

IHS Economics Series
Working Paper 270
June 2011

Nonparametric Rank Tests for Non-stationary Panels

Peter Pedroni
Timothy J. Vogelsang
Martin Wagner



INSTITUT FÜR HÖHERE STUDIEN
INSTITUTE FOR ADVANCED STUDIES
Vienna

Impressum

Author(s):

Peter Pedroni, Timothy J. Vogelsang, Martin Wagner

Title:

Nonparametric Rank Tests for Non-stationary Panels

ISSN: Unspecified

2011 Institut für Höhere Studien - Institute for Advanced Studies (IHS)

Josefstädter Straße 39, A-1080 Wien

E-Mail: office@ihs.ac.at

Web: www.ihs.ac.at

All IHS Working Papers are available online:

http://irihs.ihs.ac.at/view/ihs_series/

This paper is available for download without charge at:

<https://irihs.ihs.ac.at/id/eprint/2062/>

270

Reihe Ökonomie
Economics Series

Nonparametric Rank Tests for Non-stationary Panels

Peter Pedroni, Timothy J. Vogelsang, Martin Wagner,
Joakim Westerlund



INSTITUT FÜR HÖHERE STUDIEN
INSTITUTE FOR ADVANCED STUDIES
Vienna

270

**Reihe Ökonomie
Economics Series**

Nonparametric Rank Tests for Non-stationary Panels

**Peter Pedroni, Timothy J. Vogelsang, Martin Wagner,
Joakim Westerlund**

June 2011

**Institut für Höhere Studien (IHS), Wien
Institute for Advanced Studies, Vienna**

Contact:

Peter Pedroni
Williams College
24 Hopkins Hall Drive
Williamstown, MA 01267
email: Peter.L.Pedroni@williams.edu

Timothy J. Vogelsang
Department of Economics
Michigan State University
310 Marshall-Adams Hall
East Lansing, MI 48824-1038, USA
email: tjv@msu.edu

Martin Wagner
Department of Economics and Finance
Institute for Advanced Studies
Stumpergasse 56
1060 Vienna, Austria
☎: +43/1/599 91-150
email: martin.wagner@ihs.ac.at
and
Frisch Centre for Economic Research
Oslo

Joakim Westerlund – Corresponding author

Department of Economics
University of Gothenburg
P. O. Box 640
SE-405 30 Gothenburg, Sweden
☎: +46/31/786 5251
fax: +46/31/786 1043
email: joakim.westerlund@economics.gu.se

Founded in 1963 by two prominent Austrians living in exile – the sociologist Paul F. Lazarsfeld and the economist Oskar Morgenstern – with the financial support from the Ford Foundation, the Austrian Federal Ministry of Education and the City of Vienna, the Institute for Advanced Studies (IHS) is the first institution for postgraduate education and research in economics and the social sciences in Austria. The **Economics Series** presents research done at the Department of Economics and Finance and aims to share “work in progress” in a timely way before formal publication. As usual, authors bear full responsibility for the content of their contributions.

Das Institut für Höhere Studien (IHS) wurde im Jahr 1963 von zwei prominenten Exilösterreichern – dem Soziologen Paul F. Lazarsfeld und dem Ökonomen Oskar Morgenstern – mit Hilfe der Ford-Stiftung, des Österreichischen Bundesministeriums für Unterricht und der Stadt Wien gegründet und ist somit die erste nachuniversitäre Lehr- und Forschungsstätte für die Sozial- und Wirtschaftswissenschaften in Österreich. Die **Reihe Ökonomie** bietet Einblick in die Forschungsarbeit der Abteilung für Ökonomie und Finanzwirtschaft und verfolgt das Ziel, abteilungsinterne Diskussionsbeiträge einer breiteren fachinternen Öffentlichkeit zugänglich zu machen. Die inhaltliche Verantwortung für die veröffentlichten Beiträge liegt bei den Autoren und Autorinnen.

Abstract

This study develops new rank tests for panels that include panel unit root tests as a special case. The tests are unusual in that they can accommodate very general forms of both serial and cross-sectional dependence, including cross-unit cointegration, without the need to specify the form of dependence or estimate nuisance parameters associated with the dependence. The tests retain high power in small samples, and in contrast to other tests that accommodate cross-sectional dependence, the limiting distributions are valid for panels with finite cross-sectional dimensions.

Keywords

Nonparametric rank tests, unit roots, cointegration, cross-sectional dependence

JEL Classification

C12, C22, C23

Comments

We thank seminar participants at Brown University, Cornell University, Maastricht University, Texas AM, the University of Montreal and Williams College, and conference participants at the Econometric Society World Congress, the Midwest Econometrics Group at the University of Chicago, the Unit Root and Cointegration Conference in Faro, the Conference on Factor Models in Panels at Goethe University Frankfurt, the 11th International Panel Data Conference, the NY Econometrics Camp and Robert Kunst for helpful comments and suggestions. An earlier draft version of this paper circulated with a similar title by the first two authors. Wagner thanks the Jubiläumsfonds of the Oesterreichische Nationalbank (Project Nr.: 13398) for financial support and Westerlund would like to thank the Jan Wallander and Tom Hedelius Foundation for financial support under research grant P2009–0189:1. The usual disclaimer applies.

Contents

1	Introduction	1
2	Assumptions and Setup	3
3	The Tests	6
	3.1 The Rank Statistics and Their Limiting Distributions	6
	3.2 Critical Values for the Rank Tests	11
	3.3 First-generation Analogues as a Special Case	12
4	Local Power	13
5	Comparison to Factor Model Approaches	17
6	Small-sample Performance	18
	6.1 Simulation Design	18
	6.2 Results	19
	6.3 Comparisons to Some Existing Tests	21
7	Empirical Applications	24
	7.1 Purchasing Power Parity	24
	7.2 Income Convergence	25
8	Conclusions	27
	References	27
	Tables	30

1 Introduction

This paper develops new rank¹ tests for panels which include panel unit root tests as a special case. The tests are unusual in that they can accommodate very general forms of serial and cross-sectional dependence in panels, including cross-unit cointegration, without the need to either specify the form of dependence or to estimate nuisance parameters associated with the dependence. This is in contrast to approaches in the earlier literature on non-stationary panels, which typically attempt to accommodate the dependence by estimation, either parametrically or nonparametrically, so that the limiting distributions of the test statistics are purged of nuisance parameters. A potential disadvantage of this more conventional approach is that one must make choices in the process of estimation which can have a substantial impact on subsequent inference. More to the point, in small samples the best choices are not easily apparent, and poor choices may further aggravate problems of size distortion and loss of power.

Examples of the conventional approach to treating dependence in non-stationary panels include both the so-called first- and second-generation panel unit tests. To give some examples, among the early first-generation panel unit root tests, Levin *et al.* (2002), Im *et al.* (2003) and numerous others used an augmented Dickey–Fuller (ADF) approach to accommodate serial dependence, where the order of the augmentation was treated as being unknown and likely heterogeneous among the units of the panel. This order was then chosen by any one of a number of criteria. Nonparametric treatments of the serial dependence analogous to the Phillips–Perron approach are also possible for panels, in which case a bandwidth parameter and a kernel function must be chosen for estimation of the long-run variance.

For cross-sectional dependence, researchers who employed the first-generation tests either assumed it was absent, or assumed that the dependence could be adequately captured by time effects. In the latter case, time effects were estimated and then extracted prior to estimating the individual ADF regressions for each cross-sectional unit. However, time effects presume a specialized form of cross-sectional dependence. Accordingly, second-generation panel unit root tests sought to generalize the forms of permissible cross-sectional dependence by assuming an unobserved common factor structure, see for example Bai and Ng (2004), and Moon and Perron (2004). The basic approach required one to first choose or

¹With rank we refer to the rank of the long-run covariance matrix, i.e. to the number of non-cointegrated I(1) series (common trends) in the panel of time series.

estimate the number of common factors using some criterion. Next, conditional on having chosen the number of factors, these common factors were estimated by principle components and then extracted prior to treating the serial dependence either parametrically or nonparametrically. Whether the dependence is serial or cross-sectional, the underlying strategy shared by both the first- and second-generation tests is to estimate the source of the dependence and eliminate its effect on the limiting distribution of the test statistics.

Another important issue that pertains to most first- and second-generation tests is that the form of the null and alternative hypotheses is somewhat restricted. Specifically, these tests typically take the null hypothesis to be that all series in the panel are unit root non-stationary and the alternative hypothesis to be that at least some series are trend-stationary. This leaves a rejection of the null somewhat uninformative, because the rejection does not indicate how many series are stationary (or more generally, when cross-sectional dependence is allowed how many cointegrating relationships there are). A notable exception on this issue is the work of Ng (2008), which allows one to test the fraction of the panel with a unit root. However, the extent of cross-sectional dependence is highly restricted in that paper since cross-unit cointegration, defined below in Section 2, is not permitted.

Motivated by these issues, the current paper uses an entirely different approach to accommodate general serial and cross-sectional dependence of unknown form. In contrast to existing panel unit root tests, the tests developed in this paper are based on the rank of the long-run variance matrix of the first differences of the N -dimensional vector of stacked observations of the observed panel data, where N denotes the cross-section sample size. This makes the tests suitable both as conventional panel unit root tests with the corresponding null and alternative hypotheses, or, more generally, as flexible rank tests that allow one to determine the number of common trends in the panel.

An additional practical feature of the new tests is that they are constructed such that one does not have to make any choices regarding the treatment of the underlying serial or cross-sectional dependencies. Thus, with these tests one does not need to make any choices regarding lag augmentation, kernel, bandwidth or the number of common factors. Specifically, the tests are constructed as simple variance ratios whose limiting distributions are already invariant to both serial and cross-sectional dependence, so that there is no need to estimate these dependencies or make choices associated with their estimation. Our approach is closely related to the univariate unit root and cointegration tests of Breitung (2002), which

also avoid choices of this kind.

Despite their simplicity, the new tests are remarkably general. In fact, except for some mild regulatory conditions, there are virtually no restrictions on the forms of serial and cross-sectional dependence. Accordingly, the techniques are applicable under assumptions typically made for first- and second-generation tests, but can also accommodate more general form of dependencies. Furthermore, in contrast to most first- and second-generation approaches, the tests developed in this paper do not require letting N go to infinity, and in fact perform well even when N is relatively small. We believe that these various features make our tests some of the most widely applicable.² It has to be noted that the new tests are also very powerful, compared to first-generation tests, in case of cross-sectionally independent data, suggesting that the cost of not requiring any prior knowledge regarding possible cross-sectional dependence is very low.

The remainder of the paper is organized as follows. Section 2 presents the assumptions and discusses the setup, based upon which in Section 3 the asymptotic distributions of our rank test statistics are derived. Section 4 then discusses some of the distinctive features of the new tests, and compares them to some second-generation factor-based tests. The small-sample properties are studied by Monte Carlo simulations in Section 5. Section 6 provides two brief empirical illustrations taken from the exchange rate and growth and convergence literatures, and Section 7 briefly concludes the paper.

2 Assumptions and Setup

We consider an N -dimensional vector $\mathbf{y}_t = [y_{1t}, \dots, y_{Nt}]'$ given by

$$\mathbf{y}_t = \boldsymbol{\alpha}_p \mathbf{d}_t^p + \mathbf{u}_t, \quad (1)$$

where $\mathbf{d}_t^p = [1, t, \dots, t^p]'$ with $p \geq 0$ is a polynomial trend function satisfying $\mathbf{d}_t^0 = 1$, with $\boldsymbol{\alpha}_p$ being the associated matrix of trend coefficients. The typical specifications considered for \mathbf{d}_t^p include a constant or a constant and a linear time trend, corresponding to $p = 0$ and $p = 1$, respectively, and these are also the specifications considered in the simulation and application sections of this paper.

²It has to be noted here that the tests developed in Palm *et al.* (2009) are also applicable to panel data with quite unrestricted cross-sectional dependencies. Their tests are, however, more computationally intensive as they resort to bootstrap techniques.

The main variable of interest is $\mathbf{u}_t = [u_{1t}, \dots, u_{Nt}]'$, which represents the stochastic part of \mathbf{y}_t . In order to describe its unit root and cointegration properties, we introduce an $N \times N$ orthogonal matrix $\mathbf{C} = [\mathbf{C}_1, \mathbf{C}_2]$, which is such that $\mathbf{C}'\mathbf{C} = \mathbf{C}\mathbf{C}' = \mathbf{I}_N$ and whose component matrices \mathbf{C}_1 and \mathbf{C}_2 are of dimensions $N \times N_1$ and $N \times N_2$ with $N_2 = N - N_1$, respectively. The matrix \mathbf{C}_1 is chosen to give a basis for the cointegrating space of \mathbf{u}_t , while \mathbf{C}_2 , which is such that $\mathbf{C}_2'\mathbf{C}_1 = \mathbf{0}$ and $\mathbf{C}_1'\mathbf{C}_2 = \mathbf{0}$, gives the common (unit root or stochastic) trends. The matrix \mathbf{C} allows us to rotate \mathbf{u}_t into stationary and unit root subsystems as

$$\mathbf{w}_t = \mathbf{C}'\mathbf{u}_t = \begin{bmatrix} \mathbf{C}_1'\mathbf{u}_t \\ \mathbf{C}_2'\mathbf{u}_t \end{bmatrix} = \begin{bmatrix} \mathbf{w}_{1t} \\ \mathbf{w}_{2t} \end{bmatrix}, \quad (2)$$

where the first N_1 series \mathbf{w}_{1t} are stationary, while the remaining N_2 series \mathbf{w}_{2t} are non-stationary and non-cointegrated. The corresponding vector of stationary errors is given by

$$\mathbf{v}_t = \begin{bmatrix} \mathbf{w}_{1t} \\ \Delta\mathbf{w}_{2t} \end{bmatrix} = \begin{bmatrix} \mathbf{v}_{1t} \\ \mathbf{v}_{2t} \end{bmatrix}, \quad (3)$$

whose long-run covariance matrix will be fundamental to the testing approach used in this paper. For weakly stationary stochastic processes \mathbf{a}_t and \mathbf{b}_t with mean zero and absolutely summable covariance function we define the long-run covariance matrix as

$$\Omega_{ab} = \sum_{s=-\infty}^{\infty} \mathbb{E}(\mathbf{a}_t \mathbf{b}_{t-s}') = \Sigma_{ab} + \Gamma_{ab} + \Gamma_{ab}',$$

where $\Sigma_{ab} = \mathbb{E}(\mathbf{a}_t \mathbf{b}_t')$ and $\Gamma_{ab} = \sum_{s=1}^{\infty} \mathbb{E}(\mathbf{a}_t \mathbf{b}_{t-s}')$ are the contemporaneous and one-sided long-run covariance matrices, respectively. The long-run covariance matrix of \mathbf{v}_t is partitioned in the following way:

$$\Omega_{vv} = \begin{bmatrix} \Omega_{v_1v_1} & \Omega_{v_1v_2} \\ \Omega_{v_2v_1} & \Omega_{v_2v_2} \end{bmatrix} = \begin{bmatrix} \omega_{v_1}^2 & \omega_{v_1v_2} & \dots & \omega_{v_1v_N} \\ \omega_{v_2v_1} & \omega_{v_2}^2 & \dots & \omega_{v_2v_N} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{v_Nv_1} & \omega_{v_Nv_2} & \dots & \omega_{v_N}^2 \end{bmatrix}.$$

Assumption 1 is enough to obtain our main results.

Assumption 1. As $T \rightarrow \infty$,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor sT \rfloor} \mathbf{v}_t \rightarrow_w \mathbf{B}(s) = \Omega_{vv}^{1/2} \mathbf{W}(s) = \begin{bmatrix} \Omega_{v_1.v_2}^{1/2} & \Omega_{v_1v_2} \Omega_{v_2v_2}^{-1/2} \\ \mathbf{0} & \Omega_{v_2v_2}^{1/2} \end{bmatrix} \begin{bmatrix} \mathbf{W}_1(s) \\ \mathbf{W}_2(s) \end{bmatrix},$$

where Ω_{vv} is positive definite, \rightarrow_w and $\lfloor x \rfloor$ signify weak convergence and integer part of x , respectively, $\Omega_{v_1.v_2} = \Omega_{v_1v_1} - \Omega_{v_1v_2} \Omega_{v_2v_2}^{-1} \Omega_{v_2v_1}$ and $\mathbf{W}(s) = [W_1(s), \dots, W_N(s)]'$ is an N -dimensional vector of independent standard Brownian motions that is partitioned conformably with \mathbf{v}_t .

Assumption 1 is stated directly in terms of the required invariance principle rather than primitive regularity conditions. This is convenient because it is this result that drives the distribution theory and we are not specifically interested here in the various sets of regularity conditions under which it holds. It may be noted, however, that there are a variety of more primitive conditions that lead to Assumption 1. For example, Phillips and Durlauf (1986) give conditions requiring that \mathbf{v}_t be weakly stationary with finite moment greater than second order and that it satisfies well-known α -mixing conditions. Phillips and Solo (1992) give other sets of conditions based on linear processes. Our approach allows for general forms of cross-sectional dependence as well as series specific patterns of serial correlation.³

The unit root and cointegration behavior of \mathbf{u}_t is governed by $\Omega_{\Delta u \Delta u}$, the long-run covariance matrix of $\Delta \mathbf{u}_t$, whose rank is henceforth going to be denoted as $r = \text{rk}(\Omega_{\Delta u \Delta u})$. This matrix can be directly related to the long-run covariance matrix of the corresponding rotated vector $\Delta \mathbf{w}_t$. In particular, by using (2) and (3), and the fact that $\Delta \mathbf{v}_{1t}$ is over-differenced with zero long-run variance, we obtain

$$\Omega_{\Delta u \Delta u} = \mathbf{C} \Omega_{\Delta w \Delta w} \mathbf{C}' = \mathbf{C} \begin{bmatrix} \Omega_{\Delta v_1 \Delta v_1} & \Omega_{\Delta v_1 \Delta v_2} \\ \Omega_{\Delta v_2 \Delta v_1} & \Omega_{\Delta v_2 \Delta v_2} \end{bmatrix} \mathbf{C}' = \mathbf{C} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Omega_{v_2 v_2} \end{bmatrix} \mathbf{C}' = \mathbf{C}_2 \Omega_{v_2 v_2} \mathbf{C}_2',$$

showing again that $r = N_2$. If $r = N$, so that the rank is full, then $\mathbf{C} = \mathbf{C}_2 = \mathbf{I}_N$, meaning that now all the elements of \mathbf{u}_t are unit root non-stationary and non-cointegrated. If, in addition, the series are cross-sectionally independent and hence $\Omega_{\Delta u \Delta u}$ is diagonal, then we have the scenario for which the first-generation unit root tests were developed. If, on the other hand, the rank is reduced such that $r < N$, then there are only $N_2 < N$ unit roots, which can be due to either unit-specific stationarity, or cross-unit cointegration, or a combination of the two. The extreme case occurs when $r = 0$, in which $\mathbf{C} = \mathbf{C}_1 = \mathbf{I}_N$, suggesting that now all the elements of \mathbf{u}_t are stationary. This discussion shows that as soon as one abolishes the cross-sectional independence assumption the relevant quantity to understand the dynamic behavior of the panel of time series is not the number of unit roots in the individual series but rather the number of common trends, which equals the rank of $\Omega_{\Delta u \Delta u}$.

Let us next formally define the concept of cross-unit cointegration, already alluded to above. A more complete discussion of this concept, which becomes more relevant for panels of multivariate time series is given in Wagner and Hlouskova (2010). Denote by \mathbf{H} an

³In principle even the assumption of weak stationarity could be abandoned. It is sufficient that the long-run covariance matrix, more generally defined via $\Omega_{vv} = \lim_{T \rightarrow \infty} \mathbb{E} \frac{1}{T} (\sum_{t=1}^T v_t) (\sum_{t=1}^T v_t)'$, exists.

$N \times n_1$ diagonal selection matrix comprised of zeros and ones that picks the individually stationary units of \mathbf{u}_t . For example, if u_{it} is stationary, then \mathbf{H} has as one of its columns the vector $[0, \dots, 0, 1, 0, \dots, 0]'$ with the one located at the i^{th} position. Note also that $n_1 \leq N_1$. The cross-unit cointegrating space of \mathbf{u}_t is given by the space spanned by $\mathbf{D} = (\mathbf{I}_N - \mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}')\mathbf{C}_1$. That is, the cross-unit cointegrating space is the span of the projection of the cointegrating space \mathbf{C}_1 of \mathbf{u}_t on the orthogonal complement of \mathbf{H} , which includes all cointegrating relationships that are not made up of linear combinations of unit-specific processes that are already stationary. The cross-unit cointegrating rank is the dimension of the space spanned by \mathbf{D} . Altogether we thus have $r = N - n_1 - \text{rk}(\mathbf{D})$.

3 The Tests

The quantity of interest is r , the rank of $\Omega_{\Delta u \Delta u}$. In this section we develop tests that are designed to test $H_0 : r = r_0$ versus $H_1 : r = r_1 < r_0$. That is, the rank being r_0 is tested against a rank smaller than r_0 . Obviously, the smallest value of r_0 that can be considered is $r_0 = 1$.

3.1 The Rank Statistics and Their Limiting Distributions

The rank tests that we consider are based on two ingredients. One is an extension of regressions involving superfluous deterministic trend terms (see Park, 1990 and Park and Choi, 1988) from the time series to the panel case, and the other is long-run variance estimation based on untruncated kernels (see Kiefer and Vogelsang, 2002a).

We begin by discussing the variance components of the test statistics. In particular, consider the least squares residual

$$\hat{\mathbf{u}}_t^p = \mathbf{y}_t - \sum_{t=1}^T \mathbf{y}_t \mathbf{d}_t^{p'} \left(\sum_{t=1}^T \mathbf{d}_t^p \mathbf{d}_t^{p'} \right)^{-1} \mathbf{d}_t^p.$$

In case of stationarity of \mathbf{u}_t , the estimated long-run variance based on $\hat{\mathbf{u}}_t^p$ is given by

$$\hat{\Omega}_p = \frac{1}{T} \sum_{j=-M}^M k(j/M) \sum_{t=j+1}^T \hat{\mathbf{u}}_t^p \hat{\mathbf{u}}_{t-j}^{p'} = \begin{bmatrix} \hat{\omega}_{1p}^2 & \hat{\omega}_{12p} & \dots & \hat{\omega}_{1Np} \\ \hat{\omega}_{21p} & \hat{\omega}_{2p}^2 & \dots & \hat{\omega}_{2Np} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\omega}_{N1p} & \hat{\omega}_{N2p} & \dots & \hat{\omega}_{Np}^2 \end{bmatrix},$$

where $k(x)$ is a kernel function and M is the bandwidth. Kiefer and Vogelsang (2002a) have shown that if untruncated such that $M = T$, the estimator $\hat{\Omega}_p$ converges to a random variable

that is proportional to $\mathbf{\Omega}$, and whose precise shape depends on the choice of $k(x)$. In case of the Bartlett kernel $k(x) = 1 - \frac{|x|}{T}$, Kiefer and Vogelsang (2002b) show that the formula for $\hat{\mathbf{\Omega}}_p$ reduces to

$$\hat{\mathbf{\Omega}}_p = \frac{2}{T^2} \sum_{t=1}^T \hat{\mathbf{S}}_t^p \hat{\mathbf{S}}_t^{p'},$$

where $\hat{\mathbf{S}}_t^p = \sum_{s=1}^t \hat{\mathbf{u}}_s^p$.

Clearly, in our situation $\hat{\mathbf{u}}_t^p$ will not be asymptotically stationary, at least under the null hypothesis. Thus, let us partition $\hat{\mathbf{w}}_t^p = \mathbf{C}' \hat{\mathbf{u}}_t^p = [\hat{\mathbf{w}}_{1t}^{p'}, \hat{\mathbf{w}}_{2t}^{p'}]'$ such that the first N_1 series $\hat{\mathbf{w}}_{1t}^p$ are asymptotically stationary, while the remaining N_2 series $\hat{\mathbf{w}}_{2t}^p$ are unit root non-stationary and non-cointegrated. The required normalization matrix to take into account the different convergence rates for the stationary and integrated components is given by $\mathbf{D}_T = \text{diag}(\mathbf{I}_{N_1}, \sqrt{T} \mathbf{I}_{N_2})$. By using Assumption 1, rotation by \mathbf{C} and standard results for least squares detrended variables it follows that as $T \rightarrow \infty$

$$\begin{aligned} \frac{1}{T} \mathbf{D}_T^{-1} \hat{\mathbf{\Omega}}_p \mathbf{D}_T^{-1} &= 2\mathbf{C} \frac{1}{T^3} \sum_{t=1}^T \begin{bmatrix} \hat{\mathbf{R}}_{1t}^p \hat{\mathbf{R}}_{1t}^{p'} & T^{-1/2} \hat{\mathbf{R}}_{1t}^p \hat{\mathbf{R}}_{2t}^{p'} \\ T^{-1/2} \hat{\mathbf{R}}_{2t}^p \hat{\mathbf{R}}_{1t}^{p'} & T^{-1} \hat{\mathbf{R}}_{2t}^p \hat{\mathbf{R}}_{2t}^{p'} \end{bmatrix} \mathbf{C}' \\ &\rightarrow_w 2\mathbf{C} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \int_0^1 \mathbf{R}_2^p(s) \mathbf{R}_2^p(s)' ds \end{bmatrix} \mathbf{C}', \end{aligned} \quad (4)$$

where $\hat{\mathbf{R}}_t^p = \sum_{s=1}^t \hat{\mathbf{w}}_s^p$, $\mathbf{R}^p(s) = \int_0^s \mathbf{B}^p(r) dr$ and $\mathbf{B}^p(s) = \mathbf{\Omega}_{vv}^{1/2} \mathbf{W}^p(s)$ with

$$\mathbf{W}^p(s) = \mathbf{W}(s) - \int_0^1 \mathbf{W}(r) \mathbf{d}^p(r)' dr \left(\int_0^1 \mathbf{d}^p(r) \mathbf{d}^p(r)' dr \right)^{-1} \mathbf{d}^p(s)$$

denoting the residual from projecting $\mathbf{W}(s)$ onto $\mathbf{d}^p(s) = [1, s, \dots, s^p]'$ with $\mathbf{d}^0(s) = 1$. All vectors are partitioned conformably with \mathbf{C} . Similarly, the contemporaneous variance estimator,

$$\hat{\mathbf{\Sigma}}_p = \frac{1}{T} \sum_{t=1}^T \hat{\mathbf{u}}_t^p \hat{\mathbf{u}}_t^{p'} = \begin{bmatrix} \hat{\sigma}_{1p}^2 & \hat{\sigma}_{12p} & \dots & \hat{\sigma}_{1Np} \\ \hat{\sigma}_{21p} & \hat{\sigma}_{2p}^2 & \dots & \hat{\sigma}_{2Np} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_{N1p} & \hat{\sigma}_{N2p} & \dots & \hat{\sigma}_{Np}^2 \end{bmatrix},$$

must also be normalized accordingly in the presence of unit root non-stationarity in order to achieve convergence:

$$\begin{aligned} \mathbf{D}_T^{-1} \hat{\mathbf{\Sigma}}_p \mathbf{D}_T^{-1} &= \mathbf{C} \frac{1}{T} \sum_{t=1}^T \begin{bmatrix} \hat{\mathbf{w}}_{1t}^p \hat{\mathbf{w}}_{1t}^{p'} & T^{-1/2} \hat{\mathbf{w}}_{1t}^p \hat{\mathbf{w}}_{2t}^{p'} \\ T^{-1/2} \hat{\mathbf{w}}_{2t}^p \hat{\mathbf{w}}_{1t}^{p'} & T^{-1} \hat{\mathbf{w}}_{2t}^p \hat{\mathbf{w}}_{2t}^{p'} \end{bmatrix} \mathbf{C}' \\ &\rightarrow_w \mathbf{C} \begin{bmatrix} \mathbf{\Sigma}_{v_1 v_1} & \mathbf{0} \\ \mathbf{0} & \int_0^1 \mathbf{B}_2^p(s) \mathbf{B}_2^p(s)' ds \end{bmatrix} \mathbf{C}'. \end{aligned} \quad (5)$$

The convergence results in (4) and (5) imply that test statistics with nuisance parameter free limiting distributions can be constructed by simply using appropriately normalized matrix ratios of $\hat{\Sigma}_p$ and $\hat{\Omega}_p$. The first test statistic of this type that we will consider can be seen as a multivariate version of the \hat{q}_T statistic introduced by Breitung (2002) in the time series context. It is given by

$$MB = \frac{1}{2T} \text{tr}(\hat{\Omega}_p \hat{\Sigma}_p^{-1}).$$

The asymptotic distribution of this statistic under the null hypothesis $0 \leq r_0 \leq N$ is easily derived from the above results. Indeed, by using the cyclical property of the trace,

$$\begin{aligned} MB &= \text{tr} \left(\frac{1}{2T} \mathbf{D}_T^{-1} \hat{\Omega}_p \mathbf{D}_T^{-1} (\mathbf{D}_T^{-1} \hat{\Sigma}_p \mathbf{D}_T^{-1})^{-1} \right) \\ &\rightarrow_w \text{tr} \left(\mathbf{C} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \int_0^1 \mathbf{R}_2^p(s) \mathbf{R}_2^p(s)' ds \end{bmatrix} \mathbf{C}' \left(\mathbf{C} \begin{bmatrix} \Sigma_{v_1 v_1} & \mathbf{0} \\ \mathbf{0} & \int_0^1 \mathbf{B}_2^p(s) \mathbf{B}_2^p(s)' ds \end{bmatrix} \mathbf{C}' \right)^{-1} \right) \\ &= \text{tr} \left(\int_0^1 \mathbf{R}_2^p(s) \mathbf{R}_2^p(s)' ds \left(\int_0^1 \mathbf{B}_2^p(s) \mathbf{B}_2^p(s)' ds \right)^{-1} \right) \\ &= \text{tr} \left(\int_0^1 \mathbf{Q}_2^p(s) \mathbf{Q}_2^p(s)' ds \left(\int_0^1 \mathbf{W}_2^p(s) \mathbf{W}_2^p(s)' ds \right)^{-1} \right), \end{aligned} \quad (6)$$

where $\mathbf{Q}^p(s) = \int_0^s \mathbf{W}^p(r) dr$ is again partitioned conformably with \mathbf{C} .

The second test statistic that we consider is based on the properties of regressions that include superfluous deterministic trend regressors. Towards this end, suppose that the data are generated as before via (1) but that the trend polynomial used in the least squares detrending is now of degree $q > p$. If \mathbf{u}_t is stationary, then the coefficients corresponding to the superfluous trends t^{p+1}, \dots, t^q are estimated consistently to be zero as $T \rightarrow \infty$. Therefore, a coefficient restriction test such the Wald test will have a well-defined limiting distribution in this case, although it will not necessarily be free of nuisance parameters. On the other hand, if \mathbf{u}_t is non-stationary, then (1) becomes spurious and the coefficients corresponding to the superfluous regressors will not go to zero asymptotically. This in turn implies that the Wald statistic becomes $O_p(T)$. This led Park and Choi (1988) to consider the Wald statistic divided by T as a unit root test statistic. Our test statistic can be seen as a multivariate analog to this statistic, and is given by

$$MJ = \text{tr}(\hat{\Sigma}_p \hat{\Sigma}_q^{-1} - \mathbf{I}_N),$$

where $\hat{\Sigma}_q$ is the estimated residual variance from (1) when the fitted trend polynomial is of degree $q > p$. Vogelsang (1998) studied the Wald statistic of Park and Choi (1988) and found

that it has strongly rising power up until $q = 9$, after which the power increments drop off. In the simulations and applications contained in this paper we only report results for this test constructed for the value of $q = 9$. Similarly to before, under the null hypothesis of rank equal to r_0 ,

$$\begin{aligned}
MJ &= \text{tr} \left(\mathbf{D}_T^{-1} \hat{\Sigma}_p \mathbf{D}_T^{-1} (\mathbf{D}_T^{-1} \hat{\Sigma}_q \mathbf{D}_T^{-1})^{-1} - \mathbf{I}_N \right) \\
&\rightarrow_w \text{tr} \left(\mathbf{C} \begin{bmatrix} \Sigma_{v_1 v_1} & \mathbf{0} \\ \mathbf{0} & \int_0^1 \mathbf{B}_2^p(s) \mathbf{B}_2^p(s)' ds \end{bmatrix} \mathbf{C}' \right. \\
&\quad \times \left. \left(\mathbf{C} \begin{bmatrix} \Sigma_{v_1 v_1} & \mathbf{0} \\ \mathbf{0} & \int_0^1 \mathbf{B}_2^q(s) \mathbf{B}_2^q(s)' ds \end{bmatrix} \mathbf{C}' \right)^{-1} - \mathbf{I}_N \right) \\
&= \text{tr} \left(\begin{bmatrix} \mathbf{I}_{N_1} & \mathbf{0} \\ \mathbf{0} & \int_0^1 \mathbf{B}_2^p(s) \mathbf{B}_2^p(s)' ds \left(\int_0^1 \mathbf{B}_2^q(s) \mathbf{B}_2^q(s)' ds \right)^{-1} \end{bmatrix} - \mathbf{I}_N \right) \\
&= \text{tr} \left(\int_0^1 \mathbf{W}_2^p(s) \mathbf{W}_2^p(s)' ds \left(\int_0^1 \mathbf{W}_2^q(s) \mathbf{W}_2^q(s)' ds \right)^{-1} - \mathbf{I}_{N_2} \right) \tag{7}
\end{aligned}$$

as $T \rightarrow \infty$, with obvious definitions of $\mathbf{B}^q(s)$ and $\mathbf{W}^q(s)$.

Let us now consider the behavior of the test statistics under the alternative that the rank $r_1 = 0$, which corresponds to the conventional stationary alternative hypothesis considered for example by Levin *et al.* (2002) for cross-sectionally independent panel data. Given that $\frac{1}{\sqrt{T}} \hat{\mathbf{R}}_{1t}^p \rightarrow_w \mathbf{B}_1^p(s)$ and because $\mathbf{C} = \mathbf{C}_1 = \mathbf{I}_N$ under this alternative, we obtain $\hat{\Omega}_p = \frac{2}{T^2} \sum_{t=1}^T \hat{\mathbf{R}}_{1t}^p \hat{\mathbf{R}}_{1t}^{p'} \rightarrow_w 2 \int_0^1 \mathbf{B}_1^p(s) \mathbf{B}_1^p(s)' ds$ and $\hat{\Sigma}_p = \frac{1}{T} \sum_{t=1}^T \hat{\mathbf{W}}_{1t}^p \hat{\mathbf{W}}_{1t}^{p'} \rightarrow_p \Sigma_{v_1 v_1}$, and therefore

$$T MB \rightarrow_w \text{tr} \left(\int_0^1 \mathbf{B}_1^p(s) \mathbf{B}_1^p(s)' ds \Sigma_{v_1 v_1}^{-1} \right),$$

so that $MB = O_p(T^{-1})$, while

$$MJ \rightarrow_w \text{tr}(\Sigma_{v_1 v_1} \Sigma_{v_1 v_1}^{-1} - \mathbf{I}_N) = 0.$$

Hence, in this case both statistics degenerate to zero under the alternative. For any other alternative $0 < r_1 < r_0$, the statistics converge to the trace of similar random matrices as under the null, but with dimensions corresponding to the rank under the alternative. Accordingly, these test statistics can be used flexibly to test any value for the rank. It should be noted, however, that it is only when the alternative is taken to be $r_1 = 0$ that MB and MJ are consistent. For all other alternatives, while the tests retain power, the tests are not consistent in the sense that power does not go to one asymptotically.

These observations in turn lead us to consider a variant of the MB statistic, that is a consistent test of the null hypothesis of full rank against any alternative $0 \leq r_1 < N$. Specifically, the following multivariate analog of the inverse of the Breitung (2002) statistic will be considered:

$$MIB = 2T \operatorname{tr}(\hat{\Sigma}_p \hat{\Omega}_p^{-1}).$$

Under the full rank null hypothesis it follows directly from our previous results that

$$MIB \rightarrow_w \operatorname{tr} \left(\int_0^1 \mathbf{W}_2^p(s) \mathbf{W}_2^p(s)' ds \left(\int_0^1 \mathbf{Q}_2^p(s) \mathbf{Q}_2^p(s)' ds \right)^{-1} \right).$$

Unlike the MB statistic, the MIB statistic diverges should the rank be less than full. To show this property, we first rewrite MIB equivalently as

$$MIB = 2T \sum_{i=1}^N \hat{\lambda}_i,$$

where $\hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_N$ are the eigenvalues of the matrix $\hat{\Sigma}_p \hat{\Omega}_p^{-1}$ arranged in descending order. Suppose that the rank is less than full in which case $N_1 > 0$. Then, the eigenvalues $\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_{N_1}$ correspond to the stationary components and it follows that

$$\frac{1}{T} MIB = 2 \sum_{i=1}^{N_1} \hat{\lambda}_i + o_p(1) \rightarrow_w 2 \sum_{i=1}^{N_1} \lambda_i = \operatorname{tr} \left(\Sigma_{v_1 v_1} \left(\int_0^1 \mathbf{B}_1^p(s) \mathbf{B}_1^p(s)' ds \right)^{-1} \right),$$

where λ_i is an eigenvalue of the matrix $\Sigma_{v_1 v_1} \left(\int_0^1 \mathbf{B}_1^p(s) \mathbf{B}_1^p(s)' ds \right)^{-1}$. Thus, $MIB = O_p(T)$, suggesting that, unlike the other tests, MIB is consistent against all alternatives $r_1 < N$, and not just against $r_1 = 0$, which we can exploit.

A further modification is required in case that the null rank tested is less than N , since in this case the first N_1 eigenvalues diverge and so will the MIB statistic. To circumvent this problem we modify the statistic to

$$MMIB = 2T \sum_{i=N-r_0+1}^N \hat{\lambda}_i,$$

which in fact coincides with the Λ_q statistic studied by Breitung (2002). Note that $MIB = MMIB$ when $r_0 = N$. Under the null hypothesis that the rank is r_0 , it follows that

$$MMIB \rightarrow_w \operatorname{tr} \left(\int_0^1 \mathbf{W}_2^p(s) \mathbf{W}_2^p(s)' ds \left(\int_0^1 \mathbf{Q}_2^p(s) \mathbf{Q}_2^p(s)' ds \right)^{-1} \right),$$

whereas $MMIB = O_p(T)$ if $r_1 < r_0$.

The eigenvalue form of the *MMIB* statistic suggests a natural procedure that can be implemented to determine the rank of $\Omega_{\Delta u \Delta u}$. The idea is to proceed as is commonly done in cointegration or common trends testing, as outlined in e.g. Johansen (1995), by successively testing down the rank of $\Omega_{\Delta u \Delta u}$ using a sequence of *MMIB* statistics. Specifically, one begins by testing the full rank null hypothesis $r_0 = N$. If this null hypothesis is not rejected, one concludes that all the cross-sectional units are unit root non-stationary and furthermore non-cointegrated. On the other hand, if this initial null hypothesis is rejected, the sequential testing proceeds by testing $r_0 = N - 1$, this time using the *MMIB* statistics based on only the $N - 1$ smallest eigenvalues. The testing then continues by sequentially dropping the largest eigenvalue until the null hypothesis cannot be rejected, or until zero rank is reached.

Note that although the same sequential procedure can in principle also be applied to the *MB* and *MJ* tests, this is generally not recommended. The reason is that the resulting procedures will be only able to discriminate between full and zero rank with unit asymptotic power, but will have asymptotically diminished power for intermediate cases.⁴

3.2 Critical Values For the Rank Tests

To incorporate the dependence of the asymptotic distributions on N_2 , response surface regressions were used to obtain 5% critical values for each of the tests. We experimented with a variety of specifications and opted for a linear regression model of the form $q = \delta' \mathbf{x} + \eta$, where q is the simulated 5% critical value and η is an error term. The choice of regressors to include was dictated by overall significance subject to the requirement that the R^2 of the regression be no smaller than 0.999. The set of regressors that we retained for the *MIB*, *MMIB* and *MJ* tests is $\mathbf{x} = (1, N_2^{1/4}, \sqrt{N_2}, N_2, N_2^2, N_2^3, \frac{N_2^2}{T}, \frac{N_2^3}{T}, \frac{1}{T}, \frac{1}{T^2}, \frac{N_2}{T^2}, \frac{N_2^2}{T^2})'$, while for the *MB* test, $\mathbf{x} = (1, \frac{1}{N_2^{1/4}}, \frac{1}{\sqrt{N_2}}, \frac{1}{N_2}, \frac{1}{N_2^2}, \frac{1}{N_2^3}, \frac{1}{TN_2^2}, \frac{1}{TN_2^3}, \frac{1}{T}, \frac{1}{T^2}, \frac{1}{T^2N_2}, \frac{1}{T^2N_2^2})'$. The simulated critical values are based on 1,000 draws from the limiting distribution of each of the three test statistics with normal random walks of dimension $N_2 = 1, 2, \dots, 50$ and length $T = \max\{30, 2N_2\}, \max\{30, 2N_2\} + 5, \dots, 300$ in place of the vector Brownian motion $\mathbf{W}(s)$. This implies that there are a total of 2,165 observations available for each regression. The resulting estimated response surface coefficients are reported in the top panel of Table 1.

Unreported simulation results suggest that the fit of the response surface regressions can

⁴Another possibility is to consider maximum eigenvalue type statistics. However, unreported simulation results suggest that the trace statistics perform better in small samples, and we therefore only consider these further.

be poor when N_2 is close to the sample endpoints. To compensate for this we simulate critical values for the values of $N_2 = 1, 2, \dots, 5$ and $T = 1,000$. These are reported in the bottom panel of Table 1.

3.3 First-generation Analogues as a Special Case

Should it be reasonable to assume the data are cross-sectionally independent, we can impose this restriction on our trace statistics to obtain first-generation analogs. Although circumstances in which this assumption is expected to be justifiable are likely to be rare, exploiting cross-sectional independence leads to simplified test statistics.

As is typical for first-generation tests, the null and alternative hypotheses are formulated as $r_0 = N$ versus $r_1 < N$. Thus, because $\mathbf{C} = \mathbf{C}_2 = \mathbf{I}_N$ under this null, $\frac{1}{\sqrt{T}}\hat{\mathbf{u}}_t^p \rightarrow_w \boldsymbol{\Omega}_{vv}^{1/2}\mathbf{W}^p(s)$ as $T \rightarrow \infty$, where $\boldsymbol{\Omega}_{vv} = \boldsymbol{\Omega}_{v_2v_2} = \text{diag}(\omega_{v_1}, \dots, \omega_{v_N})$ in case that cross-sectional dependence is absent. The diagonal nature of $\boldsymbol{\Omega}_{vv}$ suggests that nuisance parameter free test statistics can be constructed simply using sums of unit-specific variance ratios. Consider as an example the *MB* statistic. An easy way to eliminate the dependence on $\omega_{v_i}^2$ is to take the ratio before summing over the cross-sectional dimension. This gives rise to the following *between* version of the *MB* statistic:

$$BMB = \frac{1}{2TN} \sum_{i=1}^N \frac{\hat{\omega}_{ip}^2}{\hat{\sigma}_{ip}^2}.$$

By using similar steps as before, the limiting distribution under the null hypothesis as $T \rightarrow \infty$ can be shown to be

$$BMB \rightarrow_w \frac{1}{N} \sum_{i=1}^N \frac{\int_0^1 (Q_i^p(s))^2 ds}{\int_0^1 (W_i^p(s))^2 ds},$$

which depends on N , but where the individual limiting random variables are otherwise independent and identically distributed with constant mean and variance, written in a generic notation as μ and σ^2 , respectively, with the dependence on p suppressed. This result is convenient because it lends itself to simple large- N asymptotics. In particular, by nontrivial extensions of the sequential limit theory discussed in Phillips and Moon (1999), one could formally establish that

$$\frac{\sqrt{N}(BMB - \mu)}{\sigma} \rightarrow_w N(0, 1)$$

as $T \rightarrow \infty$ and then $N \rightarrow \infty$. Under the alternative hypothesis the same statistic is $o_p(1)$.

Another possibility is to sum over the cross-sectional dimension before taking the ratio,

which results in the following *within* type *MB* statistic:

$$WMB = \frac{1}{2T} \frac{\sum_{i=1}^N \hat{\omega}_{ip}^2}{\sum_{i=1}^N \hat{\sigma}_{ip}^2},$$

which also attains a limiting normal distribution after appropriate mean and variance normalization. Similar arguments apply to the normalized within and between versions of the *MJ* and *MIB* statistics, which are constructed in an analogous fashion. The appropriate mean and variance adjustment terms, obtained from simulations based on 100,000 draws of scalar Brownian motions of length $T = 1,000$, are provided in Table 2.

Note that, in contrast to the general form of our tests, because we are taking N to infinity, the asymptotic distributions of the first-generation analogues do not depend N , suggesting that we have only one set of critical values. However, this advantage comes at a relatively high price. Firstly, cross-sectional independence has to be assumed, at least conditional on time effects. Secondly, the tests can only be used to test the null hypothesis of full rank. Thirdly, the large- N limiting normal distribution may provide a poor approximation in panels where N is only small to moderately large. In Section 4 we will show that there are no power losses when using the more robust trace statistics relative to existing first-generation panel unit root tests even when there is no cross-sectional dependence in the data. For these reasons, we see no real practical advantage for using the first-generation analogs of our trace statistics.

4 Local Power

In this section we consider the local asymptotic power of our tests and compare it with that of some popular first-generation panel unit root tests. In particular, we consider the local alternative of N_2 roots close to unity as discussed in Phillips (1988), which amounts to replacing $\Delta \mathbf{w}_{2t} = \mathbf{v}_{2t}$ in (2) with

$$\Delta \mathbf{w}_{2t} = \frac{1}{T} \mathbf{c} \mathbf{w}_{2t-1} + \mathbf{v}_{2t}, \quad (8)$$

where \mathbf{c} is a $N_2 \times N_2$ drift parameter matrix that measures the extent of the deviation from the null of rank N_2 . If $\mathbf{c} = \mathbf{0}$, then $\Delta \mathbf{w}_{2t} = \mathbf{v}_{2t}$ and we are back in the scenario with N_1 stationary and N_2 non-stationary and non-cointegrated series. If, on the other hand, $\mathbf{c} = \text{diag}(c_1, \dots, c_{N_2})$, then the series in \mathbf{w}_{2t} may be unit root non-stationary, locally stationary, or

even locally explosive, depending on whether c_i is zero, negative or positive, respectively. It is also possible to specify \mathbf{c} as a non-diagonal but nonzero matrix, in which case the series in \mathbf{w}_{2t} may be near-integrated of different orders. By using the invariance principle for near-integrated processes given in Phillips (1988, Lemma 3.1), we obtain

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor sT \rfloor} \mathbf{v}_{2t} \rightarrow_w \Omega_{v_2 v_2}^{1/2} \mathbf{J}_c(s)$$

as $T \rightarrow \infty$, where $\mathbf{J}_c(s) = \int_0^s \exp((s-r)\mathbf{c}) d\mathbf{W}_2(r)$ is a standard vector diffusion or Ornstein-Uhlenbeck process. This means that in order to obtain the local power functions of the rank statistics the process $\mathbf{W}_2(s)$ in the limiting null distributions should be replaced by $\mathbf{J}_c(s)$. For example, in case of the MB statistic,

$$MB \rightarrow_w \text{tr} \left(\int_0^1 \mathbf{K}_c^p(s) \mathbf{K}_c^p(s)' ds \left(\int_0^1 \mathbf{J}_c^p(s) \mathbf{J}_c^p(s)' ds \right)^{-1} \right) \quad (9)$$

as $T \rightarrow \infty$, where $\mathbf{K}_c^p(s) = \int_0^s \mathbf{J}_c^p(r) dr$ with $\mathbf{J}_c^p(r)$ being the detrended version of $\mathbf{J}_c(r)$.⁵

It is interesting to compare the power of our rank tests with the power of some of the existing first-generation panel unit root tests for cross-sectionally independent data. Intuitively, because these tests impose diagonality on $\Omega_{\Delta u \Delta u}$ we might expect them to have higher power when this restriction is correct. To examine this, suppose therefore that $\Omega_{vv} = \mathbf{I}_N$, so that the trend coefficients in α_p are the only nuisance parameters. The drift parameters are homogenous such that $\mathbf{c} = c\mathbf{I}_N$. Under these assumptions it can be shown that the Im *et al.* (2003) and Levin *et al.* (2002) statistics, henceforth denoted *IPS* and *LL*, respectively, have the following local power functions as $T \rightarrow \infty$:

$$\begin{aligned} IPS &\rightarrow_w \frac{1}{\sigma\sqrt{N}} \sum_{i=1}^N \left(c \sqrt{\int_0^1 (J_{ci}^p(s))^2 ds} + \frac{\int_0^1 J_{ci}^p(s) dW_{2i}(s)}{\sqrt{\int_0^1 (J_{ci}^p(s))^2 ds}} - \mu \right), \\ LL &\rightarrow_w \frac{1}{\sigma} \left(c \sqrt{\int_0^1 \mathbf{J}_c^p(s)' \mathbf{J}_c^p(s) ds} + \frac{\int_0^1 \mathbf{J}_c^p(s)' d\mathbf{W}_2(s)}{\sqrt{\int_0^1 \mathbf{J}_c^p(s)' \mathbf{J}_c^p(s) ds}} - \mu \right), \end{aligned}$$

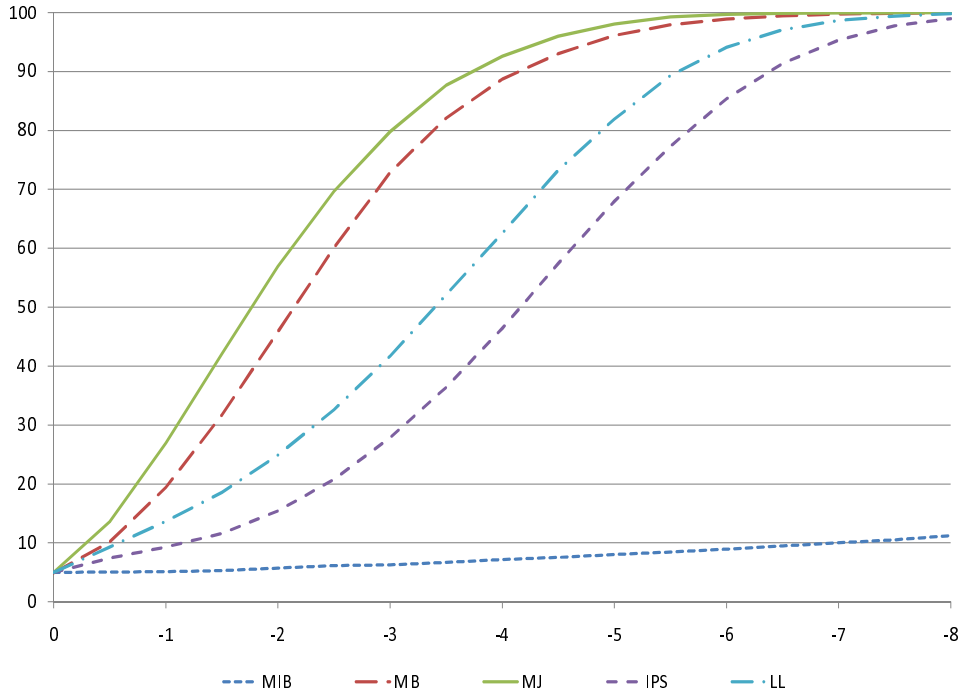
where, as before, μ and σ^2 are certain mean and variance adjustment terms, and $J_{ci}^p(s)$ and $dW_{2i}(s)$ are the units of $\mathbf{J}_c^p(s)$ and $d\mathbf{W}_2(s)$, respectively.

Given the local power functions that we have derived, the asymptotic local power can be simulated using methods similar to those used to obtain the asymptotic critical values, that

⁵Note that in the scalar case the above limiting distribution coincides with the one given in Appendix B of Breitung (2002) for his \hat{q}_T test.

is, by using simulated diffusions in place of $J_c(s)$. The results for $N = 10$ and varying c are reported in Figure 1 for the case when $p = 0$ and in Figure 2 for the case when $p = 1$.⁶

Figure 1: Local asymptotic power of five panel unit root tests for $p = 0$.



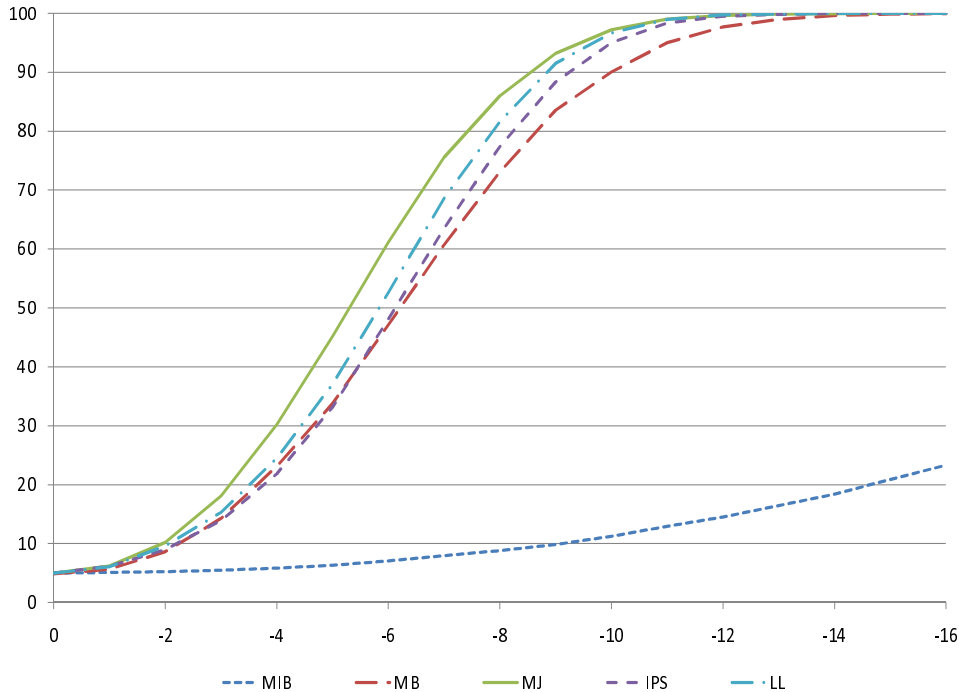
Note: The horizontal axis displays the parameter c .

The first thing to notice is that the knowledge concerning the absence of cross-sectional dependence does not seem to be very helpful in improving the relative power of the *IPS* and *LL* tests. In fact, on the contrary, we see that the *MJ* test is uniformly more powerful than the other tests even under cross-sectional independence, and that the difference in power can sometimes be substantial, especially when c is close to zero. Take for example the case when $p = 0$ and $-3 \leq c < 0$, in which the power of *MJ* is almost twice as large as that of the *LL* test and many times more than that of the *IPS* test. Of course, power gains are less impressive for more distant alternatives, but nevertheless the *MJ* test continues to dominate the others.

As for the other tests we see that while *MB* ends up in second place when $p = 0$, when $p = 1$, *LL* is more powerful. The *LL* test in turn dominates the *IPS* test, which is to be

⁶The results for other values of N were very similar, and are therefore not presented.

Figure 2: Local asymptotic power for five panel unit root tests for $p = 1$.



Note: See note to Figure 1.

expected given the homogenous specification of the alternative hypothesis used here. The *MIB* test is least powerful, and only rarely rejects more than 5% of the time. We also see that power is generally lower when there is a trend in the model, which is in agreement with the well-known incidental trends problem, see Moon *et al.* (2007).⁷

Summarizing this section, we find that, with the exception of the *MIB* test, the new rank tests generally enjoy good local asymptotic power, and that they compare favorably against the *IPS* and *LL* tests. These results appear to be quite robust, and extend to all values of N considered. It should also be noted that because the results are asymptotic, the adverse effect that lag augmentation has on power is not accounted for. The rank tests are therefore expected to compare even more favorably in small samples, especially when high augmentation lags are needed for the *IPS* and *LL* tests.

⁷Strictly speaking, since N is fixed here the theory of the incidental trends problem does not apply, and therefore the radial order of the shrinking neighborhoods around unity for which asymptotic power is nonnegligible should not be affected. However, there might still be small-sample effects.

5 Comparison to Factor Model Approaches

A very popular approach to model cross-sectional dependence in time series panels is to use approximate factor models, as put forward by Bai and Ng (2004). The underlying assumption is that the process \mathbf{u}_t has the following representation

$$\mathbf{u}_t = \mathbf{\Lambda}' \mathbf{f}_t + \mathbf{e}_t, \quad (10)$$

where \mathbf{f}_t is an m -dimensional vector of common factors with $\mathbf{\Lambda}$ being the associated matrix of loading coefficients, here assumed to be non-random. Together $\mathbf{\Lambda}' \mathbf{f}_t$ represent the common component of \mathbf{u}_t , while \mathbf{e}_t represents the idiosyncratic component. By assuming that the units of \mathbf{e}_t are independent of each other and also of the common factors, it is possible to decompose the long-run covariance matrix of $\Delta \mathbf{u}_t$ as

$$\mathbf{\Omega}_{\Delta u \Delta u} = \mathbf{\Lambda}' \mathbf{\Omega}_{\Delta f \Delta f} \mathbf{\Lambda} + \mathbf{\Omega}_{\Delta e \Delta e}, \quad (11)$$

where $\mathbf{\Omega}_{\Delta f \Delta f}$ is of dimension $m \times m$ and $\mathbf{\Omega}_{\Delta e \Delta e}$ is an $N \times N$ diagonal matrix.⁸ The above decomposition (11) highlights the main difference: factor models restrict the set of feasible long-run covariance matrices to those that can be additively decomposed into an $m \times m$ dimensional component (pre- and post-multiplied by the loadings matrix) and a diagonal variance covariance matrix.⁹ Our tests do not require to put any such constraints on the long-run covariance matrix.

On top of the restrictive dependence structure, factor models pose the problem of consistent estimation of both the factors and the loadings, which requires additional assumptions. In classical factor analysis, \mathbf{f}_t and \mathbf{e}_t are generally assumed to be serially and cross-sectionally uncorrelated, which then allows for consistent estimation of $\mathbf{\Lambda}$ as $T \rightarrow \infty$. However, when N is fixed, consistent estimation of \mathbf{f}_t is not generally possible. One way to ensure consistent estimation of both quantities is therefore to assume that N goes to infinity with T . More precisely, because $\mathbf{\Lambda}$ and \mathbf{f}_t are not separately identifiable, the best that one can hope for here is consistent estimation of the spaces spanned by these quantities. That is, instead of estimating $\mathbf{\Lambda}$ and \mathbf{f}_t , one estimates $(\mathbf{R}^{-1})' \mathbf{\Lambda}$ and $\mathbf{R} \mathbf{f}_t$, where \mathbf{R} is an $m \times m$ rotation matrix of

⁸In approximate factor models, such as the one considered by Bai and Ng (2004), the individual idiosyncratic component does not necessarily have to be cross-sectionally independent. For simplicity, however, in this section we keep the cross-sectional independence assumption for the idiosyncratic component.

⁹More generally, such a decomposition and the ensuing restrictions do not only hold for the long-run covariance matrix, but is implied by factor models for the spectral density functions, with the long-run variance, as is well-known, being proportional to the spectral density at frequency 0.

full rank.¹⁰ Identification of the whole factor structure requires not only that $N, T \rightarrow \infty$, but also that $\frac{1}{N}\Lambda'\Lambda$ converges to a positive definite matrix, suggesting that if a variable has only a finite number of nonzero loadings, then it does not qualify as a factor, but is absorbed in the idiosyncratic component.

Hence, the factor model approach not only assumes a particular structure for the cross-sectional dependence, but also imposes other restrictions to ensure that the structure is identified and hence estimable. The requirement that N should go to infinity is especially problematic in the sense that it puts a limit on the practical applicability of the factor-based tests. This is especially true in applied macro and finance, where N is typically relatively small. By contrast, the rank tests investigated in this paper are N -specific and completely nonparametric, and therefore more general in this regard.

It should also be mentioned that some other factor model approaches that are available, such as Moon and Perron (2004), Pesaran (2007), and Phillips and Sul (2003) are even more restrictive. As Bai and Ng (2010) discuss, these other approaches assume that the common and idiosyncratic components have the same order of integration, and are therefore somewhat more restrictive in this regard.

6 Small-sample Performance

In this section, we use Monte Carlo simulations to evaluate the small-sample properties of the new tests, and also compare them with those of some existing tests.

6.1 Simulation Design

The data are generated according to (1)–(3). By assuming that $\alpha_p = \mathbf{0}$, i.e. we consider the case $p = 0$, $\mathbf{C}_1 = [\mathbf{I}_{N_1}, \mathbf{0}]'$ and $\mathbf{C}_2 = [\mathbf{0}, \mathbf{I}_{N_2}]'$, so that the stationary units are ordered first, we have $\mathbf{y}_t = \mathbf{w}_t$. The vector of stationary innovations is assumed to be generated as

$$\begin{bmatrix} \mathbf{w}_{1t} \\ \Delta \mathbf{w}_{2t} \end{bmatrix} = \begin{bmatrix} \mathbf{v}_{1t} \\ \mathbf{v}_{2t} \end{bmatrix} = \begin{bmatrix} \rho \mathbf{I}_{N_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{v}_{1t-1} \\ \mathbf{v}_{2t-1} \end{bmatrix} + \boldsymbol{\eta}_t,$$

with $|\rho| < 1$. Thus, \mathbf{y}_t is generated in a similar way as in Toda (1994). The errors $\boldsymbol{\eta}_t$ are allowed to be both serially and cross-sectionally correlated through $\boldsymbol{\eta}_t = \boldsymbol{\Theta} \boldsymbol{\eta}_{t-1} + \boldsymbol{\epsilon}_t$ with $\boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ and $\boldsymbol{\Theta} = \text{diag}(\theta_1, \dots, \theta_N)$, where θ_i is either set to zero or drawn from

¹⁰Since \mathbf{R} has m^2 free elements, identification of Λ and \mathbf{f}_t requires m^2 restrictions. A common way to accomplish this is to assume that $\frac{1}{T} \sum_{t=1}^T \mathbf{f}_t \mathbf{f}_t' = \mathbf{I}_r$ and that $\Lambda' \Lambda$ is diagonal.

$U(-0.3, 0.3)$. To ensure that Σ is a symmetric positive definite matrix, we follow Chang (2004) and set $\Sigma = \mathbf{PVP}'$, where $\mathbf{V} = \text{diag}(\lambda_1, \dots, \lambda_N)$ is a matrix of eigenvalues such that $\lambda_1 = 0.1$, $\lambda_N = 1$ and $\lambda_2, \dots, \lambda_{N-1} \sim U(0.1, 1)$. Also, $\mathbf{P} = \mathbf{U}(\mathbf{U}'\mathbf{U})^{-1/2}$, where the elements of the $N \times N$ matrix \mathbf{U} are all drawn from $U(0, 1)$. The number of replications is set to 3,000, and for each unit of the panel we generate 100 pre-sample values, starting with an initial value of zero. For brevity, we only reports size and power at the 5% level.¹¹ Some results on the sequential rank selection procedure are also reported.

6.2 Results

Consider first the size results for testing $r_0 = N$, which are reported in Table 3. As expected, we see that the tests perform well with good size accuracy in most experiments. The effect of serial correlation is, however, not completely removed, and some distortions seem to remain, especially for the *MIB* test. However, in most cases that we have considered there is a significant improvement as T increases. Increasing N does not have the same effect, though, which is to be expected given our large- T , fixed- N asymptotic theory.

Table 3 also contains some results for testing $r_0 = 0.5N$ when $\rho = 0.1$. The first thing to notice is the size of the *MIB* test, which is grossly distorted in all experiments considered. The reason is that because the rank under the null is no longer full, as explained in Section 3.1, the *MIB* statistic is now divergent. Being a right-tailed test, this causes *MIB* to reject too often, which is just what we observe. The results for the other tests are, however, more encouraging. In fact, except for the tendency to under reject when N increases, the performance of *MB* and *MJ* remains just as good as before. The *MMIB* test is very conservative.

Consider next the power results when testing the full rank null against $r_1 = N_2 < N$, which are reported in Table 4 for ρ fixed and varying values of N_2 , and in Table 5 for N_2 fixed and varying ρ . Because we are testing the null of full rank, the *MIB* and *MMIB* statistics are identical. The information content of these tables may be summarized as follows:

1. The power of all three tests generally improves as T increases, as N_2 departs from its hypothesized value of N and as ρ moves away from one. We also see that although power is generally increasing in N , this is not always the case.

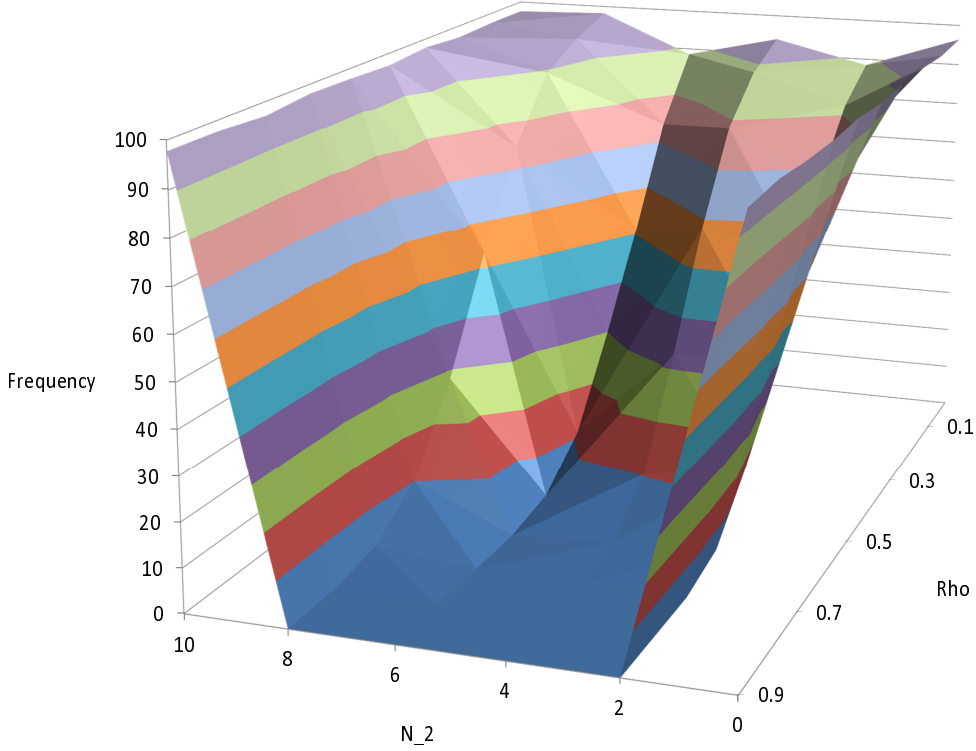
¹¹The power results are not size corrected because such a correction is generally not available in practice. Hence, a test is useful for applied work only if it respects roughly the nominal significance level.

2. *MIB* has power that is complementary to the power of *MB* and *MJ*. When N_2 is close to N and ρ is not close to one, *MIB* generally has higher power than *MB* and *MJ*, and in some cases has substantially higher power. That *MIB* test can detect small deviations in the rank away from the null is expected given its consistency properties. In contrast, when N_2 is close to or equal to zero, *MB* and *MJ* have substantially higher power than *MIB*, an expected finding given the local power depicted in Figures 1 and 2. We see from these patterns that *MIB* is good at detecting small deviations of the rank away from the null when the stationary series in the panel are not persistent, whereas the *MB* and *MJ* statistics are good at detecting small deviations of ρ from one when there are a large number of series that are stationary.
3. The fact that *MB* and *MJ* have power against small deviations of ρ from one when $N_2 = 0$ indicates that they can be used when testing the conventional hypothesis of a unit root versus a fully stationary alternative.
4. Power depends on whether there is a constant or a constant and trend in the model, but power is often higher when the trend is included. This is somewhat unexpected given the theory of the incidental trends problem, although we do see in the bottom panel of Table 5 that power against a fully stationary alternative does fall when the trend is added.¹²
5. As expected, all three tests generally have low power when N_2 is close to the value under the null and ρ is close to one.

We now explore the performance of the *MMIB* test as a rank test. Results concerning correct rank selection frequencies for the sequential *MMIB* test, starting with a rank of $r_0 = N$, are reported in Figure 3 for $T = 100$. Figure 4 gives results for $T = 200$. Both figures are for the case with $p = 0$, $\theta_i = 0$ and $N = 10$. Although we expect a reduction in the accuracy of inference as the true rank becomes more distant to the full rank null, we see that the magnitudes displayed in Figures 3 and 4 can sometimes be substantial. For example, when $T = 100$ and $\rho = 0.9$ the correct selection frequency decreases from about 95% to 0% as N_2 decreases from 10 to 8. However, these magnitudes naturally decrease with ρ and

¹²The drop in power from the inclusion of the trend regressors is also apparent when looking at local asymptotic power depicted in Figures 1 and 2.

Figure 3: Correct rank selection frequency of the *MMIB* test when $T = 100$.

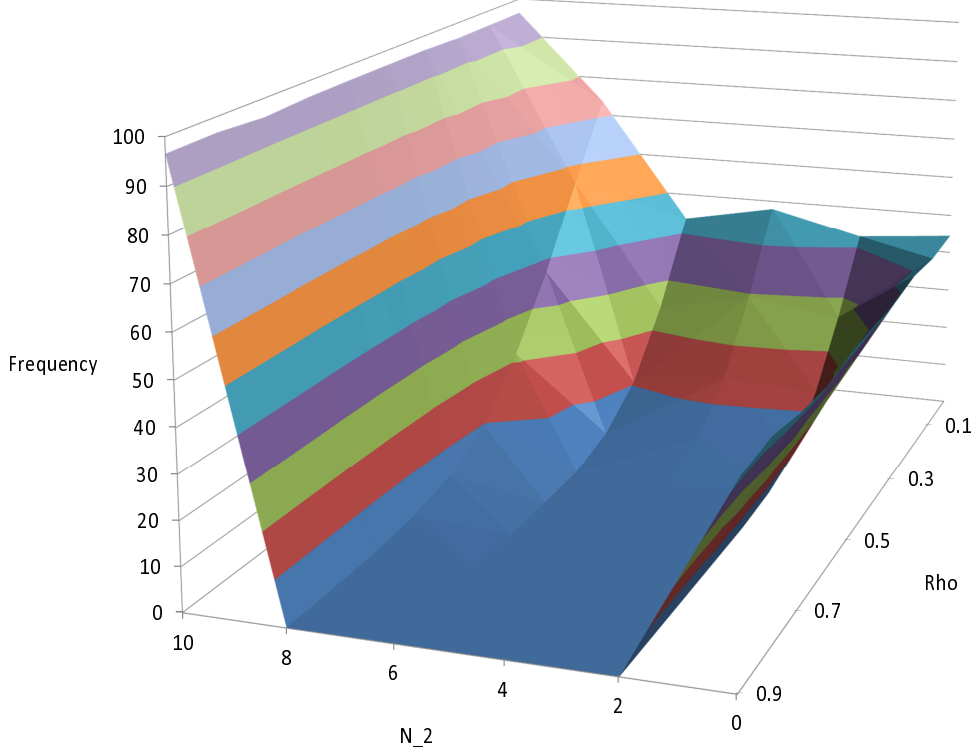


increase with T . Indeed, with $T = 200$ and $\rho = 0$ the correct selection frequency never falls below 75%.

6.3 Comparisons to Some Existing Tests

We have argued that one of the distinctive features of the new tests is that they are very flexible when it comes to the types of hypotheses that can be tested. However, there is another test that shares this feature. This is the test of Ng (2008), which is designed to infer the fraction of units with a unit root, and this makes it interesting as a comparison. However, this test is by far not as general as ours, and cannot, e.g., accommodate cross-unit cointegration. This restriction is clearly necessary, since once (cross-unit) cointegration is allowed, it is not the fraction of series that is non-stationary that is relevant for describing the dynamic properties of the panel of time series but rather the rank, i.e. the number of common trends (compare the discussion in Section 2). Furthermore, unreported simulation results suggest that the small-sample properties of this test can be extremely poor when the test is

Figure 4: Correct rank selection frequency of the *MMIB* test when $T = 200$.



constructed to account for serial correlation of the first-order autoregressive type.¹³ In this section we therefore only report the results for the most simple version of the Ng (2008) test that assumes that the data are both serially and cross-sectionally independent.

The results for the case with $\Theta = \mathbf{0}$ and $\Sigma = \mathbf{I}_N$ are summarized in Table 6. Looking first at the size results we see that even though η_t is serially uncorrelated with unit covariance matrix, the Ng (2008) test is still rather distorted and tends to over-reject. The over-rejections do tend to fall as T increases, but the rate at which this happens is very slow. The new tests perform much better in a majority of the cases considered. When testing the null that the rank is less than full we see that the *MIB* test tends to over-reject but this is to be expected because the test is not designed for this case. In these cases, the *MMIB* statistic is better suited and it has very good power relative to the other tests. The good power of *MMIB*

¹³We found that size and power is always very close to zero when the Ng (2008) test is configured to allow serial correlation. This occurs even when the data has no serial correlation, i.e. $\eta_t \sim N(\mathbf{0}, \mathbf{I}_N)$. While we have carefully checked our code and cannot find an error, because we cannot rule out a programming error, we have chosen not to report results for the Ng (2008) implemented to allow serial correlation.

occurs even though it tends to be conservative. It is very interesting to note that the Ng (2008) test tends to have lower power than the rank tests even though the Ng (2008) test tends to over-reject and power is not size-adjusted.

We now explore the performance of the rank test when there is a factor structure in the data. We compare the rank tests to the panel unit root test of Bai and Ng (2004), which is designed explicitly for the factor case. The data generating process is similar to the one described in Section 6.1. The only difference is that now we also allow for a non-stationary common factor via $\mathbf{y}_t = f_t + \mathbf{w}_t$, where $\Delta f_t \sim N(0, 1)$ is a scalar time series. We consider configurations where \mathbf{w}_t has no dependence and configurations where \mathbf{w}_t features both serial and cross-sectional correlation.

In applying the Bai and Ng factor model approach, we treat the number of factors as unknown. Following the recommendation of Bai and Ng (2002), the number of factors used in the Bai and Ng (2004) test is determined using the IC_{p2} information criterion with the maximum number of factors set to five. The lag length is determined using the Campbell–Perron sequential test rule based on the t -statistic of the last ordered lag. Consistent with the results of Ng and Perron (1995), the maximum number of lags is allowed to increase with T at the rate $4(T/100)^{2/9}$.

Size results are reported in Table 7. We report results in three panels corresponding to the dependence structure of the data. The first panel gives results for the case where the data have neither common factors nor serial or cross-sectional dependence. The second panel shows the results for the case when the data have a common factor but without serial and cross-sectional dependence. The results in the third panel correspond to the case where the data have a common factor as well as serial and cross-sectional dependence. As with the Ng (2008) test, we see that the Bai and Ng (2004) test is oversized although the over-rejections decrease as T increases. The rank tests have sizes that are close to the nominal level although the *MIB* test tends to be somewhat oversized when $N = 20$. Table 8 reports power for case where there is no dependence in the data. The patterns are similar for the other cases. Power of the rank tests is comparable to the Bai and Ng (2004) test and no test dominates the others. While the Bai and Ng (2004) test has higher power in many cases, this largely is an artifact of the over-rejections under the null. Overall the new tests are more robust to dependence in the data than the Bai and Ng (2004) test and have comparable power even for data generated from a simple factor model.

In summary, we find that the new tests show more robustness to dependence in the data and, at the same time, maintain relatively good power in small samples relative to existing tests. We believe that the new rank tests should be a valuable additions to the existing menu of panel unit root tests.

7 Empirical Applications

In this section we briefly present two empirical applications of the tests developed in this study. The first employs a multi-country panel of real exchange rate data to examine purchasing power parity (PPP). The second employs a multi-country panel of log per-capita GDP data to test whether income is converging across countries over time.

7.1 Purchasing Power Parity

In the recent panel time series literature, long run PPP is often tested by means of various panel unit root tests applied to panels of real exchange rates data. When using conventional time series tests, rejection of the unit root null is difficult to achieve with real exchange rate time series, while rejections with panel unit root tests are commonly reported. The increased rejection rate is typically attributed to the increased power of the panel tests relative to the analogous time series tests. However, in the presence of cross-sectional dependence that is not adequately accommodated, these reported results may alternatively be due to size distortions rather than increased power. This can occur when the researcher assumes a form of cross-sectional dependence that differs from the true unknown nature of cross-sectional dependence. Another difficulty may arise when the data from only a small number of countries is stationary. This will lead to a rejection in most panel unit root tests, yet it does not imply that PPP holds pervasively in the sample that is being tested.

Both of these issues are well addressed by the panel tests developed in this study. Firstly, in contrast to most other tests, the tests developed in this paper are robust to cross-sectional dependence of very general forms, and do not require us to assume a particular form of dependence. Secondly, when employed as rank test, our approach can determine whether the number of common trends in the sample is large or small, thereby giving an upper bound for the number of countries for which the PPP hypothesis can hold.

Finally, our tests retain power even when unit-specific deterministic trends are estimated. This is not directly relevant for the conventional PPP hypothesis, for which trends are not

employed. But it does become relevant if one wished to test the Balassa–Samuelson hypothesis that PPP failure is likely attributable to the fact that countries with relatively higher productivity growth in traded goods compared to non-traded goods will experience exchange rates that appreciate gradually over time. One interpretation is that this feature should be well captured by a country-specific trend in the real exchange rate, so that under the Balassa–Samuelson variant, real exchange rates should become trend stationary.

The data that we use for this empirical illustration are the same as those used in Wagner (2008), and comprise four panels of monthly bilateral real exchange rates, which are constructed from consumer price indices with the United States dollar as the numeraire currency. A brief description is provided in Table 9.¹⁴ Note that for this application N is fairly small, especially for the Euro area and CEEC panels. This suggests that factor-based approaches are likely to be biased, and that our finite- N approach is likely to be more appropriate.

The results of our tests are reported in the top panel of Table 10. The first thing to notice is that, except when we apply the *MB* test to the world wide panel, there seem to be no violations of the full rank null, suggesting that PPP fails for all countries considered. We also see that this result is the same regardless of whether there is a constant, or a constant and trend in the model, suggesting that, to the extent that productivity differences can be captured by the deterministic trends, the PPP failure cannot be attributed to the Balassa–Samuelson effect. These results are confirmed by the sequential *MMIB* test, which in all four panels leads to a conclusion of full rank, implying that we cannot reject that PPP, as well as its Balassa–Samuelson variant, fails in all countries considered.

7.2 Income Convergence

In this section we pursue a simple illustration taken from the empirical growth and convergence literature. Our analysis is following in large part the interpretation proposed in Evans (1998). Specifically, suppose that y_{it} , the income for country i at time t , is unit root non-stationary. Then the panel is said to exhibit convergence if, for any pair of countries $i \neq j$, $y_{it} - y_{jt}$ is stationary, and that y_{it} and y_{jt} are thereby cointegrated with cointegrating vector $[1, -1]'$. Therefore, by this definition, a necessary condition for convergence is that the rank of the income panel must be one. From the definition of Evans (1998) it furthermore follows

¹⁴See Wagner (2008) for a more detailed description of the data.

that convergence implies that the income data for each individual country must cointegrate with the cross-sectional average, $\bar{y}_t = \frac{1}{N} \sum_{i=1}^N y_{it}$, so that $y_{it} - \bar{y}_t$ is stationary for all i . Consequently, by this equivalent definition a necessary and sufficient condition for convergence is that the panel comprised of the demeaned series $y_{it} - \bar{y}_t$ has rank zero.

This definition for convergence is relatively strict, and one can consider a generalization that allows for the possibility of different convergence clubs among the countries of the sample. Under this interpretation, a necessary condition for a small number of convergence clubs is a rank that is small relative to N . In all cases, a rank that is close in magnitude to N constitutes a failure of both the strict form of convergence as well as club convergence.

Needless to say, the cross-unit cointegration associated with convergence is not the only form of cross-sectional dependence to be expected among countries, and to avoid size distortion, these other unknown forms of dependence must be accommodated. Our tests are well suited for this, since they allow us to test the rank of the panel without having to specify the particular form of the cross-sectional dependence that links countries.

The specific data we use to assess the convergence hypothesis along these lines are taken from Maddison (2007), and comprise annual observations on the log per-capita GDP for 22 countries over the period 1870–2001, see Table 8.¹⁵

The rank test results are reported in the bottom panel of Table 10. Since tests with country-specific trends are not sensible for a convergence test, we report only values for the case with country-specific intercepts. To begin, we test the raw data, y_{it} . While the *MIB* and *MB* tests are able to reject the full rank null, the *MJ* test does not. Thus, the *MJ* test already points to a violation of a necessary condition for convergence in the sense of Evans (1998). For completeness, we also further investigate using a variant of the *MIB* and *MB* tests. Specifically, since they have rejected the full rank 22 null in favor of a reduced rank, we next proceed to test the null of rank 21 against the alternative of rank less than 21. For this, we use the *MMIB* test, which is specifically designed for the case in which the null hypothesis is taken to be less than full rank. Based on sequentially stepping down the null value for the *MMIB* test, we eventually find that we are unable to reject the null of rank 20 against the alternative of rank less than 20 for the panel of raw data y_{it} .

Similarly, when we test for a sufficient condition for convergence based on the cross-

¹⁵The included countries are Australia, Austria, Belgium, Brazil, Canada, Denmark, Finland, France, Germany, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Spain, Sri Lanka, Sweden, United Kingdom, United States, Switzerland and Uruguay.

sectionally demeaned data, $y_{it} - \bar{y}_t$, we find that only the *MB* test rejects the null hypothesis of full rank in favor of reduced rank. The *MIB* and *MJ* tests fail to reject the null hypothesis of full rank, which contradicts both a necessary and sufficient condition for convergence in the sense of Evans (1998).

Consequently, based on these various tests, the rank of the panel appears to be at most only slightly below full rank, which is a contradiction of the conditions required for convergence. The evidence based on this data therefore points to at most only a slightly reduced rank of the panel, thereby suggesting that income convergence must be rejected, and that a small number of convergence clubs is also unlikely to be adequate in characterizing the long-run income dynamics.

8 Conclusions

In this paper we introduce new rank tests for panel data that have a number of advantages when compared to existing panel unit root tests. First, our tests are applicable for data with serial and cross-sectional dependencies with only mild restrictions on the extent of these dependencies. Second, despite this level of generality, the rank tests do not require any treatment of nuisance parameters. Hence, with these tests there is no need for lag augmentation, bandwidth and kernel selection or estimation of common factors. Implementation is therefore very simple. Third, the tests have relatively high power, even when unit-specific trends are included. Fourth, even in the absence of cross-sectional dependence, the tests still retain high power as compared to tests that were designed explicitly for cross-sectionally independent panel data. Thus, there is little or no price paid for the generality of the tests with respect to the treatment of cross-sectional dependence. Finally, because our asymptotic results do not rely on $N \rightarrow \infty$, these tests are ideally suited for typical macro and finance applications where cross-sectional dependence is pervasive yet N is often small.

References

- Bai, J., and S. Ng (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica* **70**, 191–221.
- Bai, J., and S. Ng (2004). A PANIC Attack on Unit Roots and Cointegration. *Econometrica* **72**, 1127–1178.

- Bai, J., and S. Ng (2010). Panel Unit Root Tests With Cross-Section Dependence: A Further Investigation. *Econometric Theory* **26**, 1088–1114.
- Breitung, J. (2002). Nonparametric Tests for Unit Root and Cointegration. *Journal of Econometrics* **108**, 343–363.
- Chang, Y. (2004). Bootstrap Unit Root Tests in Panels with Cross-Sectional Dependency. *Journal of Econometrics* **120**, 263–293.
- Evans, P. (1998). Using Panel Data to Evaluate Growth Theories. *International Economic Review*, **39**, 295–306.
- Im, K. S., M. H. Pesaran, and Y. Shin (2003). Testing for Unit Roots in Heterogeneous Panels. *Journal of Econometrics* **115**, 53–74.
- Johansen, S. (1995). *Likelihood Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, Oxford.
- Kiefer, N. M., and T. J. Vogelsang (2002a). Heteroskedasticity-Autocorrelation Robust Testing Using Bandwidth Equal to Sample Size. *Econometric Theory*, **18**, 1350–1366.
- Kiefer, N. M., and T. J. Vogelsang (2002b). Heteroskedasticity-Autocorrelation Robust Standard Errors Using the Bartlett Kernel without Truncation. *Econometrica* **70**, 2093–2095.
- Levin, A., C. F. Lin and C.-S. J. Chu (2002). Unit Root Tests in Panel Data: Asymptotic and Finite Sample Properties. *Journal of Econometrics* **108**, 1–22.
- Maddison, A. (2007). World Population, GDP and Per Capita GDP, 1-2003 AD, [http://www-ggdc.net/maddison](http://www.ggdc.net/maddison).
- Moon, H. R., and B. Perron (2004). Testing for a Unit Root in Panels with Dynamic Factors. *Journal of Econometrics* **122**, 81–126.
- Moon, H. R., B. Perron, and P. C. B. Phillips (2007). Incidental Trends and the Power of Panel Unit Root Tests. *Journal of Econometrics* **141**, 416–459.
- Palm, F. C., S. Smeekes and J.-P. Urbain (2009) Cross-Sectional Dependence Robust Block Bootstrap Panel Unit Root Tests. Forthcoming in *Journal of Econometrics*.

- Park, J. Y. (1990). Testing for Unit Roots and Cointegration by Variable Addition. In Fomby, T. B. and G. F. Rhodes (Eds.) *Advances in Econometrics, Volume 8: Co-Integration, Spurious Regression, and Unit Roots*, 107–133. JAI Press, Greenwich.
- Park, J. Y., and B. Choi (1988). A New Approach to Testing for a Unit Root. CAE Working Paper 88-23, Cornell University.
- Pesaran, M. H. (2007). A Simple Panel Unit Root Test in the Presence of Cross Section Dependence. *Journal of Applied Econometrics* **22**, 265–312.
- Phillips, P. C. B. (1988). Regression Theory for Near-Integrated Time Series. *Econometrica* **56**, 1021–1044.
- Phillips, P. C. B., and S. N. Durlauf (1986). Multiple Time Series Regression with Integrated Processes. *Review of Economic Studies* **53**, 473–496.
- Phillips, P. C. B., and H. R. Moon (1999). Linear Regression Limit Theory for Nonstationary Panel Data. *Econometrica* **67**, 1057–1111.
- Phillips, P. C. B., and V. Solo (1992). Asymptotics for Linear Processes. *Annals of Statistics* **20**, 971–1001.
- Phillips, P. C. B., and D. Sul (2003). Dynamic Panel Estimation and Homogeneity Testing under Cross Section Dependence. *Econometrics Journal* **6**, 217–259.
- Toda, H. Y. (1994). Finite Sample Properties of Likelihood Ratio Tests for Cointegrating Ranks when Linear Trends are Present. *Review of Economics and Statistics* **76**, 66–79.
- Vogelsang, T. J. (1998). Trend Function Hypothesis Testing in the Presence of Serial Correlation. *Econometrica* **66**, 123–148.
- Wagner, M. (2008). On PPP, Unit Roots and Panels. *Empirical Economics* **35**, 229–249.
- Wagner, M., and J. Hlouskova (2010). The Performance of Panel Cointegration Methods: Results from a Large Scale Simulation Study. *Econometric Reviews* **29**, 182–223.

Table 1: Critical values for the nonparametric tests.

Value	$p = 0$			$p = 1$		
	MIB	MB	MJ	MIB	MB	MJ
Response surface coefficients for $5 < N < 50$						
δ_1	169037.15804	0.28636	45.12093	173133.91134	0.02454	4.86511
δ_2	-375970.47567	-0.74096	-156.67769	-386036.11573	0.25342	-33.63947
δ_3	236742.44174	1.30607	88.62252	242753.06689	-0.42490	18.71568
δ_4	-28714.71985	-1.55572	-0.26967	-29227.10664	0.09250	1.66408
δ_5	462.26011	1.08594	0.68685	499.44273	0.14925	0.23377
δ_6	5.07971	-0.37162	-0.00337	4.95667	-0.09111	-0.00119
δ_7	12978.51792	0.59838	-61.11665	11680.94595	0.88073	-23.24877
δ_8	-800.55043	-0.63492	1.15489	-823.41937	-0.97470	0.44045
δ_9	-531706.18433	0.06527	3193.19031	-435669.56662	0.04085	1199.08007
δ_{10}	34932714.04854	4.42248	-23173.56365	33299454.36837	5.13046	-10069.57330
δ_{11}	-7336758.89483	-34.18132	-25072.41143	-7720478.56401	-35.78908	-9009.64022
δ_{12}	286553.86002	29.48598	3577.85398	334056.64067	32.62937	1349.07676
Specific critical values for $N \leq 5$						
$N = 5$	2261.66016	0.10545	32.75513	3397.36193	0.02813	13.62580
$N = 4$	1393.16502	0.08999	22.98579	2235.34896	0.02254	9.63480
$N = 3$	751.72488	0.06450	13.96212	1329.77879	0.01639	6.16500
$N = 2$	323.89602	0.03505	7.01714	697.89714	0.00994	3.17750
$N = 1$	101.56031	0.00985	1.90737	289.02594	0.00346	0.91449

Notes: The regression model is given by $q = \delta'x + \varepsilon$, where q is the 5% critical value and x is given in Section 4.4. The estimated 5% critical value is computed as the fitted value of the regression. The block-columns $p = 0$ and $p = 1$ refer to the model with constant, and constant and linear trend, respectively.

Table 2: Mean and variance adjustment terms for the between and within tests.

Adjustment	Between tests			Within tests		
	<i>BMIB</i>	<i>BMB</i>	<i>BMJ</i>	<i>WMIB</i>	<i>WMB</i>	<i>WMJ</i>
$p = 0$						
μ	33.35913	0.05080	13.38137	14.98190	0.13349	12.33025
σ^2	1358.83596	0.00074	198.41452	51.60810	0.00410	157.30748
$p = 1$						
μ	125.32524	0.01083	4.67848	76.51020	0.02614	4.31645
σ^2	7694.99507	0.00003	15.50459	1569.08434	0.00018	12.53802

Notes: $p = 0$ refers to the model with a constant, while $p = 1$ refers to the model with constant and linear trend. The transformation to a standard normal distribution of, e.g., the *BMB* statistic is given by $\sqrt{N}(BMB - \mu)/\sigma$.

Table 3: Size at the 5% level.

θ_i	T	N	$p = 0$				$p = 1$				
			MIB	$MMIB$	MB	MJ	MIB	$MMIB$	MB	MJ	
$r_0 = N$											
$U(-0.3, 0.3)$	0	100	10	3.4	3.4	5.9	4.4	3.5	3.5	4.5	4.2
		100	20	5.4	5.4	6.5	3.8	5.5	5.5	5.0	3.3
		200	10	2.2	2.2	5.5	5.7	2.3	2.3	4.2	6.1
		200	20	7.1	7.1	5.0	4.6	7.4	7.4	4.4	4.8
		100	10	5.6	5.6	6.3	4.8	6.8	6.8	5.1	4.5
		100	20	15.5	15.5	8.1	3.9	17.4	17.4	7.1	3.7
		200	10	3.1	3.1	5.3	6.3	2.9	2.9	4.7	6.7
		200	20	12.9	12.9	4.9	4.9	13.4	13.4	5.2	4.9
$r_0 = 0.5N$											
$U(-0.3, 0.3)$	0	100	10	100.0	0.5	3.0	4.5	100.0	0.1	2.2	3.7
		100	20	100.0	0.0	0.8	1.3	100.0	0.0	0.2	0.6
		200	10	100.0	0.0	3.0	2.9	100.0	0.1	2.8	2.4
		200	20	100.0	0.0	2.0	3.1	100.0	0.0	1.2	2.6
		100	10	100.0	0.4	3.1	4.5	100.0	0.0	2.1	3.7
		100	20	100.0	0.0	0.9	1.3	100.0	0.0	0.1	0.7
		200	10	100.0	0.0	3.1	3.0	100.0	0.1	2.9	2.7
		200	20	100.0	0.0	2.0	3.0	100.0	0.0	1.2	2.5

Notes: θ_i refers to the autoregressive coefficient in the errors. The block-columns $p = 0$ and $p = 1$ refer to the model with constant, and constant and linear trend, respectively.

Table 4: Power at the 5% level when testing $r_0 = N$ versus $r_1 = N_2 < N$ for varying values of N_2 .

N_2	T	N	$p = 0$			$p = 1$		
			MIB	MB	MJ	MIB	MB	MJ
$\rho = 0.9$								
0.1N	100	10	7.5	96.3	90.3	12.6	71.6	76.1
	100	20	8.3	99.6	98.0	23.3	90.3	90.4
	200	10	24.5	100.0	96.4	25.4	99.7	96.6
	200	20	27.0	100.0	99.4	35.3	100.0	100.0
0.3N	100	10	6.3	69.5	58.4	6.2	40.6	44.1
	100	20	7.9	86.5	74.4	8.2	61.4	59.7
	200	10	18.0	95.3	73.7	15.2	82.7	73.2
	200	20	21.7	99.8	88.7	20.0	97.8	91.4
0.7N	100	10	4.6	17.6	14.2	4.6	12.2	11.1
	100	20	6.0	23.8	16.6	6.5	15.1	12.6
	200	10	6.1	27.6	19.9	5.5	19.7	19.9
	200	20	11.3	41.0	25.5	10.8	30.9	25.9
$\rho = 0$								
0.5N	100	10	100.0	99.5	67.8	100.0	99.5	77.4
	100	20	100.0	100.0	84.9	100.0	100.0	93.3
	200	10	100.0	99.6	74.6	100.0	99.8	84.5
	200	20	100.0	100.0	92.1	100.0	100.0	97.5
0.8N	100	10	99.9	40.6	18.7	99.9	39.5	23.1
	100	20	100.0	84.6	24.5	100.0	81.3	30.2
	200	10	100.0	41.2	22.9	100.0	39.4	28.6
	200	20	100.0	88.4	32.4	100.0	88.4	41.6
0.9N	100	10	90.0	17.3	9.7	85.6	16.0	10.4
	100	20	93.3	38.1	11.1	92.1	33.6	11.4
	200	10	99.8	17.5	13.1	99.8	15.7	14.1
	200	20	100.0	37.3	14.6	100.0	35.9	16.8

Notes: N_2 refers to the number of unit roots under the alternative, while ρ refers to the autoregressive coefficient of the remaining stationary units. hypothesis. The block-columns $p = 0$ and $p = 1$ refer to the model with constant, and constant and linear trend, respectively.

Table 5: Power at the 5% level when testing $r_0 = N$ versus $r_1 = N_2 < N$ for varying values of ρ .

ρ	T	N	$p = 0$			$p = 1$		
			MIB	MB	MJ	MIB	MB	MJ
$N_2 = 0.8N$								
0.9	100	10	3.9	12.1	9.0	4.1	8.7	7.7
	100	20	5.9	15.2	10.0	6.2	9.7	8.3
	200	10	4.7	16.0	13.0	4.2	11.4	13.4
	200	20	9.8	21.6	14.9	8.9	17.5	14.9
0.7	100	10	14.9	23.4	12.4	12.9	19.0	13.8
	100	20	13.0	39.6	14.6	12.9	31.0	15.0
	200	10	53.0	29.9	18.0	45.3	25.7	20.6
	200	20	60.3	56.8	21.6	54.0	51.0	26.2
0.5	100	10	53.1	30.9	15.3	46.7	27.8	17.8
	100	20	44.9	59.4	18.2	43.3	52.2	20.3
	200	10	97.4	35.4	20.7	95.9	32.6	24.2
	200	20	99.7	74.1	26.4	99.2	72.1	33.1
$N_2 = 0$								
0.99	100	10	4.1	13.6	15.1	3.5	5.8	5.2
	100	20	5.3	16.2	18.9	5.6	6.5	4.5
	200	10	2.4	26.2	42.0	2.2	6.6	11.2
	200	20	7.8	34.4	61.1	7.5	8.0	11.8
0.95	100	10	4.6	83.6	93.4	4.5	28.6	36.2
	100	20	5.8	96.2	99.5	7.0	43.8	53.7
	200	10	5.9	100.0	100.0	5.3	84.7	97.2
	200	20	12.0	100.0	100.0	11.3	98.3	99.8
0.9	100	10	8.5	100.0	100.0	8.6	87.8	95.6
	100	20	8.4	100.0	100.0	9.6	98.9	99.8
	200	10	26.5	100.0	100.0	25.2	100.0	100.0
	200	20	30.6	100.0	100.0	27.8	100.0	100.0

Notes: See Table 4 for explanations.

Table 6: Size and power comparison with the Ng (2008) test at the 5% level when $p = 0$.

$T = 100$								$T = 200$				
N	N_2	r_0	MIB	$MMIB$	MB	MJ	t	MIB	$MMIB$	MB	MJ	t
Size												
10	10	10	3.8	3.8	5.8	5.2	20.1	2.6	2.6	5.6	6.0	12.5
20	20	20	4.7	4.7	6.5	4.1	18.8	7.8	7.8	5.3	5.1	12.8
10	7	7	100.0	0.4	4.4	5.1	17.4	100.0	0.1	4.0	5.6	10.0
20	14	14	100.0	0.0	1.8	2.0	18.1	100.0	1.5	3.4	4.6	12.1
10	4	4	100.0	16.1	4.1	6.8	14.2	100.0	0.2	4.4	2.7	5.6
20	8	8	100.0	0.0	1.1	1.8	16.5	100.0	0.0	2.3	3.9	9.8
Power												
10	9	10	90.5	90.5	17.4	12.0	21.9	99.8	99.8	18.2	13.2	14.7
20	18	20	94.9	94.9	41.1	12.4	23.5	100.0	100.0	39.3	15.8	18.1
10	7	10	100.0	100.0	72.6	36.8	30.4	100.0	100.0	74.5	42.4	22.9
20	14	20	100.0	100.0	99.5	51.0	37.4	100.0	100.0	99.9	59.9	34.4
10	6	7	100.0	47.3	13.3	13.5	20.5	100.0	88.6	13.5	14.2	13.1
20	12	14	100.0	22.5	19.1	9.6	24.0	100.0	99.9	29.9	17.0	17.7
10	4	7	100.0	100.0	70.3	52.0	30.2	100.0	100.0	74.9	54.1	22.4
20	8	14	100.0	100.0	98.8	58.0	41.5	100.0	100.0	99.9	72.7	40.1

Notes: The values reported in the columns N_2 and r_0 refer to the true and hypothesized rank under the null, respectively. t refers to the Ng (2008) t -test.

Table 7: Size comparison with the Bai and Ng (2004) test at the 5% level when $r_0 = N$.

N	T	$p = 0$				$p = 1$			
		MIB	MB	MJ	$P_{\hat{\epsilon}}$	MIB	MB	MJ	$P_{\hat{\epsilon}}$
No common factor and no error serial or cross-sectional dependence									
10	100	3.8	5.8	5.2	12.7	3.9	4.4	4.6	17.5
20	100	4.7	6.5	4.1	15.0	5.3	5.6	3.7	22.0
10	200	2.6	5.6	6.0	9.9	2.3	5.2	6.7	11.2
20	200	7.8	5.3	5.1	9.4	7.7	3.6	5.1	11.4
Common factor but no error serial or cross-sectional dependence									
10	100	3.3	5.6	5.0	13.3	3.7	4.5	4.6	17.0
20	100	5.5	6.3	3.6	14.4	5.8	4.8	3.6	20.9
10	200	2.3	5.1	6.2	9.3	2.9	4.4	6.9	10.6
20	200	8.0	5.1	5.0	9.6	8.6	4.7	5.3	11.9
Common factor and error serial and cross-sectional dependence									
10	100	4.7	6.0	4.6	15.7	5.2	5.0	4.7	21.0
20	100	13.4	6.7	3.7	19.7	15.8	5.9	3.5	32.8
10	200	2.6	5.5	6.6	12.9	2.7	4.5	6.5	14.4
20	200	12.0	6.2	5.0	12.0	13.0	5.3	5.3	16.4

Notes: $P_{\hat{\epsilon}}$ refers to the Bai and Ng (2004) test. In case of a common factor we add to the data generating process a single unit root factor with unit loadings. The error serial and cross-sectional dependence is generated as described in Section 6.1. See Table 4 for an explanation of the rest of the features.

Table 8: Power comparison with the Bai and Ng (2004) test at the 5% level when $r_0 = N$.

ρ	N	T	$p = 0$				$p = 1$			
			MIB	MB	MJ	$P_{\hat{\epsilon}}$	MIB	MB	MJ	$P_{\hat{\epsilon}}$
$N_2 = 0.5N$										
0.9	10	100	6.0	49.1	46.2	67.1	6.6	29.1	33.8	54.6
	20	100	6.1	67.4	58.9	92.6	7.1	44.8	43.2	80.3
	10	200	12.7	74.2	60.8	96.8	10.9	59.7	62.9	93.8
	20	200	18.0	93.2	75.8	100.0	16.9	85.2	79.7	100.0
0.7	10	100	52.3	90.2	62.1	99.6	46.5	82.9	68.4	99.8
	20	100	40.5	99.6	76.6	100.0	40.0	98.7	84.1	100.0
	10	200	98.8	96.8	70.9	99.9	98.2	95.3	80.2	100.0
	20	200	99.4	100.0	86.5	100.0	98.9	100.0	94.1	100.0
$N_2 = 0$										
0.95	10	100	5.4	84.3	94.3	68.2	5.4	30.5	36.7	43.6
	20	100	6.1	96.1	99.4	91.5	7.1	43.4	52.9	64.5
	10	200	5.7	99.9	100.0	99.8	5.1	84.3	97.4	88.5
	20	200	11.9	100.0	100.0	100.0	11.2	98.3	99.9	99.4
0.9	10	100	10.1	100.0	100.0	99.7	9.8	87.8	96.1	92.1
	20	100	8.8	100.0	100.0	100.0	10.2	98.5	99.8	99.5
	10	200	27.6	100.0	100.0	100.0	24.2	100.0	100.0	100.0
	20	200	30.5	100.0	100.0	100.0	28.6	100.0	100.0	100.0

Notes: The data are generated with no common factor and without error serial or cross-sectional dependence. See Table 4 for an explanation of the rest of the features.

Table 9: Data description.

Panel	Start date	End date	T	N
PPP				
Euro area	1980:1	1998:12	228	11
CEEC	1993:1	2004:6	138	11
Industrial	1980:1	1998:12	228	29
World wide	1981:1	2004:4	280	57
Income convergence				
Maddison	1870	2001	132	22

Table 10: Empirical rank test results.

Panel	$p = 0$			$p = 1$		
	<i>MIB</i>	<i>MB</i>	<i>MJ</i>	<i>MIB</i>	<i>MB</i>	<i>MJ</i>
PPP						
Euro area	14593.7	0.14422	143.1	16713.4	0.04997	71.2
CEEC	10553.7	0.14392	568.2	13389.2	0.04813	291.2
Industrial	202626.8	0.15799	645.6	219843.8	0.05861	350.8
World wide	1260710.9	0.16223*	4393.8	1307459.6	0.06252	1178.1
Income convergence						
Raw data	100010.6*	0.15483*	2724.7			
Cross-section demeaned data	87999.7	0.31084*	1210.2			

Notes: A * superscript denotes significance at the 5% level when testing the null hypothesis of full rank, whereas the block-columns $p = 0$ and $p = 1$ refer to the model with constant, and constant and linear trend, respectively.

Authors: Peter Pedroni, Timothy J. Vogelsang, Martin Wagner, Joakim Westerlund

Title: Nonparametric Rank Tests for Non-stationary Panels

Reihe Ökonomie / Economics Series 270

Editor: Robert M. Kunst (Econometrics)

Associate Editors: Walter Fisher (Macroeconomics), Klaus Ritzberger (Microeconomics)

ISSN: 1605-7996

© 2011 by the Department of Economics and Finance, Institute for Advanced Studies (IHS),
Stumpergasse 56, A-1060 Vienna • ☎ +43 1 59991-0 • Fax +43 1 59991-555 • <http://www.ihs.ac.at>
