# Nonparametric Rank Tests for Non-stationary Panels*

Peter Pedroni
Williams College
Williamstown

Timothy J. Vogelsang
Michigan State University
East Lansing

Martin Wagner
Institute for Advanced Studies
Vienna
&
The Frisch Centre for Economic Research
Oslo

Joakim Westerlund†
University of Gothenburg
Gothenburg

September 13, 2010

**Abstract**

This study develops new rank tests for panels that are unusual in that they account for very general forms of temporal and cross-sectional dependence, including cross-unit cointegration. Yet they do not require any treatment of the associated nuisance parameters, which makes them very easy to implement. The tests also retain high power in small-samples, and, unlike most other panel unit root tests, do not require infinite cross-sections.

*Keywords*: Nonparametric rank tests, unit roots, cointegration, cross-sectional dependence.

*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 World Congress, the Midwest Econometrics Group at the University of Chicago, the Unit Root and Cointegration Conference in Faro, the Conference on Factor Models at Panels at Goethe University, the International 11th International Panel Data Conference, and the NY Econometrics Camp for helpful comments and suggestions. An earlier draft version of this paper circulated with a similar title by the first two authors. Westerlund would like to thank the Jan Wallander and Tom Hedelius Foundation for financial support under research grant P2009–0189:1.

†Corresponding author: Department of Economics, University of Gothenburg, P. O. Box 640, SE-405 30 Gothenburg, Sweden. Telephone: +46 31 786 5251, Fax: +46 31 786 1043, E-mail address: joakim.westerlund@economics.gu.se.

## 1 Introduction

Most, if not all, tests for unit roots in panels are based on the idea of an underlying parametric model, which is almost always assumed to be first-order autoregressive in nature. Take for example the well-known first-generation tests developed by Levin *et al.* (2002) and Im *et al.* (2003), which are designed to test for a unit autoregressive root in the special case when the errors are cross-sectionally uncorrelated. However, they may still be serially correlated, which then calls for some kind of adjustment to get rid of the associated nuisance parameters. The test of Im *et al.* (2003) uses lag augmentation in the spirit of the augmented Dickey–Fuller test, while the test of Levin *et al.* (2002) uses both lag augmentation and long-run variance estimation as in the Phillips–Perron test. The usual rationale for carrying out the adjustment in this way is that if the errors admit to an autoregressive representation of known order, then this can be mimicked in sample.

In practice, of course, one cannot rule out the possibility that the errors are dependent also across the cross-section, in which case the size properties of the first-generation tests are known to be suspect, see for example Wagner and Hlouskova (2006). Numerous attempts have therefore been made in order to relax the rather unrealistic assumption of cross-section independence. One approach is to assume that the dependence can be represented by means of a common time effect, which can be eliminated by simply subtracting the cross-sectional mean from the data. In general, however, with differing pair-wise cross-section correlations, this approach is not expected to work.

The second-generation approach allows for more general types of cross-section dependencies, and can be seen as a response to these considerations. However, as with all parametric approaches, this greater generality creates the need to be precise about the allowable dependencies. Some tests assume that the dependence can be restricted to the contemporaneous covariance matrix of the errors. There is for example the test of O'Connell (1998), which uses seemingly unrelated regressions techniques to allow for a completely unrestricted covariance matrix. However, this requires the time series dimension $T$ being substantially larger than the cross-section dimension $N$, a condition that is rarely fulfilled in practice. Another possibility is therefore to follow for example Bai and Ng (2004), Moon and Perron (2004), Pesaran (2007), and Phillips and Sul (2003), to mention a few, and to assume that the covariances can be modeled using a small number of common factors, which, provided that the true number of factors is known, can be estimated and subtracted from the data. To also account for the presence of serial correlation, some kind of lag augmentation or long-run variance estimation is typically required, either before or after the removal of the factors.

Clearly, whether first- or second-generation, these tests all require the researcher to make an explicit assumption regarding the underlying parametric model. This poses a problem because once one deviates from this assumption, the ensuing tests will no longer be invariant with respect to the data generating process. But this is not all. There is also the problem that the true number of common factors, lag length and optimal bandwidth to use in the long-run variance estimation are never available in practice, and different choices can have a significant impact on test performance.

Motivated by these problems, the purpose of the current paper is to develop new tests, which, unlike most existing panel unit root tests, are based on the cointegrating rank of the $N$-dimensional vector of stacked observations. This makes them suitable for testing a variety of hypotheses and not just that of a common unit root. Another special feature of the new tests is that they are constructed such that one does not have to parametrically specify the underlying model, nor do the statistics involve adjustment factors for the inherent nuisance parameters, which of course significantly reduces both the number and the complexity of the computations required. This is achieved by considering simple variance ratios that eliminate these nuisance parameters from the asymptotic distributions of the test statistics.

Despite their simplicity, the new tests are remarkably general. In fact, except for some mild regulatory conditions, there are no restrictions at all. The data could be both serially and cross-sectionally independent, but they could also involve complex dynamic interrelationships including cointegration and Granger causality. These allowances make our tests some of the most widely applicable.[1] They are also very powerful, even in comparison to first-generation tests that are implemented in absence of cross-sectional dependence, suggesting that the cost of not requiring any prior knowledge regarding the dependence is very low. Moreover, the derivation of the asymptotic distributions of the test statistics as $T \to \infty$ for a fixed $N$, as opposed to joint or sequential asymptotics in which both $T$ and $N$ are taken

[1] In fact, as far as we are aware the only other tests that are comparable ours in terms of generality and ease of parametrization are those of Palm *et al.* (2009). The main difference is that while our tests are completely adjustment-free, the tests of Palm *et al.* (2009) requires bootstrapping, which makes them computationally relatively burdensome. They are also not suitable for testing hypotheses other than that of a common unit root.

to infinity (see Phillips and Moon, 1999), makes the tests valid for any $N$, thus adding to their applicability.

The rest of the paper is organized as follows. Section 2 presents the assumptions, which are used in Section 3 to derive the asymptotic distributions of our rank test statistics. Section 4 then discusses some of the distinctive features of the new tests, and sets them out against the factor-based tests of Bai and Ng (2004), which represents the closest parametric alternative around. The asymptotic properties of the tests are verified using both simulated and real data in Sections 5 and 6, respectively. Section 7 concludes.

## 2 Assumptions

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

$$\mathbf{y}_t = \alpha_p \mathbf{d}_t^p + \mathbf{u}_t, \tag{1}$$

where $\mathbf{d}_t^p = [1, t, \ldots, t^p]'$ with $p \geq 0$ is a polynomial trend function satisfying $\mathbf{d}_t^0 = 1$, with $\alpha_p$ being the associated matrix of trend coefficients. The typical elements of $\mathbf{d}_t^p$ include a constant and a linear time trend, corresponding to $p = 0$ and $p = 1$, respectively, and these are also the specifications considered here.

The main variable of interest is $\mathbf{u}_t = [u_{1t}, \ldots, u_{Nt}]'$, 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 dimension $N \times N_1$ and $N \times N_2$ with $N_2 = N - N_1$, respectively. $\mathbf{C}_1$ gives a basis for the cointegrating space of $\mathbf{u}_t$, while $\mathbf{C}_2$, which is such that is such that $\mathbf{C}_2'\mathbf{C}_1 = \mathbf{0}$ and $\mathbf{C}_1'\mathbf{C}_2 = \mathbf{0}$, gives the common trends. The matrix $\mathbf{C}$ allows us to rotate $\mathbf{u}_t$ into stationary and unit root subsystems as

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

where the first $N_1$ units in $\mathbf{w}_{1t}$ are stationary, while the remaining $N_2$ units in $\mathbf{w}_{2t}$ are non-stationary. As usual, $N_1$ is referred to as the cointegrating rank, while $N_2$ count the number of common trends. The vector of stationary innovations is given by

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

whose long-run covariance matrix is going to be key in this paper. For arbitrary stochastic processes $\mathbf{a}_t$ and $\mathbf{b}_t$ with absolutely summable covariance function we define

$$\boldsymbol{\Omega}_{ab} = \sum_{s=-\infty}^{\infty} E(\mathbf{a}_t \mathbf{b}_{t-s}') = \boldsymbol{\Sigma}_{ab} + \boldsymbol{\Gamma}_{ab} + \boldsymbol{\Gamma}_{ab}',$$

where $\boldsymbol{\Sigma}_{ab} = E(\mathbf{a}_t \mathbf{b}_t)$ and $\boldsymbol{\Gamma}_{ab} = \sum_{s=1}^{\infty} E(\mathbf{a}_t \mathbf{b}_{t-s}')$ 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:

$$\boldsymbol{\Omega}_{vv} = \begin{bmatrix} \boldsymbol{\Omega}_{v_1 v_1} & \boldsymbol{\Omega}_{v_1 v_2} \\ \boldsymbol{\Omega}_{v_2 v_1} & \boldsymbol{\Omega}_{v_2 v_2} \end{bmatrix} = \begin{bmatrix} \omega_{v_1}^2 & \omega_{v_1 v_2} & \dots & \omega_{v_1 v_N} \\ \omega_{v_2 v_1} & \omega_{v_2}^2 & \dots & \omega_{v_2 v_N} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{v_N v_1} & \omega_{v_N v_2} & \dots & \omega_{v_N}^2 \end{bmatrix}.$$

Assumption 1 is enough to obtain our main results.

**Assumption 1.** *As $T \to \infty$,*

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

*where $\boldsymbol{\Omega}_{vv}$ is positive definite, $\to_w$ and $\lfloor x \rfloor$ signify weak convergence and integer part of $x$, respectively, $\boldsymbol{\Omega}_{v_1.v_2} = \boldsymbol{\Omega}_{v_1 v_1} - \boldsymbol{\Omega}_{v_1 v_2} \boldsymbol{\Omega}_{v_2 v_2}^{-1} \boldsymbol{\Omega}_{v_2 v_1}$ and $\mathbf{W}(s) = [W_1(s), \dots, W_N(s)]'$ is an $N$-dimensional vector standard Brownian motion on $s \in [0, 1]$ 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 interested here in the regularity conditions under which it holds. It may be noted, however, that there are a variety of 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 $\alpha$-mixing conditions. Phillips and Solo (1992) give other sets of conditions based on linear processes. In any case, since we are considering the entire $N$-vector this means that we can allow for very general forms of dependencies between the elements of $\mathbf{v}_t$. It also means that we can allow for heterogeneous variances that may vary freely across the elements, but also to some extent across $t$. Needless to say, these allowances represent a much more general consideration than is usually the case in the non-stationary panel literature.

The unit root and cointegration behavior of $\mathbf{u}_t$ is governed by the rank of $\mathbf{\Omega}_{\Delta u \Delta u}$, the long-run covariance matrix of $\Delta \mathbf{u}_t$. This can be seen by studying 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

$$\mathbf{\Omega}_{\Delta u \Delta u} = \mathbf{C}\mathbf{\Omega}_{\Delta w \Delta w}\mathbf{C}' = \mathbf{C}\begin{bmatrix} \mathbf{\Omega}_{\Delta v_1 \Delta v_1} & \mathbf{\Omega}_{\Delta v_1 v_2} \\ \mathbf{\Omega}_{v_2 \Delta v_1} & \mathbf{\Omega}_{v_2 v_2} \end{bmatrix}\mathbf{C}' = \mathbf{C}\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega}_{v_2 v_2} \end{bmatrix}\mathbf{C}' = \mathbf{C}_2\mathbf{\Omega}_{v_2 v_2}\mathbf{C}_2'.$$

The rank of $\mathbf{\Omega}_{\Delta u \Delta u}$ can either full or reduced. If $\text{rk}(\mathbf{\Omega}_{\Delta u \Delta u}) = N$, 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. If, in addition, $\mathbf{\Omega}_{\Delta u \Delta u}$ is diagonal, then we are back in the first-generation unit root scenario with $N$ uncorrelated random walks. If, on the other hand, the rank is reduced, then $\text{rk}(\mathbf{\Omega}_{\Delta u \Delta u}) = N_2 < N$, suggesting that now there are only $N_2$ unit roots. This can be due to genuine stationarity, or cointegration, or a combination. The extreme case being when $\text{rk}(\mathbf{\Omega}_{\Delta u \Delta u}) = 0$, in which $\mathbf{C} = \mathbf{C}_1 = \mathbf{I}_N$, suggesting that now all the elements of $\mathbf{u}_t$ are stationary.

Clearly, the rank of $\mathbf{\Omega}_{\Delta u \Delta u}$ is related to the cointegration rank concept of Johansen (1995). However, there is a conceptual difference in that here cointegration does not refer to cointegration among variables, but rather to cointegration across units. Let us clarify what we mean by this. Denote by $\mathbf{H}$ an $N \times n_1$ 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 $[0, \ldots, 0, 1, 0, \ldots, 0]'$ with the one located at the $i^{\text{th}}$ 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 (Wagner and Hlouskova, 2010). The cross-unit cointegrating rank is the dimension of the space spanned by $\mathbf{D}$. Altogether we have $\text{rk}(\mathbf{\Omega}_{\Delta u \Delta u}) = N - n_1 - \text{rk}(\mathbf{D})$.

# 3 The tests

## 3.1 The rank statistics and their limiting distributions

The cointegration rank tests that we consider are based on two ingredients. One is the asymptotic theory of regressions involving superfluous trend terms (see Park, 1990; Park and Choi, 1988) and the other is long-run variance estimation based on untruncated kernels (see Kiefer *et al.*, 2000).

Let us start by discussing the variance components. In particular, consider the least squares residual

$$\hat{\mathbf{u}}_t^p = \mathbf{y}_t - \sum_{t=1}^{T} \mathbf{y}_t \mathbf{d}_t^{p\prime} \left( \sum_{t=1}^{T} \mathbf{d}_t^p \mathbf{d}_t^{p\prime} \right)^{-1} \mathbf{d}_t^p,$$

whose estimated long-run variance under stationarity is given by

$$\hat{\mathbf{\Omega}}_p = \frac{1}{T} \sum_{j=-M}^{M} k(j/M) \sum_{t=j+1}^{T} \hat{\mathbf{u}}_t^p \hat{\mathbf{u}}_{t-j}^{p\prime} = \begin{bmatrix} \hat{\omega}_{1p}^2 & \hat{\omega}_{12p} & \dots & \hat{\omega}_{1Np} \\ \hat{\omega}_{21p} & \hat{\omega}_{2p}^2 & \dots & \hat{\omega}_{2Np} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\omega}_{N1p} & \hat{\omega}_{N2p} & \dots & \hat{\omega}_{Np}^2 \end{bmatrix},$$

where $k(x)$ is a kernel function and $M$ is the bandwidth. It has been shown by Kiefer *et al.* (2000) that if untruncated such that $M = T$, under stationary, $\hat{\mathbf{\Omega}}_p$ converges to a random variable that is proportional to $\mathbf{\Omega}$ and whose precise shape depend on the choice of $k(x)$. In case of the Bartlett kernel $k(x) = 1 - \frac{|x|}{T}$, Kiefer and Vogelsang (2002) show that the formula for $\hat{\mathbf{\Omega}}_p$ reduces to

$$\hat{\mathbf{\Omega}}_p = \frac{2}{T^2} \sum_{t=1}^{T} \hat{\mathbf{S}}_t^p \hat{\mathbf{S}}_t^{p\prime}.$$

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

The problem is that while the first $N_1$ units of $\hat{\mathbf{w}}_t^p = \mathbf{C}'\hat{\mathbf{u}}_t^p = [\hat{\mathbf{w}}_{1t}^{p\prime}, \hat{\mathbf{w}}_{2t}^{p\prime}]'$, contained in $\hat{\mathbf{w}}_{1t}^p$, are asymptotically stationary, the remaining ones in $\hat{\mathbf{w}}_{2t}^p$ are unit root non-stationary, suggesting that a different normalization with respect to $T$ is needed. Let us therefore introduce the normalization matrix $\mathbf{D}_T = \text{diag}(\mathbf{I}_{N_1}, \sqrt{T}\mathbf{I}_{N_2})$. By using Assumption 1, rotation by $\mathbf{C}$ and standard results for least squares detrended processes it is possible to show that as $T \to \infty$

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

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 $\mathbf{d}^p(s) = [1, s, \ldots, s^p]'$ with $\mathbf{d}^0(s) = 1$. All vectors are partitioned conformably with $\mathbf{C}$.

The contemporaneous variance estimator,

$$\hat{\mathbf{\Sigma}}_p = \frac{1}{T}\sum_{t=1}^{T} \hat{\mathbf{u}}_t^p \hat{\mathbf{u}}_t^{p\prime} = \begin{bmatrix} \hat{\sigma}_{1p}^2 & \hat{\sigma}_{12p} & \ldots & \hat{\sigma}_{1Np} \\ \hat{\sigma}_{21p} & \hat{\sigma}_{2p}^2 & \ldots & \hat{\sigma}_{2Np} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_{N1p} & \hat{\sigma}_{N2p} & \ldots & \hat{\sigma}_{Np}^2 \end{bmatrix},$$

must also normalized in order to achieve convergence:

$$\begin{aligned} \mathbf{D}_T^{-1}\hat{\mathbf{\Sigma}}_p\mathbf{D}_T^{-1} &= \mathbf{C}\frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} \hat{\mathbf{w}}_{1t}^p\hat{\mathbf{w}}_{1t}^{p\prime} & T^{-1/2}\hat{\mathbf{w}}_{1t}^p\hat{\mathbf{w}}_{2t}^{p\prime} \\ T^{-1/2}\hat{\mathbf{w}}_{2t}^p\hat{\mathbf{w}}_{1t}^{p\prime} & T^{-1}\hat{\mathbf{w}}_{2t}^p\hat{\mathbf{w}}_{2t}^{p\prime} \end{bmatrix}\mathbf{C}' \\ &\to_w \mathbf{C}\begin{bmatrix} \mathbf{\Sigma}_{v_1v_1} & \mathbf{0} \\ \mathbf{0} & \int_0^1 \mathbf{B}_2^p(s)\mathbf{B}_2^p(s)'ds \end{bmatrix}\mathbf{C}'. \end{aligned} \tag{5}$$

The convergence results in (4) and (5) imply that asymptotically nuisance parameter free test statistics can be constructed using nothing but appropriately normalized ratios of $\hat{\mathbf{\Sigma}}_p$ and $\hat{\mathbf{\Omega}}_p$. The first test statistic of this type that we will consider can seen as a multivariate version of the $\hat{\varrho}_T$ statistic introduced by Breitung (2002). It is given by

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

The asymptotic distribution of this statistic under the null hypothesis $H_0 : \mathrm{rk}(\mathbf{\Omega}_{\Delta u\Delta u}) = N_2$ is easily derived from the above results. Indeed, by using the cyclical property of the trace,

$$\begin{aligned} MB &= \mathrm{tr}\left(\frac{1}{2T}\mathbf{D}_T^{-1}\hat{\mathbf{\Omega}}_p\mathbf{D}_T^{-1}(\mathbf{D}_T^{-1}\hat{\mathbf{\Sigma}}_p\mathbf{D}_T^{-1})^{-1}\right) \\ &\to_w \mathrm{tr}\left(\mathbf{C}\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \int_0^1 \mathbf{R}_2^p(s)\mathbf{R}_2^p(s)'ds \end{bmatrix}\mathbf{C}'\left(\mathbf{C}\begin{bmatrix} \mathbf{\Sigma}_{v_1v_1} & \mathbf{0} \\ \mathbf{0} & \int_0^1 \mathbf{B}_2^p(s)\mathbf{B}_2^p(s)'ds \end{bmatrix}\mathbf{C}'\right)^{-1}\right) \\ &\to_w \mathrm{tr}\left(\int_0^1 \mathbf{R}_2^p(s)\mathbf{R}_2^p(s)'ds\left(\int_0^1 \mathbf{B}_2^p(s)\mathbf{B}_2^p(s)'ds\right)^{-1}\right) \\ &= \mathrm{tr}\left(\int_0^1 \mathbf{Q}_2^p(s)\mathbf{Q}_2^p(s)'ds\left(\int_0^1 \mathbf{W}_2^p(s)\mathbf{W}_2^p(s)'ds\right)^{-1}\right), \end{aligned} \tag{6}$$

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 will consider is based on the properties of regressions including superfluous 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 $t^{p+1}, \ldots, t^q$ are estimated consistently to zero. A coefficient restriction test like the Wald test is therefore going to have a well-defined limiting distribution in this case, although not necessarily free of nuisance parameters. If $\mathbf{u}_t$ is non-stationary, however, then (1) is spurious and the coefficients corresponding to the superfluous regressors will not go to zero, and this has implications for the asymptotic behavior of the Wald statistic, which is then of order $O_p(T)$. This led Park and Choi (1988) to consider as a unit root test statistic the Wald statistic divided by $T$. Our test statistic can be seen as a multivariate version of this statistic, and is given by

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

where $\hat{\mathbf{\Sigma}}_q$ 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 dropped off. In the current paper we therefore only consider this value of $q$. In any case, similarly to before, under the rank-$N_2$ null,

$$\begin{aligned} MJ &= \text{tr}\left(\mathbf{D}_T^{-1}\hat{\mathbf{\Sigma}}_p\mathbf{D}_T^{-1}(\mathbf{D}_T^{-1}\hat{\mathbf{\Sigma}}_q\mathbf{D}_T^{-1})^{-1} - \mathbf{I}_N\right) \\ \to_w &\quad \text{tr}\left(\mathbf{C}\begin{bmatrix} \mathbf{\Sigma}_{v_1v_1} & \mathbf{0} \\ \mathbf{0} & \int_0^1 \mathbf{B}_2^p(s)\mathbf{B}_2^p(s)'ds \end{bmatrix}\mathbf{C}'\right. \\ \times &\quad \left.\left(\mathbf{C}\begin{bmatrix} \mathbf{\Sigma}_{v_1v_1} & \mathbf{0} \\ \mathbf{0} & \int_0^1 \mathbf{B}_2^q(s)\mathbf{B}_2^q(s)'ds \end{bmatrix}\mathbf{C}'\right)^{-1} - \mathbf{I}_N\right) \\ &= \text{tr}\left(\begin{bmatrix} \mathbf{I}_{N_1} & \mathbf{0} \\ \mathbf{0} & \int_0^1 \mathbf{B}_2^p(s)\mathbf{B}_2^p(s)'ds(\int_0^1 \mathbf{B}_2^q(s)\mathbf{B}_2^q(s)'ds)^{-1} \end{bmatrix} - \mathbf{I}_N\right) \\ &= \text{tr}\left(\int_0^1 \mathbf{W}_2^p(s)\mathbf{W}_2^p(s)'ds\left(\int_0^1 \mathbf{W}_2^q(s)\mathbf{W}_2^q(s)'ds\right)^{-1} - \mathbf{I}_{N_2}\right) \end{aligned} \quad (7)$$

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

A special case of the above results arises when the null is tested against $\text{rk}(\mathbf{\Omega}_{\Delta u \Delta u}) = 0$, which is the conventional stationary alternative hypothesis considered by for example Levin *et al.* (2002). Because $\frac{1}{\sqrt{T}}\hat{\mathbf{R}}_{1t}^p \to_w \mathbf{B}_1^p(s)$ and since $\mathbf{C} = \mathbf{C}_1 = \mathbf{I}_N$ under this alternative, we get $\hat{\mathbf{\Omega}}_p = \frac{2}{T^2}\sum_{t=1}^T \hat{\mathbf{R}}_{1t}^p \hat{\mathbf{R}}_{1t}^{p\prime} \to_w 2\int_0^1 \mathbf{B}_1^p(s)\mathbf{B}_1^p(s)'ds$ and $\hat{\mathbf{\Sigma}}_p = \frac{1}{T}\sum_{t=1}^T \hat{\mathbf{w}}_{1t}^p \hat{\mathbf{w}}_{1t}^{p\prime} \to_p \mathbf{\Sigma}_{v_1v_1}$, and therefore

$$T\,MB \to_w \text{tr}\left(\int_0^1 \mathbf{B}_1^p(s)\mathbf{B}_1^p(s)'ds\,\mathbf{\Sigma}_{v_1v_1}^{-1}\right), \quad (8)$$

or $MB = O_p(T^{-1})$, whereas

$$MJ \to_w \operatorname{tr}(\mathbf{\Sigma}_{v_1v_1}\mathbf{\Sigma}_{v_1v_1}^{-1} - \mathbf{I}_N) = 0. \tag{9}$$

Hence, in this case both statistics degenerate to zero under the alternative. For any other alternative $0 < \operatorname{rk}(\mathbf{\Omega}_{\Delta u\Delta u}) < N_2$, the statistics converge to the trace of the same random matrices as under the null but with a different dimension. Thus, given this randomness, $MB$ and $MJ$ will generally not be consistent, only unbiased.

This leads us to consider the following multivariate inverse Breitung (2002) statistic:

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

whose asymptotic distribution under the null hypothesis is given by

$$MIB \to_w \operatorname{tr}\left(\int_0^1 \mathbf{W}_2^p(s)\mathbf{W}_2^p(s)'ds \left(\int_0^1 \mathbf{Q}_2^p(s)\mathbf{Q}_2^p(s)'ds\right)^{-1}\right), \tag{10}$$

which follows directly from our previous results. The analysis of this statistic under the alternative hypothesis is, however, not as straightforward. In particular, consider testing the null against $\operatorname{rk}(\mathbf{\Omega}_{\Delta u\Delta u}) < N_2$. The problem here is that, unlike $\hat{\mathbf{\Sigma}}_p$, $\hat{\mathbf{\Omega}}_p$ is singular and therefore cannot be inverted. However, note that under the full rank hypothesis, $MIB$ can also be written as $MIB = 2T\sum_{i=1}^N \hat{\lambda}_i$, where $\hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_N$ are the eigenvalues of the matrix $\hat{\mathbf{\Sigma}}_p\hat{\mathbf{\Omega}}_p^{-1}$ arranged in descending order. Suppose now that $\operatorname{rk}(\mathbf{\Omega}_{\Delta u\Delta u}) = N_2 < N$. Then,

$$\frac{1}{T} MIB = 2\sum_{i=1}^{N_1} \hat{\lambda}_i + o_p(1) \to_w 2\sum_{i=1}^{N_1} \lambda_i = \operatorname{tr}\left(\mathbf{\Sigma}_{v_1v_1}\left(\int_0^1 \mathbf{B}_1^p(s)\mathbf{B}_1^p(s)'ds\right)^{-1}\right), \tag{11}$$

where $\lambda_i$ is an eigenvalue of the matrix $\mathbf{\Sigma}_{v_1v_1}(\int_0^1 \mathbf{B}_1^p(s)\mathbf{B}_1^p(s)'ds)^{-1}$. Thus, $MIB = O_p(T)$, suggesting that, unlike the other tests, $MIB$ is consistent against all alternatives $\operatorname{rk}(\mathbf{\Omega}_{\Delta u\Delta u}) < N$, and not just against $\operatorname{rk}(\mathbf{\Omega}_{\Delta u\Delta u}) = 0$, which is of course a great advantage. The problem is that because the first $N_1$ eigenvalues are diverging this statistic cannot be used when the rank under the null is not full. To circumvent this we may use the following modified $MIB$ statistic:

$$MMIB = 2T\sum_{i=N_1+1}^{N} \hat{\lambda}_i,$$

which coincides with the $\Lambda_q$ statistic studied by Breitung (2002).

This eigenvalue interpretation of $MMIB$ suggests a natural procedure that can be applied to determine the rank of $\mathbf{\Omega}_{\Delta u\Delta u}$ from the data. The idea is to proceed as in Johansen (1995) by successively testing down the rank of $\mathbf{\Omega}_{\Delta u \Delta u}$. We begin by applying $MMIB$. If the null of $\text{rk}(\mathbf{\Omega}_{\Delta u \Delta u}) = N$ is not rejected, we conclude that all the cross-sectional units are non-stationary and non-cointegrated, and proceed no further. If the null is rejected, however, the testing proceeds by testing whether $\text{rk}(\mathbf{\Omega}_{\Delta u \Delta u}) = N - 1$ using $MMIB$ based on the $N - 1$ smallest eigenvalues. The testing then continues by sequentially dropping the largest eigenvalue until the null cannot be rejected or zero rank is reached.

The advantage of doing the testing in this way is that if the significance level at each stage is $\alpha$, while the probability of selecting the true rank approaches to $1 - \alpha$, the probability of selecting a smaller rank converges to zero. This property is generic of successive testing procedures, including the Johansen (1995) trace test for the cointegrating rank. The result is a consequence of the fact that the test is applied to the same variable and thus not independent across stages.

Note also that although the same sequential procedure can in principle be applied also to the $MB$ and $MJ$ tests, this is generally not recommended. The reason is that the resulting procedures will only be able to discriminate between full and zero rank, and not between intermediate cases.[2]

### 3.2 Critical values

Response surface regressions are to obtain the 5% critical values for 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 $\eta$ 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 $MIB$ and $MJ$ tests is $\mathbf{x} = (1, N^{1/4}, \sqrt{N}, N, N^2, N^3, \frac{N^2}{T}, \frac{N^3}{T}, \frac{1}{T}, \frac{1}{T^2}, \frac{N^2}{T^2})'$, while for the $MB$ test, $\mathbf{x} = (1, \frac{1}{N^{1/4}}, \frac{1}{\sqrt{N}}, \frac{1}{N}, \frac{1}{N^2}, \frac{1}{N^3}, \frac{1}{TN^2}, \frac{1}{TN^3}, \frac{1}{T}, \frac{1}{T^2}, \frac{1}{T^2 N^2})'$. The simulated critical values are based on making $1,000$ draws from the limiting distribution of each of the three test statistics with normal random walks of dimension $N = 1, 2, \dots, 50$ and length $T = \max\{30, 2N\}, \max\{30, 2N\} + 5, \dots, 300$ in place of the vector Brownian motion $\mathbf{W}(s)$. This means that there is 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.

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

Unreported simulation results suggest that the fit of the response surface regressions can be poor when $N$ is close to the sample endpoints. To compensate for this we simulate critical values for specific values of $N \leq 5$ when $T = 1,000$. These are reported in the bottom panel of Table 1.

### 3.3 First-generation analogues

In this section we ask how the absence of cross-section dependence is going to affect the results obtained so far. Intuitively, independence is expected to simplify not only the asymptotic analysis but also the calculations needed in order to obtain nuisance parameter free test statistics. As usual with first-generation tests, the null and alternative hypotheses are formulated as $H_0 : \text{rk}(\mathbf{\Omega}_{\Delta u \Delta u}) = N$ versus $H_1 : \text{rk}(\mathbf{\Omega}_{\Delta u \Delta u}) < N$. Thus, since $\mathbf{C} = \mathbf{C}_2 = \mathbf{I}_N$ under this null, $\frac{1}{\sqrt{T}}\hat{\mathbf{u}}_t^p \to_w \mathbf{\Omega}_{vv}^{1/2}\mathbf{W}^p(s)$ as $T \to \infty$, where $\mathbf{\Omega}_{vv} = \mathbf{\Omega}_{v_2 v_2} = \text{diag}(\omega_{v_1}, ..., \omega_{v_N})$, suggesting that nuisance parameter free test statistics can be constructed using nothing but simple sums of unit-specific variance estimates. Consider as an example the $MB$ statistic. An easy way to get rid of $\omega_{v_i}^2$ 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 the same steps as before, this statistic has the following limiting null distribution as $T \to \infty$:

$$BMB \to_w \frac{1}{N} \sum_{i=1}^{N} \frac{\int_0^1 (Q_i^p(s))^2 ds}{\int_0^1 (W_i^p(s))^2 ds}, \tag{12}$$

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 $\mu$ and $\sigma^2$, respectively, with the dependence on $p$ suppressed. This result is very convenient, because it lends itself to simple large-$N$ asymptotics. In particular, by using the sequential limit theory discussed in Phillips and Moon (1999),

$$\frac{\sqrt{N}(BMB - \mu)}{\sigma} \to_w N(0, 1) \tag{13}$$

as $T \to \infty$ and then $N \to \infty$. 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. The same applies 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 rank tests, with these test there is just one critical value for all values of $N$. These tests are therefore even simpler to implement. However, this advantage comes at a relatively high price. Firstly, cross-sectional independence has to be assumed. 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.

# 4 Distinctive features

## 4.1 Local power

Instead of investigating the behavior of the test statistics under a fixed formulation of the alternative hypothesis, we can consider the local asymptotic power of the tests. In particular, consider the local alternative of $N_2$ roots close to unity given in Phillips (1988), which amounts to replacing $\Delta\mathbf{w}_{2t} = \mathbf{v}_{2t}$ in (2) with

$$\Delta\mathbf{w}_{2t} = \frac{1}{T}\mathbf{c}\mathbf{w}_{2t-1} + \mathbf{v}_{2t}, \tag{14}$$

where $\mathbf{c}$ is a $N_2 \times N_2$ drift parameter matrix that measure the extent of the deviation from the rank-$N_2$ null. 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 unit root units. If, on the other hand, $\mathbf{c} = \text{diag}(c_1, \ldots, c_{N_2})$, then the units 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 units of $\mathbf{w}_{2t}$ may be near-integrated of different orders. In any case, by using the invariance principle for near-integrated processes given in Phillips (1988, Lemma 3.1), we obtain

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor sT\rfloor}\mathbf{v}_{2t}\to_w \mathbf{\Omega}_{v_2v_2}^{1/2}\mathbf{J}_c(s)$$

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

$$MB\to_w \mathrm{tr}\left(\int_0^1 \mathbf{K}_c^p(s)\mathbf{K}_c^p(s)'ds\left(\int_0^1 \mathbf{J}_c^p(s)\mathbf{J}_c^p(s)'ds\right)^{-1}\right) \tag{15}$$

as $T\to\infty$, 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 $\mathbf{J}_c(r)$.[3]

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 when the cross-section dependence is absent. Intuitively, since these tests make use of the fact that in this case $\mathbf{\Omega}_{\Delta u\Delta u}$ is a diagonal matrix, they should have higher power.

Bowman (1999) studies the exact power of the popular first-generation unit root tests of Im *et al.* (2003) and Levin and Lin (1992) under a fixed alternative. He characterizes the class of admissible panel unit root tests, and shows that in absence of unknown nuisance parameters, while the Im *et al.* (2003) test is inadmissible, the Levin and Lin (1992) test is admissible, and in fact uniformly most powerful when alternative is homogenous. He also shows, via simulations, that these results are not substantially altered by the presence of unknown nuisance parameters, such as deterministic constant and trend terms. When the alternative is heterogeneous Bowman (1999) shows that while the Levin and Lin (1992) test tends to perform better for near-homogenous alternatives, for more heterogeneous alternatives the Im *et al.* (2003) test performs best.

This discussion suggests that the Im *et al.* (2003) and Levin and Lin (1992) tests possess some optimality property, which makes them interesting as a comparison.[4] Suppose therefore that $\mathbf{\Omega}_{vv}=\mathbf{I}_N$, so that the trend coefficients in $\alpha_p$ are the only nuisance parameters. The drift parameters are homogenous such that $\mathbf{c}=c\mathbf{I}_N$. Under these assumptions it can be

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

[4] Of course, being sample-specific, the results of Bowman (1999) are not expected to hold in the present $T$-asymptotic context. However, this is not necessary.

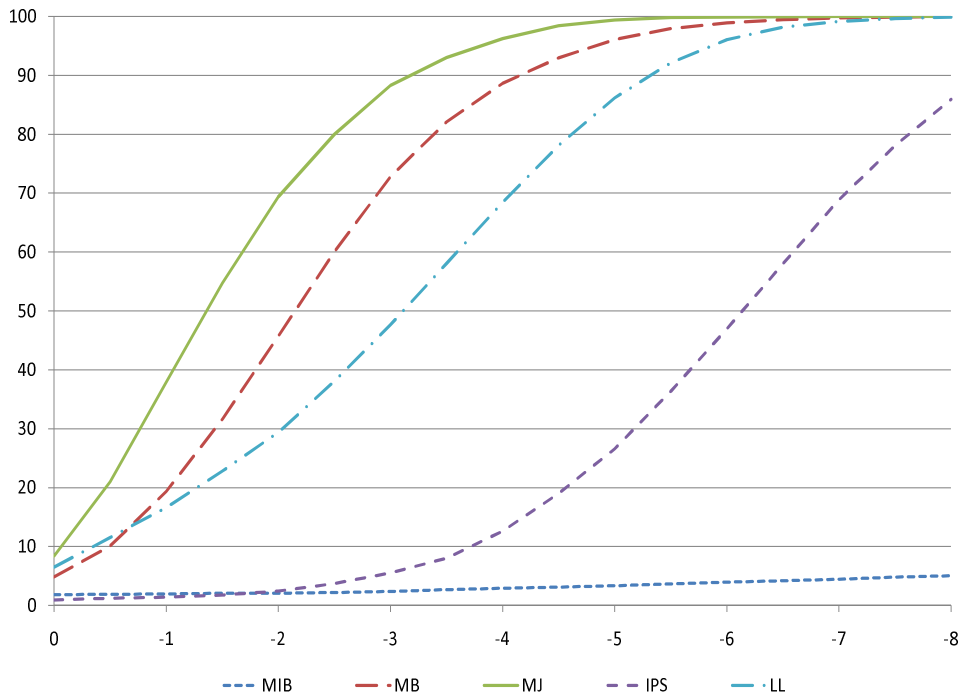
shown that the Im *et al.* (2003) and Levin and Lin (1992) statistics, henceforth denoted $IPS$ and $LL$, respectively, have the following local power functions as $T \to \infty$:

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

where, as before, $\mu$ and $\sigma^2$ 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.

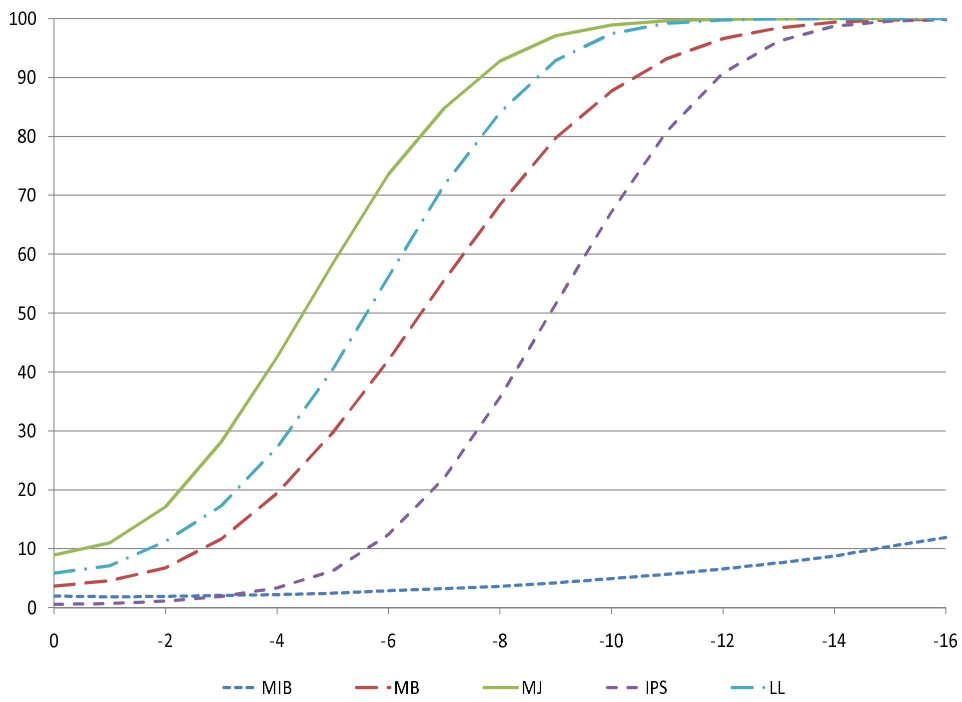
Given the local power functions that we have derived, the asymptotic local power can be simulated using methods similar to those used to obtain the asymptotic critical values, that is, by using simulated diffusions in place of $\mathbf{J}_c(s)$. The results for $N = 10$ and varying $c$ are reported in Figure 1 for the case when $p = 0$ and in Figure 2 for the case when $p = 1$.

Figure 1: Local power for different values of $c$ when $p = 0$.




The first thing to notice is that prior knowledge regarding the cross-section 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,

Figure 2: Local power for different values of $c$ when $p = 1$.




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 $LL$, the best performing first-generation test. Of course, power gains are less impressive for more distant alternatives, but at least it is not possible to do better than the $MJ$ test.

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 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).[5]

Summarizing this section, we find that, except for $MIB$, the 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

[5]Strictly 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.

also be noted that since 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 a high augmentation lag is needed.

### 4.2 Generality

Most panel unit root tests around are designed to test the null hypothesis that all the units are unit root non-stationary versus the alternative that there is at least some units that are trend-stationary, which equivalent to testing the hypothesis of $\text{rk}(\mathbf{\Omega}_{\Delta u \Delta u}) = N$ versus $\text{rk}(\mathbf{\Omega}_{\Delta u \Delta u}) < N$. Clearly, while reasonable in some situations, this hypothesis must be considered as a rather limited consideration. The new rank tests are much more flexible and is suitable for testing a variety of hypotheses, including that of full rank.

Another drawback of most existing tests is the way they handle cross-section dependence. At the one end of the scale we have the first-generation tests, which assume that the dependence is absent altogether. But this approach is not expected to work in general, and second-generation tests that relax the independence assumption have therefore been developed. The factor-based approach of Bai and Ng (2004) is one of the most general. It assumes that $\mathbf{u}_t$ admits to the following components representation:

$$\mathbf{u}_t = \Lambda' \mathbf{f}_t + \mathbf{e}_t, \tag{16}$$

where $\mathbf{f}_t$ is an $r$-dimensional vector of common factors with $\Lambda$ being the associated matrix of loading coefficients, here assumed to be non-random. Together $\mathbf{f}_t$ and $\Lambda$ 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

$$\mathbf{\Omega}_{\Delta u \Delta u} = \Lambda' \mathbf{\Omega}_{\Delta f \Delta f} \Lambda + \mathbf{\Omega}_{\Delta e \Delta e} \tag{17}$$

where $\mathbf{\Omega}_{\Delta f \Delta f}$ is of dimension $r \times r$ and $\mathbf{\Omega}_{\Delta e \Delta e}$ is an $N \times N$ diagonal matrix. This illustrates the main difference when compared to our approach; factor models achieve complexity reduction by assuming that $\mathbf{f}_t$ is the only source of dependence.[6]

[6] In 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.

In terms of allowances, however, the two approaches are very similar. Consider for example the cross-unit cointegrating space. Let us denote by $\mathbf{G} = [\mathbf{G}_1, \mathbf{G}_2]$ an orthogonal $r \times r$ matrix that rotates the factor space into its stationary and non-stationary components. In particular, suppose that the $r_1$ units in $\mathbf{G}_1'\mathbf{f}_t = \mathbf{f}_{1t}$ are stationary, while the remaining $r_2 = r - r_1$ units contained in $\mathbf{G}_2'\mathbf{f}_t = \mathbf{f}_{2t}$ are non-stationary. The integratedness of $\mathbf{e}_t$ is also allowed to differ amongst the cross-sectional units. However, since $\mathbf{\Omega}_{\Delta e \Delta e}$ is diagonal, $\mathbf{C}_1'\mathbf{e}_t$ cannot be stationary unless the elements of $\mathbf{e}_t$ are already stationary. The long-run covariance matrix of $\Delta \mathbf{w}_{1t} = \mathbf{C}_1'\Delta \mathbf{u}_t = \mathbf{C}_1'\Lambda'\mathbf{G}\mathbf{G}'\Delta \mathbf{f}_t + \mathbf{C}_1'\Delta \mathbf{e}_t$ can therefore be written as

$$\mathbf{\Omega}_{\Delta w_1 \Delta w_1} = \mathbf{C}_1'\Lambda'\mathbf{G}\mathbf{\Omega}_{\Delta f \Delta f}\mathbf{G}'\Lambda\mathbf{C}_1 = \mathbf{C}_1'\Lambda'\mathbf{G}_2\mathbf{\Omega}_{\Delta f_2 \Delta f_2}\mathbf{G}_2'\Lambda\mathbf{C}_1,$$

where $\mathbf{\Omega}_{\Delta f_2 \Delta f_2}$ is the long-run covariance matrix of $\Delta \mathbf{f}_{1t}$. It follows that $\text{rk}(\mathbf{\Omega}_{\Delta w_1 \Delta w_1}) = r_1$. The cointegrating space $\mathbf{C}_1$ is given by the orthogonal complement of $\Lambda'\mathbf{G}_2$, because only then will it be true that $\mathbf{C}_1'\Lambda'\mathbf{G}_2 = \mathbf{0}$, which in turn implies $\mathbf{\Omega}_{\Delta w_1 \Delta w_1} = \mathbf{0}$. Of course, this does not rule out the possibility that the elements of $\Lambda$ may be zero for some units, which would then be stationary but non-cointegrated. It is therefore convenient to let $\mathbf{P}$ be an $N \times r_1$ matrix that selects those units that correspond to zero loadings. The cross-unit cointegrating space can now be defined as the space spanned by $(\mathbf{I}_N - \mathbf{P}(\mathbf{P}'\mathbf{P})^{-1}\mathbf{P}')\mathbf{C}_1$.

Clearly, these reduced rank restrictions correspond exactly to those that apply to our nonparametric model. Thus, as far as the long-run unit root and cointegration properties are concerned, the factor and nonparametric models are the same, see Banerjee and Wagner (2009, Appendix B) for a further discussion. In fact, there is very little to separate the two models, even in terms of short-run dynamics.

The problem here is that the common factor representation in (16) may not exist. But even if it does, there is the problem of consistent estimation of both factors and 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 $\Lambda$ as $T \to \infty$. However, since $N$ is fixed, consistent estimation of $\mathbf{f}_t$ is usually not possible. The only way to ensure consistent estimation of both quantities is therefore to assume that $N$ goes to infinity with $T$. More precisely, since $\Lambda$ 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 $\Lambda$ and $\mathbf{f}_t$, we estimate $(\mathbf{R}^{-1})'\Lambda$ and $\mathbf{R}\mathbf{f}_t$, where $\mathbf{R}$ is an $r \times r$

rotation matrix of full rank.[7] Identification of the whole factor structure requires not only that $N$, $T \to \infty$, 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 parametric structure for the cross-section 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 applicability of the factor-based tests. This is especially true in applied macro and finance, where $N$ is typically rather small. The rank tests are $N$-specific and completely nonparametric, and therefore more general in this regard.

It should also be mentioned that the particular factor model considered here, the one of Bai and Ng (2004), is the most general one out there, and that most factor-based test approaches are even more restrictive. For example, the tests of Moon and Perron (2004), Pesaran (2007), and Phillips and Sul (2003) all assume that the common and idiosyncratic components have the same order of integration, see Bai and Ng (2010) for a detailed discussion.

### 4.3 Simplicity

It is interesting to compare our rank statistics with those that would arise had the previously discussed second-generation factor-based approach been used. Suppose therefore that (1) holds and that the errors have the factor structure in (16). The idea of Bai and Ng (2004) is to first estimate the common and idiosyncratic components in (16), and then to test for unit roots in both, which can be done using the following step-wise procedure:

1. Let us rewrite (1) in first differences as
$$\Delta \mathbf{y}_t = \alpha_p \mathbf{D}'(\mathbf{D}\mathbf{D}')^{-1}\mathbf{D}\Delta\mathbf{d}_t^p + \Delta\mathbf{u}_t = \alpha_{p-1}\mathbf{d}_t^{p-1} + \Delta\mathbf{u}_t,$$
where $\alpha_{p-1} = \alpha_p\mathbf{D}'(\mathbf{D}\mathbf{D}')^{-1}$, $\mathbf{d}_t^{p-1} = \mathbf{D}\Delta\mathbf{d}_t^{p-1}$ and $\mathbf{D}$ is a $p \times (p+1)$ matrix chosen to exclude the first element in $\Delta\mathbf{d}_t^p$, which is equal to zero as the constant in $\mathbf{d}_t^p$ is eliminated when taking differences. The first step in the Bai and Ng (2004) approach is to obtain the least squares residuals, $\Delta\hat{\mathbf{u}}_t^{p-1}$ say, from the above regression.

[7] Since $\mathbf{R}$ has $r^2$ free elements, identification of $\Lambda$ and $\mathbf{f}_t$ requires $r^2$ 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.

2. The first-step residuals are then used to obtain estimates $\Delta\hat{\mathbf{f}}_t$ and $\hat{\Lambda}$ of $\Delta\mathbf{f}_t$ and $\Lambda$, respectively, which can be done by using the method of principal components. But before this method can be applied, we need to estimate the number of common factors, $r$, and for this reason we may use any of the information criteria developed by Bai and Ng (2002).

3. From $\Delta\hat{\mathbf{f}}_t$ and $\hat{\Lambda}$ we construct $\Delta\hat{\mathbf{e}}_t^{p-1} = \Delta\hat{\mathbf{u}}_t^{p-1} - \hat{\Lambda}'\Delta\hat{\mathbf{f}}_t$, which can be cumulated to obtain $\hat{\mathbf{e}}_t^{p-1} = \sum_{s=2}^{T} \Delta\hat{\mathbf{e}}_s^{p-1}$. This variable is then regressed onto $\mathbf{d}_t^p$, which yields the $p^{\text{th}}$-order detrended residuals, $\hat{\mathbf{e}}_t^p$.

4. The final step is to test for unit roots in $\mathbf{e}_t$ and $\mathbf{f}_t$. To test for unit roots in $\mathbf{e}_t$ we can apply to $\hat{\mathbf{e}}_t^p$ the same within and between tests as discussed in Section 3.1. Similarly, to determine the rank of $\mathbf{f}_t$ we can apply to $\hat{\mathbf{f}}_t$ any of the rank tests developed in Section 3.1.

Clearly, when compared to our rank test statistics, the statistics discussed here have no apparent computational advantage. On the contrary, the factor-based test statistics are actually quite difficult to compute. Then there is also the problem of small-sample bias, which in a step-wise approach like the one just described, can be rather serious as the bias from one step gets imported in subsequent steps.[8]

# 5 Small Sample Performance

In this section we report the findings of a small set of simulations using (1)–(3) to generate the data. By assuming that $\alpha_p = \mathbf{0}$, $\mathbf{C}_1 = [\mathbf{I}_{N_1}, \mathbf{0}]'$ 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

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

with $|\rho| < 1$, which implies that $\mathbf{y}_t$ has the canonical data generating process of Toda (1994). The error $\eta_t$ is allowed to be both serially and cross-sectionally correlated through $\eta_t =$

[8]Tests like those of Moon and Perron (2004), Pesaran (2007), and Phillips and Sul (2003), which restrict the order of integration of the common and idiosyncratic components to be the same, are simpler to implement because there is only one source of potential non-stationarity. These tests are therefore designed to test for a unit root in the idiosyncratic component only.

$\Theta\eta_{t-1} + \epsilon_t$ with $\epsilon_t \sim N(\mathbf{0}, \mathbf{\Sigma})$ and $\Theta = \text{diag}(\theta_1, \ldots, \theta_N)$, where $\theta_i$ is either set to zero or made a draw from $U(-0.3, 0.3)$. To ensure that $\mathbf{\Sigma}$ is a symmetric positive definite matrix, we follow Chang (2004) and set $\mathbf{\Sigma} = \mathbf{PVP}'$, where $\mathbf{V} = \text{diag}(\lambda_1, \ldots, \lambda_N)$ is a matrix of eigenvalues such that $\lambda_1 = 0.1$, $\lambda_N = 1$ and $\lambda_2, \ldots, \lambda_{N-1} \sim U(0.1, 1)$. 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, where $N$ and $T$ are chosen to reflect roughly the sample sizes considered in our empirical applications. In addition, for each cross-section we generate 100 presample values, starting with an initial value of zero. For brevity, we only report the size and power at the 5% level.[9] Some results on the sequential rank selection procedure are also reported.

Consider first the size results for testing $H_0 : \text{rk}(\mathbf{\Omega}_{\Delta u \Delta u}) = N$, 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 the 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 $H_0 : \text{rk}(\mathbf{\Omega}_{\Delta u \Delta u}) = \frac{N}{2}$ when $\rho = 0.1$. The first thing to notice is the size of the $MIB$ test, which 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 underreject when $N$ increases, the performance of $MB$ and $MJ$ remain just as good as before.

Consider next the power results when testing the full rank null against $H_1 : \text{rk}(\mathbf{\Omega}_{\Delta u \Delta u}) = N_2 < N$, which are reported in Tables 4 and 5 for $\rho$ fixed and varying values of $N_2$, and in Table 6 for $N_2 = 0.8N$ fixed and varying $\rho$. The information content of these tables may be summarized as follows:

1. The power of $MB$ and $MJ$ generally improves as $T$ increases, and as $N_2$ departs from its hypothesized value of $N$, which is presumably a reflection of the unbiasedness of these tests. We also see that although the power is generally increasing in $N$, this is not

[9] The 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.

always the case, which is also not expected.

2. In agreement with the consistency of $MIB$, we see that the rate at which its power increases with $T$ is generally much faster than for the other tests. The power is also increasing in $\rho$ but not necessarily in $N_2$ and $N$.

3. The $MB$ and $MJ$ tests generally outperform the $MIB$ test. One exception is when the deviation of $\rho$ from one is relatively large, in which case the latter test is more powerful. The extreme case being when $\rho = 0$, in which the $MIB$ test is vastly superior.

4. There is little difference in power depending on whether there is a constant or a constant and trend in the model, which is somewhat unexpected given the theory of the incidental trends problem. Of course, since we are considering a fixed alternative here, this theory, which refers to the local power, is not really applicable.[10]

5. As expected, all three tests generally perform rather poorly when $N_2$ and $\rho$ are close to their respective values under the null.

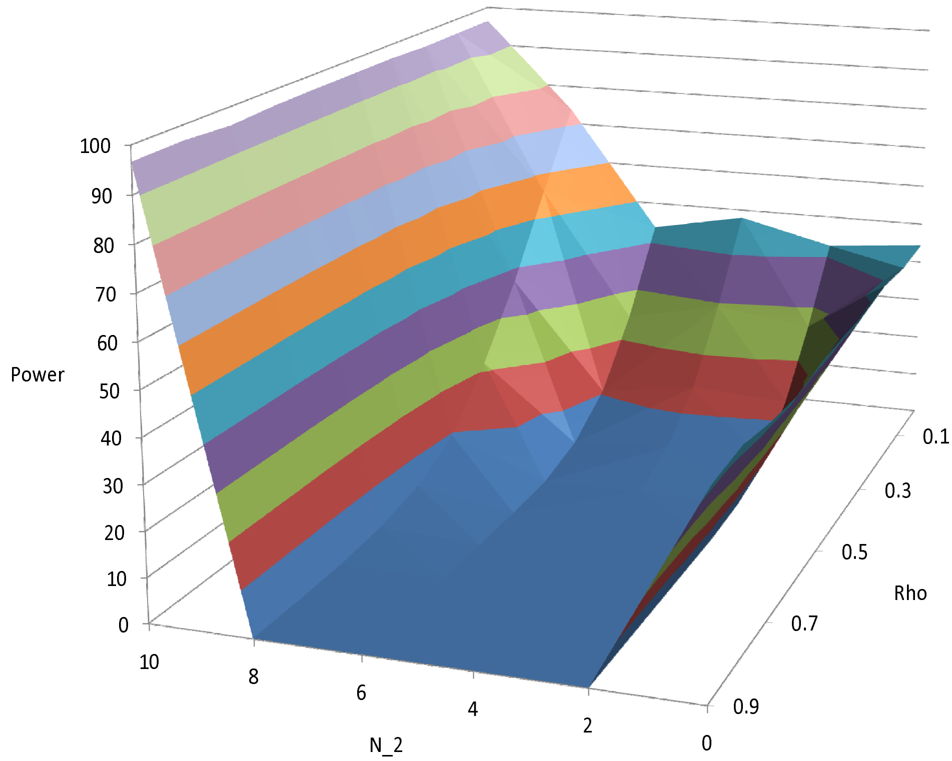
Some results of the correct rank selection frequencies for the sequential $MMIB$ test are reported in Figure 3 when $T = 100$ and in Figure 4 when $T = 200$. Both figures are for the case with $p = 0$, $\theta_i = 0$ and $N = 10$. Although we expect a reduction in the accuracy of inference as the true 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 eight. However, these magnitudes naturally decrease with $\rho$ and inversely with $T$. Indeed, with $T = 200$ and $\rho = 0$ the correct selection frequency never falls below 75%.

# 6 Empirical applications

In this section we examine 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 over time.

[10]The difference in the results when compared to a local alternative hypothesis is easily seen from Figures 1 and 2, where the effect of the trend is more pronounced.

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

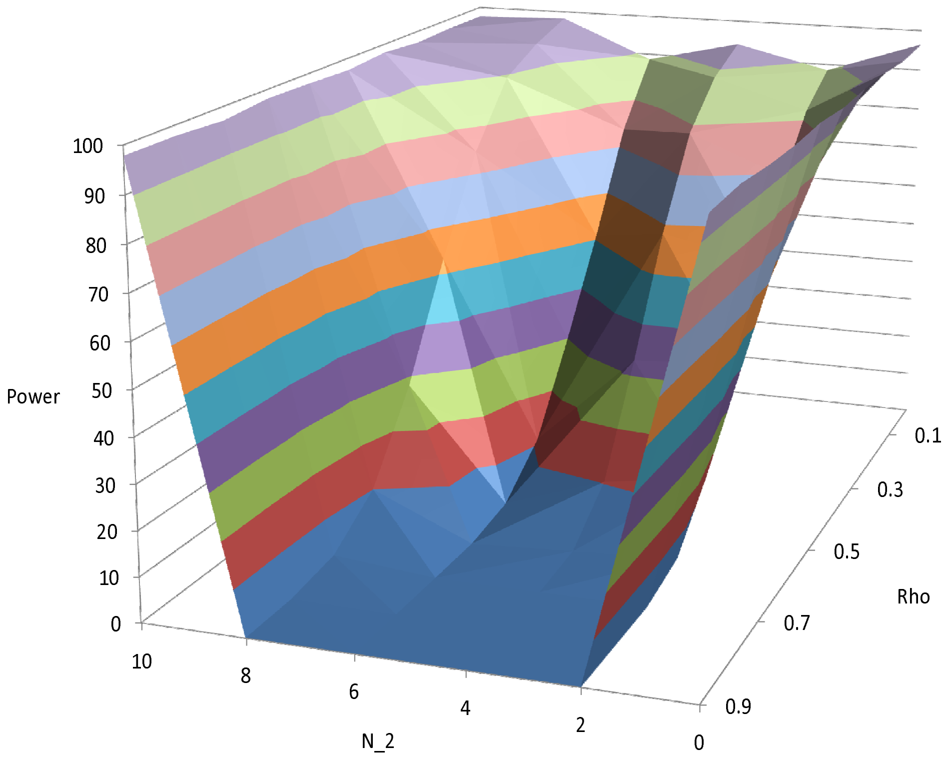



## 6.1 PPP

A common way of testing PPP is to apply to the real exchange rate any conventional first-generation panel unit root test that allows for a non-zero mean. A time trend is usually not included, as this is deemed inconsistent with PPP theory. Panel tests of this kind were initially motivated by their potential for power gains over univariate tests. However, because of the cross-country dependence, some of these power gains are more likely to reflect size distortions.

A large part of the dependence typically originates with the numeraire country, which can be given a common factor interpretation, and this has motivated the use of factor-based second-generation tests. However, such studies typically end up estimating a large number of factors, suggesting that the factor model might not be appropriate, see Wagner (2008) for a discussion and some empirical results. Then there is also the issue of the omitted time trend, which, in view of the well-known Balassa–Samuelson argument that countries with high

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




productivity in traded goods will have appreciating exchange rates, is equally problematic.

In this section, we try to address both these problems simultaneously by using our rank tests, which, in addition to not requiring that the cross-country dependence is of the common factor type, have good power when there are deterministic trends present under the stationary alternative. The data that we use are the same as 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 7.[11] It is seen that in this application $N$ is rather small, especially in the Euro area and CEEC panels. This means that factor-based approaches are likely to be biased, and that our finite-$N$ approach might be more appropriate.

The results of the rank tests are reported in Table 8. 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

[11]See Wagner (2008) for a more detailed description of the data.

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 full rank estimate.

### 6.2 Income convergence

Our analysis of the convergence of income is rooted in the definition of Evans (1998). To formalize the idea, suppose that $y_{it}$, the income for country $i$ at time $t$, is non-stationary. Then the panel is said to 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. Hence, if the convergence hypothesis holds, the cross-unit cointegrating rank is one, whereas if it fails, the rank is $N$.

As Evans (1998) points out this definition is fairly general, and even allows for the possibility of different convergence clubs. However, this is not what empirical researchers tend to focus on. Indeed, most researchers assume that the countries have the same stochastic trend. The main reason for this is that if the trend can be well-measured by the overall cross-sectional average, $\overline{y}_t$ say, then the definition of convergence is equivalent to the condition that $y_{it} - \overline{y}_t$ is stationary for all $i$, which is easily tested by using any first-generation panel unit root test. However, if one would like to entertain the possibility of convergence clubs, then one needs to allow for more than one cross-country cointegrating relationship. Our rank tests are ideally suited for this.

The data that we are going to use to assess the convergence hypothesis are taken from Maddison (2007), and comprise annual observations on the log per-capita GDP for 22 countries over the period 1870–2001.[12] The rank test results are reported in Table 9. We see that while $MIB$ and $MB$ are able to reject the full rank null, $MJ$ is not. The results for this null are thus inconclusive, and we therefore proceed to test the null of rank 21. In this case it is only the $MIB$ test that rejects. But since the rank under the null is not necessarily full, this result should be interpreted with caution. It is better to use the $MMIB$ test, which leads to a rank estimate of 20 in case of a constant, and 21 in case of a constant and trend. A majority of the evidence therefore leans towards a reduced rank that is close to being full, suggesting that

[12] The 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.

the convergence must be rejected.

## 7 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, since the serial and cross-sectional correlations do not affect the asymptotic null distributions of the test statistics, the tests are robust against deviations from the usual assumption of a factor model with linear short-run dynamics. This property is important not only in theory but also in small samples, where deviations from these parametric assumptions may have a substantial effect on the behavior of the parametric test statistic. Second, despite these allowances, the rank tests do not require any treatment of nuisance parameters. Hence, with these tests there is no need for any lag augmentation, or estimation of common factors. Implementation is therefore very simple. Third, the tests have relatively high power, even when compared to first-generation tests in cases when the cross-sectional independence restriction holds. Finally, since the asymptotic results do not require $N \to \infty$, the tests are ideally suited for applications with the typical macro or finance panel, in which $N$ is rather small.

## References


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.

Bai, J., and S. Ng (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica* **70**, 191–221.

Banerjee, A., and M. Wagner. Panel Methods to Test for Unit Roots and Cointegration. In Mills, T.C. and K. Patterson (Eds.) *Palgrave Handbook of Econometrics, Volume II*, Chapter 13, 632–726. Palgrave MacMillan.

Bowman, D. (1999). Efficient Tests for Autoregressive Unit Roots in Panel Data. International Finance Discussion Paper 646, Board of Governors of the Federal Reserve System.

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

Breitung, J., and M. H. Pesaran (2008). Unit Roots and Cointegration in Panels. In Matyas, L. and P. Sevestre (Eds.) *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*, Chapter 9, 279-322. Kluwer Academic Publishers, Boston.

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., T. J. Vogelsang, and H. Bunzel (2000). Simple Robust Testing of Regression Hypotheses. *Econometrica* **68**, 695–714.

Kiefer, N. M., and T. J. Vogelsang (2000). Heteroskedasticity-Autocorrelation Robust Standard Errors Using the Bartlett Kernel without Truncation. *Econometrica* **70**, 2093–2095.

Levin, A., and C. Lin (1992). Unit Root Tests in Panel Data: Asymptotic and Finite-sample Properties. University of California working paper 9223.

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.

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.

O'Connell, P. G. J. (1998). The Overvaluation of Purchasing Power Parity. *Journal of International Economics* **44**, 1–19.

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., and B. Choi (1988). A New Approach to Testing for a Unit Root. CAE Working Paper 88-23, Cornell University.

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 P. C. B. Phillips (1988). Statistical Inference in Regressions with Integrated Processes: Part 1. *Econometric Theory* **4**, 468–497.

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 V. Solo (1992). Asymptotics for Linear Processes. *Annals of Statistics* **20**, 971–1001.

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 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.

Hlouskova, J., and M. Wagner (2006). The Performance of Panel Unit Root and Stationarity Tests: Results from a Large Scale Simulation Study. *Econometric Reviews* **25**, 85-116.

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.

| | $p = 0$ | | | $p = 1$ | | |
|---|---|---|---|---|---|---|
| Value | $MIB$ | $MB$ | $MJ$ | $MIB$ | $MB$ | $MJ$ |
| Response surface coefficients for $5 < N < 50$ | | | | | | |
| $\delta_1$ | 169037.15804 | 0.28636 | 45.12093 | 173133.91134 | 0.02454 | 4.86511 |
| $\delta_2$ | −375970.47567 | −0.74096 | −156.67769 | −386036.11573 | 0.25342 | −33.63947 |
| $\delta_3$ | 236742.44174 | 1.30607 | 88.62252 | 242753.06689 | −0.42490 | 18.71568 |
| $\delta_4$ | −28714.71985 | −1.55572 | −0.26967 | −29227.10664 | 0.09250 | 1.66408 |
| $\delta_5$ | 462.26011 | 1.08594 | 0.68685 | 499.44273 | 0.14925 | 0.23377 |
| $\delta_6$ | 5.07971 | −0.37162 | −0.00337 | 4.95667 | −0.09111 | −0.00119 |
| $\delta_7$ | 12978.51792 | 0.59838 | −61.11665 | 11680.94595 | 0.88073 | −23.24877 |
| $\delta_8$ | −800.55043 | −0.63492 | 1.15489 | −823.41937 | −0.97470 | 0.44045 |
| $\delta_9$ | −531706.18433 | 0.06527 | 3193.19031 | −435669.56662 | 0.04085 | 1199.08007 |
| $\delta_{10}$ | 34932714.04854 | 4.42248 | −23173.56365 | 33299454.36837 | 5.13046 | −10069.57330 |
| $\delta_{11}$ | −7336758.89483 | −34.18132 | −25072.41143 | −7720478.56401 | −35.78908 | −9009.64022 |
| $\delta_{12}$ | 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 $q = \delta' \mathbf{x} + \varepsilon$, where $q$ is the 5% critical value and $\mathbf{x}$ is given in Section 4.4. The estimated 5% critical value is computed as the fitted value of the regression. $p = 0$ and $p = 1$ refer to the model with constant, and constant and trend, respectively.

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

| | Between tests | | | Within tests | | |
|---|---|---|---|---|---|---|
| Adjustment | $BMIB$ | $BMB$ | $BMJ$ | $WMIB$ | $WMB$ | $WMJ$ |
| | | | $p = 0$ | | | |
| $\mu$ | 33.35913 | 0.05080 | 13.38137 | 14.98190 | 0.13349 | 12.33025 |
| $\sigma^2$ | 1358.83596 | 0.00074 | 198.41452 | 51.60810 | 0.00410 | 157.30748 |
| | | | $p = 1$ | | | |
| $\mu$ | 125.32524 | 0.01083 | 4.67848 | 76.51020 | 0.02614 | 4.31645 |
| $\sigma^2$ | 7694.99507 | 0.00003 | 15.50459 | 1569.08434 | 0.00018 | 12.53802 |

*Notes*: $p = 0$ refers to the model with an intercept, while $p = 1$ refers to the model with intercept and trend. The standard normal transformation of for example the $BMB$ statistic is given by $\sqrt{N}(BMB - \mu)/\sigma$.

Table 3: Size at the 5% level.

| | | $H_0 : \mathrm{rk}(\mathbf{\Omega}_{\Delta u \Delta u}) = N$ | | | | | | $H_0 : \mathrm{rk}(\mathbf{\Omega}_{\Delta u \Delta u}) = \frac{N}{2}$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $p = 0$ | | | $p = 1$ | | | $p = 0$ | | | $p = 1$ | | |
| $T$ | $N$ | $MIB$ | $MB$ | $MJ$ | $MIB$ | $MB$ | $MJ$ | $MIB$ | $MB$ | $MJ$ | $MIB$ | $MB$ | $MJ$ |
| | | | | | | | $\theta_i = 0$ | | | | | | |
| 100 | 10 | 3.4 | 5.9 | 4.4 | 3.5 | 4.5 | 4.2 | 100.0 | 3.0 | 4.5 | 100.0 | 2.2 | 3.7 |
| 100 | 20 | 5.4 | 6.5 | 3.8 | 5.5 | 5.0 | 3.3 | 100.0 | 0.8 | 1.3 | 100.0 | 0.2 | 0.6 |
| 100 | 40 | 5.2 | 0.8 | 5.4 | 4.7 | 0.5 | 4.7 | 100.0 | 0.0 | 0.0 | 100.0 | 0.0 | 0.0 |
| 200 | 10 | 2.2 | 5.5 | 5.7 | 2.3 | 4.2 | 6.1 | 100.0 | 3.0 | 2.9 | 100.0 | 2.8 | 2.4 |
| 200 | 20 | 7.1 | 5.0 | 4.6 | 7.4 | 4.4 | 4.8 | 100.0 | 2.0 | 3.1 | 100.0 | 1.2 | 2.6 |
| 200 | 40 | 5.8 | 5.3 | 5.1 | 6.4 | 11.3 | 4.7 | 100.0 | 0.0 | 1.1 | 100.0 | 0.0 | 0.4 |
| | | | | | | | $\theta_i \sim U(-0.3, 0.3)$ | | | | | | |
| 100 | 10 | 5.6 | 6.3 | 4.8 | 6.8 | 5.1 | 4.5 | 100.0 | 3.1 | 4.5 | 100.0 | 2.1 | 3.7 |
| 100 | 20 | 15.5 | 8.1 | 3.9 | 17.4 | 7.1 | 3.7 | 100.0 | 0.9 | 1.3 | 100.0 | 0.1 | 0.7 |
| 100 | 40 | 31.9 | 1.9 | 4.6 | 31.9 | 2.2 | 4.7 | 100.0 | 0.0 | 0.0 | 100.0 | 0.0 | 0.0 |
| 200 | 10 | 3.1 | 5.3 | 6.3 | 2.9 | 4.7 | 6.7 | 100.0 | 3.1 | 3.0 | 100.0 | 2.9 | 2.7 |
| 200 | 20 | 12.9 | 4.9 | 4.9 | 13.4 | 5.2 | 4.9 | 100.0 | 2.0 | 3.0 | 100.0 | 1.2 | 2.5 |
| 200 | 40 | 31.2 | 9.4 | 5.6 | 33.3 | 17.4 | 5.3 | 100.0 | 0.0 | 1.1 | 100.0 | 0.1 | 0.3 |

*Notes*: $\theta_i$ refers to the autoregressive coefficient in the errors. $p = 0$ and $p = 1$ refer to the model with constant, and constant and trend, respectively.

Table 4: Power at the 5% level for $\rho = 0.9$ fixed and varying values of $N_2$.

| | | $p=0$ | | | $p=1$ | | |
|---|---|---|---|---|---|---|---|
| $T$ | $N$ | $MIB$ | $MB$ | $MJ$ | $MIB$ | $MB$ | $MJ$ |
| | | $N_2 = 0.1N$ | | | | | |
| 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 |
| 100 | 40 | 6.6 | 98.9 | 99.7 | 36.1 | 79.9 | 97.8 |
| 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 |
| 200 | 40 | 14.8 | 100.0 | 100.0 | 47.0 | 100.0 | 100.0 |
| | | $N_2 = 0.3N$ | | | | | |
| 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 |
| 100 | 40 | 6.4 | 67.6 | 89.6 | 6.1 | 38.4 | 81.2 |
| 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 |
| 200 | 40 | 11.9 | 100.0 | 96.9 | 11.5 | 99.9 | 98.2 |
| | | $N_2 = 0.7N$ | | | | | |
| 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 |
| 100 | 40 | 5.4 | 4.9 | 25.0 | 5.2 | 3.4 | 20.4 |
| 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 |
| 200 | 40 | 7.9 | 50.5 | 32.1 | 7.9 | 59.5 | 34.5 |

*Notes*: $N_2$ refers to the number of unit roots under the alternative, while $\rho$ refers to the autoregressive coefficient of the remaining stationary units. hypothesis. $p = 0$ and $p = 1$ refer to the model with constant, and constant and trend, respectively. The null hypothesis is that of full rank.

Table 5: Power at the 5% level for $\rho = 0$ fixed and varying values of $N_2$.

| | | $p=0$ | | | $p=1$ | | |
|---|---|---|---|---|---|---|---|
| $T$ | $N$ | $MIB$ | $MB$ | $MJ$ | $MIB$ | $MB$ | $MJ$ |
| | | | $N_2 = 0.5N$ | | | | |
| 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 |
| 100 | 40 | 100.0 | 100.0 | 95.7 | 100.0 | 100.0 | 98.9 |
| 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 |
| 200 | 40 | 100.0 | 100.0 | 99.2 | 100.0 | 100.0 | 100.0 |
| | | | $N_2 = 0.8N$ | | | | |
| 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 |
| 100 | 40 | 99.8 | 92.1 | 38.2 | 99.7 | 91.4 | 46.9 |
| 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 |
| 200 | 40 | 100.0 | 100.0 | 47.8 | 100.0 | 100.0 | 58.8 |
| | | | $N_2 = 0.9N$ | | | | |
| 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 |
| 100 | 40 | 77.9 | 25.5 | 16.6 | 73.2 | 24.8 | 19.6 |
| 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 |
| 200 | 40 | 100.0 | 80.1 | 19.2 | 100.0 | 90.8 | 22.2 |

*Notes*: See Table 4 for an explanation.

Table 6: Power at the 5% level for $N_2 = 0.8N$ fixed and varying values of $\rho$.

| | | $p=0$ | | | $p=1$ | | |
|---|---|---|---|---|---|---|---|
| $T$ | $N$ | $MIB$ | $MB$ | $MJ$ | $MIB$ | $MB$ | $MJ$ |
| | | $\rho = 0.9$ | | | | | |
| 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 |
| 100 | 40 | 5.5 | 2.5 | 15.7 | 4.9 | 1.6 | 12.6 |
| 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 |
| 200 | 40 | 7.4 | 26.7 | 18.9 | 7.5 | 38.6 | 18.8 |
| | | $\rho = 0.7$ | | | | | |
| 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 |
| 100 | 40 | 9.1 | 15.2 | 21.5 | 7.9 | 11.3 | 24.1 |
| 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 |
| 200 | 40 | 30.7 | 81.8 | 28.9 | 29.1 | 89.7 | 33.9 |
| | | $\rho = 0.5$ | | | | | |
| 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 |
| 100 | 40 | 24.8 | 42.1 | 26.9 | 22.6 | 37.8 | 31.5 |
| 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 |
| 200 | 40 | 95.4 | 97.8 | 36.0 | 93.9 | 99.1 | 43.6 |

*Notes*: See Table 4 for an explanation.

Table 7: PPP panels.

| Panel | Start date | End date | $T$ | $N$ |
|---|---|---|---|---|
| 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 |

Table 8: Rank test results for the PPP panels.

| | $p=0$ | | | $p=1$ | | |
|---|---|---|---|---|---|---|
| Panel | $MIB$ | $MB$ | $MJ$ | $MIB$ | $MB$ | $MJ$ |
| 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 |

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

Table 9: Rank test results for the income panel.

| $p$ | $MIB$ | $MB$ | $MJ$ |
|---|---|---|---|
| 0 | 100010.6* | 0.15483* | 2724.7 |
| 1 | 109378.4* | 0.05558* | 252.0 |

*Notes*: See Table 8 for an explanation.