Testing for cointegration in high-dimensional systems

Jörg Breitung
University of Bonn

Gianluca Cubadda
University of Rome “Tor Vergata”

December 11, 2008

highly preliminary and incomplete

Abstract

This paper considers cointegration tests for dynamic systems where the number of variables is large relative to the sample size. Typical examples include tests for unit roots in panels where the units are linked by complicated dynamic relationships. It is well known that conventional cointegration tests based on a parametric (vector autoregressive) representation of the system break down if the number of variables approaches the number of time periods. To sidestep this difficulty we propose nonparametric cointegration tests based on eigenvalue problems that are asymptotically free of nuisance parameters. It turns out that the nonparametric tests outperform their parametric (likelihood-ratio based) counterparts by a clear margin.

1 Introduction

During the last two decades, various tests for cointegration were proposed. Among them, the residual Dickey-Fuller test suggested by Engle and Granger (1987) and Phillips and Ouliaris (1990) as well the system cointegration test suggested by Johansen (1988, 1991) have become most popular. In a macroeconomic context, these tests are applied for data sets including a fairly large number of time periods (usually $100 - 300$) and a small number of variables (less than 6). More recently, however, unit root or cointegration tests are applied to panel data sets with a smaller number of time periods (often less than 100) and a moderate number of countries (typically between 10 and 50). For a recent review of the literature see, e.g., Choi (2006) and Breitung and Pesaran (2008). In the first generation of panel unit root and cointegration tests, the panel units are treated as independent and a central limit theorem is applied to the standardized sum of the individual test statistics. The second generation panel data tests allow for a contemporaneous correlation among the errors of the panel units. This contemporaneous dependence is often represented by common dynamic factors (e.g. Phillips and Moon (2004) and Pesaran (2007)). However, if the panel units are correlated due to a more complex dynamic relationship (for example due to (Granger) causality among the panel units), then the introduction of common factors is not sufficient to obtain a pooled test statistic that is free of nuisance parameters. For instance, Cubadda et al. (2008) document that common factor corrections are useless in estimating common parameters when common cycles are present in the data. This possibility calls for the development of "third-generation" panel unit roots or cointegration tests that account for a complex dynamic relationship among the panel units. In this paper, we consider test procedures that are valid in a very general dynamic framework, where first and second-generation test are invalid.

However, the tests considered in this paper are not only valid in the typical panel data framework, where country specific variables are contemporaneously correlated. As argued by Banerjee et al. (2005) there are convincing reasons to assume cointegration among variables from different panel units. This situation is sometimes called "cross cointegration" (e.g. Banerjee et al. 2005 and Wagner 2008). In this situation the usual panel unit roots and cointegration tests are invalid (see, inter alia, Banerjee et al. 2005, Breitung and Das 2008 and Gengenbach et al. 2007).

Our paper is closely related to the work of Pedroni and Vogelsang (2005) who have proposed tests for Panel unit and cointegration in a very general linear dynamic model. However, as we argue in this paper, the finite sample properties of their test deteriorates dramatically if the number of panel units are large relative to the number of time series observations. The main reason is that the null hypothesis of this test implies n^2 parameter restrictions, where n denotes the number of variables in the data set. Therefore, in typical sample sizes for panel unit root tests, the number of restrictions implied by the panel unit root hypothesis may even exceed the number of time periods which gives rise to severe size distortions of the test. Furthermore, the tests lacks power if only a few restrictions in the large set of restrictions are violated.

The tests suggested in this paper have two major features. First, the test statistics are "nonparametric" in the sense that no multivariate parametric model needs to be specified and estimated in order to derive the test statistics. This is an important feature as for the usual models (like the vector autoregression) the number of parameters grow with the square of n. Second, the number of restrictions is reduced to the set of n necessary conditions, leaving out all restrictions that are not essential for the hypothesis under test. Our Monte Carlo simulation suggest that the resulting tests can be adopted in dynamic systems where the number of variables is moderately large (20, say). In contrast all other tests exhibit severe size distortions and very low power in these situations.

2 Motivation

To highlight the problems involved by testing for cointegration in high dimensional systems, consider the simple case of a cointegrated VAR(1) model

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

where $E(y_t) = 0$ and $\varepsilon_t \sim N(0, \Sigma)$. Let us first consider the maximum eigenvalue (henceforth: *maxEV*) statistic for the hypothesis that y_t is not cointegrated, i.e., $H_0 : r = rk(\Pi) = 0$. Define

$$S_{00} = \frac{1}{T}\sum_{t=1}^{T} \Delta y_t \Delta y_t', \ S_{11} = \frac{1}{T}\sum_{t=1}^{T} y_{t-1} y_{t-1}' \ \text{ and } \ S_{01} = \frac{1}{T}\sum_{t=1}^{T} \Delta y_t y_{t-1}'. \quad (1)$$

The *maxEV* statistic is given by

$$T\lambda_1 = \frac{v_1' S_{01}' S_{00}^{-1} S_{01} v_1}{v_1' S_{11} v_1 / T}, \quad (2)$$

where λ_1 is the largest eigenvalue and v_1 is the associated eigenvector of the eigenvalue problem

$$|\lambda S_{11} - S_{01}' S_{00}^{-1} S_{01}| = 0. \quad (3)$$

Accordingly, the *maxEV* statistic is equivalent to $T \cdot R^2$ of the regression of $v_1' y_{t-1}$ on the $n \times 1$ vector Δy_t. Therefore, the test is equivalent to an LM test of the n-dimensional hypothesis $\varphi = 0$ in the regression $v_1' y_{t-1} = \varphi' \Delta y_t + u_t$.

It is interesting to compare this test statistic to the residual Dickey-Fuller (*resDF*) test suggested by Engle and Granger (1987). Let $\widehat{\gamma} = [1, -\widehat{\beta}']'$, where $\widehat{\beta}$ denotes the OLS estimator of the coefficients from a regression of $y_{1,t}$ on $y_{2,t}, \ldots, y_{n,t}$. The residual DF statistic is the t-statistic of the one-dimensional hypothesis $\varrho = 0$ in the regression

$$(\widehat{\gamma}' \Delta y_t) = \varrho (\widehat{\gamma}' y_{t-1}) + e_t \ . \quad (4)$$

Note that the two-sided LM version of this test statistic is equivalent to the LM statistic of a regression of $(\widehat{\gamma}' y_{t-1})$ on $(\widehat{\gamma}' \Delta y_t)$ since the R^2 statistic of a bivariate regression is equivalent to the R^2 of the reverse regression. Accordingly, there are three important differences between the *maxEV* statistic and the *resDF* statistic:

(i) The *maxEV* statistic is a two-sided test statistic, whereas the *resDF* statistic is one-sided.

(ii) The *maxEV* test employs the ML estimator v_1 to construct the "most stationary" linear combination $v_1' y_{t-1}$, whereas the *resDF* test is based on the OLS estimator of the cointegration relationship $\widehat{\gamma}' y_{t-1}$.

(iii) The null hypothesis of the *maxEV* test is n-dimensional, whereas the *resDF* test is based on an one-dimensional hypothesis.

It is the last property of the *resDF* approach that makes this test particularly attractive in large dimensional systems. In what follows we therefore construct test statistics that share properties (i) and (iii) with the *resDF* test. To construct a test statistic that is based on the r "most stationary" linear combinations the test can employs the r eigenvectors $v_1, \ldots, v_r$ associated with the r largest eigenvalues of problem (3). Since under the null hypothesis all linear combinations are non-stationary, our test is based on the sum of r Dickey-Fuller t-statistics

$$\xi(r) = \sum_{j=1}^{r} \tau_j \ ,$$

where

$$\tau_j = \frac{\sum_{t=1}^{T} (v_j' \Delta y_t)(v_j' y_{t-1})}{\widehat{\sigma}_j \sqrt{\sum_{t=1}^{T} (v_j' y_{t-1})^2}},$$

where $\widehat{\sigma}_j^2$ is the usual estimator of the residuals variance of the component specific Dickey-Fuller regressions. However to obtain a test statistic that is free from nuisance parameters we replace the eigenvectors v_j from (3) by eigenvectors of an eigenvalue problem that is shown to be invariant to the short-run dynamics of the system.

Asymptotic distribution of the test statistics

In this section we study the asymptotic properties of alternative test statistics that are based on the sum of r unit root statistics computed from the r "most stationary" linear combinations. Under the null hypothesis we assume that the vector y_t is an $n \times 1$ vector of not cointegrated $I(1)$ variables. The complete set of assumptions is summarized in

Assumption 1: *(i) Under the null hypothesis $r = 0$, the $n \times 1$ vector Δy_t has the moving average representation $\Delta y_t = \sum_{j=0}^{\infty} C_j u_{t-j}$ with $C_0 = I$ and $\sum_{j=0}^{\infty} j^2 ||C_j|| < \infty$. (ii) u_t is i.i.d. with $E(u_t) = 0$, $E(u_t u_t') = \Sigma$ (positive definite) and finite fourth moments and $u_s = 0$ for $s \leq 0$.*

As shown by Phillips and Solo (1992) this assumption ensures that

$$\frac{1}{\sqrt{T}} y_{[aT]} \;\Rightarrow\; B(a),$$

where $B(a)$ is an n-dimensional vector of Brownian motion with covariance matrix $E[B(1)B(1)]' = \Omega \equiv C(1)\Sigma C(1)'$. It is important to note that we do not assume a particular parametric model for Δy_t, such as a finite order vector autoregressive model. The main reason is that if n is large, parametric models imply a large numbers of additional parameters. For example, a VAR(p) model requires to estimate $n^2 p$ autoregressive parameters that are unreliably estimated if n is large relative to T. We therefore consider nonparametric approaches to deal with short-run dynamics.

The first test statistic is based on the eigenvectors of the adjusted eigenvalue problem

$$|\lambda S_{11} - (S_{01} - \widehat{\Psi})' \widehat{\Omega}^{-1} (S_{01} - \widehat{\Psi})| = 0. \tag{5}$$

where $\widehat{\Psi}$ is a consistent estimator of $\Psi = \sum_{j=1}^{\infty} E(\Delta y_t y_{t-j})$, $\widehat{\Omega}$ is a consistent estimator of $\Omega = \Sigma + \Psi + \Psi'$ and S_{jk} are defined in (1). As has been shown by Hassler (2006), the asymptotic distribution of the eigenvalues and eigenvectors of (5) do not depend on nuisance parameters and, therefore, the eigenvalues and eigenvectors obtained from this eigenvalue problem are asymptotically free of nuisance parameters. Unfortunately, for large n the small sample distributions of the test statistic are affected from the sampling variability of the estimators $\widehat{\Psi}$ and $\widehat{\Omega}$, which is in particular problematic as the inverse of $\widehat{\Omega}$ enters the eigenvalue problem. We therefore consider two further approaches that are shown to have superior small sample properties in typical sample sizes.

Following Shintani (2001) we consider the (inverse) principal component (PC) estimator that solves the eigenvalue problem

$$|\eta S_{11}^* - \widehat{\Omega}| = 0, \tag{6}$$

where $S_{11}^* = T^{-1} \sum_{t=1}^{T} y_t y_t'$. It is not difficult to see that the asymptotic distribution of the eigenvalues of this problem do not depend on nuisance parameters. Note that η_j is the eigenvalue of the matrix $(\widehat{\Omega}^{-1/2} S_{11}^* \widehat{\Omega}^{-1/2})^{-1}$, and therefore the eigenvalues are the inverse of the ordinary PC eigenvalue problem.

Finally, we adapt the nonparametric approach suggested by Breitung (2002) that is based on the eigenvalue problem

$$|\mu S_{22} - S_{11}^*| = 0, \tag{7}$$

where $S_{22} = T^{-1} \sum_{t=1}^{T} S_t S_t'$ and $S_t = \sum_{j=1}^{t} y_j$. Note that we have modified the original eigenvalue problem $|\mu^* S_{11} - S_{22}| = 0$ of Breitung (2002) in order to retain the same ordering of eigenvalues as in the other two approaches. Again, as $T \to \infty$ the eigenvalues and eigenvectors of (7) do not depend on nuisance parameters. An important advantage of the latter statistic is that no estimator of nuisance parameters is required that depends on some suitably chosen truncation lag.

Let v_j^k denote the j'th eigenvector associated with the ordered eigenvalues of (5) – (7) with $k \in \{LR, PC, NP\}$. The test statistic for testing the hypothesis H_0: $rk(\Omega) = n$ against H_1: $rk(\Omega) = n - r$ is

$$\tau_k(r) = \sum_{j=1}^{r} \tau_j^k \ ,$$

where τ_j^k is the ADF t-statistic for $\varrho = 0$ from the regression

$$\Delta z_{jt}^k = \varrho^k z_{j,t-1}^k + \sum_{\ell=1}^{p_j} \vartheta_\ell^k \Delta z_{j,t-\ell}^k + e_{jt}^k$$

and $z_{jt}^k = y_t' v_j^k$.

For estimating the matrices Ω and Ψ we make the following assumption:

Assumption 2: $\widehat{\Psi} = \sum_{j=1}^{T-1} w(j/M)\widehat{\Gamma}_j$, $\widehat{\Omega} = \Gamma_0 + \widehat{\Psi} + \widehat{\Psi}'$, $\widehat{\Gamma} = T^{-1}\sum_{t=j+1}^{T} \Delta y_t \Delta y_{t-j}'$. *The kernel function it is a twice continuously differentiable even function with* $w(0) = 1$, $w'(0) = 0$, $w''(0) = 0$, $w(|1+\delta|) = 0$ *for* $\delta \geq 0$ *and (a)* $w(x)/(1-|x|)^2$ *converges to a constant as* $|x| \to 1$ *or (b)* $w(x) = O(x^{-2})$. *For the bandwidth it is assumed that* $M = O(T^m)$, *where* $m \in (0, 1/2)$.

As shown by Phillips (1995) Assumption 2 ensures that $\widehat{\Psi}$ and $\widehat{\Omega}$ are consistent estimators of Ψ and Ω. In the appendix we show that the test statistics have the asymptotic null distributions as stated in the following Theorem:

Theorem 1: *Let* y_t *be a* n*-dimensional vector of random walks. Under Assumption 1 – 2 and as* $T \to \infty$

$$\tau_k(r) \Rightarrow \sum_{j=1}^{r} \frac{\int \xi_j^{k\prime} W dW' \xi_j^k}{\sqrt{\int \xi_j^{k\prime} W W' \xi_j^k}},$$

where $W = [W_1, \ldots, W_n]'$ *is an* $n \times 1$ *vector of independent standard Brownian motions, with* $k \in \{LR, PC, N\}$ *and* ξ_j^k *is an* $n \times 1$ *function of* W *given in the appendix.*

Remark A: For the derivation of the limiting distribution it is assumed that $E(y_t) = 0$. In many empirical applications the mean is specified as $E(y_t) = \delta' x_t$ where δ is a vector of coefficients and x_t is a vector of deterministic variables. Examples include $x_t = 1$ (constant) and $x_t = [1, t]'$ (linear trend). Let

$$\Psi_T x_{[aT]} \;\rightarrow\; \zeta(a),$$

where Ψ_T is a diagonal matrix with appropriate normalization factors (such as Ψ_T =diag$(1, T^{-1})$ for the linear trend) on the leading diagonal, then the residual from a regression of y_t on x_t yields

$$\frac{1}{\sqrt{T}}\Big(y_{[aT]} - \widehat{\delta}' x_{[aT]}\Big) \;\Rightarrow\; \widetilde{W}_\zeta \equiv W - \left(\int W\zeta'\right)\left(\int \zeta\zeta'\right)^{-1}\zeta.$$

When using the demeaned variables $\widetilde{y}_t = y_t - \widehat{\delta}' x_t$ instead of y_t, the limiting distributions result from replacing W in Theorem 1 by $\widetilde{W}_\zeta$.

Remark B: It is interesting to note that the test statistics are similar to the t-bar statistic proposed by Im, Pesaran, and Shin (2003). Accordingly, one may assert that for tests against the alternative $r = n$ (stationary time series) the standardized version

$$\tau_k^*(r) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (\tau_j^k - \mu_\tau^k)/\sigma_\tau^k \;, \tag{8}$$

with $\mu_\tau^k = E[\tau_k(n)]$ and $\sigma_\tau^k = \sqrt{\text{var}[\tau_k(n)]}$, converges to a standard normal limit distribution as $N \to \infty$ and $T \to \infty$. Table 1 presents the critical values for selected values of N and T, where the moments of $\tau_k(n)$ are approximated by the mean of 5000 Monte Carlo replication with $T = 1000$. Under a standard limiting distribution the critical values with respect to a significance level of 0.05 are expected to converge to a limiting value of –1.645. For reference we also report the critical values of the standardized versions of the *maxLR* and the Pedroni-Vogelsang (*Ped-Vog*) statistic. Since the latter two statistics rejects for large values of the test statistic, the critical values are supposed to converge to 1.645. The results presented in Table 1 suggest that the critical values indeed converge to critical values of the standard normal distribution. However for the *maxLR* and (*Ped-Vog*) the convergence to the limiting values appears to be much smaller. Unfortunately, we are not able to establish these results analytically. First, the asymptotic distributions presented in Theorem 1 are based on the assumption that N is fixed. Assuming $N \to \infty$ would require to derive the limiting distributions of the eigenvalues and eigenvectors for a infinite dimensional matrix. Furthermore, the linear combinations $v_i^{k\prime} y_t$ and $v_j^{k\prime} y_t$ are not independent and, therefore, a direct application of the central limit theorem employed in Im et al. (2003) is not possible. Nevertheless, the observed convergence of the critical values useful to interpolate the actual critical values, and therefore, the tables presented in Appendix B report the critical values of the statistics with a constant or a trend for a few sample sizes only.

3 A panel unit root statistic

For the special case $H_1 : rk(\Omega) = 0$ (i.e., all elements of y_t are stationary) it is possible to construct a much simpler test statistic. Since all components enter the test statistic, we do not need to order the components according to their largest autoregressive root. Since under the null hypothesis $T^{-1/2}\Omega^{-1/2}y_{[Ta]} \Rightarrow W(a)$ it is possible to remove the nuisance parameters from the limiting distribution by just applying the "long-run GLS transformation" $z_t = \Omega^{-1/2}y_{[Ta]}$. The ADF tests applied to the i'th elements of z_t are asymptotically distributed as $\int W_i dW_i / \sqrt{\int W_i^2}$, where W_i is the i'th element of $W \equiv W(a)$. Thus, as $T \to \infty$ the individual tests are independent and it is therefore natural to pool the individual test statistics as in the test suggested by Im et al. (2003). The resulting statistic is related to the second generation panel unit root test suggested by Harvey and Bates (2005) who suggest to apply a panel unit root test to $\widetilde{z}_t = \Sigma^{-1/2}y_t$, where Σ is the ordinary covariance matrix of the errors. There are however two important differences. First, we use the (estimated) long-run covariance matrix instead of the ordinary covariance matrix to remove the nuisance parameters from the limiting distribution. Second, whereas Harvey and Bates employ a pooled regression similar to Levin et al. (2002), we pool the test statistics. The latter approach has the advantage that the individual tests do not involve the eigenvalues of Ω. Since the individual ADF test is invariant to scaling, the test is equivalent to a test based on the transformed vector $z_t^* = V'y_t$, where V is the matrix of eigenvectors of Ω.

The details of the test statistic are provided in the following theorem.

Theorem 2: *Let $\widehat{\Omega}$ denote a long-run covariance matrix estimator obeying Assumption 2 and $\widehat{V}$ is the associated $n \times n$ matrix of eigenvectors. Under the null hypothesis $r = 0$ and $T \to \infty$ we have*

$$\sum_{i=1}^{n} \widetilde{\tau}_i \;\Rightarrow\; \sum_{i=1}^{n} \frac{\int W_i dW_i}{\sqrt{W_i^2}},$$

where $\widetilde{\tau}_j$ denotes the ADF statistic of the i'th component of $\widehat{z}_t^* = \widehat{V}'y_t$.

Note that the limiting distribution of this test statistic is much simpler than the ones involved by Theorem 1. This is due to the fact that the matrix $\widehat{\Omega}$ converges in distribution to Ω. Accordingly, the matrix of eigenvectors converge to a fixed matrix V, whereas in Theorem 1 the matrix of eigenvectors converge to the vector of random variables ξ_i^k that can be written as a (rather complicated) function of $W(a)$.

4 Empirical example

— to be completed —

5 Small sample properties

To investigate the small sample properties of the alternative test statistics we first consider the framework of a second generation unit root test where, under the null hypothesis, the time series are correlated random walks. In this situation the long-run covariance matrix Ω is identical to the ordinary covariance matrix $\Sigma = E(\Delta y_t \Delta y_t')$. Accordingly, apart from a contemporaneous correlation, there exists no intertemporal dynamic relationships among the variables. Since all test statistics are invariant to Ω, the critical values presented in Table 1 (see Appendix B) yield actual sizes very close to the nominal ones. Therefore, we omit the presentation of the empirical sizes in this case.

Under the alternative we assume that all time series are stationary AR(1) processes generated by $y_{i,t} = \varrho y_{i,t-1} + \varepsilon_{i,t}$, where $\varepsilon_{i,t} \sim N(0,1)$ and common autocorrelation parameter $\varrho < 1$. Table 2 presents the power of the alternative test statistics for $T = 200$ and different numbers of variables n. Since under the null hypothesis Δy_t is a vector of white noise, we can set the truncation parameter for the LR and PC statistics equal to zero and all DF tests of the statistic $\tau_k(r)$ can be performed without lag augmentation (e.g. $p_j = 0$).[1] Table 2 presents the empirical powers of the alternative test statistics. LR, PC and NP represent the test statistics considered in Theorem 1. *maxEval* and *Ped-Vog* refer to the maximum eigenvalue statistic of Johansen (1988) and the test statistic of Pedroni and Vogelsang (2005).

The results clearly indicate that the *lr-GLS* test of Section 4 outperform all competitors. Among the nonparametric tests of Theorem 1 the PC test performs best followed by the NP test and the LR test. In contrast, the LR and *Ped-Vog* statistics have virtually no power and may even have smaller

[1] In practice the truncation lag and the augmentation can be chosen by using the procedures suggest by Andrews (1991) and Perron and Ng (1996), for example. In our simulation exercise we abstract from the additional uncertainty that is due to employing selection procedures for the truncation and augmentation lags. Note that the NP and *Ped-Vog* statistics do not require to choose these additional parameters.

power than the size. For $\varrho = 0.9$ the latter test have some nontrivial power for $n = 5$ but for larger numbers of n the power disappears, whereas the tests suggested in Section 3 and 4 reject in all cases.

Next we consider the more interesting case of a scenario with "system dynamics", that is, the time series exhibit (Granger) causal relationships among the variables. In this case, the 2nd generation unit root tests suggested by Harvey and Bates (2005), Moon and Perron (2004) and Pesaran (2007), for example, are invalid. The vector of time series is generated as

$$\Delta y_t = \Pi y_t + A\Delta y_{t-1} + \varepsilon_t \quad \text{where } \Pi = (\varrho - 1)I_n$$

Under the null hypothesis $\varrho = 1$ all components are $I(1)$, whereas under the alternative $|\varrho| < 1$, the vector y_t is stationary. The autoregressive matrix A is generated randomly. To control for the persistence of the dynamics we set all eigenvalues of the matrix A equal to $-1 < \theta < 1$. This is done by generating randomly matrices A subject to the normalization $A'A = \theta I$. Table 3 reports the empirical sizes for the tests by using critical values as suggested in Remark B. Note that these critical values are obtained from time series without short-run dynamics. Therefore, it is interesting to investigate the actual sizes of the test statistics under short run correlation. The lag length p for the ADF tests used to construct $\tau_k(n)$ and the truncation lags for the kernel estimates $\widehat{\Omega}$ and $\widehat{\Psi}$ are set equal to $T^{1/3}$. Note that the linear combination $v'\Delta y_t$ does not have a (univariate) AR(1) representation. Therefore a higher order augmentation lag is required for the DF test applied to the linear combinations. Our results suggest that for moderately large values of θ the size distortions of the tests suggested in section 3 are small. Only if n and θ is large, then the size distortions become more severe. In contrast, the LR test and also the *Ped-Vog* test are less robust to short-run dynamics although the size distortions tends to zero as $T \to \infty$ (not reported).

Finally, we consider the (size adjusted) powers of the tests under system dynamics. By and large the results confirm our findings of the experiment

Table 1: Critical values for $\alpha = 0.05$ (standardized)

T	LR	PC	NP	maxEV	Ped-Vog
			$N = 10$		
100	-1.847	-1.861	-1.773	1.013	1.499
200	-1.867	-1.939	-1.802	1.052	1.578
300	-1.832	-1.825	-1.772	1.300	1.637
500	-1.729	-1.768	-1.707	1.597	1.715
1000	-1.692	-1.656	-1.671	1.800	1.703
			$N = 20$		
100	-3.285	-4.440	-3.519	-2.934	–0.979
200	-2.334	-2.797	-2.332	-0.713	0.814
300	-2.065	-2.287	-2.039	0.285	1.261
500	-1.863	-1.918	-1.819	1.055	1.569
1000	-1.644	-1.714	-1.710	1.707	1.719

without system dynamics. The three statistics suggested in Section 3 and the panel statistic of Section 4 are much more powerful than the *maxEV* and *Ped-Vog* statistics. For $\varrho = 0.95$ the latter two statistics fail to have power no matter of the parameter θ, whereas for $\varrho = 0.9$ the *Ped-Vog* statistic exhibit some nontrivial power but much lower than the other three statistics.

Table 2: Empirical power (correlated random walks), $T = 200$)

n	lr-GLS	LR	PC	NP	maxEV	Ped-Vog
	$\varrho = 0.99$					
5	0.498	0.141	0.296	0.136	0.041	0.044
10	0.827	0.074	0.353	0.102	0.047	0.053
15	0.940	0.045	0.317	0.099	0.056	0.053
20	0.983	0.052	0.299	0.089	0.037	0.054
25	0.998	0.043	0.312	0.068	0.033	0.046
	$\varrho = 0.95$					
5	1.000	0.974	0.999	0.939	0.052	0.147
10	1.000	0.911	0.999	0.916	0.003	0.061
15	1.000	0.743	1.000	0.841	0.000	0.042
20	1.000	0.581	1.000	0.843	0.000	0.008
25	1.000	0.572	1.000	0.873	0.000	0.001
	$\varrho = 0.9$					
5	1.000	1.000	1.000	1.000	0.239	0.490
10	1.000	1.000	1.000	1.000	0.006	0.289
15	1.000	1.000	1.000	1.000	0.000	0.149
20	1.000	1.000	1.000	1.000	0.000	0.050
25	1.000	1.000	1.000	1.000	0.000	0.019

Table 3: Empirical sizes (system dynamics)

n	lr-GLS	LR	PC	NP	maxEV	Ped-Vog
	$T = 200,\ N = 10$					
$\theta = 0.8$	0.065	0.194	0.203	0.171	0.341	0.820
$\theta = 0.5$	0.053	0.064	0.065	0.066	0.075	0.180
$\theta = 0.2$	0.047	0.064	0.068	0.068	0.054	0.050
$\theta = 0.0$	0.052	0.057	0.056	0.067	0.052	0.064
$\theta = -0.2$	0.048	0.081	0.078	0.091	0.054	0.079
$\theta = -0.5$	0.036	0.081	0.082	0.069	0.055	0.176
$\theta = -0.8$	0.060	0.202	0.193	0.148	0.294	0.791
	$T = 200,\ N = 20$					
$\theta = 0.8$	0.036	0.289	0.050	0.144	0.146	1.000
$\theta = 0.5$	0.037	0.079	0.033	0.057	0.047	0.662
$\theta = 0.2$	0.040	0.045	0.043	0.047	0.039	0.094
$\theta = 0.0$	0.051	0.051	0.040	0.043	0.040	0.082
$\theta = -0.2$	0.050	0.052	0.047	0.058	0.049	0.102
$\theta = -0.5$	0.033	0.092	0.030	0.060	0.063	0.672
$\theta = -0.8$	0.042	0.277	0.045	0.143	0.174	1.000

Table 4: Empirical power (size corrected)

n	lr-GLS	LR	PC	NP	maxEV	Ped/Vog
	$T = 200,\ N = 10$					
$\rho = 0.95$						
$\theta = 0.8$	1.000	0.538	0.809	0.528	0.050	0.144
$\theta = 0.5$	1.000	0.757	0.960	0.686	0.045	0.120
$\theta = -0.5$	1.000	0.758	0.959	0.680	0.025	0.118
$\theta = -0.8$	1.000	0.493	0.814	0.470	0.030	0.112
	$T = 200,\ N = 20$					
$\rho = 0.90$						
$\theta = 0.9$	1.000	0.222	0.534	0.207	0.021	0.087
$\theta = 0.5$	1.000	0.413	0.708	0.221	0.041	0.092
$\theta = -0.5$	1.000	0.506	0.715	0.229	0.030	0.082
$\theta = -0.9$	1.000	0.246	0.519	0.251	0.039	0.067

Table 5: Test statistics for real exchange rates

test	$p = 0$	$p = 1^{**}$	$p = 2$	crit.val.
LR	–5.422*	–3.147*	–4.049*	–1.678
PC	–5.137*	–3.268*	–3.171*	–1.641
NP	–5.424*	–2.828*	–2.707*	–1.676
Ped-Vog	–0.376	–0.376	–0.376	1.715
maxEV	2.643*	1.870*	6.326*	1.774

Note: Test statistics are normalized by subtracting the mean and dividing by the variance of the test statistic. The last column presents the (small sample) critical values with respect to a significance level of 0.05. *∗ Optimal lag length according to the AIC criterion.

References

[1] Banerjee, A., Marcellino, M. and C. Osbat (2005), Testing for PPP: Should we use panel methods?, Empirical Economics, 30, 77-91.

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

[3] Breitung, J. and S Das (2008), Testing for Unit Roots in Panels with a Factor Structure, *Econometric Theory*, 24, 88-108.

[4] Breitung, J. and M. Pesaran (2008) Unit Roots and Cointegration in Panels, in: L. Matyas and P. Sevestre (eds), *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*, Kluwer Academic Publishers, 279-322.

[5] Choi, I. (2006), Unit Root Tests for Cross-Sectionally Correlated Panels, *Econometric Theory and Practice: Frontiers of Analysis and Applied Research*, 311-333, Cambridge University Press.

[6] Cubadda, G., Hecq, A. and F.C. Palm (2008), Macro-panels and Reality, *Economics Letters*, 99, 537-540.

[7] Engle, R. F. and C. W. J. Granger (1987), Cointegration and error correction: representation, estimation and testing, *Econometrica*, 55, 251-278.

[8] Gengenbach, C., Palm, F.C., and J.P. Urbain (2008), Panel Unit Root Tests in the Presence of cross-sectional Dependencies: Comparison and Implications for Modelling, *Econometric Reviews*, forthcoming.

[9] Harvey, A. and D. Bates (2005), Multivariate Unit Root Tests, Stability and Convergence, University of Cambridge, DAE Working Paper No. 301, University of Cambridge, England.

[10] Hassler, U. (2006), A Note on Phillips-Perron-type Statistics for Cointegration Testing, *Economics Bulletin*, 3, 1-7.

[11] Im, K.S., M.H. Pesaran, and Y. Shin (2003), Testing for Unit Roots in Heterogenous Panels, *Journal of Econometrics*, 115, 53–74.

[12] Johansen, S. (1988), Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control*, 12, 231-254.

[13] Johansen, S., (1991), Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, 59, 1551–1580.

[14] Moon, R. and B. Perron (2004), Testing for Unit Root in Panels with Dynamic Factors, *Journal of Econometrics*, 122, 81–126.

[15] Pedroni, P. and T. Vogelsang (2005), Robust Tests for Unit Roots in Heterogeneous Panels, Mimeo, Williams College.

[16] Pesaran, M.H. (2007), A simple panel unit root test in the presence of cross section dependence, *Journal of Applied Econometrics*, 22, 265–312.

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

[18] Phillips, P.C.B. and V. Solo (1992), Asymptotics for linear processes, *Annals of Statistics*, 20, 971-1001.

[19] Shintani, M. (2001), A simple cointegrating rank test without vector autoregression, *Journal of Econometrics,* 105, 337–362.

[20] Wagner, M. (2008), On PPP, Unit Roots and Panels, *Empirical Economics*, 35, 229-249.

Appendix A:

Proof of Theorem 1:

First, we consider the nonparametric version of the LR statistic that is based on the eigenvalue problem (2), which is equivalent to

$$|\lambda\Omega^{-1/2}S_{11}\Omega^{-1/2} - \Omega^{-1/2}(S_{01} - \widehat{\Psi})'\widehat{\Omega}^{-1}(S_{01} - \widehat{\Psi})\Omega^{-1/2}| = 0 \qquad (9)$$

in the sense that the eigenvalues of both problems are identical and the eigenvectors $\widetilde{v}_j^{LR}$ result as $\widetilde{v}_j^{LR} = \Omega^{1/2}v_j^{LR}$, where $\Omega_{11}^{1/2}$ is a symmetric matrix obeying $\Omega_{11}^{1/2}\Omega_{11}^{1/2} = \Omega$. As $T \to \infty$ the eigenvalue problem is equivalent to solving

$$|\lambda \int WW' - \int WdW' \int dWW'| = 0,$$

where W is a vector of standard Brownian motions. Since this eigenvalue problem does not depend on unknown parameters, it follows that $v_j^{LR'}y_t = \widetilde{v}_j^{LR'}(\Omega^{-1/2}y_t)$ are functions of Brownian motions and free of nuisance parameters.

Let $[I_r, -B_{LR}]'$ denote the $n \times r$ matrix of the re-nomalized eigenvectors obeying

$$\left(\lambda \int WW' - \int WdW' \int dWW'\right) \begin{bmatrix} I_r \\ B_{LR} \end{bmatrix} = 0. \qquad (10)$$

Using the partitioning $W = [W_1', W_2']'$, where W_1 and W_2 are r and $(n-r)$ dimensional subvectors, respectively, it follows from the lower $n-r$ equations of (10)

$$\lambda_j \int W_2 W_1' - \lambda_j \int W_2 W_2' B_{LR} - \int W_2 dW' \int dW W_1' + \int W_2 dW' \int dW W_2' B_{LR} = 0.$$

Solving for B_{LR} yields

$$B_{LR} = \left(\lambda_j \int W_2 W_2' - \int W_2 dW' \int dW W_2'\right)^{-1} \left(\lambda_j \int W_2 W_1' - \int W_2 dW' \int dW W_1'\right)$$

and λ_j is the j'th largest eigenvalue of the matrix $\mathrm{tr}[\int dWW'(\int WW')^{-1} \int WdW']$. These result can be used to derive the limiting process of the linear combination $v_j^{LR\prime} y_t = \widetilde{v}_j^{LR\prime} \Omega^{-1/2} y_t$. To this end we have to re-normalize the eigenvectors, that is, we have to find a matrix R such that $R[I_r, -B_{LR}']' = \widetilde{V}_r = [\widetilde{v}_1, \ldots, \widetilde{v}_r]$. From the Jordan decomposition of a positive definite symmetric matrix we have

$$[I_r, -B_{LR}'](\int WW')[I_r, -B_{LR}']' = H_{LR} \Psi_{LR} H_{LR}',$$

where Ψ_{LR} is a diagonal matrix with the ordered eigenvalues of the main diagonal and $H_{LR}' H_{LR} = I_r$. Accordingly, we obtain the desired normalization as $\widehat{V}_r^{LR} = \widetilde{V}_r^{LR} H_{LR} \Psi_{LR}^{-1/2}$. Hence,

$$\frac{1}{\sqrt{T}} \widetilde{V}_r^{LR\prime} y_t \Rightarrow \Psi_{LR}^{-1/2} H_{LR}' (W_1 - B' W_2) \equiv \Xi_{LR} W$$

and $T^{-1/2} \widetilde{v}_j^{LR\prime} y_t \Rightarrow \xi_j^{LR\prime} W$, where $\xi_j^{LR\prime}$ is the j'th row of Ξ^{LR}.

(ii) As $T \to \infty$ the PC eigenvalue problem (6) is equivalent to solving

$$|\nu \int WW' - I_n| = 0.$$

The eigenvectors associated with the r largest eigenvalues can be normalized as $\widetilde{V}_{PC}^* = [I, -B_{PC}']'$. Similar as for the LR statistic it can be shown that

$$B_{PC} = \left(\lambda_j \int W_2 W_2' - I_r\right)^{-1} \left(\lambda_j \int W_2 W_1' - J_{21}\right)$$

where J_{21} is the lower left $(n-r)\times r$ block of the $n\times n$ identity matrix and λ_j is the j'th eigenvalue of the matrix $\left(\int WW'\right)^{-1}$. Using the Jordan decomposition

$$[I_r, -B'_{PC}]\left(\int WW'\right)[I_r, -B'_{PC}] = H_{PC}\Psi_{PC}H'_{PC}$$

we obtain

$$\frac{1}{\sqrt{T}}\widetilde{V}_r^{PC'}y_t \Rightarrow \Psi_{PC}^{-1/2}H'_{PC}(W_1 - B'W_2) \equiv \Xi_{PC}W.$$

(iii) The NP eigenvalue problem (6) is asymptotically equivalent to solving

$$|\mu\int ZZ' - \int WW'| = 0,$$

where $Z \equiv Z(a) = \int_0^a W(a)$. It follows that

$$B_{NP} = \left(\lambda_j\int V_2V_2' - \int W_2W_2'\right)^{-1}\left(\lambda_j\int V_2V_1' - \int W_2W_1'\right)$$

and μ_j is the j'th eigenvalue of $\int WW'(\int VV')^{-1}$. Letting

$$[I_r, -B'_{NP}]\left(\int WW'\right)[I_r, -B'_{NP}] = H_{NP}\Psi_{NP}H'_{NP}$$

yields

$$\frac{1}{\sqrt{T}}\widetilde{V}_r^{NP'}y_t \Rightarrow \Psi_{NP}^{-1/2}H'_{NP}(W_1 - B'W_2) \equiv \Xi_{NP}W.$$