C \hat{\alpha}_\perp' \hat{\Gamma}(1) \otimes d_{t-1}'$ for $t = 1, \ldots, T$, and to do so we set $y_0 = \ldots = y_{-k} = 0$ and $d_0 = \ldots = d_{-k} = 0$. Next, the likelihood function (8) should now be maximized, and again this could be done in several ways. We found it easiest to concentrate the likelihood function (8) with respect to μ and therefore maximize

$$\hat{\mathcal{L}}_T^d(C) = -\frac{1}{2} \hat{Q}_{YY,T}(C) + \frac{1}{2} \hat{Q}_{DY,T}(C)' \hat{Q}_{DD,T}(C)^{-1} \hat{Q}_{DY,T}(C)$$

with respect to C over the parameter space $\mathcal{C}$, where $\hat{Q}_{YY,T}(C) = \sum_{t=1}^T \hat{Y}_{Tt}(C)' \hat{Y}_{Tt}(C)$, $\hat{Q}_{DY,T}(C) = \sum_{t=1}^T \hat{D}_{Tt}(C)' \hat{Y}_{Tt}(C)$, and $\hat{Q}_{DD,T}(C) = \sum_{t=1}^T \hat{D}_{Tt}(C)' \hat{D}_{Tt}(C)$. In the case where $\mathcal{C} = \mathcal{M}_p(r)$, this can be done by unrestricted maximization over $a \in \mathbb{R}^{p \times r}$ and $b = [I_r, b_2]$ with $b_2 \in \mathbb{R}^{(p-r) \times r}$, setting $C = ab'$. Thus, if for example $\mathcal{C} = \mathbb{R}^{p \times p}$, as will often be the case when trace-type tests are considered, the maximization is unrestricted.

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. $q = p - r_0$ refers to the number of integrated linear combinations under the null hypothesis. LR($\mathcal{M}^-$) and LR($\mathcal{M}$) refer to the likelihood ratio tests allowing for a linear trend, with or without the "sign" restriction imposed, JT refers to Johansen's trace test with a restricted linear trend, and SL refer to the test proposed by Saikkonen and Lütkepohl (2000b).

On the other hand, when $\mathcal{C} = \mathcal{M}_p^-(r)$, the maximization should be carried out under the appropriate eigenvalue restriction. In practice we used Ox, and, as in the asymptotic local power analysis, we applied the unrestricted maximization routine `MaxBFGS` in the former case, and the restricted maximization routine `MaxSQP` (sequential quadratic programming) in the latter case. Software implementing our proposed procedure is available from the authors upon request.

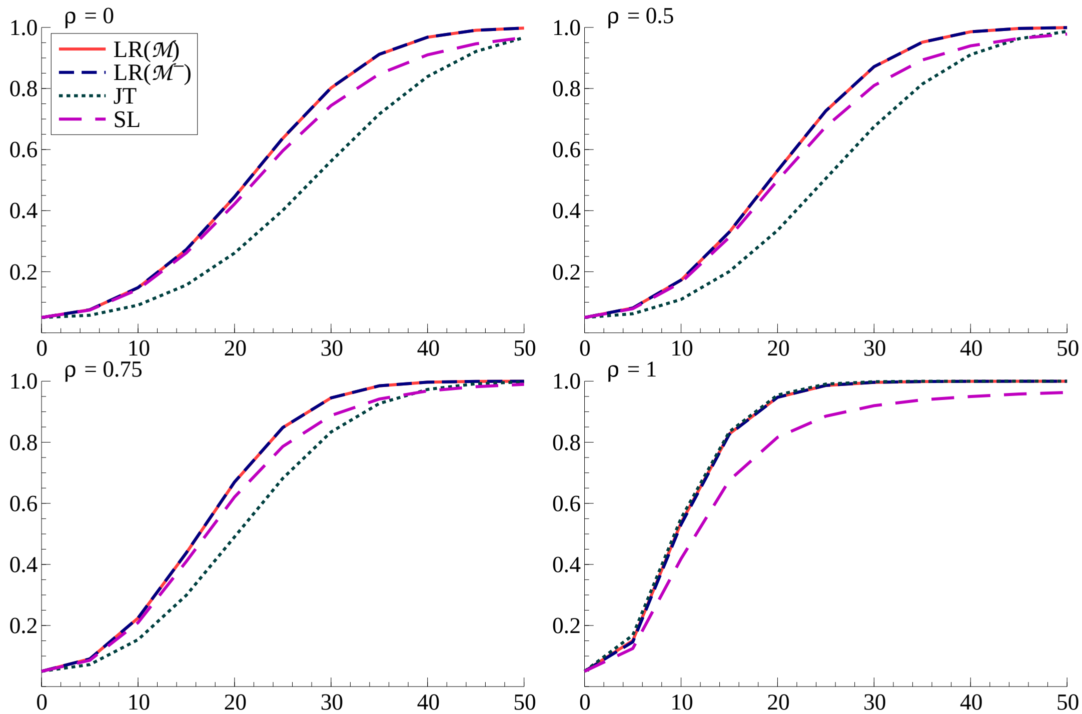
Finally, given the value of the maximized likelihood function, we subtract the value under the null hypothesis, $C = 0$, to calculate the LR statistic, noting that there is no multiplication by two. The value of the LR statistic is compared with the appropriate critical value obtained from Table 1, and as usual the test rejects if the LR statistic exceeds the critical value.

4.2. Monte Carlo Setup. We consider the two-dimensional VAR model,

$$y_t = \mu_1 + \mu_2 t + v_t, \; \Delta v_t = \Pi v_{t-1} + \Gamma \Delta v_{t-1} + \varepsilon_t, \tag{9}$$

where $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0, I_2)$. The 2×1 parameters μ_1, μ_2 are set equal to zero in the data generating process, so that $y_t = v_t$, but either μ_1 or (μ_1, μ_2), in the constant mean and trend cases, respectively, are estimated as part of the testing procedure. We set $\Gamma = \gamma I_2$ with

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. $q = p - r_0$ refers to the number of integrated linear combinations under the null hypothesis. LR($\mathcal{M}^-$) and LR($\mathcal{M}$) refer to the likelihood ratio tests allowing for a linear trend, with or without the "sign" restriction imposed, JT refers to Johansen's trace test with a restricted linear trend, and SL refer to the test proposed by Saikkonen and Lütkepohl (2000b).

$\gamma \in \{0, 1/2\}$, and, following the setup in the asymptotic local power analysis, we set

$$\Pi = \eta(1-\gamma)\begin{bmatrix} -\sqrt{1-\rho^2} & 0 \\ \rho & 0 \end{bmatrix}, \tag{10}$$

where $\eta \geq 0$ is a scale parameter, ρ is an angle parameter, and the factor $(1-\gamma)$ arises from $\Gamma(1) = I_2(1-\gamma)$ as in (7). Thus, when $\eta = 0$ the cointegration rank is zero and when $\eta > 0$ the cointegration rank is one. We simulate with $\rho \in \{0, 1/2, 3/4\}$ as in the asymptotic local power study.

To examine the sensitivity of our results to the initial values assumption, we initialize the process in two different ways. In one set of results, the process is initialized with zeros, i.e. $v_0 = \cdots = v_{-k} = 0$. In the other set of results, we initialize the process from the stationary distribution of $(v_{1,0}, \Delta v_{2,0}), \ldots, (v_{1,-k}, \Delta v_{2,-k})$ together with $v_{2,-k-1} = 0$ (when the cointegrating rank is one) and the stationary distribution of $\Delta v_0, \ldots, \Delta v_{-k}$ together with $v_{-k-1} = 0$ (when the cointegrating rank is zero).

We simulate 10,000 independent replications from this data generating process with sample sizes $T \in \{250, 500, 750\}$, reflecting, e.g., quarterly data since about 1950 ($T = 250$) or monthly data since about 1970 and 1950 ($T = 500$ and $T = 750$, respectively). When $\gamma = 0$, a VAR(1) is applied, i.e., $k = 0$ (no lag augmentation), and when $\gamma = 1/2$, a VAR(2) is applied, i.e., $k = 1$, so that correct specification is assumed. We simulate the LR test with $\mathcal{C} = \mathcal{M}_2^-(2)$ and $\mathcal{C} = \mathcal{M}_2(2) = \mathbb{R}^{2\times 2}$; that is, both with the "sign" restriction imposed and without. These tests are denoted LR_T^- and LR_T, respectively, in our tables. We also report the corresponding results for the Johansen trace test, denoted JT_T, which are implemented with a restricted constant term (for the constant mean case) or with a restricted trend term (for the linear trend case). Finally, we report results for the Saikkonen and Luukkonen (1997) tests (in the constant mean case) or the Saikkonen and Lütkepohl (2000b) tests (in the linear trend case), and in either case these are denoted SL_T in our tables.

4.3. Monte Carlo Results. The Monte Carlo simulation results with zero initialization are presented in Tables 2–5. The tables present the (percentage) empirical rejection frequencies for three tests. In Panel A of each table, we examine the size of the tests by testing the (true) null hypotheses $H_0 : r = 1$ with $\eta = 1/30$ (1/20 in the linear trend case) and $H_0 : r = 0$ with $\eta = 0$. In Panel B of each table we examine power by testing the (false) null hypothesis $H_0 : r = 0$ with $\eta = 1/30$ (1/20 in the linear trend case). Both the raw rejection frequencies and the corresponding size-corrected powers are reported. Throughout the tables, the nominal size of the tests is 5%.

Tables 2–5 about here

First, consider the results for the model allowing for a constant mean (but no trend) presented in Tables 2–3. All the tests demonstrate good size control for the null $H_0 : r = 1$. For the null $H_0 : r = 0$, the SL_T test is rather over-sized, whereas the newly proposed LR_T and LR_T^- tests have very good size properties. The (unadjusted) power of the tests reflect the relatively poor size control of the SL_T test. After size-correction, the powers of the SL_T test and the LR_T and LR_T^- tests are almost identical, and all three tests have much higher rejection frequencies than the JT_T test.

Next, consider the results for the model that allows for a linear trend. These are presented in Tables 4–5, which are laid out exactly as the previous tables, but with a slightly larger scale parameter, η. For this model, all tests show excellent size control (recall that the SL_T test in the model with a linear trend is different from the SL_T test in the model with a constant mean), although all tests are slightly undersized for the null $H_0 : r = 1$. In Panel B of both tables, the asymptotic local power results are reflected very clearly, in the sense that the newly proposed LR_T and LR_T^- tests have higher (both unadjusted and size-corrected) power than the JT_T test as well as the SL_T test. The differences in size-corrected power are in many cases quite substantial. For $T = 500$ and $T = 750$, the size-corrected power of the LR_T and LR_T^- tests are five to eight percentage points higher than that of the SL_T test throughout Tables 4–5.

Tables 6–9 about here

Finally, to examine the robustness of the tests towards the initial values assumption, Tables 6–9 present results corresponding to those in Tables 2–5, but with stationary initialization. It is clear from the tables that the empirical rejection frequencies of the JT_T test are almost identical for the two different initializations. Furthermore, the size properties of the SL_T, LR_T, and LR_T^- tests under the stationary initialization are also almost identical to those under the zero initialization. Specifically, the SL_T test is slightly oversized in the model with a constant mean (recall again that it is different from the SL_T test in the model with a linear trend), whereas the other tests show excellent size control.

In terms of power, both the SL_T test and the LR_T and LR_T^- tests show some loss in power compared with the zero initialization. In fact, the (unadjusted and size-corrected) power of the JT_T test is often higher than that of the SL_T test in the model that allows for a linear trend, see Panel B in Tables 8 and 9. However, even under the stationary initialization, the newly proposed LR_T and LR_T^- tests still have superior (unadjusted and size-corrected) power compared with the JT_T test throughout all of Tables 6–9.

5. Discussion and Conclusions

In this paper, we have suggested improved tests for cointegration rank in the vector autoregressive model and developed relevant asymptotic distribution theory and local power results. The tests are (quasi-)likelihood ratio tests based on a Gaussian likelihood, but as usual the asymptotic results do not require normally distributed innovations. 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 when implemented using the procedure described in Section 4.1.

To deal with the nuisance parameters, we use a plug-in approach for those parameters that are irrelevant to the asymptotic distributions (and asymptotic local power). On the other hand, the likelihood is maximized with respect to those parameters that are important for 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.

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 asymptotic local power. In the univariate case, our tests reduce to those of Jansson and Nielsen (2012) 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 optimality properties.

Appendix: Proofs

Proof of Theorem 1 We use a method of proof similar to that of Jansson and Nielsen (2012). Expanding $L_T(C)$ around $C=0$, we have

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

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

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

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

implying in particular that $F\left(\bar{C}, S_T, H_T\right) \to_d F\left(\bar{C}, \mathcal{S}_C, \mathcal{H}_C\right) = \Lambda_{p,C}\left(\bar{C}\right)$ 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 $F^*(C, H) = -\frac{1}{2}\operatorname{tr}(C'CH)$, 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)) shows that $\max_{\bar{C}\in\mathcal{C}} F\left(\bar{C}, \cdot\right)$ is continuous at (S_0, H_0).

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

To do so, note first that

$$\begin{aligned}\mathcal{L}_T^d\left(\bar{C}\right)-\mathcal{L}_T^d\left(0\right)&=L_T\left(\bar{C}\right)-L_T\left(0\right)\\&\quad+\frac{1}{2}Q_{DY,T}\left(\bar{C}\right)'Q_{DD,T}\left(\bar{C}\right)^{-1}Q_{DY,T}\left(\bar{C}\right)-\frac{1}{2}Q_{DY,T}\left(0\right)'Q_{DD,T}\left(0\right)^{-1}Q_{DY,T}\left(0\right),\end{aligned}$$

where $L_T\left(\bar{C}\right)-L_T\left(0\right)\rightarrow_d\Lambda_{p,C}\left(\bar{C}\right)$ because $v_t=y_t$. Next, let $d_0=0$ and $y_0=0$ and define $\Psi_T=I_p\otimes\mathrm{diag}(1,1/\sqrt{T})$ and $\tilde{d}_{Tt}=\mathrm{diag}(1,1/\sqrt{T})d_t$. For any $\bar{C}\in\mathcal{C}$, we have

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

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

Also, using Phillips (1988) and the identity $\int_0^1W_C\left(u\right)du=W_C\left(1\right)-\int_0^1udW_C\left(u\right)$,

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

where $\mathcal{Y}$ is a random variable independent of $W_C\left(\cdot\right)$. The desired conclusion follows from the preceding displays and simple algebra.

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 $\int_0^1 W_C(u)\, du = W_C(1) - \int_0^1 u dW_C(u)$, 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

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

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 $\Lambda_{p,C}^{\tau}(\bar{C}) = \Lambda_{p,C}^{GLS}\left(\bar{C}; \bar{C}\right)$, implying in particular that $\mathrm{LR}_T^d(\mathcal{C}) \to_d \max_{\bar{C} \in \mathcal{C}} \Lambda_{p,C}^{GLS}\left(\bar{C}; \bar{C}\right)$ as claimed in the main text.

Proof of Theorem 3 Because $(\alpha, \alpha_\perp)$ is orthogonal and replacing y_t by $y_t^* = \left(y_{1,t}^{*\prime}, y_{2,t}^{*\prime}\right)' = (y_t'\alpha, y_t'\alpha_\perp)'$ 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 (5) with $\Pi = T^{-1}C \in \mathbb{R}^{q \times q}$ (as in (3)). Moreover, it follows from simple algebra that, for any $\bar{C}$,

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

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

$$\mathcal{L}_T^d(C; r_0) = -\frac{1}{2} Q_{YY,T}(C; r_0) + \frac{1}{2} Q_{DY,T}(C; r_0)' Q_{DD,T}(C; r_0)^{-1} Q_{DY,T}(C; r_0), \qquad (11)$$

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 $D_{Tt}(C) = I_{p-r_0} \otimes \Delta d_t' - T^{-1}C \otimes d_{t-1}'$,

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

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

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

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

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 $\varepsilon^*_t = (\varepsilon^{*\prime}_{1t}, \varepsilon^{*\prime}_{2t})' = (\varepsilon'_t\Sigma^{-1}\alpha, \varepsilon'_t\alpha_\perp)'$ 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 $\bar{C}$,

$$\max\nolimits_\mu L^d_T\left(\bar{C}, \mu; r_0, \theta\right) - \max\nolimits_\mu L^d_T\left(0, \mu; r_0, \theta\right) = \mathcal{L}^d_T\left(\bar{C}; r_0\right) - \mathcal{L}^d_T\left(0; r_0\right),$$

where $\mathcal{L}^d_T\left(C; r_0\right)$ is defined as in (11), 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) can be used to show that

$$\left(T^{-1/2}w_{\lfloor T\cdot\rfloor}, T^{-1/2}w^*_{\lfloor T\cdot\rfloor}\right) \to_d \left(W_C(\cdot), W_C(\cdot)\right),$$

where $W_C(u) = \int_0^u \exp(C(u-s))dW(s)$ and $W(\cdot)$ is a q-dimensional Wiener process, obtained as the weak limit of $T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor}\alpha'_\perp\varepsilon_t$. (The derivation further replaces $\alpha_1\beta'_1$ in the notation of Hansen and Johansen (1998) by $\Sigma\alpha_\perp C\alpha'_\perp\Gamma(1)$, implying that their "standardized" mean-reversion parameter 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

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

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

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

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

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

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

References

Ahn, S. K., and G. C. Reinsel (1990): "Estimation for Partially Nonstationary Multivariate Autoregressive Models," *Journal of the American Statistical Association*, 85, 813–823.

Canjels, E., and M. W. Watson (1997): "Estimating Deterministic Trends in the Presence of Serially Correlated Errors," *Review of Economics and Statistics*, 79, 184–200.

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.

Fountis, N. G., and D. A. Dickey (1989): "Testing for a Unit Root Nonstationarity in Multivariate Autoregressive Time Series," *Annals of Statistics*, 17, 419–428.

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.

HORVATH, M. T. K., AND M. W. WATSON (1995): "Testing for Cointegration When some of the Cointegrating Vectors are Prespecified," *Econometric Theory*, 11, 984–1014.

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 (2012): "Nearly Efficient Likelihood Ratio Tests of the Unit Root Hypothesis," *Econometrica*, 80, 2321–2332.

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.

MÜLLER, U. K., AND G. ELLIOTT (2003): "Tests for Unit Root and the Initial Condition," *Econometrica*, 71, 1269–1286.

PERRON, P., AND G. RODRIGUEZ (2012): "Residual Based Tests for Cointegration with GLS Detrended Data," Working Paper, Pontificia Universidad Católica del Perú.

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

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 C. C. LEE (1996): "Efficiency Gains from Quasi-Differencing Under Nonstationarity," in *Essays in Memory of E.J. Hannan*, ed. by P. M. Robinson, and M. Rosenblatt. New York: Springer-Verlag, 300-314.

SAIKKONEN, P., AND H. LÜTKEPOHL (2000a): "Testing for the Cointegrating Rank of a VAR Process with an Intercept," *Econometric Theory*, 16, 373–406.

——— (2000b): "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.

Stock, J. H., and M. W. Watson (1988): "Testing for Common Trends," *Journal of the American Statistical Association*, 83, 1097–1107.

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.

Table 2: Simulation results, VAR(1) model, $\gamma_0 = 0$, constant mean, initialized with zeros

Panel A: Simulated size									
		$\eta = 1/30, H_0 : r = 1$				$\eta = 0, H_0 : r = 0$			
ρ	T	JT_T	SL_T	LR_T	LR_T^-	JT_T	SL_T	LR_T	LR_T^-
0.00	250	0.9	4.6	4.3	5.5	5.6	8.8	5.9	5.8
0.00	500	2.6	6.2	5.9	6.8	5.3	7.3	5.3	5.2
0.00	750	4.4	6.1	5.3	6.2	5.7	7.0	5.4	5.3
0.50	250	1.2	3.9	3.3	4.2	5.5	8.7	5.8	5.7
0.50	500	2.9	5.5	4.6	5.5	5.2	7.8	5.6	5.9
0.50	750	4.6	5.6	4.2	4.9	5.4	7.0	5.4	5.2
0.75	250	2.1	4.7	2.2	2.8	5.5	8.3	5.7	5.9
0.75	500	4.8	5.2	2.7	3.1	5.7	7.4	5.6	5.2
0.75	750	5.1	5.7	2.7	3.2	5.3	6.9	5.4	5.2
Panel B: Simulated power with $\eta = 1/30, H_0 : r = 0$									
		Not size-corrected				Size-corrected			
0.00	250	8.7	23.7	17.5	18.3	7.5	14.7	14.8	16.2
0.00	500	25.5	56.9	48.0	50.1	24.4	47.3	47.0	48.8
0.00	750	54.3	88.0	83.5	84.7	50.7	82.3	82.2	83.5
0.50	250	11.8	26.2	21.1	22.2	10.8	17.3	18.7	19.8
0.50	500	33.7	64.1	58.5	59.8	32.6	53.2	56.0	57.3
0.50	750	65.6	92.2	89.8	90.2	63.6	88.5	89.0	90.0
0.75	250	17.9	33.8	30.1	31.1	16.7	24.5	27.5	28.5
0.75	500	50.6	76.9	74.4	75.6	48.1	68.1	72.0	74.3
0.75	750	82.4	96.5	96.0	96.0	81.4	94.7	95.4	95.8

Note: The table presents (percentage) empirical rejection frequencies for the tests and models described in Section 4.2. The nominal size of the tests is 5% and all entries are based on 10,000 Monte Carlo replications.

Table 3: Simulation results, VAR(2) model, $\gamma_0 = 1/2$, constant mean, initialized with zeros

Panel A: Simulated size									
		$\eta = 1/30, H_0 : r = 1$				$\eta = 0, H_0 : r = 0$			
ρ	T	JT_T	SL_T	LR_T	LR_T^-	JT_T	SL_T	LR_T	LR_T^-
0.00	250	1.1	3.8	4.0	5.2	6.6	7.9	4.6	4.5
0.00	500	2.4	5.0	5.5	6.7	5.5	6.2	4.4	4.4
0.00	750	4.4	5.4	5.1	6.0	5.9	5.8	4.6	4.6
0.50	250	1.1	3.4	3.1	4.0	6.2	7.1	4.3	4.2
0.50	500	3.1	5.0	4.0	4.9	5.8	6.5	4.9	4.7
0.50	750	4.8	5.4	4.2	4.9	5.9	5.9	4.9	4.7
0.75	250	1.9	4.3	2.1	2.6	6.3	7.4	4.2	4.0
0.75	500	4.5	4.9	2.5	2.9	5.9	5.9	4.5	4.2
0.75	750	5.2	5.6	2.7	3.2	5.6	6.1	4.9	4.7
Panel B: Simulated power with $\eta = 1/30, H_0 : r = 0$									
		Not size-corrected				Size-corrected			
0.00	250	9.6	20.5	14.0	14.8	7.3	13.5	14.9	15.9
0.00	500	25.1	51.7	44.7	46.5	23.1	47.0	47.4	49.5
0.00	750	52.8	84.8	81.1	82.3	48.0	82.4	82.3	83.9
0.50	250	12.1	23.1	15.7	16.6	9.8	17.4	17.9	18.8
0.50	500	32.3	59.4	52.3	53.9	29.2	52.5	53.0	55.1
0.50	750	63.3	89.9	87.4	87.7	60.1	87.5	87.7	88.3
0.75	250	17.7	31.4	21.4	22.3	15.0	24.4	24.3	25.1
0.75	500	47.5	73.7	67.4	68.5	44.5	69.7	70.0	72.5
0.75	750	79.3	95.6	94.4	94.5	77.5	94.4	94.5	95.1

Note: The table presents (percentage) empirical rejection frequencies for the tests and models described in Section 4.2. The nominal size of the tests is 5% and all entries are based on 10,000 Monte Carlo replications.

Table 4: Simulation results, VAR(1) model, $\gamma_0 = 0$, linear trend, initialized with zeros

Panel A: Simulated size									
		$\eta = 1/20, H_0 : r = 1$				$\eta = 0, H_0 : r = 0$			
ρ	T	JT_T	SL_T	LR_T	LR_T^-	JT_T	SL_T	LR_T	LR_T^-
0.00	250	1.4	2.5	3.4	3.4	5.9	5.0	5.9	6.0
0.00	500	3.3	4.2	5.4	5.4	5.2	4.5	5.6	5.5
0.00	750	5.1	4.9	5.7	5.7	5.9	5.0	5.9	5.9
0.50	250	1.4	2.2	2.3	2.3	5.4	4.6	5.8	5.8
0.50	500	4.3	3.2	4.2	4.2	5.4	4.8	5.6	5.6
0.50	750	5.5	3.7	4.5	4.5	5.9	5.1	6.3	6.2
0.75	250	2.0	1.4	1.3	1.3	5.6	5.1	6.1	6.1
0.75	500	5.1	2.1	1.8	1.8	5.7	5.1	6.0	5.9
0.75	750	5.7	2.4	1.9	1.9	5.6	5.0	5.8	5.8
Panel B: Simulated power with $\eta = 1/20, H_0 : r = 0$									
		Not size-corrected				Size-corrected			
0.00	250	12.8	17.8	21.6	21.7	11.3	17.7	19.5	19.6
0.00	500	41.8	55.4	64.5	64.4	41.2	57.5	62.4	62.3
0.00	750	80.3	86.6	95.0	95.0	77.7	86.8	93.6	93.5
0.50	250	15.6	21.8	26.3	26.3	14.7	23.1	23.7	23.7
0.50	500	51.5	64.9	73.2	73.2	50.0	65.6	71.0	71.1
0.50	750	87.8	90.5	97.2	97.2	85.5	90.2	95.9	96.0
0.75	250	22.6	28.4	33.4	33.5	20.7	28.0	30.0	30.0
0.75	500	69.0	76.6	85.3	85.2	66.2	76.3	82.7	82.8
0.75	750	95.8	94.9	99.3	99.2	95.1	94.9	99.1	99.0

Note: The table presents (percentage) empirical rejection frequencies for the tests and models described in Section 4.2. The nominal size of the tests is 5% and all entries are based on 10,000 Monte Carlo replications.

Table 5: Simulation results, VAR(2) model, $\gamma_0 = 1/2$, linear trend, initialized with zeros

Panel A: Simulated size									
		$\eta = 1/20, H_0 : r = 1$				$\eta = 0, H_0 : r = 0$			
ρ	T	JT_T	SL_T	LR_T	LR_T^-	JT_T	SL_T	LR_T	LR_T^-
0.00	250	1.1	1.5	2.7	2.7	6.9	5.1	4.2	4.2
0.00	500	3.1	3.3	5.1	5.1	5.4	4.8	4.7	4.7
0.00	750	5.0	4.4	5.4	5.4	6.3	5.0	5.4	5.4
0.50	250	1.4	1.3	2.0	2.0	6.6	5.1	4.4	4.4
0.50	500	4.0	2.6	3.6	3.6	6.1	4.9	4.8	4.7
0.50	750	5.4	3.2	4.3	4.3	6.0	5.3	5.9	5.8
0.75	250	2.2	1.2	1.4	1.4	6.7	5.2	4.5	4.5
0.75	500	5.0	1.7	1.6	1.6	6.3	5.1	5.4	5.3
0.75	750	5.7	1.7	1.8	1.8	5.8	4.9	5.1	5.1
Panel B: Simulated power with $\eta = 1/20, H_0 : r = 0$									
		Not size-corrected				Size-corrected			
0.00	250	13.4	17.3	17.3	17.4	10.1	17.2	19.7	19.8
0.00	500	40.1	53.2	59.8	59.6	38.2	53.7	61.0	60.9
0.00	750	77.8	84.8	92.8	92.8	72.6	84.8	92.3	92.1
0.50	250	16.2	20.0	18.6	18.7	13.2	19.7	20.3	20.4
0.50	500	49.4	61.6	66.5	66.5	44.7	62.2	68.0	68.1
0.50	750	84.9	88.3	95.7	95.7	81.4	87.7	94.5	94.5
0.75	250	22.4	25.9	21.3	21.4	18.4	25.2	23.2	23.2
0.75	500	64.9	72.1	77.6	77.8	59.5	71.6	76.4	76.6
0.75	750	94.0	92.7	98.3	98.3	92.7	92.9	98.2	98.2

Note: The table presents (percentage) empirical rejection frequencies for the tests and models described in Section 4.2. The nominal size of the tests is 5% and all entries are based on 10,000 Monte Carlo replications.

Table 6: Simulation results, VAR(1) model, $\gamma_0 = 0$, constant mean, initialized with stat.

Panel A: Simulated size									
		$\eta = 1/30, H_0 : r = 1$				$\eta = 0, H_0 : r = 0$			
ρ	T	JT_T	SL_T	LR_T	LR_T^-	JT_T	SL_T	LR_T	LR_T^-
0.00	250	1.1	3.5	3.7	4.8	5.6	8.8	5.9	5.8
0.00	500	2.9	5.3	5.6	6.6	5.3	7.2	5.3	5.2
0.00	750	4.3	5.8	5.4	6.0	5.7	7.0	5.4	5.3
0.50	250	1.3	3.3	3.0	3.9	5.5	8.6	5.8	5.7
0.50	500	3.1	4.7	4.2	5.0	5.2	7.7	5.6	5.9
0.50	750	4.9	5.2	4.3	5.0	5.4	6.9	5.4	5.2
0.75	250	2.2	4.3	2.3	2.8	5.5	8.3	5.7	5.9
0.75	500	5.0	5.0	2.9	3.3	5.7	7.3	5.6	5.2
0.75	750	5.3	5.3	2.9	3.4	5.3	6.9	5.4	5.2
Panel B: Simulated power with $\eta = 1/30, H_0 : r = 0$									
		Not size-corrected				Size-corrected			
0.00	250	9.7	20.4	13.1	13.9	8.5	12.4	10.9	12.1
0.00	500	27.2	48.7	34.6	36.3	25.9	40.4	33.8	35.2
0.00	750	56.2	80.5	63.2	64.6	52.5	73.5	61.9	63.4
0.50	250	12.7	21.8	15.4	16.2	12.0	14.3	13.7	14.5
0.50	500	35.6	53.6	42.6	44.0	34.6	42.8	40.1	41.7
0.50	750	67.5	83.0	71.5	72.5	65.7	77.9	70.4	72.0
0.75	250	20.4	29.0	24.2	24.9	19.2	20.4	22.0	22.4
0.75	500	54.1	64.2	58.5	59.0	51.8	55.4	56.3	57.7
0.75	750	84.5	87.8	82.1	82.1	83.7	84.1	81.2	81.7

Note: The table presents (percentage) empirical rejection frequencies for the tests and models described in Section 4.2. The nominal size of the tests is 5% and all entries are based on 10,000 Monte Carlo replications.

Table 7: Simulation results, VAR(2) model, $\gamma_0 = 1/2$, constant mean, initialized with stat.

Panel A: Simulated size									
		$\eta = 1/30, H_0 : r = 1$				$\eta = 0, H_0 : r = 0$			
ρ	T	JT_T	SL_T	LR_T	LR_T^-	JT_T	SL_T	LR_T	LR_T^-
0.00	250	1.3	2.4	3.2	4.0	6.6	7.8	4.6	4.5
0.00	500	2.6	3.6	5.2	6.1	5.5	6.2	4.5	4.3
0.00	750	4.4	4.4	5.1	5.8	5.9	5.9	4.6	4.6
0.50	250	1.4	2.3	2.6	3.2	6.2	7.0	4.3	4.2
0.50	500	3.1	3.8	3.7	4.7	5.8	6.4	4.8	4.7
0.50	750	4.8	4.7	4.2	4.9	5.9	5.9	4.9	4.7
0.75	250	2.3	3.2	1.8	2.3	6.3	7.3	4.1	4.2
0.75	500	4.9	4.1	2.5	2.8	5.9	6.0	4.5	4.2
0.75	750	5.1	4.9	2.8	3.2	5.6	6.0	4.9	4.7
Panel B: Simulated power with $\eta = 1/30, H_0 : r = 0$									
		Not size-corrected				Size-corrected			
0.00	250	10.4	16.6	10.1	10.7	8.0	10.9	10.9	11.5
0.00	500	26.7	38.1	29.6	30.8	24.8	34.1	32.0	33.0
0.00	750	54.4	67.5	57.9	58.9	49.8	64.4	59.4	60.4
0.50	250	13.6	18.6	10.8	11.4	11.1	13.4	12.7	13.7
0.50	500	34.7	44.0	35.6	36.8	31.5	38.0	36.3	37.8
0.50	750	65.8	72.9	65.1	65.9	62.2	69.7	65.6	66.8
0.75	250	20.1	24.9	15.6	16.0	17.0	19.2	17.7	18.7
0.75	500	50.9	56.6	49.2	49.9	47.6	52.6	51.5	53.7
0.75	750	81.6	81.1	76.0	76.4	79.9	79.0	76.3	77.2

Note: The table presents (percentage) empirical rejection frequencies for the tests and models described in Section 4.2. The nominal size of the tests is 5% and all entries are based on 10,000 Monte Carlo replications.

Table 8: Simulation results, VAR(1) model, $\gamma_0 = 0$, linear trend, initialized with stat.

Panel A: Simulated size									
		$\eta = 1/20, H_0 : r = 1$				$\eta = 0, H_0 : r = 0$			
ρ	T	JT_T	SL_T	LR_T	LR_T^-	JT_T	SL_T	LR_T	LR_T^-
0.00	250	1.4	2.1	3.1	3.1	5.9	5.0	5.9	6.0
0.00	500	3.4	3.4	5.5	5.5	5.2	4.5	5.6	5.6
0.00	750	5.1	4.2	5.8	5.8	5.9	5.0	5.9	5.9
0.50	250	1.4	1.6	2.1	2.1	5.4	4.6	5.8	5.8
0.50	500	4.3	2.2	4.0	4.0	5.4	4.8	5.6	5.6
0.50	750	5.5	2.5	4.5	4.5	5.9	5.1	6.3	6.2
0.75	250	2.1	1.2	1.2	1.2	5.6	5.1	6.1	6.1
0.75	500	5.2	1.5	1.9	1.9	5.7	5.1	6.0	5.9
0.75	750	5.7	1.2	1.9	1.9	5.6	5.0	5.8	5.8
Panel B: Simulated power with $\eta = 1/20, H_0 : r = 0$									
		Not size-corrected				Size-corrected			
0.00	250	13.3	15.3	18.7	18.8	11.8	15.2	16.8	17.0
0.00	500	42.9	46.4	55.1	55.1	42.5	48.3	52.9	52.9
0.00	750	81.6	76.6	87.2	87.1	78.5	76.8	85.1	85.0
0.50	250	16.4	18.2	22.2	22.3	15.5	19.2	19.7	19.7
0.50	500	52.9	54.8	63.5	63.5	51.5	55.7	61.3	61.3
0.50	750	88.7	82.0	91.2	91.2	86.6	81.7	88.8	88.9
0.75	250	25.0	24.9	29.3	29.2	22.7	24.3	25.9	25.9
0.75	500	71.2	67.7	77.6	77.7	68.2	67.4	74.9	74.9
0.75	750	96.3	89.7	96.4	96.4	95.7	89.7	96.0	96.0

Note: The table presents (percentage) empirical rejection frequencies for the tests and models described in Section 4.2. The nominal size of the tests is 5% and all entries are based on 10,000 Monte Carlo replications.

Table 9: Simulation results, VAR(2) model, $\gamma_0 = 1/2$, linear trend, initialized with stat.

Panel A: Simulated size									
		$\eta = 1/20, H_0 : r = 1$				$\eta = 0, H_0 : r = 0$			
ρ	T	JT_T	SL_T	LR_T	LR_T^-	JT_T	SL_T	LR_T	LR_T^-
0.00	250	1.2	1.3	2.4	2.3	7.0	5.1	4.3	4.3
0.00	500	3.2	2.7	5.1	5.1	5.4	4.8	4.7	4.7
0.00	750	5.0	3.4	5.6	5.6	6.3	4.9	5.3	5.3
0.50	250	1.7	1.0	1.8	1.8	6.6	5.1	4.3	4.3
0.50	500	4.0	1.7	3.3	3.3	6.1	4.8	4.8	4.8
0.50	750	5.4	1.7	4.1	4.1	6.0	5.3	5.8	5.8
0.75	250	2.5	0.8	1.2	1.2	6.7	5.2	4.4	4.4
0.75	500	5.1	1.1	1.6	1.6	6.3	5.2	5.4	5.4
0.75	750	5.7	0.9	1.9	1.9	5.8	4.9	5.1	5.1
Panel B: Simulated power with $\eta = 1/20, H_0 : r = 0$									
		Not size-corrected				Size-corrected			
0.00	250	14.4	14.2	14.4	14.5	10.8	14.0	16.0	16.0
0.00	500	41.4	43.0	48.2	48.2	39.6	43.5	49.5	49.5
0.00	750	79.0	74.0	82.0	81.8	73.9	74.7	80.9	80.8
0.50	250	17.3	16.8	15.0	15.1	14.3	16.6	16.7	17.0
0.50	500	51.2	51.0	54.8	54.8	46.8	51.7	56.1	56.1
0.50	750	85.8	79.0	86.7	86.7	82.4	78.1	84.2	84.4
0.75	250	24.3	22.4	18.2	18.1	20.1	21.9	20.2	20.3
0.75	500	66.9	61.8	66.9	66.9	61.9	61.1	65.4	65.5
0.75	750	94.4	85.3	92.6	92.6	93.3	85.6	92.4	92.5

Note: The table presents (percentage) empirical rejection frequencies for the tests and models described in Section 4.2. The nominal size of the tests is 5% and all entries are based on 10,000 Monte Carlo replications.