Journal of Time Series Econometrics

Volume 2, *Issue* 1 2010 *Article* 5

A Nonlinear IV Likelihood-Based Rank Test for Multivariate Time Series and Long Panels

J. Isaac Miller, *University of Missouri*

Recommended Citation:
Miller, J. Isaac (2010) "A Nonlinear IV Likelihood-Based Rank Test for Multivariate Time Series and Long Panels," *Journal of Time Series Econometrics*: Vol. 2: Iss. 1, Article 5.
DOI: 10.2202/1941-1928.1057
Available at: http://www.bepress.com/jtse/vol2/iss1/art5

A Nonlinear IV Likelihood-Based Rank Test for Multivariate Time Series and Long Panels

J. Isaac Miller

Abstract

A test for the rank of a vector error correction model (VECM) or panel VECM based on the well-known trace test is proposed. The proposed test employs instrumental variables (IV's) generated by a class of nonlinear functions of the estimated stochastic trends of the VECM under the null. The test improves on the standard trace test by replacing the non-standard critical values with chi-squared critical values. Extending the result to the panel VECM case, the test is robust to cross-sectional correlation of the disturbances. The nonlinear IV rank test also extends earlier research on nonlinear IV unit root tests. However, the optimal instrument in the univariate case is not admissible in the more general multivariate case. The chi-squared result suggests that IV tests may be used to replace limits of other standard tests with integrated time series that are given by nonstandard stochastic integrals, even without a panel with which to pool test statistics.

KEYWORDS: VECM, panel VECM, cointegrating rank, trace test, nonlinear instruments

Author Notes: The author is very grateful to Ted Juhl and two anonymous referees for insightful comments on this research. The author is also grateful to the Department of Economics at Texas A&M University, where a large portion of the research for this paper was conducted as a visitor during the Fall 2008 semester. In particular, the author benefitted from discussions with Yoosoon Chang, Joon Park, and Hwansik Choi. Furthermore, the author is grateful to participants of the 2009 Midwest Econometrics Group Meeting at Purdue and a brown bag seminar at MU. The usual caveat applies.

1. Introduction

The problem of rank estimation in rank-deficient regressions became particularly important in analyzing multivariate time series data with the birth of cointegration in the 1980's. Johansen (1988) and Ahn and Reinsel (1990) brought cointegration of vector autoregressive processes to the forefront of research at the time. Johansen (1988) and Johansen and Juselius (1990) introduced likelihood ratio (LR) tests for the rank of a cointegrated vector error correction model (VECM), but the limits of these test statistics are nonstandard, involving stochastic integrals and nuisance parameters.

Improvements on these tests have been introduced in the literature. For example, Shintani (2001) developed a nonparametric test based on that of Phillips and Ouliaris (1990), which utilized degeneracy in the rank of the long-run variance matrix. Breitung (2002) suggested a generalized variance ratio statistic to test the rank.

An alternative strand of the literature has sought to replace rank tests with rank selection using information criteria. Gonzalo and Pitarakis (1998), Chao and Phillips (1999), Aznar and Salvador (2002), Kapetanios (2004), and Wang and Bessler (2005) have favored this approach, since consistent estimation of the rank should outperform tests in large samples. Cheng and Phillips (2009) have recently shown that the large-sample properties of information criteria do not require lag specification.

In a separate strand of the time series literature, the notion of instrumental variables based on nonlinear functions of integrated time series (NIV's) has roots in the theoretical contributions of Park and Phillips (1999, 2001).[1] As is well-known, the asymptotic limit of the sample covariance of a vector of integrated series and a scalar-valued stationary series is given by a vector of stochastic integrals. The primary intuition underlying NIV tests is that when a nonlinear transformation of the integrated series is employed, the analogous limiting vector is mixed *normal* with a *diagonal* covariance matrix.

NIV's have been used primarily in testing for unit roots in cross-sectionally correlated panels, as first introduced by Chang (2002). Using NIV's to instrument out cross-sectional correlation has been further explored by Chang (2006), Demetrescu and Tarcolea (2008), Chang and Song (2009), Demetrescu (2009), and Chang and Nguyen (2010). These tests primarily exploit the *diagonality* of the covariance matrix in panels with large cross-sectional dimensions.

[1] Additional theoretical contributions have been made along these lines by, *inter alia*, de Jong (2004), Pötscher (2004), de Jong and Wang (2005), Jeganathan (2008), and Wang and Phillips (2009).

Phillips *et al.* (2004) used NIV's in unit root tests in order to instrument out the non-normality associated with standard unit root tests, such as the Phillips-Perron tests (Phillips and Perron, 1988). These tests exclusively benefit from the (mixed) *normality* of covariance asymptotics, since *diagonality* of a univariate covariance matrix is moot. Those authors found that the optimal instrument for unit root testing is the sign function, which had been considered previously in unit root testing by So and Shin (1999).

The present analysis aims to further promote the use of nonlinear functions to instrument out non-normality by extending the use of NIV's from univariate unit root tests to multivariate rank tests. In this light, the main benefit of the proposed testing strategy is (mixed) *normality*. However, the rank deficiency of the system and potentially of the instruments presents nontrivial complications, and *diagonality* plays a key role in preserving rank through the IV transformations. In contrast with the results of Phillips *et al.* (2004), the optimal instrument in the univariate case is not even admissible in the multivariate case, as it leads to a critical singularity.

The basic mechanism of the proposed testing strategy consists of two steps. The first step involves using a consistent estimate of the long-run variance to create linear combinations of the series orthogonal to the cointegrating space (and in the direction of the stochastic trends) and projecting the linearly transformed series onto the space of the instruments. The projection is identical to the familiar first step of a standard 2SLS procedure. The second step involves running the familiar reduced rank regression and calculating the trace test statistic. Implementation of the new test in the second step differs slightly from the standard LR test, since a different number of eigenvalues are calculated. Under the null, the eigenvalues are all zeros after the initial linear combination of the series in the direction of the stochastic trends.

I extend the nonlinear IV test to the panel VECM case. The full benefits of nonlinear IV's become apparent in a panel, as the *diagonality* of the covariance matrix of the sample moment discussed above is critical with a large correlated cross-section. Much of the literature on cointegration in panels focuses on residual-based tests for cointegration in single-equation models for each cross-sectional unit. These include the tests of Pedroni (2004) based on the Phillips-Ouliaris (1990) cointegration test and variance ratio test, the LM test of McCoskey and Kao (1998), the DF/ADF cointegration tests of Kao (1999), and the recent NIV tests of Chang and Nguyen (2010).[2] Larsson *et al.* (2001) extended cointegration tests of single-equation panel models to rank tests of panel VECM's. Their test (the LR bar test) averages the Johansen LR

[2]Baltagi and Kao (2000) provide an excellent survey of the early literature on panel unit root tests and panel cointegration tests.

DOI: 10.2202/1941-1928.1057

test to obtain a normal limiting distribution as the cross-sectional dimension increases. Groen and Kleibergen (2003) developed an alternative LR test, and showed that both their LR test and that of Larsson *et al.* (2001) are robust to cross-sectional correlation in the variance, which is generally not the case with the earlier tests. The extension of the proposed test in this analysis is also robust to cross-sectional correlation as a direct result of the *diagonality* discussed above.

The rest of the paper proceeds as follows. I establish the basic models and assumptions and discuss some preliminary results in Section 2. In Sections 3, I discuss the mechanics and the chi-squared asymptotic limits of the test. I extend the test to more general settings in Sections 4 and 5. The proposed trace test is compared with the standard trace test and LR bar test using simulations in Section 6. Section 7 briefly concludes. Mathematical proofs are contained in an appendix.

I use the following notational conventions throughout the paper. e_i is a column vector of zeros with a single unit in the i^{th} row. In particular, for some matrix or row vector B, Be_i selects the i^{th} column of B, and e_i is assumed to be conformable depending on the context of usage. The Euclidean norm of a matrix B or vector b is denoted by $\|B\|$ or $\|b\|$. $B^{1/2}$ denotes the lower Cholesky decomposition of a positive definite symmetric matrix B, and $B^{-1/2}$ denotes the inverse of $B^{1/2}$. *vec* denotes the vectorization operator and *dg* represents a diagonal (or block-diagonal) matrix with diagonal elements (or blocks) given by its arguments.

2. Model, RR Estimation, and Instruments

Consider an $m \times 1$ VECM given by

$$\triangle y_t = \Gamma A' y_{t-1} + \varepsilon_t, \tag{1}$$

where A is an $m \times r$ matrix of cointegrating vectors, Γ is an $m \times r$ error correction matrix, and

[A1] $(\varepsilon_t) \sim \mathbf{iidN}\,(0, \Sigma)$

for $t = 1, \ldots, T$. Normality may be relaxed in the theoretical results below. However, normality is convenient in formulating the likelihood function.

Consider also a more general model given by

$$\triangle y_t = \mu + \Gamma A' y_{t-1} + \Pi w_t + \varepsilon_t \tag{2}$$

with the addition of a vector of non-zero means μ and stationary covariates (w_t), which may include lags of $(\triangle y_t)$. Π is a matrix of nuisance parameters for the purposes of testing. As is typical for this type of model, (w_t) and (ε_t) are assumed to be contemporaneously uncorrelated.

For expositional simplicity, I focus on the simpler model in (1) in Sections 2 and 3. Extending the results to accommodate the model in (2) is the central focus of Section 4.

2.1 Wold Representation and Long-Run Variance

The following set of assumptions, identical to that of Cheng and Phillips (2009), characterizes the cointegrating properties of the model.

[A2] *Γ and A are $m \times r$ matrices of rank r for $0 \leq r \leq m$, such that*

(a) *The determinantal equation $|I - (I + \Gamma A')x| = 0$ has roots on or outside the unit circle;*

(b) *If $r = 0$, then $\Gamma A' = 0$, and if $r = m$, then (y_t) is (asymptotically) stationary; and*

(c) *$I + A'\Gamma$ has eigenvalues within the unit circle.*

Part (a) rules out explosive roots, part (b) codifies the two extreme cases for r, and part (c) ensures invertibility of $I - (I + A'\Gamma)x$.

Define the $m \times (m - r)$ orthogonal complements of A and Γ to be $A_\perp$ and $\Gamma_\perp$, so that $A'A_\perp = \Gamma'\Gamma_\perp = 0$ and $(A, A_\perp)$ and $(\Gamma, \Gamma_\perp)$ are invertible. Two key results are collected in the following lemma.

Lemma 1 *Let assumption [A1]-[A2] hold for the model given by (1). The process $(\triangle y_t)$ has*

(a) *a Wold representation given by*

$$\triangle y_t = C(L)\varepsilon_t,$$

where $C(x) \equiv \sum_{k=0}^{\infty} C_k x^k = I + \Gamma B(x) A'$ and $B(x) \equiv \sum_{k=0}^{\infty} B_k x^k = (I - (I + A'\Gamma)x)^{-1}$ with $\sum_{k=0}^{\infty} k^2 \|C_k\| < \infty$, and

(b) *a long-run variance given by*

$$\Xi \equiv \mathbf{lrvar}(C(L)\varepsilon_t) = A_\perp(\Gamma_\perp' A_\perp)^{-1}\Gamma_\perp'\Sigma\Gamma_\perp(A_\perp'\Gamma_\perp)^{-1}A_\perp'$$

of rank $m - r$.

DOI: 10.2202/1941-1928.1057

2.2 Reduced Rank (RR) Regression

For notational simplicity and in keeping with Johansen's notation, let

$$r_{0t} \equiv \triangle y_t \qquad \text{and} \qquad r_{1t} \equiv y_{t-1}. \tag{3}$$

(These will be redefined subsequently for the model in (2).) For known A, RR regression reduces to a simple least squares regression to estimate Γ. The LS estimator $\hat{\Gamma}$ of Γ and the variance estimator $\hat{\Sigma}$ of Σ are simply $\hat{\Gamma}_{LS}(A) = S_{01}A\,(A'S_{11}A)^{-1}$ and $\hat{\Sigma}_{LS}\,(A) = S_{00} - S_{01}A\,(A'S_{11}A)^{-1}\,A'S_{10}$, where $S_{gh} \equiv T^{-1}\sum r_{gt}r'_{ht}$ for $g, h = 0, 1$ denotes sample moments using r_{0t} and r_{1t}.

The likelihood function may be concentrated so that the maximal value (up to an irrelevant constant) is given by

$$\begin{aligned} L_{\max}^{-2/n} &= |S_{00} - S_{01}A\,(A'S_{11}A)^{-1}\,A'S_{10}| \\ &= |S_{00}||A'(S_{11} - S_{10}S_{00}^{-1}S_{01})A|/|A'S_{11}A|, \end{aligned}$$

and A is chosen to minimize the right-hand side in order to maximize the likelihood. As is well-known, A may be estimated by finding the r largest eigenvalues of $S_{11} - S_{10}S_{00}^{-1}S_{01}$ subject to $A'S_{11}A = I$. The m ordered eigenvalues $\hat{\lambda}_{RR,1}, \ldots, \hat{\lambda}_{RR,r}, \hat{\lambda}_{RR,r+1}, \ldots, \hat{\lambda}_{RR,m}$ are the same as those obtained by solving the determinantal equation $|\lambda I - S_{11}^{-1}S_{10}S_{00}^{-1}S_{01}| = 0$. The eigenvectors corresponding to the r largest of these form the columns of $\hat{A}_{RR}$. The reader is referred to Johansen (1995) for very detailed exposition on the RR technique.

2.3 Standardization and Transformation

It is straightforward to conceptualize estimating a reduced rank regression in a simple IV or 2SLS framework. In place of $\hat{\Gamma}_{LS}(A)$, the estimator would be

$$\hat{\Gamma}_{IV}(A) = \sum r_{0t}w'_{t-1}\left(A'\sum r_{1t}w'_{t-1}\right)^{-1}$$

where either $w_{t-1} \equiv z_{t-1}$ for simple IV or $w_{t-1} \equiv A'\sum r_{1t}z'_{t-1}(\sum z_{t-1}z'_{t-1})^{-1}$ z_{t-1} for 2SLS. The instrument vector $z_{t-1} = z(y_{t-1})$ would be created from integrable transformations of the m integrated series, similar to Chang's (2002, 2006) approach. However, such an estimator suffers from several deficiencies. (a) The contemporaneous variance and long-run variance of (r_{0t}) differ, so that the sample moments $\sum r_{0t}z'_{t-1}$ and $\sum r_{0t}r'_{0t}$ (properly scaled by the sample size) have limits involving different variances. (b) The rank of $\sum r_{1t}z'_{t-1}$, which is necessary to ascertain the rank of the limiting chi-squared, is not well-known. (c) The long-run variances of individual elements of r_{0t} in the asymptotic limit of any sample moment involving z_{t-1} are unknown. The latter deficiency poses no problem in the asymptotic theory, since the function may be arbitrarily scaled. However, instrument selection involves a choice of functions, and arbitrary scaling creates a small-sample estimation problem.

These deficiencies may be overcome by a standardization and transformation of (r_{0t}) in the direction of the common stochastic trends, using its long-run variance Ξ defined above. It will shortly become clear that deficiencies (a) and (c) are remedied by this transformation. That deficiency (b) is remedied is evident in the proof of Lemma A.2 (in the appendix).

Define the $m \times (m-r)$ matrix $E \equiv (I_{m-r}, 0)'$ and note that $E'E = I_{m-r}$. The spectral decomposition of the real symmetric matrix Ξ may be denoted by $P\Lambda P'$ where $P' = P^{-1}$. The non-zero eigenvalues coincide with the singular values, since Ξ is positive semidefinite. Thus, Λ has exactly $m-r$ non-zero diagonal elements, so that $P\Lambda P' = PEE'\Lambda^{1/2}EE'\Lambda^{1/2}EE'P'$ and $(E'\Lambda E)^{-1/2}E'P'\Xi PE(E'\Lambda E)^{-1/2} = I_{m-r}$. By defining $\Xi^{+1/2} \equiv (E'\Lambda E)^{-1/2}$ $E'P'$, note that $\Xi^{+1/2}\Xi\Xi^{+1/2\prime} = I_{m-r}$. From the definition of Ξ in Lemma 1, it follows that $\Xi^{+1/2} = (\Gamma_\perp'\Sigma\Gamma_\perp)^{-1/2}\Gamma_\perp'$, so that $\Xi^{+1/2}\Sigma\Xi^{+1/2\prime} = I_{m-r}$ also.

Rewrite (1) as

$$r_{0t} = \Gamma A' r_{1t} + \Gamma_\perp A_\perp' r_{1t} + \varepsilon_t, \tag{4}$$

where the variance of $\Gamma_\perp A_\perp' r_{1t}$ is zero under the null. RR estimates m eigenvalues corresponding to both A and $A_\perp$, but only the first r eigenvalues corresponding to A are non-zero under the null.

Premultiplying by $\Xi^{+1/2}$ yields

$$\Xi^{+1/2} r_{0t} = \Xi^{+1/2}\Gamma_\perp A_\perp' r_{1t} + \Xi^{+1/2}\varepsilon_t, \tag{5}$$

since $\Xi^{+1/2}\Gamma = 0$. Since the first term has degenerate variance under the null,

$$\Xi^{+1/2} r_{0t} = \Xi^{+1/2}\varepsilon_t, \tag{6}$$

and the long-run and contemporaneous variances of both sides are clearly I_{m-r}. For notational simplicity, let $\varepsilon_t^0 \equiv \Xi^{+1/2}\varepsilon_t$ and $r_{ht}^0 \equiv \Xi^{+1/2} r_{ht}$ for $h = 0, 1$. Note that (r_{1t}^0) is a vector of the unique stochastic trends of the model.

2.4 Instrument Selection

My approach to instrument selection closely mirrors Chang's (2002, 2006) approach. She uses nonlinear IV's for a panel of integrated series, transforming each series separately using a (possibly different) nonlinear function to generate each instrument. The vector series (r_{1t}^0) of stochastic trends may be viewed

DOI: 10.2202/1941-1928.1057

as a panel, and instruments may be chosen in exactly the same fashion. Specifically, I individually transform each element $e_i' r_{1t}^0$ for $i = 1, \ldots, m - r$, so that $z(r_{1t}^0)$ represents an element-wise vector function with a vector argument – its elements are simply $z_i(e_i' r_{1t}^0)$. Letting $z_{t-1} \equiv z(r_{1t}^0)$ denote this vector, z is called the *instrument generating function* (IGF) following Chang (2002).

The following definition, due to Chang (2002), delineates the class of regularly integrable functions introduced by Park and Phillips (1999, 2001).

Definition (Chang, 2002). A transformation g on $\mathbb{R}$ is said to be *regularly integrable* if g is a bounded integrable function such that for some constants $c > 0$ and $k > 3$, $|g(x) - g(y)| \leq c|x - y|^k$ on each piece A_i of its support $A = \cup_{i=1}^{\ell} A_i \subset \mathbb{R}$.

The definition allows for functions that are not continuous but still reasonably smooth.

Letting $B(\mathbb{R})$ denote the Borel σ-field on the real line, assume that

[Z] *$z(x)$ is an $(m - r)$-vector such that for $i = 1, \ldots, m - r$,*

(a) $z_i : \mathbb{R} \mapsto \mathbb{R}$,

(b) *$z_i(x_i)$ is regularly integrable and satisfies* $\int_{-\infty}^{\infty} z_i(x_i) x_i dx_i \neq 0$, and

(c) *The inverse image under z of a set on $B(\mathbb{R})$ with Lebesgue measure zero also has Lebesgue measure zero.*

Assuming an element-by-element functional mapping that generates exactly $m - r$ instruments does not sacrifice generality, but the assumption greatly simplifies the degrees of freedom of the limiting chi-squared distribution below. The integrability assumption is identical to Chang's (2002). As she points out, this assumption avoids instrument failure, which would otherwise result from uncorrelatedness of the instrument and the corresponding regressor.

Part (c) of the assumption rules out functions with inverses that concentrate mass. The functions must induce continuous distributions. Functions of the type $x^a \exp(-|x|^b)$ for positive odd powers a and $b > 0$ are acceptable. However, variations of an indicator function of the type $1\{|x| \leq K\}$ discussed by Chang (2002) do not satisfy part (c), since they concentrate probability mass at zero and one. The moment matrix of instruments may not be invertible in that case. This restriction rules out the sign function considered by Phillips *et al.* (2004) to be the optimal instrument in the univariate case.

2.5 Preliminary Asymptotic Results

Some preliminary asymptotic results help to frame subsequent exposition.

Lemma 2 *Let assumptions* [A1]-[A2] *hold for empirical moments constructed with instruments satisfying assumption* [Z][3] *and with* (r_{0t}) *and* (r_{1t}) *satisfying* (3). *Under the null hypothesis,*

(a) $T^{-1}\sum \varepsilon_t^0 \varepsilon_t^{0\prime} \to_p I_{m-r}$,

(b) $T^{-1/2}\sum z(r_{1t}^0) z(r_{1t}^0)' \to_d dg(L_i(1,0)\int z_i^2(s)\,ds)$ for $i = 1, \ldots, m-r$,

(c) $T^{-1/4}\sum vec(z(r_{1t}^0)\varepsilon_t^{0\prime})$
$\to_d (I_{m-r} \otimes dg(L_i(1,0)\int z_i^2(s)\,ds))^{1/2}\mathbf{N}(0, I_{(m-r)^2})$, and

(d) $T^{-1}\sum z(r_{1t}^0) r_{1t}^{0\prime} = O_p(1)$

as $T \to \infty$. *The convergences in parts* (b) *and* (c) *are joint.*

The mixed normality of the result in part (c) of the lemma suggests that a procedure may be constructed to instrument out non-normality from the standard trace test, which is precisely the aim of this analysis. The diagonality of the limits in parts (a)-(c) of the lemma implies that such a procedure also instruments out cross-sectional correlation. The intuition is only implicit, because (possibly non-diagonal) Ξ is still explicitly estimated. The subsequent procedures would not be effective for high-dimensional systems (large m).

Under the null, (ε_t^0) may be replaced with (r_{0t}^0) using the relationship in (6). The standardization and transformation thus clearly remedies deficiencies (a) and (c) discussed in Section 2.3 above.

2.6 Estimated Long-Run Variance Matrix

Lemma 2 relies on known long-run variance Ξ, which must be estimated in practice. This variance may be estimated consistently using standard techniques.[4] However, the estimation error may have detrimental effects on the testing strategy due to the nonlinearity of the IGF. The concern arises from the nonlinear transformation of an integrated trend premultiplied by the error itself. Specifically, $e_i'\hat{\Xi}^{+1/2} r_{1t} = e_i'\Xi^{+1/2} r_{1t} + e_i'(\hat{\Xi}^{+1/2} - \Xi^{+1/2}) r_{1t}$, and the impact of $e_i'(\hat{\Xi}^{+1/2} - \Xi^{+1/2}) r_{1t}$ on the IGF may be non-negligible.

[3]Corollary 2.2 of Wang and Phillips (2009) suggests that regularity can be relaxed in the proof of Lemma 2(b), which may be extended beyond this result.

[4]See, for example, Andrews (1991), Hansen (1992), or de Jong and Davidson (2000).

DOI: 10.2202/1941-1928.1057

Estimation error from $(\hat{\Xi}^{+1/2} - \Xi^{+1/2})$ may lie in three important subspaces of $\mathbb{R}^m$. Error may lie in the span of $e_i'\Xi^{+1/2}$, which is one of the $(m-r)$ stochastic trends, it may lie in the $(m-r-1)$-dimensional space spanned by the stochastic trends orthogonal to $e_i'\Xi^{+1/2}$, and it may also lie in the cointegrating space. To formalize the decomposition of the estimation error, let

$$P_\Xi = \Xi^{+1/2\prime}(\Xi^{+1/2}\Xi^{+1/2\prime})^{-1}\Xi^{+1/2} = PEE'P$$

be an orthogonal projection onto the space of the trends and let

$$P_{\Xi_i} = \Xi^{+1/2\prime}e_i(e_i'\Xi^{+1/2}\Xi^{+1/2\prime}e_i)^{-1}e_i'\Xi^{+1/2} = PEe_ie_i'E'P'$$

be an orthogonal projection onto the univariate space spanned by $e_i'\Xi^{+1/2}$.

Then

$$r_{1t} = P_{\Xi_i}r_{1t} + \sum\nolimits_{j\neq i} P_{\Xi_j}r_{1t} + (I - P_\Xi)r_{1t} \tag{7}$$

decomposes r_{1t} into three terms that project the vector series (r_{1t}) in these three directions.

In order to address the asymptotic contribution of estimation error of the long-run variance, it is convenient to assume differentiability of $z(x)$. Let $z^{(k)}(x)$ denote the k^{th} derivative of $z(x)$. Assume that

[Z'] *$z(x)$ is an $(m-r)$-vector such that for $i = 1, \ldots, m-r$,*

- (a) *assumption* [Z] *holds,*
- (b) *$z_i(x_i)$ is infinitely differentiable, and*
- (c) *$x_i^k z_i^{(k)}(x_i)$ also satisfies assumption* [Z] *for any $k \geq 0$.*

Although this assumption places restrictions on the class of IGF's admissible under [Z], the class of exponential functions suggested by Chang (2002) is still admissible under assumption [Z'].

The following result gives conditions under which the asymptotics of Lemma 2 are relevant using estimated Ξ.

Lemma 3 *Let assumptions* [A1]-[A2] *hold for empirical moments constructed with instruments satisfying assumption* [Z'] *and with (r_{0t}) and (r_{1t}) defined by* (3)*. Define* $\hat{\varepsilon}_t^0 \equiv \hat{\Xi}^{+1/2}\varepsilon_t$ and $\hat{r}_{1t}^0 \equiv \hat{\Xi}^{+1/2}r_{1t}$*. Under the null hypothesis,*

(a) $\sum(\hat{\varepsilon}_t^0\hat{\varepsilon}_t^{0\prime} - \varepsilon_t^0\varepsilon_t^{0\prime}) = o_p(T)$

as $T \to \infty$. *Moreover, if either* (i) *the null hypothesis is* $m - r = 1$ *and* $\hat{\Xi}^{+1/2} - \Xi^{+1/2} = o_p(1)$, *or* (ii) *the null hypothesis is* $m - r > 1$ *and* $\hat{\Xi}^{+1/2} - \Xi^{+1/2} = o_p\left(T^{-1/2}\right)$,

(b) $\sum(z(\hat{r}^0_{1t})z(\hat{r}^0_{1t})' - z(r^0_{1t})z(r^0_{1t})') = o_p\left(T^{1/2}\right)$

(c) $\sum(z(\hat{r}^0_{1t})\hat{\varepsilon}^{0\prime}_t - z(r^0_{1t})\varepsilon^{0\prime}_t) = o_p\left(T^{1/4}\right)$, *and*

(d) $\sum(z(\hat{r}^0_{1t})\hat{r}^{0\prime}_{1t} - z(r^0_{1t})r^{0\prime}_{1t}) = o_p(T)$

as $T \to \infty$.

The stricter requirement for the case with more than one stochastic trend arises because of the sample covariance between one stochastic trend and a regularly integrable nonlinear function of a *different* stochastic trend. The non-negligibility occurs because such terms have a larger asymptotic order (due to Lemma 1 of Chang and Park, 2003) than a sample covariance involving the *same* stochastic trend.

If $m - r > 1$ but the estimation error is not $o_p\left(T^{-1/2}\right)$, the limiting distributions of Lemma 2 are not obtained. Similarly to Zivot's (2000) analysis of misspecified predetermined cointegrating vectors, it is useful to note the implications of recalcitrant estimation error, which will be discussed after the test is presented.

3. A Nonlinear IV Rank Test

In this section, I put rank testing into the perspective of testing the number of non-zero eigenvalues after the r^{th} ordered eigenvalue from (4). Under the null, there are *no* non-zero eigenvalues beyond the r^{th}. I discuss the relative merits of a 2SLS-type estimator over a simple IV estimator in this context. I then introduce the test by way of a two-stage reduced rank (2SRR) regression. Finally, the simple chi-squared limit of the NIV trace test is presented.

3.1 A Different Perspective on Rank Testing

Before IV estimation, consider a new perspective on rank testing using the transformed series (r^0_{0t}). The transformed system in (5) may be written as

$$r^0_{0t} = \Xi^{+1/2}\Gamma_\perp \begin{bmatrix} (\Gamma'_\perp \Sigma \Gamma_\perp)^{1/2\prime}(A'_\perp\Gamma_\perp)^{-1}A'_\perp A_\perp \\ (\Gamma' A)^{-1}\Gamma' A_\perp \end{bmatrix}' \begin{bmatrix} (\Gamma'_\perp \Sigma \Gamma_\perp)^{-1/2}\Gamma'_\perp r_{1t} \\ A' r_{1t} \end{bmatrix} + \varepsilon^0_t,$$

since $A_\perp(\Gamma'_\perp A_\perp)^{-1}\Gamma'_\perp + \Gamma(A'\Gamma)^{-1}A' = I$, or more simply as

$$r^0_{0t} = \Gamma^0_\perp A^{*\prime}_\perp r^*_{1t} + \varepsilon^0_t, \tag{8}$$

where $\Gamma^0_\perp \equiv \Xi^{+1/2}\Gamma_\perp$, $A^*_\perp \equiv (A^{0\prime}_\perp, A'_\perp\Gamma(A'\Gamma)^{-1})'$, $A^0_\perp \equiv (\Gamma'_\perp\Sigma\Gamma_\perp)^{1/2\prime}(A'_\perp\Gamma_\perp)^{-1} A'_\perp A_\perp$, and $r^*_{1t} \equiv (r^{0\prime}_{1t}, r'_{1t}A)'$.

Allowing for reduced rank under the null, the variance estimator is

$$\hat\Sigma(A^*_\perp) = M_{00} - M^*_{01}A^*_\perp\left(A^{*\prime}_\perp M^*_{11}A^*_\perp\right)^+ A^{*\prime}_\perp M^*_{10},$$

where $M_{00} \equiv T^{-1}\sum r^0_{0t}r^{0\prime}_{0t}$, $M^*_{11} \equiv T^{-1}\sum r^*_{1t}r^{*\prime}_{1t}$, $M^*_{01} \equiv T^{-1}\sum r^0_{0t}r^{*\prime}_{1t}$, $M^*_{10} \equiv M^{*\prime}_{01}$, and B^+ denotes the Moore-Penrose inverse of a matrix B. Alternatively,

$$\hat\Sigma(A^*_\perp) = R'_0(I - R_1A^*_\perp(A^{*\prime}_\perp R'_1R_1A^*_\perp)^+A^{*\prime}_\perp R'_1)R_0$$

by defining $R_1 \equiv T^{-1/2}(r^{*\prime}_{1t})^T_{t=2}$ and $R_0 \equiv T^{-1/2}(r^{0\prime}_{0t})^T_{t=2}$ to be $(T-1)\times(m-r)$ matrices. Basic matrix results (e.g., Lütkepohl, 1996, pg. 49) allow

$$\begin{aligned}&|R'_0(I - R_1A^*_\perp(A^{*\prime}_\perp R'_1R_1A^*_\perp)^+A^{*\prime}_\perp R'_1)R_0|\\ &= \frac{|R'_0R_0|}{|A^{*\prime}_\perp R'_1R_1A^*_\perp|}|A^{*\prime}_\perp R'_1(I - R_0(R'_0R_0)^{-1}R'_0)R_1A^*_\perp|\end{aligned}$$

so that

$$\begin{aligned}L^{-2/n}_{\max} &= |M_{00} - M^*_{01}A^*_\perp(A^{*\prime}_\perp M^*_{11}A^*_\perp)^+A^{*\prime}_\perp M^*_{10}|\\ &= |M_{00}||A^{*\prime}_\perp(M^*_{11} - M^*_{10}M^{-1}_{00}M^*_{01})A^*_\perp|/|A^{*\prime}_\perp M^*_{11}A^*_\perp|\end{aligned}$$

which is very similar to the reduced rank case.

In place of (1), (5), or (8), I consider

$$r^0_{0t} = \Gamma^0_\perp A^{0\prime}_\perp r^0_{1t} + \varepsilon^0_t \tag{9}$$

in order to construct the test. As in (5) and (8), the first term has degenerate variance under the null, so that (6) holds. This procedure replaces $S^{-1}_{11}S_{10}S^{-1}_{00}S_{01}$ or $(M^*_{11})^{-1}M^*_{10}M^{-1}_{00}M^*_{01}$ with $M^{-1}_{11}M_{10}M^{-1}_{00}M_{01}$, where $M_{11} \equiv T^{-1}\sum r^0_{1t}r^{0\prime}_{1t}$, $M_{01} \equiv T^{-1}\sum r^0_{0t}r^{0\prime}_{1t}$, and $M_{10} \equiv M'_{01}$. The result yields $m-r$ rather than m eigenvalues, which are all zeros under the null.

3.2 IV Estimation

Nonlinear instruments are introduced to alleviate non-normality in the limiting distributions of rank tests based on the above procedures. Given the testing framework just introduced, a natural estimator of $\Gamma^0_\perp$ is

$$\hat\Gamma^0_{\perp,IV}(A^0_\perp) = \sum r^0_{0t}w'_{t-1}\left(A^{0\prime}_\perp\sum r^0_{1t}w'_{t-1}\right)^{-1}$$

where either $w_{t-1} \equiv z_{t-1}$ for simple IV or $w_{t-1} \equiv A_\perp^{0\prime} \sum r_{1t}^0 z_{t-1}' (\sum z_{t-1} z_{t-1}')^{-1} z_{t-1}$ for 2SLS. The estimator may contain the same number of of instruments as regressors without loss of generality, so these are identical estimators of $\Gamma_\perp^0$ under reasonable rank assumptions. For hypothesis testing, however, 2SLS has two advantages over simple IV. (a) The orthogonal projection in the 2SLS estimator is convenient, because it allows a simple quadratic form that yields the chi-squared distribution below. (b) Power of the test converges more rapidly to unity in the 2SLS case than in the simple IV case, because the variance of only the second stage is used.

The second point is illustrated by considering a parameter matrix $\Upsilon_\perp^0 = \Gamma_\perp^0 A_\perp^{0\prime}$ (*cf.* Johansen 1995, pg. 94) for the transformed model in (9). In the RR case, the variance estimator is

$$\hat{\Sigma}(\Upsilon_\perp^0) = M_{00} - M_{01} M_{11}^{-1} M_{10}.$$

Under the null, this is simply M_{00}, but the distribution of the second term collapses at the rate of T^{-1} because of the linear transformation in the direction of the trends. In this direction, $M_{01}, M_{10} = O_p(1)$ and $M_{11} = O_p(T)$. Under the alternative, both terms converge to a stable limit distribution. The null and alternative variances thus diverge at the rate of T.

Consider now the simple IV case. The variance estimator is

$$\begin{aligned}\hat{\Sigma}(\Upsilon_\perp^0) = {} & M_{00} + M_{01}^{rz}(M_{11}^{rz})^{-1} M_{11} (M_{11}^{zr})^{-1} M_{10}^{zr} \\ & - M_{01}(M_{11}^{zr})^{-1} M_{10}^{zr} - M_{01}^{rz}(M_{11}^{rz})^{-1} M_{10}\end{aligned}$$

using the notation $M_{h1}^{rz} \equiv T^{-1} \sum r_{ht}^0 z_{t-1}'$ for $h = 0, 1$, and $M_{1h}^{zr} \equiv M_{h1}^{rz\prime}$. Under the null, using Lemma 2,

$$\begin{aligned}\hat{\Sigma}(\Upsilon_\perp^0) = {} & M_{00} + T^{-1/2}(T^{3/4} M_{01}^{rz}(M_{11}^{rz})^{-1} T^{-1} M_{11} (M_{11}^{zr})^{-1} T^{3/4} M_{10}^{zr}) \\ & - T^{-3/4}(M_{01}(M_{11}^{zr})^{-1} T^{3/4} M_{10}^{zr}) - T^{-3/4}(T^{3/4} M_{01}^{rz}(M_{11}^{rz})^{-1} M_{10}),\end{aligned}$$

so that terms after the first are $O_p\left(T^{-1/2}\right)$. The null and alternative variances thus diverge at the slower rate of $T^{1/2}$. The simple IV test is therefore less powerful.

As shown below, a 2SLS-type estimator restores the rate of T.

3.3 Two-Stage Reduced Rank (2SRR) Regression

The second stage of a 2SLS regression is a linear regression onto the space of the regressors projected onto the instrument space. In the VECM case, the second-stage regression may be written as

$$r_{0t}^0 = \Gamma_\perp^0 A_\perp^{0\prime} \sum r_{1t}^0 z_{t-1}' (\sum z_{t-1} z_{t-1}')^{-1} z_{t-1} + \varepsilon_t^0,$$

DOI: 10.2202/1941-1928.1057

or more succinctly as

$$r^0_{0t} = \Gamma^0_\perp A^{0\prime}_\perp M^{rz}_{11}(M^{zz}_{11})^{-1} z_{t-1} + \varepsilon^0_t \tag{10}$$

using the notation $M^{zz}_{11} \equiv T^{-1} \sum z_{t-1} z'_{t-1}$.

The 2SLS estimator is

$$\Gamma(A^0_\perp) = Q_{01} A^0_\perp (A^{0\prime}_\perp Q_{11} A^0_\perp)^+$$

with variance estimator in the *second* stage regression given by

$$\hat{\Sigma}(A^0_\perp) = M_{00} - Q_{01} A^0_\perp (A^{0\prime}_\perp Q_{11} A^0_\perp)^+ A^{0\prime}_\perp Q_{10},$$

where $Q_{gh} \equiv M^{rz}_{g1}(M^{zz}_{11})^{-1} M^{zr}_{1h}$ for $g, h = 0, 1$. The likelihood at the second stage is

$$\begin{aligned} L^{-2/n}_{\max} &= |M_{00} - Q_{01} A^0_\perp (A^{0\prime}_\perp Q_{11} A^0_\perp)^+ A^{0\prime}_\perp Q_{10}| \\ &= |M_{00}||A^{0\prime}_\perp (Q_{11} - Q_{10} M^{-1}_{00} Q_{01}) A^0_\perp| / |A^{0\prime}_\perp Q_{11} A^0_\perp|, \end{aligned}$$

using similar arguments as above.

Let $\hat{\lambda}_{NIV,i}$ for $i = 1, \ldots, m - r$ refer to the $m - r$ eigenvalues of $Q_{11} - Q_{10} M^{-1}_{00} Q_{01}$ subject to $A^{0\prime}_\perp Q_{11} A^0_\perp = I$. These may be estimated by solving $|\lambda I - Q^{-1}_{11} Q_{10} M^{-1}_{00} Q_{01}| = 0$. In other words, running a reduced rank regression on (10) instead of on (1) or (9) is a two-stage reduced rank (2SRR) regression. These eigenvalues are all zeros under the null.

Finally, note that the variance estimator may be written as

$$\hat{\Sigma}(\Upsilon^0_\perp) = M_{00} - Q_{01} Q^{-1}_{11} Q_{10}$$

where $Q_{01}, Q_{10} = O_p\left(T^{-1/4}\right)$ and $Q_{11} = O_p(T^{1/2})$ by Lemma 2 under the null, so that $Q_{01} Q^{-1}_{11} Q_{10} = O_p(T^{-1})$. Like the RR case, but unlike the simple IV case, the rate of divergence between the null and alternative is T.

The reason for the improvement of 2SRR over simple IV (even though the 2SLS and IV estimators of $\Gamma^0_\perp$ are the same when the number of instruments equals the number of regressors) is that the variance estimator for the second stage is not the same as the variance estimator for the original model.

3.4 Proposed Rank Test

Johansen (1995) details a battery of tests that may be run for various types of restrictions on the cointegrating vectors in A. While many of these tests have chi-squared distributions, the most important of these – tests for the cointegrating rank – have nonstandard distributions.

The well-known trace test for the cointegrating rank of a VECM is derived as a likelihood ratio test. The null is $H_0 : r = r_0$, and the alternative is $H_A : r = m$ (stationary, no cointegration). The likelihoods under the null and alternative are given by $L_{\max}^{-2/n} = |S_{00}|\,|A'(S_{11} - S_{10}S_{00}^{-1}S_{01})A|$. The second determinant equals the product of $(1 - \lambda_{RR,i})$ corresponding to the first r_0 eigenvalues under the null, or it equals the product corresponding to all m eigenvalues under the alternative. The familiar trace test is therefore

$$-2\log Q_{RR}(H_{r_0|m}) = -T\sum_{i=r_0+1}^{m}\log(1 - \hat{\lambda}_{RR,i})$$

since the common factor $|S_{00}|$ cancels.

For the NIV test, the likelihood is given simply by $L_{\max}^{-2/n} = |M_{00}|$ under the null, because (9) reduces to (6). Under the alternative that $r = m$, the likelihood is given by $L_{\max}^{-2/n} = |M_{00}||A_{\perp}^{0\prime}(Q_{11} - Q_{10}M_{00}^{-1}Q_{01})A_{\perp}^{0}|$, so that the common factor $|M_{00}|$ cancels, and

$$-2\log Q_{NIV}(H_{r_0|m}) = -T\sum_{i=1}^{m-r_0}\log(1 - \hat{\lambda}_{NIV,i}) \tag{11}$$

is the analog to the standard trace test. Note that the summation is across all of the $m - r_0$ eigenvalues estimated by 2SRR. I refer to the proposed test as the *NIV trace test*.

3.5 Limiting Distribution of the Test Statistic

It is straightforward to show that the NIV trace statistic is $T\sum_{i=1}^{m-r_0}\lambda_{NIV,i}$ up to an asymptotically negligible term, using a Taylor-series expansion of $\lambda_{NIV,i}$ around zero. Letting $U_T \equiv (M_{11}^{zz})^{-1/2}M_{11}^{zr}(M_{11}^{rz}(M_{11}^{zz})^{-1}M_{11}^{zr})^{-1/2\prime}$ allows

$$T\sum_{i=1}^{m-r_0}\lambda_{NIV,i} = T\ tr\left\{(M_{11}^{zz})^{-1/2\prime}U_TU_T'(M_{11}^{zz})^{-1/2}M_{10}^{zr}M_{00}^{-1}M_{01}^{rz}\right\}$$

by expanding Q_{11}, etc., and using the equality of the sum of the eigenvalues and the trace of a matrix. Using further properties of the trace,

$$T\sum_{i=1}^{m-r_0}\lambda_{NIV,i} = \upsilon_T'(I_{m-r}\otimes U_TU_T')\upsilon_T \tag{12}$$

by defining $\upsilon_T \equiv (M_{00}\otimes T^{1/2}M_{11}^{zz})^{-1/2}vec(T^{3/4}M_{10}^{zr})$.

As I show in the following theorem, the quadratic form in (12), and therefore the trace test statistic in (11), has a limiting chi-squared distribution.

DOI: 10.2202/1941-1928.1057

Theorem 4 *Let the conditions for Lemma 2 or 3 hold for the model in* (10) *with* (r_{0t}) *and* (r_{1t}) *defined by* (3). *The LR test of* $H_0 : r = r_0$ *against* $H_A : r = m$ *has a limiting distribution given by*

$$-2\log Q_{NIV}(H_{r_0|m}) \to_d \chi^2_{(m-r_0)^2}$$

as $T \to \infty$.

Convenient critical values of the chi-squared limiting distribution replace the nonstandard critical values of the usual trace test.

The implications of $m-r > 1$ and $\hat{\Xi}^{+1/2} - \Xi^{+1/2} = o_p(1)$ but not $o_p(T^{-1/2})$ may now be considered. In this case, from the proof of Lemma 3, the asymptotically dominant terms of the k^{th}-order term of expansions of the nonlinear functions in M_{11}^{zz} and M_{10}^{zr} are

$$\sum\nolimits_{j\neq i} e_i'(\hat{\Xi}^{+1/2} - \Xi^{+1/2})\Xi^{+1/2\prime} e_j (e_j'\Xi^{+1/2}\Xi^{+1/2\prime}e_j)^{-1} O_p(T^{1/2+k/2})$$

and

$$\sum\nolimits_{j\neq i} e_i'(\hat{\Xi}^{+1/2} - \Xi^{+1/2})\Xi^{+1/2\prime} e_j (e_j'\Xi^{+1/2}\Xi^{+1/2\prime}e_j)^{-1} O_p(T^{1/4+k/2})$$

respectively. Thus

$$\upsilon_T = T^{k/4}(M_{00} \otimes T^{1/2-k/2}M_{11}^{zz})^{-1/2} vec(T^{3/4-k/2}M_{10}^{zr})$$

so that the quadratic in (12) is explosive. The test should over-reject.

4. Extension: Mean and Covariates

Extending the rank test from the model in (1) to that in (2) requires additional steps. The first step of the standard ML procedure of a VECM along the lines of Johansen (1988) is to use residuals from regressing out $(1, w_t')'$ from both $(\triangle y_t)$ and (y_{t-1}) in order to focus on the term $\Gamma A' y_{t-1}$. If A were known, this would be exactly the first step of a standard partitioned regression to estimate Γ with unknown μ and Π.

More care must be used with the NIV strategy, however, because the nonlinear nonstationary asymptotics require the argument of the IGF to have the martingale property – at least up to a negligible term if the IGF is sufficiently smooth. The NIV framework therefore necessitates a mixed approach such that $(\triangle y_t)$ and (y_{t-1}) are handled differently.

4.1 Handling $(\triangle y_t)$

The main appeal of regressing out $(1, w_t')'$ from $(\triangle y_t)$ is to orthogonalize the regressand. This may be accomplished by subtracting $A_\perp(\Gamma_\perp' A_\perp)^{-1}\Gamma_\perp'(\mu + \Pi w_t)$ from both sides, a transformation of the mean and covariates in the direction of the stochastic trends. The modified model becomes

$$r_{0t} = \Gamma A' y_{t-1} + \eta_t$$

where

$$\begin{aligned} r_{0t} &\equiv \triangle y_t - A_\perp(\Gamma_\perp' A_\perp)^{-1}\Gamma_\perp'(\mu + \Pi w_t), \text{ and} \\ \eta_t &\equiv \varepsilon_t + \Gamma(A'\Gamma)^{-1}A'(\mu + \Pi w_t). \end{aligned} \tag{13}$$

Note that (6) holds under the null, as in the case with no additional regressors.

Consistently estimating $A_\perp(\Gamma_\perp' A_\perp)^{-1}\Gamma_\perp'(\mu + \Pi w_t)$ is straightforward, so that μ and Π need not be known in practice.

Lemma 5 The matrix $A_\perp(\Gamma_\perp' A_\perp)^{-1}\Gamma_\perp'(\mu, \Pi)$ may be $\sqrt{T}$-consistently estimated by regressing $(\triangle y_t)$ onto $(1, w_t')'$.

The lemma confirms the intuitive appeal of using residuals from this regression, so that $(\triangle y_t)$ may be handled in exactly the same way as in the standard Johansen ML procedure.

4.2 Handling (y_{t-1})

On the other hand, (y_{t-1}) must be handled differently from the standard ML procedure. Rather than regressing (y_{t-1}) onto $(1, w_t')'$, there are three differences: (i) (w_t) may be ignored, (ii) linear detrending using $(1, t)'$ rather than demeaning must be used, and (iii) the detrending must be adaptive.

Chang (2002) used asymptotic arguments along the lines of Phillips and Solo (1992) to suggest that (y_{t-1}) need not be regressed onto lagged $(\triangle y_t)$ in the univariate unit root case. Chang's arguments extend to the multivariate case. The same logic may be applied to additional stationary covariates in (w_t), as long as they have moving average representations with absolutely summable coefficients.

In order for the argument of the IGF to retain the martingale property, Chang suggested adaptively detrending the argument of the IGF. In the case of a VECM, the mean in differences becomes a mean and linear trend in levels, so $(1, t)'$ should be used in adaptively detrending (y_{t-1}), rather than simply

DOI: 10.2202/1941-1928.1057

demeaning. Specifically, letting $d_t \equiv (1,t)'$ and $d_{Tt} = \kappa_T d_t$ with $\kappa_T = dg(1,T)$, adaptively detrended (y_{t-1}) is given by

$$r_{1t} \equiv y_{t-1} - \sum\nolimits_{i=1}^{t-1} y_i d_{Ti}' \left(\sum\nolimits_{i=1}^{t-1} d_{Ti} d_{Ti}'\right)^{-1} d_{T,t-1}, \tag{14}$$

which Chang (2002) showed converges to detrended Brownian motion when properly normalized.

4.3 Modified Model and Results

Using this detrending strategy, the model in (2) may be rewritten as

$$r_{0t} = \Gamma A' r_{1t} + \eta_t^*, \tag{15}$$

where (r_{0t}) and (r_{1t}) are defined by (13) and (14) and

$$\eta_t^* \equiv \varepsilon_t + \Gamma(A'\Gamma)^{-1}A'(\mu + \Pi w_t) + \Gamma \sum\nolimits_{i=1}^{t-1} A' y_i d_{Ti}' \left(\sum\nolimits_{i=1}^{t-1} d_{Ti} d_{Ti}'\right)^{-1} d_{T,t-1}.$$

Note *again* that (6) holds under the null.

In this case, the NIV model to be estimated is

$$r_{0t}^0 = \Gamma_\perp^0 A_\perp^{0\prime} M_{11}^{rz} (M_{11}^{zz})^{-1} z_{t-1} + \eta_t^{*0} \tag{16}$$

with $\eta_t^{*0} \equiv \Xi^{+1/2}\eta_t^*$ in place of (ε_t^0) in (10). Lemmas 2 and 3 are no longer directly useful. However, using Chang's (2002) asymptotic arguments, the results of parts (b) and (d) extend to the adaptively detrended case. The following lemma allows the extension of the results of part (a) and (c) of those lemmas to the model in (16).

Lemma 6 *Define $\hat{\eta}_t^{*0} \equiv \hat{\Xi}^{+1/2}\eta_t^*$ with (r_{0t}) and (r_{1t}) defined by* (13) *and* (14).

(a) $\sum(\hat{\eta}_t^{*0}\hat{\eta}_t^{*0\prime} - \hat{\varepsilon}_t^0\hat{\varepsilon}_t^{0\prime}) = o_p(T)$

(b) $\sum z(r_{1t}^0)(\hat{\eta}_t^{*0\prime} - \hat{\varepsilon}_t^{0\prime}) = o_p(T^{1/4})$

as $T \to \infty$.

The following theorem replaces Theorem 4 for the more general model with non-zero mean and stationary covariates.

Theorem 7 *Let the conditions for Lemma 2 or 3 hold for the model in* (16) *with* (r_{0t}) *and* (r_{1t}) *defined by* (13) *and* (14)*. The LR test of* $H_0 : r = r_0$ *against* $H_A : r = m$ *has a limiting distribution given by*

$$-2 \log Q_{NIV}(H_{r_0|m}) \to_d \chi^2_{(m-r_0)^2}$$

as $T \to \infty$.

The resulting distribution is exactly the same as in the case of the simpler model.

5. Extension: Panel VECM

In a single VECM, nonlinear instruments may be used to instrument out non-normality. Correlation across the vector are explicitly estimated, but implicitly instrumented out by the diagonality of the results of Lemma 2(a)-(c). The advantage of instrumenting out correlation becomes more apparent in a panel with a potentially large cross-sectional dimension N. In large-T and large-N panels, normal limiting distributions are the norm rather than the exception. In this case, the main advantage of nonlinear instruments is robustness to cross-sectional correlation.

Let the panel model be denoted by (1) with $\Gamma \equiv dg(\Gamma_n)$ and $A \equiv dg(A_n)$ for $n = 1, \ldots, N$. For simplicity, assume that $rk(\Gamma_n A_n') = r$ for all n under the null, so that $rk(\Gamma A') = Nr$. The system may be written as

$$\begin{bmatrix} \triangle y_{t1} \\ \vdots \\ \triangle y_{tN} \end{bmatrix} = \begin{bmatrix} \Gamma_1 A_1' & & 0 \\ & \ddots & \\ 0 & & \Gamma_N A_N' \end{bmatrix} \begin{bmatrix} y_{t-1,1} \\ \vdots \\ y_{t-1,N} \end{bmatrix} + \begin{bmatrix} \varepsilon_{t1} \\ \vdots \\ \varepsilon_{tN} \end{bmatrix}$$

using this convention. Assuming zero cross-sectional correlation in the mean (block-diagonality of $\Gamma A'$) is a common feature of such models for parsimony (e.g., Groen and Kleibergen, 2003). All cross-sectional correlation is therefore relegated to the variance Σ of (ε_t). No restrictions on the off-diagonal blocks of this variance are assumed. The nonlinear IV procedure instruments out this cross-sectional correlation, similarly to the panel unit root tests of Chang (2002, 2006).

In principle, a test may be constructed by simply running a trace test or NIV trace test on the entire system. However, this requires estimating the variance (ε_t) for the entire system, which is infeasible for large N.

It is more practical to estimate the individual variances of (ε_{tn}), thus calculating a rank test for each cross-sectional unit. Larsson *et al.* (2001) suggested this approach for the standard trace test (the LR bar test). An NIV trace test may be similarly performed on each cross-sectional unit. After the transformation in the first step, all of the diagonal blocks of the asymptotic variances are themselves diagonal. It is not obvious, however, that the off-diagonal blocks may be ignored before the transformation. In other words, it is not obvious that only the diagonal blocks $\Xi_{11}, \ldots, \Xi_{NN}$ of Ξ must be estimated.

To see this, define $\varepsilon^0_{tn} \equiv \Xi^{+1/2}_{nn}\varepsilon_{tn}$ and $r^0_{htn} \equiv \Xi^{+1/2}_{nn} r_{htn}$ for $h = 0, 1$ and $n = 1, \ldots, N$, where $\Xi^{+1/2}_{nn}$ is defined as above for the n^{th} diagonal block of Ξ. Redefine $\varepsilon^0_t \equiv (\varepsilon^{0\prime}_{t1}, \ldots, \varepsilon^{0\prime}_{tN})'$ and $r^0_{ht} \equiv (r^{0\prime}_{ht1}, \ldots, r^{0\prime}_{htN})'$. The series (ε^0_t) and (r^0_{ht}) are thus defined using only the diagonal blocks Ξ_{nn} of Ξ. Only these blocks must be estimated. Using these definitions,

$$
\begin{aligned}
\Omega &\equiv \mathbf{var}(\varepsilon^0_t \varepsilon^{0\prime}_t) \\
&= \begin{bmatrix} \Xi^{+1/2}_{11} & & 0 \\ & \ddots & \\ 0 & & \Xi^{+1/2}_{NN} \end{bmatrix} \begin{bmatrix} \Sigma_{11} & \cdots & \Sigma_{1N} \\ \vdots & \ddots & \vdots \\ \Sigma_{N1} & \cdots & \Sigma_{NN} \end{bmatrix} \begin{bmatrix} \Xi^{+1/2\prime}_{11} & & 0 \\ & \ddots & \\ 0 & & \Xi^{+1/2\prime}_{NN} \end{bmatrix}
\end{aligned}
$$

which by construction has diagonal blocks of I_{m-r}.

An NIV test of the whole panel is

$$
-2 \log Q_{NIV}(H_{r_0|m}) = -T \sum\nolimits_{i=1}^{N(m-r_0)} \log(1 - \hat{\lambda}_{NIV,i})
$$

for $i = 1, \ldots, N(m - r_0)$, but this requires calculating particularly large covariance matrices as the cross-sectional dimension grows. Similarly to the non-panel case, the test is equal to $T \sum_{i=1}^{N(m-r_0)} \hat{\lambda}_{NIV,i}$ up to an asymptotically negligible term. By properties of eigenvalues, this sum is equal to the trace of a matrix with those eigenvalues, which is equal to the trace of another matrix with the same diagonals. Specifically,

$$
T\ tr\left\{(M^{zz}_{11})^{-1/2\prime} U_T U'_T (M^{zz}_{11})^{-1/2} M^{zr}_{10} M^{-1}_{00} M^{rz}_{01}\right\}
$$

for the whole panel is equivalent to

$$
\sum\nolimits_{n=1}^{N} T\ tr \left\{ \begin{array}{c} E'_n (M^{zz}_{11})^{-1/2\prime} U_T U'_T (M^{zz}_{11})^{-1/2} M^{zr}_{10} \Omega^{-1/2\prime} \\ \times \Omega^{1/2\prime} M^{-1}_{00} \Omega^{1/2} \Omega^{-1/2} M^{rz}_{01} E_n \end{array} \right\},
$$

where E_n is the $Nm \times m$ matrix that selects the n^{th} set of m columns of the matrix preceding it.

Since all of the factors M^{zz}_{11}, $U_T U'_T$, $M^{zr}_{10}\Omega^{-1/2\prime}$, and $\Omega^{1/2\prime} M^{-1}_{00} \Omega^{1/2}$ are asymptotically block diagonal, looking at the diagonal blocks of the whole expression inside the trace is asymptotically equivalent to looking at the same expression created from diagonal blocks of each factor. The latter approach may be denoted by

$$\sum\nolimits_{n=1}^{N} T\ tr\left\{(M_{11,n}^{zz})^{-1/2\prime}U_{T,n}U_{T,n}^{\prime}(M_{11,n}^{zz})^{-1/2}M_{10,n}^{zr}M_{00,n}^{-1}M_{01,n}^{rz}\right\} \tag{17}$$

where $M_{11,n}^{zz} \equiv E_n^{\prime}M_{11}^{zz}E_n$, $M_{10,n}^{zr} \equiv E_n^{\prime}M_{10}^{zr}E_n$, and $M_{00,n}^{-1} \equiv E_n^{\prime}M_{00}^{-1}E_n$. The asymptotic equality of these expressions holds because the limits of $E_n^{\prime}M_{10}^{zr}\Omega^{-1/2\prime}E_n$ and $M_{10,n}^{zr}$ coincide, as do those of $E_n^{\prime}\Omega^{1/2\prime}M_{00}^{-1}\Omega^{1/2}E_n$ and $M_{00,n}^{-1}$, by the fact that Ω has identity diagonal blocks.

The expression in (17) is simply the sum of the traces of the individual blocks (cross-sectional units). The test statistic is therefore equivalent to

$$-2\log\tilde{Q}_{NIV}(H_{r_0|m}) = -\sum\nolimits_{n=1}^{N} T\sum\nolimits_{i=1}^{(m-r_0)}\log(1-\hat{\lambda}_{NIV,i,n}) \tag{18}$$

which is easier to compute in practice, as it does not require computing any covariance matrices larger than $m \times m$. I refer to the test in (18) as the *panel NIV trace test*. Clearly, the panel NIV trace test reduces to the NIV trace test when $N = 1$.

Theorem 8 *Let the conditions for Lemma 2 or 3 hold for the model in* (10) *with* $n = 1, \ldots, N$ *and with notation as defined in Section 5. The LR test of* $H_0 : r = r_0$ *against* $H_A : r = m$ *has a limiting distribution given by*

$$-2\log\tilde{Q}_{NIV}(H_{r_0|m}) \to_d \chi^2_{N(m-r_0)^2}$$

as $T \to \infty$ *for fixed* N.

Any cross-sectional correlation is effectively eliminated by the nonlinear IV testing strategy, just as in the case of panel unit root tests (Chang, 2002, 2006).

In the case without instruments, the simple trace statistic may still have the stochastic integral limit derived by Johansen (1988), but with a large number of degrees of freedom $N(m-r_0)$ from the panel dimension. This distribution has been tabulated up to 11 degrees of freedom by Osterwald-Lenum (1992) and 12 degrees of freedom by Johansen (1995), but $N(m-r_0) \leq 12$ is unrealistic for most panels.

For large N, an approach similar to Chang (2002, 2006) or Larsson *et al.* (2001) is reasonable. Letting $Q_n(H_{r_0}|H_m)$ denote the test for one cross-sectional unit n, tests of this type take the form

$$\bar{Q}(H_{r_0|m}) = \sqrt{N}\frac{\frac{1}{N}\sum_{n=1}^{N} Q_n(H_{r_0|m}) - \mathbf{E}(Q_n(H_{r_0|m}))}{\sqrt{\mathbf{var}(Q_n(H_{r_0|m}))}} \tag{19}$$

DOI: 10.2202/1941-1928.1057

for large N. A central limit theorem allows the distribution of this statistic to approximate a standard normal. Larsson *et al.* (2001) called this test the LR bar test when Q_n is $Q_{RR,n}$.

I consider only the test in (18) and not an analogous test of the type in (19). A principal advantage of a test of the latter type lies in replacing the non-standard distribution with a normal distribution for large N. However, this approach relies on a CLT approximation. The test proposed in this paper may be extended in the same way for large N, but such an extension is unnecessary. Similarly to the case without instruments, a straightforward extension of the non-panel test in (11) to the panel test in (18) necessitates finding the critical value from a distribution with a large number of degrees of freedom. However, since this limiting distribution is a chi-squared rather than a non-standard stochastic integral, critical values for large degrees of freedom may be ascertained easily.

6. Small-Sample Results

Empirical size and power from Monte Carlo experiments are presented Tables 1-3, using the usual LR trace test of Johansen (1988), the LR bar test of Larsson *et al.* (2001), and the NIV trace and panel NIV trace tests with consistently estimated long-run variances, conducted on the model in (1).

The experiments used sample sizes of $T = (24, 60, 120, 360, 600)$ and $N = (1, 5, 10, 25, 50, 100)$ with pseudo-true values of $r = 0, \ldots, m$ for $m = 2$. For simulations, I chose parameters to mimic macroeconomic data while satisfying assumptions [A1] and [A2]. Specifically, I set Σ to be an $Nm \times Nm$ matrix with ones on the diagonal and 0.9 elsewhere.[5] I set $A_n = (I_r, -A_{m-r})'$ for $n = 1, \ldots, N$ with A_{m-r} an $r \times (m-r)$ matrix with all elements equal to $(r-m)^{-1}$. And, I set $\Gamma_n = (\Gamma_r, \Gamma_{m-r})'$ for $n = 1, \ldots, N$ with $\Gamma_r = \delta_2 \iota\iota' - (\delta_1 + \delta_2) I_r$ (ι denotes a vector of ones) and with Γ_{m-r} an $r \times (m-r)$ matrix with all elements equal to $\delta_1 (m-r)/r$. I set $\delta_1 = 0.1$ and $\delta_2 = 0.05$. This experimental design generates time series with at least one root of $|I - (I + \Gamma A')\, x| = 0$ (in each block) outside the unit circle. Some results are dissimilar to those of Larsson *et al.* (2001), because the pseudo-true parameters they employed appear to contain some explosive roots.

For each of these specifications, I conducted $5,000$ repetitions. I employed the usual strategy of starting with $r_0 = 0$ and increasing until rejection fails or until $m - 1$, in which case $r = m$ is chosen if all $r_0 < m$ are rejected. For

[5] I also tried the same set of simulations with off-diagonals geometrically decreasing by 0.9, with qualitatively similar results.

the standard trace test, I used the critical values given by Osterwald-Lenum (1992), with a nominal size of 0.05.

After extensive experimentation with different instrument types, I found

$$z_i\left(x_i\right) = x_i \exp(-T^{-1/2} x_i^2)$$

(in the class suggested by Chang, 2002) to work reasonably well in small samples. Since the argument of the function is normalized by the long-run variance in the NIV procedure, using this function is robust to changes in the long-run variance of actual data.

When the true rank is full ($r = 2$), all of the tests enjoy high power. Since all of the tests are designed with an alternative of full rank, good power is expected.

Similarly, when $r = 0$, the first test of $r_0 = 0$ usually fails to reject, because there is the most contrast between the two hypotheses of zero rank and full rank. In the non-panel case, the LR test does not have very much size distortion even for T as small as 24. The size distortion of the LR bar and NIV trace tests are not unreasonable for $N = 1$. As N increases, the size increases for all of the panel tests, but is especially bad for the NIV tests with relatively small T. A formal requirement of Larsson *et al.* (2001) is that $\sqrt{N}/T \to 0$, which suggests that T should be much larger than $\sqrt{N}$ in practice. Similarly, the result for the NIV panel test is valid for finite N. Clearly, all tests perform poorly when N is too large relative to T, such as $N = 100$ and $T = 24$.

The reason why the NIV trace test performs worse in this situation is because the rate of divergence $T^{1/4}$ in Lemma 2(c) translates into a relatively slow rate of convergence to the chi-squared. This slowness is the price paid for (mixed) normality. The NIV trace test should not be used in samples with $T < 120$ if $N > 1$.

In contrast, when $r = 1$, size distortion is extreme for all of the tests when $T \leq 120$. The size improves substantially for all tests as T increases, and the NIV trace test is competitive for $T = 360$ and 600 for all N. As in the case of $r = 0$, the size distortion with both panel tests increases as N increases.

Overall, the NIV trace test appears to be competitive with Johansen's trace test when $N = 1$ even when T is as small as 24. In a panel, the NIV trace test is competitive with the LR bar test for a relatively large time dimension, say $T \geq 120$. Even though an "off-label" use of the LR bar test, the LR bar test outperforms the NIV trace test for N substantially larger than T.

DOI: 10.2202/1941-1928.1057

			LR (Trace)						LR Bar					NIV				
	T	N	$r_0 = 0$	$r_0 = 1$	$r_0 = 2$	H(0)	H(1)	Size	$r_0 = 0$	$r_0 = 1$	$r_0 = 2$	H(0)	H(1)	$r_0 = 0$	$r_0 = 1$	$r_0 = 2$	H(0)	H(1)
r = 2	24	1	0.000	0.001	**0.999** *	1.000	0.999	**0.001**	0.000	0.001	**0.999** *	1.000	0.999	0.002	0.009	**0.989** *	0.998	0.991
	24	5							0.000	0.000	**1.000** *	1.000	1.000	0.000	0.001	**0.999** *	1.000	0.999
	24	10							0.000	0.000	**1.000** *	1.000	1.000	0.000	0.000	**1.000** *	1.000	1.000
	24	25							0.000	0.000	**1.000** *	1.000	1.000	0.000	0.000	**1.000** *	1.000	1.000
	24	50							0.000	0.000	**1.000** *	1.000	1.000	0.000	0.000	**1.000** *	1.000	1.000
	24	100							0.000	0.000	**1.000** *	1.000	1.000	0.000	0.000	**1.000** *	1.000	1.000
	60	1	0.000	0.000	**1.000** *	1.000	1.000	**0.000**	0.000	0.000	**1.000** *	1.000	1.000	0.000	0.000	**1.000** *	1.000	1.000
	60	5							0.000	0.000	**1.000** *	1.000	1.000	0.000	0.000	**1.000** *	1.000	1.000
	60	10							0.000	0.000	**1.000** *	1.000	1.000	0.000	0.000	**1.000** *	1.000	1.000
	60	25							0.000	0.000	**1.000** *	1.000	1.000	0.000	0.000	**1.000** *	1.000	1.000
	60	50							0.000	0.000	**1.000** *	1.000	1.000	0.000	0.000	**1.000** *	1.000	1.000
	60	100							0.000	0.000	**1.000** *	1.000	1.000	0.000	0.000	**1.000** *	1.000	1.000
	120	1	0.000	0.000	**1.000** *	1.000	1.000	**0.000**	0.000	0.000	**1.000** *	1.000	1.000	0.000	0.000	**1.000** *	1.000	1.000
	120	5							0.000	0.000	**1.000** *	1.000	1.000	0.000	0.000	**1.000** *	1.000	1.000
	120	10							0.000	0.000	**1.000** *	1.000	1.000	0.000	0.000	**1.000** *	1.000	1.000
	120	25							0.000	0.000	**1.000** *	1.000	1.000	0.000	0.000	**1.000** *	1.000	1.000
	120	50							0.000	0.000	**1.000** *	1.000	1.000	0.000	0.000	**1.000** *	1.000	1.000
	120	100							0.000	0.000	**1.000** *	1.000	1.000	0.000	0.000	**1.000** *	1.000	1.000
	360	1	0.000	0.000	**1.000** *	1.000	1.000	**0.000**	0.000	0.000	**1.000** *	1.000	1.000	0.000	0.000	**1.000** *	1.000	1.000
	360	5							0.000	0.000	**1.000** *	1.000	1.000	0.000	0.000	**1.000** *	1.000	1.000
	360	10							0.000	0.000	**1.000** *	1.000	1.000	0.000	0.000	**1.000** *	1.000	1.000
	360	25							0.000	0.000	**1.000** *	1.000	1.000	0.000	0.000	**1.000** *	1.000	1.000
	360	50							0.000	0.000	**1.000** *	1.000	1.000	0.000	0.000	**1.000** *	1.000	1.000
	360	100							0.000	0.000	**1.000** *	1.000	1.000	0.000	0.000	**1.000** *	1.000	1.000
	600	1	0.000	0.000	**1.000** *	1.000	1.000	**0.000**	0.000	0.000	**1.000** *	1.000	1.000	0.000	0.000	**1.000** *	1.000	1.000
	600	5							0.000	0.000	**1.000** *	1.000	1.000	0.000	0.000	**1.000** *	1.000	1.000
	600	10							0.000	0.000	**1.000** *	1.000	1.000	0.000	0.000	**1.000** *	1.000	1.000
	600	25							0.000	0.000	**1.000** *	1.000	1.000	0.000	0.000	**1.000** *	1.000	1.000
	600	50							0.000	0.000	**1.000** *	1.000	1.000	0.000	0.000	**1.000** *	1.000	1.000
	600	100							0.000	0.000	**1.000** *	1.000	1.000	0.000	0.000	**1.000** *	1.000	1.000

Table 1: $r = 2$, $T = \{24, 60, 120, 360, 600\}$, $N = \{1, 5, 10, 25, 50, 100\}$, $5,000$ repetitions.

			LR (Trace)						LR Bar						NIV					
	T	N	$r_0 = 0$	$r_0 = 1$	$r_0 = 2$	H(0)	Size	H(2)	$r_0 = 0$	$r_0 = 1$	$r_0 = 2$	H(0)	Size	H(2)	$r_0 = 0$	$r_0 = 1$	$r_0 = 2$	H(0)	Size	H(2)
r = 1	24	1	0.893 *	**0.089**	0.019	0.107	**0.911**	0.981	0.847 *	**0.126**	0.026	0.153	**0.874**	0.974	0.843 *	**0.110**	0.046	0.157	**0.890**	0.954
	24	5							0.684 *	**0.243**	0.073	0.316	**0.757**	0.927	0.624 *	**0.211**	0.165	0.376	**0.789**	0.835
	24	10							0.553 *	**0.321**	0.126	0.447	**0.679**	0.874	0.489 *	**0.277**	0.234	0.511	**0.723**	0.766
	24	25							0.365	**0.437** *	0.199	0.635	**0.563**	0.801	0.271	**0.409** *	0.320	0.729	**0.591**	0.680
	24	50							0.245	**0.515** *	0.240	0.755	**0.485**	0.760	0.143	**0.493** *	0.364	0.857	**0.507**	0.636
	24	100							0.150	**0.575** *	0.275	0.850	**0.425**	0.725	0.059	**0.547** *	0.394	0.941	**0.453**	0.606
	60	1	0.717 *	**0.247**	0.036	0.283	**0.753**	0.964	0.621 *	**0.327**	0.052	0.379	**0.673**	0.948	0.699 *	**0.250**	0.051	0.301	**0.750**	0.949
	60	5							0.122	**0.728** *	0.150	0.878	**0.272**	0.850	0.260	**0.575** *	0.166	0.740	**0.425**	0.834
	60	10							0.016	**0.782** *	0.202	0.984	**0.218**	0.798	0.093	**0.692** *	0.215	0.907	**0.308**	0.785
	60	25							0.000	**0.749** *	0.251	1.000	**0.251**	0.749	0.012	**0.707** *	0.282	0.988	**0.293**	0.718
	60	50							0.000	**0.714** *	0.286	1.000	**0.286**	0.714	0.001	**0.677** *	0.322	0.999	**0.323**	0.678
	60	100							0.000	**0.688** *	0.312	1.000	**0.312**	0.688	0.000	**0.649** *	0.351	1.000	**0.351**	0.649
	120	1	0.180	**0.761** *	0.059	0.820	**0.239**	0.941	0.108	**0.824** *	0.068	0.892	**0.176**	0.932	0.333	**0.606** *	0.061	0.667	**0.394**	0.939
	120	5							0.000	**0.839** *	0.161	1.000	**0.161**	0.839	0.019	**0.829** *	0.152	0.981	**0.171**	0.848
	120	10							0.000	**0.799** *	0.201	1.000	**0.201**	0.799	0.002	**0.803** *	0.195	0.998	**0.197**	0.805
	120	25							0.000	**0.745** *	0.255	1.000	**0.255**	0.745	0.000	**0.742** *	0.258	1.000	**0.258**	0.742
	120	50							0.000	**0.715** *	0.285	1.000	**0.285**	0.715	0.000	**0.703** *	0.297	1.000	**0.297**	0.703
	120	100							0.000	**0.693** *	0.307	1.000	**0.307**	0.693	0.000	**0.666** *	0.334	1.000	**0.334**	0.666
	360	1	0.000	**0.939** *	0.061	1.000	**0.061**	0.939	0.000	**0.929** *	0.071	1.000	**0.071**	0.929	0.034	**0.900** *	0.066	0.966	**0.100**	0.934
	360	5							0.000	**0.837** *	0.163	1.000	**0.163**	0.837	0.000	**0.860** *	0.140	1.000	**0.140**	0.860
	360	10							0.000	**0.797** *	0.203	1.000	**0.203**	0.797	0.000	**0.816** *	0.184	1.000	**0.184**	0.816
	360	25							0.000	**0.741** *	0.259	1.000	**0.259**	0.741	0.000	**0.762** *	0.238	1.000	**0.238**	0.762
	360	50							0.000	**0.712** *	0.288	1.000	**0.288**	0.712	0.000	**0.721** *	0.279	1.000	**0.279**	0.721
	360	100							0.000	**0.691** *	0.309	1.000	**0.309**	0.691	0.000	**0.687** *	0.313	1.000	**0.313**	0.687
	600	1	0.000	**0.941** *	0.059	1.000	**0.059**	0.941	0.000	**0.930** *	0.070	1.000	**0.070**	0.930	0.015	**0.922** *	0.063	0.985	**0.078**	0.937
	600	5							0.000	**0.846** *	0.154	1.000	**0.154**	0.846	0.000	**0.866** *	0.134	1.000	**0.134**	0.866
	600	10							0.000	**0.798** *	0.202	1.000	**0.202**	0.798	0.000	**0.819** *	0.181	1.000	**0.181**	0.819
	600	25							0.000	**0.752** *	0.248	1.000	**0.248**	0.752	0.000	**0.768** *	0.232	1.000	**0.232**	0.768
	600	50							0.000	**0.723** *	0.277	1.000	**0.277**	0.723	0.000	**0.725** *	0.275	1.000	**0.275**	0.725
	600	100							0.000	**0.703** *	0.297	1.000	**0.297**	0.703	0.000	**0.689** *	0.311	1.000	**0.311**	0.689

Table 2: $r = 1$, $T = \{24, 60, 120, 360, 600\}$, $N = \{1, 5, 10, 25, 50, 100\}$, 5,000 repetitions.

http://www.bepress.com/jtse/vol2/iss1/art5
DOI: 10.2202/1941-1928.1057

			LR (Trace)						LR Bar						NIV					
	T	N	$r_0 = 0$	$r_0 = 1$	$r_0 = 2$	Size	H(1)	H(2)	$r_0 = 0$	$r_0 = 1$	$r_0 = 2$	Size	H(1)	H(2)	$r_0 = 0$	$r_0 = 1$	$r_0 = 2$	Size	H(1)	H(2)
$r = 0$	24	1	**0.934** *	0.060	0.007	**0.066**	0.940	0.993	**0.898** *	0.091	0.011	**0.102**	0.909	0.989	**0.859** *	0.099	0.042	**0.141**	0.901	0.958
	24	5							**0.832** *	0.156	0.012	**0.168**	0.844	0.988	**0.701** *	0.146	0.153	**0.299**	0.854	0.847
	24	10							**0.789** *	0.193	0.018	**0.211**	0.807	0.982	**0.586** *	0.189	0.225	**0.414**	0.811	0.775
	24	25							**0.712** *	0.264	0.023	**0.288**	0.736	0.977	**0.420** *	0.268	0.312	**0.580**	0.732	0.688
	24	50							**0.652** *	0.316	0.031	**0.348**	0.684	0.969	**0.303**	0.336	0.361 *	**0.697**	0.664	0.639
	24	100							**0.587** *	0.378	0.035	**0.413**	0.622	0.965	**0.197**	0.408 *	0.396	**0.803**	0.592	0.604
	60	1	**0.942** *	0.052	0.006	**0.058**	0.948	0.994	**0.917** *	0.074	0.009	**0.083**	0.926	0.991	**0.904** *	0.063	0.032	**0.096**	0.937	0.968
	60	5							**0.880** *	0.111	0.009	**0.120**	0.889	0.991	**0.799** *	0.085	0.116	**0.201**	0.915	0.884
	60	10							**0.855** *	0.135	0.010	**0.145**	0.865	0.990	**0.725** *	0.103	0.171	**0.275**	0.897	0.829
	60	25							**0.802** *	0.184	0.013	**0.198**	0.816	0.987	**0.611** *	0.131	0.258	**0.389**	0.869	0.742
	60	50							**0.751** *	0.230	0.019	**0.249**	0.770	0.981	**0.529** *	0.164	0.307	**0.471**	0.836	0.693
	60	100							**0.710** *	0.263	0.027	**0.290**	0.737	0.973	**0.454** *	0.198	0.348	**0.546**	0.802	0.652
	120	1	**0.949** *	0.044	0.007	**0.051**	0.956	0.993	**0.928** *	0.064	0.009	**0.072**	0.936	0.991	**0.917** *	0.053	0.030	**0.083**	0.947	0.970
	120	5							**0.899** *	0.095	0.007	**0.101**	0.905	0.993	**0.841** *	0.058	0.100	**0.159**	0.942	0.900
	120	10							**0.871** *	0.121	0.008	**0.129**	0.879	0.992	**0.786** *	0.068	0.145	**0.214**	0.932	0.855
	120	25							**0.830** *	0.158	0.012	**0.170**	0.842	0.988	**0.697** *	0.088	0.216	**0.303**	0.912	0.784
	120	50							**0.787** *	0.196	0.017	**0.213**	0.804	0.983	**0.630** *	0.098	0.272	**0.370**	0.902	0.728
	120	100							**0.747** *	0.232	0.021	**0.253**	0.768	0.979	**0.556** *	0.127	0.317	**0.444**	0.873	0.683
	360	1	**0.952** *	0.044	0.005	**0.048**	0.956	0.995	**0.925** *	0.066	0.008	**0.075**	0.934	0.992	**0.924** *	0.048	0.028	**0.076**	0.952	0.972
	360	5							**0.897** *	0.096	0.007	**0.103**	0.904	0.993	**0.867** *	0.050	0.082	**0.133**	0.950	0.918
	360	10							**0.872** *	0.119	0.009	**0.128**	0.881	0.991	**0.819** *	0.056	0.125	**0.181**	0.944	0.875
	360	25							**0.829** *	0.160	0.012	**0.171**	0.840	0.988	**0.744** *	0.067	0.190	**0.256**	0.933	0.810
	360	50							**0.792** *	0.189	0.018	**0.208**	0.811	0.982	**0.700** *	0.064	0.236	**0.300**	0.936	0.764
	360	100							**0.755** *	0.222	0.022	**0.245**	0.778	0.978	**0.633** *	0.085	0.282	**0.367**	0.915	0.718
	600	1	**0.953** *	0.042	0.005	**0.047**	0.958	0.995	**0.929** *	0.064	0.008	**0.071**	0.936	0.992	**0.928** *	0.046	0.025	**0.072**	0.954	0.975
	600	5							**0.902** *	0.092	0.007	**0.098**	0.908	0.993	**0.877** *	0.046	0.076	**0.123**	0.954	0.924
	600	10							**0.877** *	0.115	0.007	**0.123**	0.885	0.993	**0.835** *	0.049	0.116	**0.165**	0.951	0.884
	600	25							**0.830** *	0.157	0.013	**0.170**	0.843	0.987	**0.767** *	0.057	0.176	**0.233**	0.943	0.824
	600	50							**0.800** *	0.182	0.018	**0.200**	0.818	0.982	**0.718** *	0.058	0.224	**0.282**	0.942	0.776
	600	100							**0.763** *	0.216	0.020	**0.237**	0.784	0.980	**0.666** *	0.071	0.263	**0.334**	0.929	0.737

Table 3: $r = 0$, $T = \{24, 60, 120, 360, 600\}$, $N = \{1, 5, 10, 25, 50, 100\}$, $5,000$ repetitions.

7. Concluding Remarks

I have introduced rank tests alternative to the standard reduced rank trace test for a VECM and LR bar test for a panel VECM. These tests employ the nonlinear instrument approach that Chang (2002) applied to testing for unit roots in panel data, and the nonlinearity instruments out both non-normality and cross-sectional correlation. In a non-panel setting, this innovation may be viewed as an extension of the nonlinear IV unit root tests of Phillips *et al.* (2004). However, the rank testing strategy is vastly different from the unit root testing strategy. The optimal instrument of Phillips *et al.* is not admissible in the multivariate case. The class of regularly integrable instrument generating functions discussed by Chang (2002) (but restricted so that its inverse image does not concentrate mass) is more appropriate in this context.

Based on small-sample results, the main advantages of NIV tests lie not in increasing the power or controlling the size of standard tests, although the NIV tests appear to be competitive with extant tests, particularly when T is relatively large and N is relatively small. Rather, the asymptotic results show that the desirability of these tests lies in standard chi-squared critical values, rather than the critical values of model-dependent stochastic integrals.

Appendix: Mathematical Proofs

Before proceeding to the proofs of the main results, I present two ancillary lemmas. The relevant parts of Lemma 5 of Chang *et al.* (2001) are presented as Lemma A.1 simply for the reader's convenience and no proof is given here.

Lemma A.1 (Chang *et al.*, 2001) *Let z_i be the i^{th} element of the function z defined by assumption* [Z]. *Let (x_{it}) be the i^{th} element of an I(1) vector series (x_t) with increments satisfying an invariance principle, such that the limiting Brownian motion has unit variance and local time $L_i(1,0)$. Let (u_t) be a univariate I(0) series satisfying an invariance principle with unit variance. Then*

(a) $T^{-1/2}\sum z_i^2(x_{it}) \to_d L_i(1,0)\int_{-\infty}^{\infty} z_i(s)^2 ds,$

(b) $T^{-1/2}\sum z_i(x_{it})z_j(x_{jt}) \to_p 0$ *for* $i \neq j,$

(c) $T^{-1/4}\sum z_i(x_{it})u_t \to_d (L_i(1,0)\int_{-\infty}^{\infty} z_i(s)^2 ds)^{1/2}\ \mathbf{N}(0,1),$ *and*

(d) $T^{-1}\sum z_i(x_{it})x_{jt} = O_p(1)$ *for* $i \neq j$

as $T \to \infty$. The convergences in (a) and (c) hold jointly.

DOI: 10.2202/1941-1928.1057

Lemma A.2 *Let assumptions* [A1]-[A2] *and* [Z] *hold. Then* $U_T U_T' = I_{m-r}$ *with probability* 1.

Proof of Lemma A.2 Let $Q_T(r) \equiv T^{-1/2} \sum_{t=1}^{[Tr]} \varepsilon_t^0$ and let "$=_d$" denote distributional equivalence. There exists some stochastic process $Q_T^0(r)$ such that $Q_T(r) =_d Q_T^0(r)$ and $Q_T^0(r) \to_{a.s.} Q(r)$, according to the Skorokhod representation theorem, where $Q(r)$ is standard Brownian motion. The $T \times (m-r)$ matrix defined by $Q \equiv (Q_T'(r))_{t=1}^T$, has full column rank with probability 1,[6] because its rows consist of continuously distributed random variables that are not perfectly correlated, even in the limit.

Let $Z \equiv (z_{t-1}')_{t=2}^T$. Since $z(x)$ is assumed to be an element-wise transformation of x, Z may preserve the rank of Q if constructed properly. Z is a random matrix with full column rank and imperfectly correlated elements if their distributions are also continuous. Using a change of variables theorem (e.g., Theorem 4.1.11 of Dudley, 2002), the distribution of one such element is $P_{z_{i,t-1}} = \int p_{z_{i,t-1}} d(\mu \circ z^{-1})$ where μ is Lebesgue measure. This distribution is continuous if the measure $\mu \circ z^{-1}$ is absolutely continuous w.r.t. Lebesgue measure. Absolute continuity holds if the inverse image under z of any set with measure zero also has measure zero. Then the Borel-Cantelli lemma applies to the distribution of $z_{i,t-1}$. Thus, Z also has full column rank with probability 1. Letting $R \equiv (r_{1t}^{0\prime})_{t=2}^T$, note that R also has full column rank $m-r$, using the same arguments as above for Q.

The matrix in the lemma may be rewritten as

$$U_T U_T' = (Z'Z)^{-1/2} Z'R(R'Z(Z'Z)^{-1}Z'R)^{-1}R'Z(Z'Z)^{-1/2\prime},$$

which is a square matrix of dimension $m-r$, having full rank with probability 1. Since it is also idempotent, it must be equal to I_{m-r} with probability 1. □

Proof of Lemma 1 $A'y_{t-1} = B(L)A'\varepsilon_t$ using (1) and invertibility due to assumption [A2](c). Plugging this back into (1), $\triangle y_t = (I + \Gamma B(L)A')\varepsilon_t$ easily follows. To establish that this is in fact a Wold representation, write C_k explicitly in terms of k. By defining $C_k \equiv I + \Gamma B_k A'$, the representation in the lemma follows. Summability is straightforward to verify from assumption [A2](c). Finally, the derivation of long-run variance is given by Cheng and Phillips (2009), for example. □

Proof of Lemma 2 The proof of part (a) follows from a standard law of large numbers for iid sequences and by construction of (ε_t^0). The instrument

[6] See, e.g., Feng and Zhang (2007).

is defined to be $z(r_{1t}^0) = z(\Xi^{+1/2}y_{t-1})$, where $T^{-1/2}y_{[Tr]}$ converges to Brownian motion under assumptions [A1]-[A2]. The stated result in part (b) therefore follows directly from Lemma A.1(a) and (b). The limiting distribution of part (c) follows from a straightforward generalization of Lemma A.1(c) to account for the vectorization operation. The proof of part (d) follows from Lemma A.1(d). □

Proof of Lemma 3 The proof of part (a) is straightforward and is omitted.

For the proof of part (b), first consider the diagonals of $z(\hat{r}_{1t}^0)z(\hat{r}_{1t}^0)'$, which are $z_i^2(e_i'\hat{r}_{1t}^0)$. The first-order term of a Taylor series expansion of $\sum(z_i^2(e_i'\hat{r}_{1t}^0) - z_i^2(e_i'r_{1t}^0))$ around $e_i'r_{1t}^0$ is

$$a_{iiT}\sum \tilde{z}_i^2(e_i'r_{1t}^0)e_i'r_{1t}^0 + \sum\nolimits_{j\neq i} a_{ijT}\sum \tilde{z}_i^2(e_i'r_{1t}^0)e_j'r_{1t}^0 + \sum \tilde{z}_i^2(e_i'r_{1t}^0)b_{iT}(I-P_\Xi)r_{1t} \tag{20}$$

where

$$\begin{aligned} a_{iiT} &\equiv e_i'(\hat{\Xi}^{+1/2} - \Xi^{+1/2})\Xi^{+1/2\prime}e_i(e_i'\Xi^{+1/2}\Xi^{+1/2\prime}e_i)^{-1} \\ a_{ijT} &\equiv e_i'(\hat{\Xi}^{+1/2} - \Xi^{+1/2})\Xi^{+1/2\prime}e_j(e_j'\Xi^{+1/2}\Xi^{+1/2\prime}e_j)^{-1} \\ b_{iT} &\equiv e_i'(\hat{\Xi}^{+1/2} - \Xi^{+1/2}) \end{aligned}$$

and $\tilde{z}_i^2(x_i) \equiv \frac{\partial}{\partial x_i}z_i^2(x_i)$, using the decomposition in (7). Now, assuming that $\tilde{z}_i^2(x)x$ is regularly integrable, the first term of (20) is $O_p(a_{iiT}T^{1/2}) = o_p(T^{1/2})$. The third term of (20) is $O_p(\|b_{iT}\|\,T^{1/4}) = o_p(T^{1/2})$ with the rate $T^{1/4}$ from Theorem 5 of Jeganathan (2008).[7] If $m-r=1$, then the second term of (20) is zero and $z(\hat{r}_{1t}^0)z(\hat{r}_{1t}^0)'$ is a scalar, since there is only one stochastic trend in that case. The proof of part (b) is completed by noting that higher-order terms in the expansion are negligible.

On the other hand, if $m-r>1$, then the second term of (20) is $O_p(a_{ijT}T)$ by Lemma 1 of Chang and Park (2003). Moreover, in this case, the matrix $z(\hat{r}_{1t}^0)z(\hat{r}_{1t}^0)'$ contains off-diagonal terms of the form $z_i(e_i'\hat{r}_{1t}^0)z_k(e_k'\hat{r}_{1t}^0)$. The first-order term of a Taylor series expansion of $\sum(z_i(e_i'\hat{r}_{1t}^0)z_k(e_k'\hat{r}_{1t}^0) - z_i(e_i'r_{1t}^0)$ $z_k(e_k'r_{1t}^0))$ around $e_i'r_{1t}^0$ and $e_k'r_{1t}^0$ is given by

$$\begin{aligned} &a_{iiT}\sum z_{ik}e_i'r_{1t}^0 + a_{kkT}\sum z_{ki}e_k'r_{1t}^0 \\ &+\sum\nolimits_{j\neq i,k} a_{ijT}\sum z_{ik}e_j'r_{1t}^0 + \sum\nolimits_{j\neq i,k} a_{kjT}\sum z_{ki}e_j'r_{1t}^0 \\ &+\sum z_{ik}b_{iT}(I-P_\Xi)r_{1t} + \sum z_{ki}b_{kT}(I-P_\Xi)r_{1t} \end{aligned}$$

[7]Jeganathan's (2008) result requires generalized linear processes with summable coefficients, as Lemma 1 shows is the case in the present context. Also, his relevant moment conditions are satisfied by the assumptions about (ε_t).

where $z_{ik}(x_i, x_k) = \frac{\partial}{\partial x_i} z_i(x_i) z_k(x_k)$. The summands in the first and second terms may be written as products of two regularly integrable functions, since $e_i' r_{1t}^0$ and $e_k' r_{1t}^0$ are arguments of z_{ik} and z_{ki}. These terms are $o_p(a_{iiT} T^{1/2})$ and $o_p(a_{kkT} T^{1/2})$ by Lemma A.1(b) and thus pose no problem. The fifth and sixth terms also pose no problem, since the Cauchy-Schwarz inequality allows

$$\begin{aligned} &T^{-3/8} \sum z_{ki} b_{iT} (I - P_\Xi) r_{1t} \\ &\leq \sqrt{T^{-1/2} \sum z_i^{(1)} (e_i' r_{1t}^0) T^{-1/4} \sum z_k (e_k' r_{1t}^0) b_{iT} (I - P_\Xi) r_{1t}} \end{aligned}$$

so that these terms are $O_p(\|b_{iT}\| T^{3/8})$ and $O_p(\|b_{kT}\| T^{3/8})$. The third and fourth terms are more problematic. Note that

$$\sum z_{ik} e_j' r_{1t}^0 \leq \sqrt{T^{-1/2} \sum z_i^{(1)} (e_i' r_{1t}^0) T^{-1} \sum z_k (e_k' r_{1t}^0) e_j' r_{1t}^0}$$

so that these terms are $O_p(a_{ijT} T^{3/4})$ and $O_p(a_{kjT} T^{3/4})$. The largest term is $O_p(a_{ijT} T)$, which is $o_p(T^{1/2})$ only if the estimation error is $o_p(T^{1/2})$ in the direction of the trends. In this case, higher order terms will also be negligible.

For parts (c) and (d), the first-order term of a Taylor series expansion of $\sum (z_i(e_i' \hat{r}_{1t}^0) - z_i(e_i' r_{1t}^0))$ around $e_i' r_{1t}^0$ is

$$a_{iiT} \sum \tilde{z}_i (e_i' r_{1t}^0) e_i' r_{1t}^0 + \sum_{j \neq i} a_{ijT} \sum \tilde{z}_i (e_i' r_{1t}^0) e_j' r_{1t}^0 + \sum \tilde{z}_i (e_i' r_{1t}^0) b_{iT} (I - P_\Xi) r_{1t} \tag{21}$$

where $\tilde{z}_i(x_i) \equiv \frac{\partial}{\partial x_i} z_i(x_i)$.

Consider multiplying each term of (21) by $\varepsilon_t' \hat{\Xi}^{+1/2\prime}$ for part (c). The first term becomes

$$a_{iiT} \sum \tilde{z}_i (e_i' r_{1t}^0) e_i' r_{1t}^0 \varepsilon_t' \hat{\Xi}^{+1/2\prime} = O_p(a_{iiT} T^{1/4}),$$

and the third term similarly becomes

$$\sum \tilde{z}_i (e_i' r_{1t}^0) b_{iT} (I - P_\Xi) r_{1t} \varepsilon_t' \hat{\Xi}^{+1/2\prime} = O_p(\|b_{iT}\| T^{1/4}),$$

since $(I - P_\Xi) r_{1t} \varepsilon_t'$ is an MDS with finite conditional variance, due to the stationarity of $(I - P_\Xi) r_{1t}$. The second term is zero if $m - r = 1$. Higher order terms are similarly negligible.

If $m - r > 1$, then the second term of (21) multiplied by $\varepsilon_t' \hat{\Xi}^{+1/2\prime}$ is a sum of $a_{ijT} \sum \tilde{z}_i (e_i' r_{1t}^0) e_j' r_{1t}^0 \varepsilon_t' \hat{\Xi}^{+1/2\prime}$, which is $O_p(a_{ijT} T^{3/4})$ by Lemma 1 of Chang and Park (2003). The result for $m - r > 1$ in part (c) again requires the estimation error to be $o_p(T^{1/2})$ in the direction of the trends.

Consider first the diagonal elements of the matrix in the summand of part (d). The first-order term of a Taylor series expansion of $\sum(z_i(e_i'\hat{r}_{1t}^0)e_i'\hat{r}_{1t}^0 - z_i(e_i'r_{1t}^0)e_i'r_{1t}^0)$ around $e_i'r_{1t}^0$ is

$$a_{iiT}\sum z_i^*(e_i'r_{1t}^0)e_i'r_{1t}^0 + \sum_{j\neq i} a_{ijT}\sum z_i^*(e_i'r_{1t}^0)e_j'r_{1t}^0 + \sum z_i^*(e_i'r_{1t}^0)b_{iT}(I-P_\Xi)r_{1t}$$

where $z_i^*(x_i) \equiv x_i \frac{\partial}{\partial x_i} z_i(x_i) + z_i(x_i)$. The first and third term have the same asymptotics as the respective terms of (20), as long as $z_i^*(x_i)$ is regularly integrable, which completes the proof for $m-r=1$.

For $m-r>1$, the second term is non-zero, but $O_p(a_{ijT}T)$, similarly to part (b). For the off-diagonal elements of the summand in the case of $m-r>1$, the first-order term of a Taylor series expansion of $\sum(z_i(e_i'\hat{r}_{1t}^0)e_k'\hat{r}_{1t}^0 - z_i(e_i'r_{1t}^0)e_k'r_{1t}^0)$ around $e_i'r_{1t}^0$ and $e_k'r_{1t}^0$ is given by

$$\begin{aligned}
&a_{iiT}\sum z_{ik}^* e_i'r_{1t}^0 + a_{kkT}\sum z(e_i'r_{1t}^0)e_k'r_{1t}^0 \\
&+\sum_{j\neq i,k} a_{ijT}\sum z_{ik}^* e_j'r_{1t}^0 + \sum_{j\neq i,k} a_{kjT}\sum z(e_i'r_{1t}^0)e_j'r_{1t}^0 \\
&+\sum z_{ik}^* b_{iT}(I-P_\Xi)r_{1t} + \sum z(e_i'r_{1t}^0)b_{kT}(I-P_\Xi)r_{1t}
\end{aligned}$$

where $z_{ik}^*(x_i,x_k) = \frac{\partial}{\partial x_i} z(x_i)x_k$. The asymptotics of each term follow from Lemma 1 of Chang and Park (2003). The first and second terms are $O_p(a_{iiT}T)$ and $O_p(a_{kkT}T)$. The third and fourth are $O_p(a_{ijT}T^{3/2})$ and $O_p(a_{kjT}T)$. Finally, the last two terms are $O_p(\|b_{iT}\|\, T^{3/4})$ and $O_p(\|b_{kT}\|\, T^{1/4})$. As the largest term is $O_p(a_{ijT}T^{3/2})$, the estimation error must again be $o_p(T^{1/2})$ in the direction of the trends when $m-r>1$. □

Proof of Theorem 4 The proof follows by noting from Lemma A.1 that the convergences are joint, so that the local times, though random, may be canceled directly. As a result, $\upsilon_T \to_d \mathbf{N}(0, I_{(m-r)^2})$, so that $\upsilon_T'(I\otimes U_T U_T')\upsilon_T \to_d \chi^2_{(m-r)^2}$ by the continuous mapping theorem and Lemma A.2. □

Proof of Lemma 5 Let $\Pi^* \equiv (\mu, \Pi)$ and $w_t^* = (1, w_t')'$. The Wold representation in Lemma 1 may be modified to $\triangle y_t = C(L)(\Pi^* w_t^* + \varepsilon_t)$, although it is no longer an MA if (w_t) contains lagged $(\triangle y_t)$'s. A Beveridge-Nelson decomposition of $C(L)$ allows

$$\triangle y_t = C(1)\Pi^* w_t^* + C(1)\varepsilon_t + \tilde{C}(L)(\Pi\triangle w_t + \varepsilon_t)$$

DOI: 10.2202/1941-1928.1057

where $\tilde{C}(z)$ is defined by $\sum_{k=0}^{\infty}\sum_{s=k+1}^{\infty} C_s z^k$ as in Phillips and Solo (1992). The least squares estimator of $C(1)\Pi^*$ is given by

$$\widehat{C(1)\Pi}^* - C(1)\Pi^* = \left(\sum C(1)\varepsilon_t w_t^{*\prime} + \sum \tilde{C}(L)(\Pi\triangle w_t + \triangle\varepsilon_t)w_t^{*\prime}\right) \times \left(\sum w_t^* w_t^{*\prime}\right)^{-1}$$

The first term is $O_p(T^{-1/2})$ due to the fact that (w_t^*) includes a constant and stationary covariates, and $(\varepsilon_t w_t^{*\prime})$ is a martingale difference sequence due to the serial independence of (ε_t) and contemporaneous independence of (ε_t) and (w_t). The second term is $o_p(T^{-1/2})$ along the lines of the proof of Theorem 3.4 of Phillips and Solo (1992). Finally, the definition of $C(1)$ in terms of $A_\perp(\Gamma_\perp' A_\perp)^{-1}\Gamma_\perp'$ gives the stated result. □

Proof of Lemma 6 From the definitions,

$$\hat{\eta}_t^{*0} - \hat{\varepsilon}_t^0 = \hat{\Xi}^{+1/2}(C(1)\Pi^* - \widehat{C(1)\Pi}^*)w_t^* + (\hat{\Xi}^{+1/2} - \Xi^{+1/2})\Gamma(A'\Gamma)^{-1}A'\Pi^* w_t^* \tag{22}$$
$$+ (\hat{\Xi}^{+1/2} - \Xi^{+1/2})\Gamma\sum_{i=1}^{t-1} A' y_i d_{Ti}'\left(\sum_{i=1}^{t-1} d_{Ti}d_{Ti}'\right)^{-1} d_{T,t-1}$$

may be verified, using the facts that $\Xi^{+1/2}\Gamma = 0$ and

$$I = \Gamma(A'\Gamma)^{-1}A' + A_\perp(\Gamma_\perp' A_\perp)^{-1}\Gamma_\perp',$$

and using an estimate of $C(1)\Pi^*$. Now, from the definition of d_{Ti},

$$\Gamma\sum_{i=1}^{t-1} A' y_i d_{Ti}'\left(\sum_{i=1}^{t-1} d_{Ti}d_{Ti}'\right)^{-1} d_{T,t-1} = 2\sum_{i=1}^{t-1}\frac{3i-t}{t(t-1)}\Gamma A' y_i$$

and it is straightforward to verify that

$$2\sum_{i=1}^{t-1}\frac{3i-t}{t(t-1)}\Gamma A' y_i = -\Gamma(A'\Gamma)^{-1}A'\mu + 2\sum_{i=1}^{t-1}\frac{3i-t}{t(t-1)}\Gamma B(L)A'(\Pi w_i + \varepsilon_i)$$

Now, the common term $(\hat{\Xi}^{+1/2} - \Xi^{+1/2})\Gamma(A'\Gamma)^{-1}A'\mu$ cancels, so that (22) may be rewritten as

$$\hat{\eta}_t^{*0} - \hat{\varepsilon}_t^0 = \hat{\Xi}^{+1/2}(C(1)\Pi^* - \widehat{C(1)\Pi}^*)w_t^* + (\hat{\Xi}^{+1/2} - \Xi^{+1/2})\Gamma(A'\Gamma)^{-1}A'\Pi w_t$$
$$+ (\hat{\Xi}^{+1/2} - \Xi^{+1/2})\sum_{i=1}^{t-1}\Gamma B(L)A'(\Pi w_i + \varepsilon_i)$$
$$\times d_{Ti}'\left(\sum_{i=1}^{t-1} d_{Ti}d_{Ti}'\right)^{-1} d_{T,t-1}$$

or more simply as

$$\hat{\eta}_t^{*0} - \hat{\varepsilon}_t^0 = O_p(T^{-1/2})\hat{\Xi}^{+1/2}w_t^* + O_p(T^{-1/2})(\hat{\Xi}^{+1/2} - \Xi^{+1/2})d_{T,t-1} \qquad (23)$$
$$+ (\hat{\Xi}^{+1/2} - \Xi^{+1/2})\Gamma(A'\Gamma)^{-1}A'\Pi w_t$$

where the $O_p(T^{-1/2})$ in the first term comes directly from Lemma 5 and the $O_p(T^{-1/2})$ in the second term may be deduced from a FCLT for weakly dependent processes (e.g., Davidson, 1994).

An expansion of the summation in part (a) of the lemma reveals that all terms are $o_p(1)$, as long as $(\hat{\Xi}^{+1/2} - \Xi^{+1/2}) = o_p(1)$. Many of these terms involve a product of a stationary series and a properly scaled deterministic trend, and their asymptotic orders may be deduced from CLT's for martingale difference sequences for $(\hat{\varepsilon}_t^0)$ (e.g., McLeish, 1974) or weakly dependent processes for (w_t) (e.g., de Jong, 1997). The asymptotic orders of the remaining terms are standard.

The summation in part (b) may be written as

$$O_p(T^{-1/2})\sum z(r_{1t}^0)w_t^{*\prime}\hat{\Xi}^{+1/2\prime} + o_p(T^{-1/2})\sum z(r_{1t}^0)d_{T,t-1}' \qquad (24)$$
$$+ o_p(1)\sum z(r_{1t}^0)w_t'\Pi'A(\Gamma'A)^{-1}\Gamma'$$

using (23) and the consistency of $\hat{\Xi}^{+1/2} - \Xi^{+1/2}$. The first column of the first term is $O_p(1)$ due to the unit in the first element of w_t^*. As the remaining columns involve stationary (w_t), they are $O_p(T^{-1/4})$. The second term is $o_p(1)$ using Lemma 5(g) of Chang *et al.* (2001). The last term is $o_p(T^{1/4})$ using standard Park and Phillips (1999) asymptotics. The stated result holds since all terms in (24) are thus $o_p(T^{1/4})$. □

Proof of Theorem 7 By Lemma 6 and the convergence of an adaptively detrended I(1) process to detrended Brownian motion (Chang, 2002), the asymptotic results of the appropriate sample moments reduce to those of Lemma 2. The same result as that of Theorem 4 is thus obtained. □

Proof of Theorem 8 The proof follows directly from Theorem 4 by noting that the sum of independent χ^2 variates is also a χ^2 variate with degrees of freedom given by the sum of the degree of freedom. □

DOI: 10.2202/1941-1928.1057

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.

Andrews, D.W.K. (1991). "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica* 59, 817-858.

Aznar, A. and M. Salvador (2002). "Selecting the Rank of Cointegration and the Form of the Intercept Using Information Criterion," *Econometric Theory* 18, 926-947.

Baltagi, B.H. and C. Kao (2000). "Nonstationary Panels, Cointegration in Panels, and Dynamic Panels: A Survey," *Advances in Econometrics* 15, 7-52.

Breitung, J. (2002). "Nonparametric Tests for Unit Roots and Cointegration," *Journal of Econometrics* 108, 343-363.

Chang, Y. (2002). "Nonlinear IV Unit Root Tests in Panels with Cross-Sectional Dependency," *Journal of Econometrics* 110, 261-292.

Chang, Y. (2006). "Nonlinear IV Panel Unit Root Tests," in *Econometric Theory and Practice: Frontiers of Analysis and Applied Research* (D. Corbae, S.N. Durlauf, and B.E. Hansen, eds.). New York: Cambridge University Press, 334-358.

Chang, Y. and C.M. Nguyen (2010). "Residual Based Tests for Cointegration in Dependent Panels," mimeograph, Indiana University.

Chang, Y. and J.Y. Park (2003). "Index Models with Integrated Time Series," *Journal of Econometrics* 114, 73-106.

Chang, Y., J.Y. Park, and P.C.B. Phillips (2001). "Nonlinear Econometrics Models with Cointegrated and Deterministically Trending Regressors," *The Econometrics Journal* 4, 1-36.

Chang, Y. and W. Song (2009). "Testing for Unit Roots in Small Panels with Short-Run and Long-Run Cross-Sectional Dependencies," *Review of Economic Studies* 76, 903-935.

Chao, J.C. and P.C.B. Phillips (1999). "Model Selection in Partially Nonstationary Vector Autoregressive Processes with Reduced Rank Structure," *Journal of Econometrics* 91, 227-271.

Cheng, X. and P.C.B. Phillips (2009). "Semiparametric Cointegrating Rank Selection," *The Econometrics Journal* 12, S83-S104.

Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians*. Oxford: Oxford University Press.

Demetrescu, M. (2009). "Panel Unit Root Testing with Nonlinear Instruments for Infinite-Order Autoregressive Processes," *Journal of Time Series Econometrics* 1, Article 3.

Demetrescu, M. and A.I. Tarcolea (2008). "Panel Cointegration Testing Using Nonlinear Instruments," mimeograph, Goethe-University Frankfurt.

Dudley, R.M. (2002). *Real Analysis and Probability*. Cambridge: Cambridge University Press.

Feng, X. and Z. Zhang (2007). "The Rank of a Random Matrix," *Applied Mathematics and Computation* 185, 689-694.

Gonzalo, J. and J.Y. Pitarakis (1998). "Specification via Model Selection in Vector Error Correction Models," *Economics Letters* 60, 321-328.

Groen, J.J.J. and Kleibergen, F. (2003). "Likelihood-Based Cointegration Analysis in Panels of Vector Error-Correction Models," *Journal of Business and Economic Statistics*, 21, 295-318.

Hansen, B.E. (1992). "Consistent Covariance Matrix Estimation for Dependent Heterogeneous Processes," *Econometrica* 60, 967-972.

Jeganathan, P. (2008). "Limit Theorems for Functionals of Sums that Converge to Fractional Brownian and Stable Motions," Discussion Paper No. 1649, Cowles Foundation, Yale University.

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

Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.

DOI: 10.2202/1941-1928.1057

Johansen, S. and K. Juselius (1990). "Maximum Likelihood Estimation and Inference on Cointegration – With Applications to the Demand for Money," *Oxford Bulletin of Economics and Statistics* 52, 169-210.

de Jong, R.M. (1997). "Central Limit Theorems for Dependent Heterogeneous Random Variables," *Econometric Theory* 13, 353-367.

de Jong, R.M. (2004). "Addendum to 'Asymptotics for Nonlinear Transformation of Integrated Time Series,'" *Econometric Theory* 20, 627-635.

de Jong, R.M. and J. Davidson (2000). "Consistency of Kernel Estimators of Heteroscedastic and Autocorrelated Covariance Matrices," *Econometrica* 68, 407-423.

de Jong, R.M. and C.-H. Wang (2005). "Further Results on the Asymptotics for Nonlinear Transformations of Integrated Time Series," *Econometric Theory* 21, 413-430.

Kao, C. (1999). "Spurious Regression and Residual-Based Tests for Cointegration in Panel Data," *Journal of Econometrics* 90, 1-44.

Kapetanios, G. (2004). "The Asymptotic Distribution of the Cointegrating Rank Estimator Under the Akaike Information Criterion," *Econometric Theory* 20, 735-742.

Larsson, R., J. Lyhagen, and M. Löthgren (2001), "Likelihood-Based Cointegration Tests in Heterogeneous Panels," *The Econometrics Journal* 4, 109-122.

Lütkepohl, H. (1996). *Handbook of Matrices.* New York: John Wiley & Sons.

McCoskey, S. and C. Kao (1998). "A Residual-Based Test of the Null of Cointegration in Panel Data," *Econometric Reviews* 17, 57-84.

McLeish, D.L. (1974). "Dependent Central Limit Theorems and Invariance Principles," *Annals of Probability* 2 620-628.

Osterwald-Lenum, M. (1992). "A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics," *Oxford Bulletin of Economics and Statistics* 54, 461-472.

Park, J.Y. and P.C.B. Phillips (1999). "Asymptotics for Nonlinear Transformation of Integrated Time Series," *Econometric Theory* 15, 269-298.

Park, J.Y. and P.C.B. Phillips (2001). "Nonlinear Regressions with Integrated Time Series," *Econometrica* 69, 117-161.

Pedroni, P. (2004), "Panel Cointegration: Asymptotic and Finite-Sample Properties of Pooled Time Series Tests with an Application to PPP," *Econometric Theory* 20, 597-625.

Phillips, P.C.B., Y. Chang, and J.Y. Park (2004). "Nonlinear Instrumental Variable Estimation of an Autoregression," *Journal of Econometrics* 118, 219-246.

Phillips, P.C.B. and S. Ouliaris (1990). "Asymptotic Properties of Residual Based Tests for Cointegration," *Econometrica* 58, 165-193.

Phillips, P.C.B. and P. Perron (1988). "Testing for a Unit Root in Time Series Regressions," *Biometrika* 75, 335-346.

Phillips, P.C.B. and V. Solo (1992). "Asymptotics for Linear Processes," *Annals of Statistics* 20, 971-1001.

Pötscher, B. (2004). "Nonlinear Functions and Convergence to Brownian Motion: Beyond the Continuous Mapping Theorem," *Econometric Theory* 20, 1-22.

Shintani, M. (2001). "A Simple Cointegration Rank Test Without Vector Autoregression," *Journal of Econometrics* 105, 337-362.

So, B.S. and D.W. Shin (1999). "Cauchy Estimators for Autoregressive Processes with Applications to Unit Root Tests and Confidence Intervals," *Econometric Theory* 15, 165-176.

Wang, Q. and P.C.B. Phillips (2009). "Asymptotic Theory for Local Time Density Estimation and Nonparametric Cointegrating Regression," *Econometric Theory* 25, 710-738.

Wang, Z. and D.A. Bessler (2005). "A Monte Carlo Study on the Selection of Cointegrating Rank Using Information Criteria," *Econometric Theory* 21, 593-620.

Zivot, E. (2000). "The Power of Single Equation Tests for Cointegration When the Cointegrating Vector Is Prespecified," *Econometric Theory* 16, 407-439.

DOI: 10.2202/1941-1928.1057