



HAL
open science

A Meta Analytic Approach to Testing for Panel Cointegration

Christoph Hanck

► **To cite this version:**

Christoph Hanck. A Meta Analytic Approach to Testing for Panel Cointegration. Communications in Statistics - Simulation and Computation, 2009, 38 (05), pp.1051-1070. 10.1080/03610910902750039 . hal-00514347

HAL Id: hal-00514347

<https://hal.science/hal-00514347>

Submitted on 2 Sep 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



A Meta Analytic Approach to Testing for Panel Cointegration

Journal:	<i>Communications in Statistics - Simulation and Computation</i>
Manuscript ID:	LSSP-2008-0161.R2
Manuscript Type:	Original Paper
Date Submitted by the Author:	12-Jan-2009
Complete List of Authors:	Hanck, Christoph; Universiteit Maastricht, Quantitative Economics
Keywords:	Panel cointegration tests, Monte Carlo study, Meta Analysis
Abstract:	We propose new tests for panel cointegration by extending the panel unit root tests of Choi (JIMF 2001) and Maddala and Wu (OBES 1999) to the panel cointegration case. The tests are flexible, intuitively appealing and relatively easy to compute. We investigate the finite sample behavior in a simulation study. Several variants of the tests compare favorably in terms of both size and power with other widely used panel cointegration tests.
<p>Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.</p> <p>panel ci meta approach commstat rev3.tex panel ci meta approach commstat rev3.bbl</p>	



A Meta Analytic Approach to Testing for Panel Cointegration*

Christoph Hanck[†]

January 12, 2009

Abstract

We propose new tests for panel cointegration by extending the panel unit root tests of Choi [2001] and Maddala and Wu [1999] to the panel cointegration case. The tests are flexible, intuitively appealing and relatively easy to compute. We investigate the finite sample behavior in a simulation study. Several variants of the tests compare favorably in terms of both size and power with other widely used panel cointegration tests.

Keywords: Panel cointegration tests, Monte Carlo study, Meta Analysis

*Partial support by DFG under Sonderforschungsbereich 475 and Ruhr Graduate School in Economics is gratefully acknowledged. I am indebted to an anonymous referee whose suggestions greatly helped improve the paper.

[†]Universiteit Maastricht, Tongersestraat 53, 6211 LM Maastricht, The Netherlands. Tel. (+31) 43-3883815, c.hanck@ke.unimaas.nl.

1 Introduction

There is wide consensus in economics that cointegration is an important statistical concept which is implied by many economic models. In practice, however, evidence of cointegration or non-cointegration is often weak because of the rather small sample sizes typically available in macroeconometrics. To overcome this problem, the cointegration methodology has recently been extended to panel data. This allows the researcher to work with larger samples, thereby improving the performance of tests and estimators.

Pedroni [2004] and Kao [1999] generalize the residual-based tests of Engle and Granger [1987] and Phillips and Ouliaris [1990], Larsson et al. [2001] extend the Johansen [1988] tests to panel data while McCoskey and Kao [1998] propose a test for the null of panel cointegration in the spirit of Shin [1994].

The present paper studies some new tests for panel cointegration, extending the p -value combination panel unit root tests of Maddala and Wu [1999] and Choi [2001] to the cointegration setting. In this framework, it is straightforward to account for unbalanced panels and arbitrary heterogeneity in the serial correlation structure of the series. Moreover, the tests are simple to implement and intuitively appealing. We explore the finite sample performance of the tests in a simulation study. Certain variants of the tests compare favorably with many of the previously proposed panel cointegration tests.

The next section introduces the tests. Section 3 presents the finite sample study. Section 4 concludes.

2 P -Value Combination Tests for Panel Cointegration

The present section develops the tests for panel cointegration. The following notation is used throughout. \mathbf{x}_{ik} is a $(T_i \times 1)$ column vector collecting the observations on the k th variable of unit i of the panel, where $i = 1, \dots, N$ and $k = 1, \dots, K$. To the K

variables we may add time polynomials of order up to 2, i.e. constants, trend and squared trend terms. The number of observations T_i per unit may depend on i , i.e. the panel may be unbalanced. Denote by p_i the marginal significance level, or p -value, of a time series cointegration test applied to the i th unit of the panel. Let θ_{i,T_i} be a time series cointegration test statistic on unit i for a sample size of T_i . F_{T_i} denotes the exact, finite T_i null cumulative distribution function (cdf) of θ_{i,T_i} . Since the tests we shall consider here are one-sided, $p_i = F_{T_i}(\theta_{i,T_i})$ if the test rejects for small values of θ_{i,T_i} and $p_i = 1 - F_{T_i}(\theta_{i,T_i})$ if the test rejects for large values of θ_{i,T_i} . However, F_{T_i} is unknown in practice, such that we need to work with a suitable approximation which will be described in more detail below.

We are interested in testing the following null hypothesis

$$H_0 : \text{There is no (within-unit) cointegration for any } i, i = 1, \dots, N, \quad (1)$$

against the alternative

$$H_1 : \text{There is (within-unit) cointegration for at least one } i, i = 1, \dots, N.$$

The alternative H_1 states that a rejection is evidence of 1 to N cointegrated units in the panel. That is, a rejection neither allows to conclude that the entire panel is cointegrated nor does it provide information about the number of units of the panel that exhibit cointegrating relationships.

The main idea of the suggested testing principle has been used in meta analytic studies for a long time [cf. Fisher, 1932; Hedges and Olkin, 1985]. Consider the testing problem on the panel as consisting of N testing problems, one for each unit of the panel. That is, conduct N separate time series cointegration tests and obtain the corresponding p -values of the test statistics.¹ We make the following assumptions [see Pedroni, 2004].

ASSUMPTION 1 (*Continuity*)

As $T_i \rightarrow \infty$, θ_{i,T_i} has a continuous asymptotic cdf F_i under H_0 for all $i =$

¹Both Maddala and Wu [1999] and Choi [2001] suggest extending their panel unit root tests to the cointegration case. However, to the best of our knowledge, they do not provide an actual implementation nor do they investigate the performance of the tests. Furthermore, our approach is more general and likely to be more accurate in some respects to be discussed below.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

$1, \dots, N$.

ASSUMPTION 2 (*Cross-Sectional Uncorrelatedness*)

$x_{ik,t} = x_{ik,t-1} + \xi_{ik,t}$, $t = 1, \dots, T_i$, $i = 1, \dots, N$, $k = 1, \dots, K$. Let $\boldsymbol{\xi}_{i,t} \equiv (\xi_{i1,t}, \dots, \xi_{iK,t})'$. We require $E[\boldsymbol{\xi}_{i,t}\boldsymbol{\xi}_{j,s}'] = \mathbf{0} \forall s, t = 1, \dots, T_i$ and $i \neq j$. The error process $\boldsymbol{\xi}_{i,t}$ is generated as a linear vector process $\boldsymbol{\xi}_{i,t} = C_i(L)\boldsymbol{\eta}_{i,t}$, where L is the lag operator and C_i are coefficient matrices. $\boldsymbol{\eta}_{i,t}$ is vector white noise.

Remark 1. Assumption 1 asymptotically ensures, among other things, a uniform p -value distribution of the time series test statistics under H_0 on the unit interval: $p_i \sim \mathcal{U}[0, 1]$ ($i \in \mathbb{N}_N$) [see, e.g., Bickel and Doksum, 2001, Sec. 4.1]. It is satisfied by the time series cointegration tests considered in this paper.

Remark 2. The second assumption is strong [see, e.g., Banerjee et al., 2005]. It implies that the different units of a panel must not be linked to each other beyond relatively simple forms of correlation such as common time effects which can be eliminated by demeaning across the cross sectional dimension. This assumption is likely to be violated in many typical macroeconomic panel data sets. We will return to this issue below.

We now present the test statistics. Combine the N p -values of the individual time series cointegration tests, p_i , $i = 1, \dots, N$, as follows to obtain three test statistics for panel cointegration:

$$P_{\chi^2} = -2 \sum_{i=1}^N \ln(p_i) \tag{2a}$$

$$P_{\Phi^{-1}} = N^{-\frac{1}{2}} \sum_{i=1}^N \Phi^{-1}(p_i) \tag{2b}$$

$$P_t = \sqrt{\frac{3(5N+4)}{\pi^2 N(5N+2)}} \sum_{i=1}^N \ln\left(\frac{p_i}{1-p_i}\right) \tag{2c}$$

When considered together we refer to Eqs. (2a) to (2c) as P tests from now on. The P tests, via pooling p -values, provide convenient tests for panel cointegration by imposing minimal homogeneity restrictions on the panel. For instance, the different units of the panel can be unbalanced. Furthermore, the evidence for (non-)cointegration is first investigated for each unit of the panel and then compactly expressed with the p -value of the

time series cointegration test. Hence, the coefficients describing the relationship between the different variables for each unit of the panel can be heterogeneous across i . Thus, the availability of large- T time series allows for pooling the data into a panel without having to impose strong homogeneity restrictions on the slope coefficients as in traditional panel data analysis.² Under Assumptions 1 and 2, as $T_i \rightarrow \infty$ for all i , the test statistics are asymptotically distributed as

$$\begin{aligned} P_{\chi^2} &\rightarrow_d \chi_{2N}^2 \\ P_{\Phi^{-1}} &\rightarrow_d \mathcal{N}(0, 1) \\ P_t &\xrightarrow{d, \text{approx.}} \mathcal{T}_{5N+4}, \end{aligned} \quad (3)$$

where χ^2 is a chi-squared distributed random variable and \mathcal{T} denotes Student's t distribution. The subscripts give the degrees of freedom.

Remark 3. We require that $T_i \rightarrow \infty$ for all i , as, unlike in Fisher's [1932] original contribution, the P tests are not exact when dealing with nonstationary data. This is because, as F_{T_i} is unknown, we need to employ an approximation to F_{T_i} for finite T . This matters because the null hypothesis (1) is not a simple one and the available test statistics are not pivotal in finite samples. H_0 is satisfied by a wide class of multivariate nonstationary processes. See, for instance, the fairly general framework of Phillips and Ouliaris [1990]. Hence, for finite T , the p -values of the test need no longer be uniformly distributed on the unit interval, even if the true Data Generating Process (DGP) of the time series is from the null hypothesis set of nonstationary but non-cointegrated processes. Thus, we need $T_i \rightarrow \infty$ for all i to ensure that $p_i \sim \mathcal{U}[0, 1]$ under H_0 so as to be able to invoke the limiting distribution in (3). We do not need $N \rightarrow \infty$ to obtain (3). **Of course, these asymptotic distributions are, as usual, only approximations to the finite-sample ones. Section 3 evaluates their accuracy via extensive simulation experiments.**

Using consistent time series cointegration tests, $p_i \rightarrow_p 0$ under the alternative of cointegration. Hence, quite intuitively, the smaller p_i , the more it acts towards rejecting the null of no panel cointegration. The decision rule therefore is to reject the null of no panel

²For an overview of panel data models relying on $N \rightarrow \infty$ asymptotics see Hsiao [2003].

1
2
3
4
5 cointegration when P_{χ^2} exceeds the critical value from a χ^2_{2N} distribution at the desired
6 significance level. For (2b) and (2c) one would reject for large negative values of the panel
7 test statistics $P_{\Phi-1}$ and P_t , respectively.
8
9

10
11 We now discuss how to obtain the p -values required for computation of the P test statis-
12 tics. Hanck [2008] shows that using accurate p -values is crucial to achieve a precise control
13 of the type I error rate in meta analytic panel testing. The null distributions of both resid-
14 ual and system-based time series cointegration tests converge to functionals of Brownian
15 motion. Hence, analytic expressions of the asymptotic cdfs F_i are hard to obtain, and
16 p -values of the tests cannot simply be obtained by evaluating F_i . A remedy frequently
17 adopted in the literature is to approximate F_i (which, in turn, is typically used to approx-
18 imate F_{T_i}) by Monte Carlo simulation. However, this approach is unsatisfactory for (at
19 least) the following reason. These simulations are typically only performed for one sam-
20 ple size which is meant to provide an approximation to the asymptotic distribution. This
21 sample size need neither be large enough to be useful as an asymptotic approximation
22 nor does it generally yield an accurate estimate of F_{T_i} for other sample sizes. MacKinnon
23 et al. [1999] show for certain special cases where analytic expressions of the asymptotic
24 distribution functions F_i are available that this approach may deliver fairly inaccurate es-
25 timates of F_i . In the time series case, it is now fairly standard practice to report p -values
26 of unit root and cointegration tests using the results of the response surface regressions
27 introduced by MacKinnon [1991]. We follow this approach here.
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43

44 The null hypothesis (1) formulates no precise econometric characterization of (non-) coin-
45 tegration. This is to allow for generality in testing the long-run equilibrium properties
46 of the series, enabling the researcher to use whichever time series tests seem suitable to
47 test for time series (non-)cointegration in the different units of the panel. We use p -values
48 of the Augmented Dickey-Fuller (*ADF*) cointegration tests [Engle and Granger, 1987] as
49 provided by MacKinnon [1996].³ That is, the p -values are obtained from the t -statistic of
50
51
52
53
54
55

56 ³MacKinnon improves upon his prior work by using a heteroscedasticity and serial correlation robust
57 technique to approximate between the estimated quantiles of the response surface regressions. Our
58 application is based on a translation of James MacKinnon's Fortran code into a GAUSS procedure which
59 is available upon request. The procedure implements all panel data tests developed in this section.
60

$\gamma_i - 1$ in the *OLS* regression

$$\Delta \hat{u}_{i,t} = (\gamma_i - 1) \hat{u}_{i,t-1} + \sum_{p=1}^P \nu_p \Delta \hat{u}_{i,t-p} + \epsilon_{i,t}.$$

Here, $\hat{u}_{i,t}$ is the usual residual from a first stage *OLS* regression of one of the \mathbf{x}_{ik} on the remaining $\mathbf{x}_{i,-k}$. Alternatively, one could capture serial correlation by the semiparametric approach of Phillips and Ouliaris [1990]. Finally, we obtain the p -values for the Johansen [1988] λ_{trace} and λ_{max} tests provided in MacKinnon et al. [1999]. That is, we test for the presence of $h = 0$ cointegrating relationships by estimating the number of significantly non-zero eigenvalues of the matrix $\hat{\Pi}_i$ estimated from the Vector Error Correction Model

$$\Delta \mathbf{x}_{i,t} = -\Pi_i \mathbf{x}_{i,t-P} + \sum_{p=1}^{P-1} \Gamma_{i,p} \Delta \mathbf{x}_{i,t-p} + \epsilon_{i,t}$$

by the λ_{trace} -test

$$\lambda_{trace,i}(h) = -T \sum_{k=h+1}^K \ln(1 - \hat{\pi}_{k,i}) \quad (4)$$

and the λ_{max} -test

$$\lambda_{max,i}(h|h+1) = -T \ln(1 - \hat{\pi}_{h+1,i}). \quad (5)$$

Here, $\hat{\pi}_{k,i}$ denotes the k th largest eigenvalue of $\hat{\Pi}_i$. In (4), the alternative is a general one, while one tests against $h + 1$ cointegration relationships in (5). For small T , the Johansen [1988] tests are known to be oversized [see e.g. Cheung and Lai, 1993]. To remove this size distortion, we employ Cheung and Lai's [1993] small sample degree of freedom correction. As mentioned above, this is particularly important in our framework as the size distortions in the time series tests would (via unduly small p -values) otherwise "add up" when combined over N to form the panel test statistics [Hanck, 2008], to yield an arbitrarily severely size-distorted panel test.

To summarize, we obtain the p -values required for performing the P tests from the most widely used time series cointegration tests.

3 Finite Sample Performance

We now present a Monte Carlo study of the finite sample performance of the tests proposed in the previous section. The Data Generating Process (DGP) is similar to the one used by Engle and Granger [1987]. The extension to the panel data setting is discussed in Kao [1999]. For simplicity, only consider the bivariate case, i.e. $K = 2$:

DGP A

$$x_{i,1t} - \alpha_i - \beta x_{i,2t} = z_{i,t}, \quad a_1 x_{i,1t} - a_2 x_{i,2t} = w_{i,t}$$

where

$$z_{i,t} = \rho z_{i,t-1} + e_{zi,t}, \quad \Delta w_{i,t} = e_{wi,t}$$

and

$$\begin{pmatrix} e_{zi,t} \\ e_{wi,t} \end{pmatrix} \stackrel{iid}{\sim} \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \psi\sigma \\ \psi\sigma & \sigma^2 \end{bmatrix} \right)$$

Remark 4. When $|\rho| < 1$ the equilibrium error in the first equation is stationary such that $x_{i1,t}$ and $x_{i2,t}$ are cointegrated with $\beta_i = (1 \quad -\alpha_i \quad -\beta)'$.

Remark 5. When writing the above DGP as an error correction model [see, e.g., Gonzalo, 1994] it is immediate that $x_{i2,t}$ is weakly exogenous when $a_1 = 0$.

We investigate all combinations of the following values for the parameters of the model: $\beta = 2$, $a_1 \in \{0, 1\}$, $a_2 = -1$, $\sigma \in \{0.5, 1\}$, $\rho \in \{0.9, 0.99, 1\}$ and $\psi \in \{-0.5, 0, 0.5\}$. The fraction of cointegrated series in the panel is increased from 0 to 1 in steps of 0.1, i.e. $\delta \in \{0, 0.1, \dots, 1\}$. The dimensions of the panel are $N \in \{10, 20, 50, 100, 150\}$ and, after having discarded 150 initial observations, $T \in \{10, 30, 50, 100, 250, 500\}$, for a total of $2 \times 1 \times 2 \times 3 \times 3 \times 11 \times 5 \times 6 = 11,880$ experiments. For a given cross-sectional dimension, the unit specific intercepts are drawn as $\alpha_i \sim \mathcal{U}[0, 10]$ and kept fixed for all T_i . Each experiment involves $M = 5,000$ replications.⁴ We choose a common β for all i in order to be able to compare the performance of our tests with results for other

⁴Uniform random numbers are generated using the KM algorithm from which Normal variates are created with the fast acceptance-rejection algorithm, both implemented in GAUSS. Part of the calculations are performed with COINT 2.0 by Peter Phillips and Sam Ouliaris.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

panel cointegration tests as reported by Gutierrez [2003]. The p -values are from the Engle and Granger [1987] ADF test, holding the number of lagged differences fixed at 1. We further test for cointegration using the λ_{trace} -test for $h = 0$ vs. an unrestricted number of cointegrating relationships. **As the null hypothesis (1) is tested against an unspecified number of cointegrating relationships, we employ λ_{trace} rather than λ_{max} . The setup of course alternatively allows for using λ_{max} if one is concretely interested in testing against exactly one cointegrating relationship.**

For brevity, we only give the results for $\psi = 0$, $a_1 = 0$ and $\sigma = 1$.⁵ Table I shows the empirical size of the tests ($\rho = 1$) at the nominal 5% level using the ADF - and λ_{trace} -tests as the underlying time series tests. The main conclusions are as follows. First, the Engle/Granger-based tests control size rather well. Only for very small T do we observe an (albeit tolerable) size distortion. As predicted analytically by Hanck [2008], the tests' size distortions "add up" over N to become more size distorted as N increases. These mild distortions vanish quickly with increasing T . The P_{χ^2} test seems to have slightly better size than the other two.⁶

Second, the Johansen-based tests are oversized in panels of small time series dimensions. As pointed out above, this is because the underlying λ_{trace} -test overrejects for short time series when using asymptotic critical values. Apparently, the Cheung and Lai [1993] correction factors employed here do not completely eliminate that distortion. **This is not entirely surprising as Cheung and Lai [1993] conduct their response surface regressions for $T \geq 33$, such that the corrections used here for $T = 10$ are effectively only extrapolations.** (For $T = 10$, we found the corrected univariate λ_{trace} -test to have a finite-sample size of around 10%. The uncorrected finite-sample size exceeds 30%.) This flaw then inevitably carries over to the panel tests via erroneously

⁵The full set of results of the finite sample study are available upon request. Broadly speaking, a lower σ seems to have little, if any, systematic effect. Correlation in the error processes ($\psi \neq 0$) has a slightly negative effect on power.

⁶We also investigate whether using MacKinnon's [1996] p -values improves the behavior of the tests relative to obtaining quantiles by generating only one set of replicates. For smaller panels, the latter approach (with 50,000 replications) exhibits non-negligible upward size distortions even when using quantiles specifically generated for the sample sizes considered. Interestingly, however, there does not seem to be a trend towards higher distortions with increasing N . For medium- and large-dimensional panels neither approach has a clear advantage over the other.

TABLE I—EMPIRICAL SIZE OF THE P TESTS

T	N	ADF					λ_{trace}				
		10	20	50	100	150	10	20	50	100	150
<i>(i) P_{χ^2}</i>											
10		0.058	0.062	0.068	0.069	0.072	0.241	0.345	0.604	0.829	0.932
30		0.049	0.047	0.044	0.048	0.042	0.050	0.046	0.044	0.036	0.032
50		0.055	0.051	0.044	0.044	0.044	0.051	0.044	0.039	0.036	0.034
100		0.047	0.050	0.046	0.047	0.051	0.052	0.046	0.047	0.043	0.031
250		0.050	0.046	0.057	0.049	0.049	0.052	0.044	0.044	0.048	0.050
500		0.053	0.048	0.051	0.048	0.050	0.049	0.050	0.051	0.047	0.045
<i>(ii) $P_{\Phi^{-1}}$</i>											
10		0.051	0.047	0.039	0.030	0.026	0.203	0.277	0.484	0.720	0.842
30		0.045	0.043	0.038	0.037	0.029	0.049	0.045	0.036	0.033	0.025
50		0.054	0.044	0.041	0.040	0.037	0.046	0.039	0.033	0.030	0.030
100		0.049	0.052	0.048	0.047	0.045	0.051	0.048	0.043	0.041	0.026
250		0.053	0.044	0.052	0.047	0.044	0.048	0.047	0.048	0.048	0.048
500		0.052	0.051	0.049	0.048	0.051	0.053	0.048	0.050	0.050	0.050
<i>(iii) P_t</i>											
10		0.052	0.047	0.038	0.027	0.025	0.215	0.293	0.509	0.740	0.860
30		0.046	0.043	0.038	0.037	0.028	0.051	0.045	0.034	0.035	0.024
50		0.055	0.045	0.043	0.040	0.035	0.047	0.038	0.034	0.031	0.030
100		0.049	0.050	0.050	0.048	0.045	0.051	0.048	0.042	0.041	0.026
250		0.054	0.045	0.052	0.048	0.044	0.048	0.046	0.048	0.047	0.050
500		0.053	0.050	0.050	0.046	0.051	0.052	0.048	0.051	0.050	0.049

Note: $\rho = 1$, $\psi = 0$, $\sigma = 1$ and $a_1 = 0$. $M = 5,000$ replications.

5% nominal level. ADF and λ_{trace} are the underlying time series tests.

small p -values. (Note, though, that time series as short as $T \approx 10$ are rather uncommon in typical macroeconomic applications.) This size distortion vanishes for $T \geq 30$, such that the λ_{trace} -based P tests can be recommended for application at least for $T \geq 50$.

We now relate our results to those of some other recently proposed panel cointegration tests.⁷ We first give the key statistics of the various tests that are considered. For more details refer to the original contributions. Furthermore, Banerjee [1999], Baltagi and Kao [2000] or Breitung and Pesaran [2008] provide surveys of the literature.

Pedroni [2004]

Pedroni [2004] derives seven different tests for panel cointegration. These may be categorized according to what information on the different units of the panel is pooled. The

⁷Gutierrez [2003] provides a power study of these tests. He does however not analyze the finite sample size, whence our study should be viewed as complementary to his.

“Group-Mean” Statistics are essentially means of the conventional time series tests [see Phillips and Ouliaris, 1990]. The “Within” Statistics separately sum the numerator and denominator terms of the corresponding time series statistics. Let $A_i = \sum_{t=1}^T \tilde{e}_{i,t} \tilde{e}'_{i,t}$, where $\tilde{e}_{i,t} = (\Delta \hat{e}_{i,t}, \hat{e}_{i,t-1})'$. The $\hat{e}_{i,t}$ are obtained from heterogenous Engle/Granger-type first stage *OLS* regressions of an \mathbf{x}_{ik} on the remaining $\mathbf{x}_{i,-k}$, possibly including some deterministic regressors. We consider the “Group- ρ ”, “Panel- ρ ” and (nonparametric) “Panel- t ”-test statistics which are given by, respectively,

$$\begin{aligned} \tilde{Z}_{\hat{\rho}_{NT-1}} &= \sum_{i=1}^N A_{22i}^{-1} (A_{21i} - T \hat{\lambda}_i), \\ Z_{\hat{\rho}_{NT-1}} &= \left(\sum_{i=1}^N A_{22i} \right)^{-1} \sum_{i=1}^N (A_{21i} - T \hat{\lambda}_i) \quad \text{and} \\ Z_{\hat{t}_{NT}} &= \left(\tilde{\sigma}_{NT}^2 \sum_{i=1}^N A_{22i} \right)^{-1/2} \sum_{i=1}^N (A_{21i} - T \hat{\lambda}_i). \end{aligned}$$

The expressions $\hat{\lambda}_i$ and $\tilde{\sigma}_{NT}^2$ estimate nuisance parameters from the long-run conditional variances. After proper standardization, all statistics have a standard normal limiting distribution. The decision rule is to reject the null hypothesis of no panel cointegration for large negative values.

Kao [1999]

Kao [1999] proposes five different panel extensions of the time series (A)DF-type tests. We focus on those that do not require strict exogeneity of the regressors. More specifically,

$$\begin{aligned} DF_{\rho}^* &= \frac{\sqrt{NT}(\hat{\rho} - 1) + \frac{3\sqrt{N}\hat{\sigma}_{\nu}^2}{\hat{\sigma}_{0\nu}^2}}{\sqrt{3 + \frac{36\hat{\sigma}_{\nu}^4}{5\hat{\sigma}_{0\nu}^4}}} \quad \text{and} \\ DF_t^* &= \frac{t_{\rho} + \frac{\sqrt{6N}\hat{\sigma}_{\nu}^2}{2\hat{\sigma}_{0\nu}}}{\sqrt{\frac{\hat{\sigma}_{0\nu}^2}{2\hat{\sigma}_{\nu}^2} + \frac{3\hat{\sigma}_{\nu}^2}{10\hat{\sigma}_{\nu}^2}}}. \end{aligned}$$

Here, $\hat{\rho}$ is the estimate of the AR(1) coefficient of the residuals from a fixed effects panel regression and t_{ρ} is the associated t -statistic. The remaining terms play a role similar

to the nuisance parameter estimates in the Pedroni [2004] tests. Again, both tests are standard normal under the null of no panel cointegration and reject for large negative values.

Larsson et al. [2001]

The panel cointegration test of Larsson et al. [2001] applies a Central Limit Theorem to (4). Defining $\bar{\lambda}_{trace} = N^{-1} \sum_{i=1}^N \lambda_{trace,i}$, their panel cointegration test statistic is given by

$$\Upsilon_{LR} = \sqrt{N} \left(\frac{\bar{\lambda}_{trace} - E[\bar{\lambda}_{trace}]}{\sqrt{\text{Var}[\bar{\lambda}_{trace}]}} \right).$$

Under some conditions, including $\sqrt{NT}^{-1} \rightarrow 0$, Larsson et al. [2001] can show that $\Upsilon_{LR} \xrightarrow{T,N} \mathcal{N}(0, 1)$. The moments are obtained by stochastic simulation and are tabulated in the paper. The null hypothesis of no cointegration at a level α is rejected if the test statistic exceeds the $(1 - \alpha)$ -quantile of the standard normal distribution, i.e. for large values.

Table II reports the empirical sizes of the other panel cointegration tests presented above. All the other tests have difficulty to control size for small T .⁸ In addition, Kao's tests also seem to require a larger N to work well. In contrast, the empirical size of Pedroni's tests improves quite rapidly with T for any N . The small size distortion of the λ_{trace} test seems to carry over to a corresponding severe size distortion for the system-based Υ_{LR} test, which however vanishes quite fast with T . Comparing these results with those for the P tests, we find the P tests to compare quite favorably with existing panel cointegration tests in terms of finite-sample size.

Table III shows the size-adjusted power of the P tests at $\rho = 0.9$. The major findings are as follows. Both the Engle/Granger- and Johansen-based tests behave consistently in that power for all variants eventually grows with both dimensions. The use of panel data

⁸These size distortions are well in line with results found by Kao [1999] and Larsson et al. [2001] (cf. their Tables 4 and 2, resp.). Pedroni [2004] conducts experiments for $T \geq 40$ only. Any remaining differences are due to differences in the underlying DGP. As these results suggest that the asymptotic approximations should be used with care for very small T , future research might attempt to provide correction factors similar to those of Cheung and Lai [1993].

TABLE II—EMPIRICAL SIZE OF THE OTHER COINTEGRATION TESTS

T	N	10	20	50	100	150	10	20	50	100	150
		$(i) Z_{\hat{\rho}_{NT-1}}$					$(ii) Z_{\hat{t}_{NT}}$				
10		0.022	0.025	0.050	0.084	0.131	0.471	0.680	0.941	0.998	1.000
30		0.096	0.109	0.147	0.224	0.276	0.139	0.187	0.303	0.460	0.597
50		0.101	0.090	0.119	0.152	0.177	0.102	0.101	0.147	0.203	0.250
100		0.097	0.100	0.103	0.133	0.154	0.070	0.076	0.090	0.111	0.126
250		0.095	0.083	0.082	0.084	0.087	0.066	0.061	0.067	0.073	0.071
500		0.088	0.084	0.078	0.072	0.083	0.059	0.064	0.062	0.065	0.068
		$(iii) \tilde{Z}_{\hat{\rho}_{NT-1}}$					$(iv) DF_t^*$				
10		0.000	0.000	0.000	0.000	0.000	0.020	0.020	0.043	0.097	0.177
30		0.014	0.006	0.002	0.001	0.000	0.080	0.064	0.061	0.059	0.066
50		0.027	0.018	0.010	0.006	0.002	0.105	0.079	0.073	0.062	0.064
100		0.042	0.036	0.028	0.020	0.022	0.124	0.100	0.084	0.072	0.066
250		0.054	0.045	0.053	0.046	0.036	0.144	0.110	0.081	0.072	0.072
500		0.063	0.054	0.054	0.056	0.056	0.154	0.105	0.084	0.072	0.067
		$(v) DF_\rho^*$					$(vi) \Upsilon_{LR}$				
10		0.093	0.108	0.195	0.315	0.433	0.260	0.362	0.614	0.828	0.932
30		0.074	0.073	0.084	0.091	0.105	0.063	0.058	0.047	0.045	0.036
50		0.081	0.066	0.072	0.078	0.079	0.058	0.051	0.046	0.043	0.044
100		0.084	0.078	0.072	0.070	0.067	0.065	0.058	0.058	0.053	0.039
250		0.087	0.075	0.074	0.059	0.060	0.063	0.055	0.057	0.062	0.063
500		0.087	0.071	0.064	0.057	0.062	0.062	0.059	0.063	0.058	0.060

Note: $\rho = 1$, $\psi = 0$, $\sigma = 1$ and $a_1 = 0$. $M = 5,000$ replications.

5% nominal level.

is therefore justified, as it yields a power gain over univariate approaches. The Johansen-based tests appear to require somewhat larger T to achieve high power under this DGP.⁹ Second, the $P_{\Phi-1}$ and the P_t tests outperform the P_{χ^2} test at least for the ADF variant. This finding is in line with the results reported by Choi [2001] for his panel unit root tests. Whether to choose the $P_{\Phi-1}$ or the P_t in any application would be a matter of taste. Third, in each of the cases, power tends to grow faster along the time series dimension. More specifically, the power of the tests rises quickly between $T = 50$ and $T = 100$. The simulation evidence therefore suggests that the P tests are particularly useful in medium to relatively long panels. Figure I plots the power of the Engle/Granger-based tests for $N = 100$ as the fraction of cointegrated variables in the system, δ , increases. Panels (a) and (b) depict the cases $T = 50$ and $T = 100$, respectively. It can be seen that the power of the P tests rises to one substantially quicker when the underlying time series are longer.

⁹See below for their performance under other DGPs.

TABLE III—SIZE-ADJUSTED POWER OF THE P TESTS

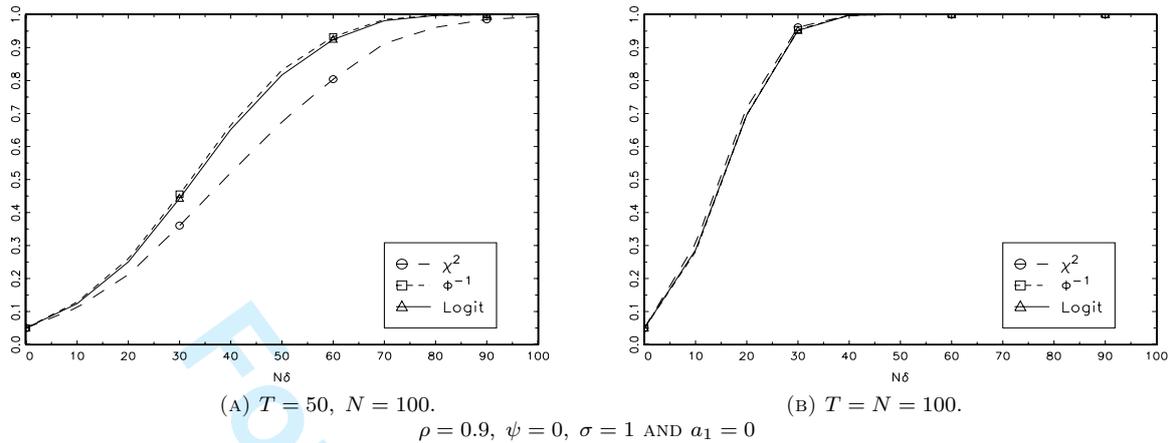
T	N	ADF					λ_{trace}				
		10	20	50	100	150	10	20	50	100	150
<i>(i) P_{χ^2}</i>											
10		0.061	0.065	0.069	0.081	0.083	0.049	0.043	0.044	0.042	0.034
30		0.085	0.124	0.185	0.276	0.372	0.037	0.036	0.031	0.019	0.022
50		0.138	0.236	0.454	0.717	0.843	0.052	0.038	0.051	0.039	0.040
100		0.502	0.765	0.990	1.000	1.000	0.091	0.098	0.219	0.234	0.348
250		1.000	1.000	1.000	1.000	1.000	0.547	0.748	0.994	1.000	1.000
500		1.000	1.000	1.000	1.000	1.000	0.989	1.000	1.000	1.000	1.000
<i>(ii) $P_{\Phi^{-1}}$</i>											
10		0.059	0.068	0.085	0.102	0.111	0.048	0.041	0.043	0.037	0.036
30		0.095	0.151	0.250	0.380	0.552	0.035	0.037	0.031	0.022	0.019
50		0.191	0.284	0.575	0.851	0.948	0.050	0.042	0.052	0.047	0.042
100		0.508	0.805	0.992	1.000	1.000	0.100	0.104	0.234	0.295	0.383
250		0.995	1.000	1.000	1.000	1.000	0.440	0.591	0.982	0.999	1.000
500		1.000	1.000	1.000	1.000	1.000	0.898	0.986	1.000	1.000	1.000
<i>(iii) P_t</i>											
10		0.060	0.070	0.086	0.103	0.114	0.048	0.044	0.043	0.038	0.037
30		0.092	0.148	0.244	0.382	0.555	0.035	0.035	0.033	0.019	0.020
50		0.185	0.286	0.561	0.835	0.939	0.052	0.042	0.051	0.045	0.043
100		0.500	0.789	0.991	1.000	1.000	0.096	0.101	0.236	0.290	0.378
250		0.997	1.000	1.000	1.000	1.000	0.476	0.648	0.987	0.999	1.000
500		1.000	1.000	1.000	1.000	1.000	0.950	0.997	1.000	1.000	1.000

Note: $\rho = 0.9$, $\psi = 0$, $\sigma = 1$, $\delta = 0.5$ and $a_1 = 0$. $M = 5,000$ replications. 5% nominal level. ADF and λ_{trace} are the underlying time series tests.

Now, let us compare the power results of Table III and Figure I with those obtained for the other panel cointegration tests considered here, presented in Table IV. The ADF -based P tests are somewhat less powerful than the residual-based panel tests $\tilde{Z}_{\hat{\rho}_{NT-1}}$, $Z_{\hat{t}_{NT}}$ and $Z_{\hat{\rho}_{NT-1}}$ for shorter panels. Overall, the $\tilde{Z}_{\hat{\rho}_{NT-1}}$ test performs best.¹⁰ However, the P tests' power for longer panels is similar or even better than that of Pedroni's tests. The ADF -based P tests' power exceeds that of DF_{ρ}^* and DF_t^* . Finally, the P tests always outperform the system-based Υ_{LR} test by Larsson et al. [2001].

We think that the power experiments for DGP A are somewhat restrictive. Apart from

¹⁰It is however not clear whether this attractive performance would be available in practice, as these numbers are based on size-adjusted critical values and as the $\tilde{Z}_{\hat{\rho}_{NT-1}}$ test appears to be rather undersized at least for small T (cf. Table II).

FIGURE I—POWER OF THE P PANEL COINTEGRATION TESTS

the unit specific intercepts, no heterogeneity is allowed for. But, in many practical applications, the units of a panel, say, countries, differ in their short-run dynamic adjustment behavior. We therefore elicit how the performance of the tests changes when we introduce heterogeneity in the serial correlation properties. Since, to the best of our knowledge, no power comparison of the different panel cointegration tests under these circumstances is available in the literature, we also include the tests presented above.

Consider the following modification of DGP A to introduce higher order serial correlation in the equilibrium error z_t .¹¹ Following Said and Dickey [1984], define

DGP B

$$\begin{aligned}
 x_{i1,t} - \alpha_i - \beta x_{i2,t} &= z_{i,t}, & a_1 x_{i1,t} - a_2 x_{i2,t} &= w_{i,t}, \\
 z_{i,t} &= \rho_i z_{i,t-1} + e_{zi,t}, & e_{zi,t} &= \sum_{p_i=1}^{\tilde{\zeta}_i} \varphi_{i,p_i} e_{i,t-p_i}, & \Delta w_{i,t} &= e_{wi,t}, \\
 \left(\begin{array}{c} e_{zi,t} \\ e_{wi,t} \end{array} \right) &\stackrel{iid}{\sim} \mathcal{N} \left(\left[\begin{array}{c} 0 \\ 0 \end{array} \right], \left[\begin{array}{cc} 1 & \psi\sigma \\ \psi\sigma & \sigma^2 \end{array} \right] \right)
 \end{aligned}$$

We draw, for each series in the panel, the order of the AR -process according to $\tilde{\zeta}_i = [\zeta_i]$, where $\zeta_i \sim \mathcal{U}[1, 2]$, $i = 1, \dots, N$ and $[y]$ rounds to the nearest integer. We then generate the AR -coefficients from $\varphi_{i,p} \sim \mathcal{U}[0.1, 0.35]$, $p_i = 1, \dots, \tilde{\zeta}_i$. In the size study, $\rho_i = 1$ for $i = 1, \dots, N$. To study power, we let $\rho_i \sim \mathcal{U}[0.9, 1]$ for $i = 1, \dots, \delta N$, while $\rho_i = 1$ for

¹¹I am grateful to an anonymous referee for suggesting this extension.

TABLE IV—SIZE-ADJUSTED POWER OF OTHER COINTEGRATION TESTS

T	N	10	20	50	100	150	10	20	50	100	150
		$(i) Z_{\hat{\rho}_{NT-1}}$					$(ii) Z_{\hat{t}_{NT}}$				
10		0.061	0.072	0.094	0.126	0.165	0.063	0.070	0.092	0.106	0.126
30		0.121	0.189	0.356	0.547	0.731	0.098	0.156	0.289	0.440	0.619
50		0.232	0.331	0.573	0.844	0.935	0.186	0.264	0.487	0.736	0.863
100		0.399	0.616	0.873	0.986	0.999	0.347	0.509	0.799	0.964	0.995
250		0.617	0.785	0.967	0.997	1.000	0.558	0.728	0.939	0.994	1.000
500		0.716	0.851	0.989	1.000	1.000	0.669	0.816	0.976	0.999	1.000
		$(iii) \tilde{Z}_{\hat{\rho}_{NT-1}}$					$(iv) DF_t^*$				
10		0.061	0.072	0.086	0.102	0.124	0.051	0.042	0.029	0.031	0.025
30		0.111	0.171	0.291	0.461	0.645	0.120	0.195	0.361	0.575	0.726
50		0.215	0.365	0.664	0.919	0.979	0.236	0.365	0.631	0.870	0.957
100		0.630	0.892	1.000	1.000	1.000	0.392	0.594	0.862	0.980	0.996
250		1.000	1.000	1.000	1.000	1.000	0.554	0.717	0.939	0.996	1.000
500		1.000	1.000	1.000	1.000	1.000	0.610	0.768	0.965	0.998	1.000
		$(v) DF_\rho^*$					$(vi) \Upsilon_{LR}$				
10		0.064	0.055	0.047	0.063	0.058	0.048	0.043	0.042	0.042	0.035
30		0.094	0.149	0.240	0.371	0.498	0.037	0.035	0.031	0.019	0.020
50		0.177	0.263	0.428	0.685	0.815	0.052	0.037	0.049	0.042	0.040
100		0.306	0.469	0.705	0.905	0.972	0.095	0.101	0.229	0.266	0.375
250		0.456	0.596	0.843	0.975	0.997	0.526	0.717	0.993	1.000	1.000
500		0.531	0.656	0.897	0.982	0.999	0.978	1.000	1.000	1.000	1.000

Note: $\rho = 0.9$, $\psi = 0$, $\sigma = 1$ and $a_1 = 0$. $M = 5,000$ replications.
5% nominal level.

$$i = \delta N + 1, \dots, N.^{12}$$

Tables V to VIII report results on size and size-adjusted power of the tests for $\sigma = 1$, $\psi = 0$, $a_1 = 0$ and $\delta = 0.5$. The number of lagged differences for the *ADF* regression is chosen according to the automatic procedure suggested by Ng and Perron [2001]. As regards size, Tables V shows that the *P* tests to maintain their good performance, with the exception of very small T . The Kao and Υ_{LR} tests are slightly, whereas Pedroni's are quite severely oversized (Table VI).

It is not possible to compare the size-adjusted power with the results from Table III because the alternative is now different. But, Tables VII and VIII show that the *ADF*-based *P* tests as well as Pedroni's tests generally outperform Kao's and the Υ_{LR} test. This is intuitive as the *P* and Pedroni [2004] tests are designed to accommodate cross-sectional

¹²We also parameterized DGP B with other combinations of distributions for the $\tilde{\zeta}_i$ and φ_{i,p_i} . The additional results, which were qualitatively very similar, are available upon request.

TABLE V—SIZE OF THE P TESTS WITH AR(P) ERRORS

T	N	ADF					λ_{trace}				
		10	20	50	100	150	10	20	50	100	150
<i>(i) P_{χ^2}</i>											
10		0.104	0.113	0.086	0.069	0.053	0.559	0.781	0.977	0.999	1.000
30		0.050	0.035	0.029	0.019	0.010	0.042	0.043	0.036	0.034	0.037
50		0.045	0.042	0.032	0.025	0.016	0.048	0.047	0.042	0.043	0.045
100		0.042	0.042	0.033	0.027	0.024	0.054	0.054	0.051	0.055	0.058
250		0.042	0.048	0.034	0.037	0.034	0.065	0.068	0.073	0.087	0.097
500		0.046	0.043	0.038	0.040	0.034	0.066	0.081	0.090	0.107	0.122
<i>(ii) $P_{\Phi^{-1}}$</i>											
10		0.018	0.006	0.000	0.000	0.000	0.481	0.696	0.943	0.997	1.000
30		0.028	0.017	0.007	0.003	0.001	0.043	0.039	0.031	0.025	0.026
50		0.030	0.029	0.015	0.009	0.002	0.043	0.038	0.032	0.036	0.034
100		0.036	0.032	0.025	0.017	0.013	0.051	0.048	0.043	0.044	0.049
250		0.041	0.044	0.028	0.028	0.025	0.064	0.067	0.069	0.077	0.087
500		0.045	0.043	0.038	0.034	0.030	0.056	0.077	0.081	0.090	0.101
<i>(iii) P_t</i>											
10		0.043	0.021	0.003	0.000	0.000	0.509	0.720	0.954	0.998	1.000
30		0.033	0.021	0.009	0.004	0.002	0.043	0.039	0.031	0.026	0.027
50		0.031	0.032	0.018	0.011	0.003	0.046	0.039	0.034	0.035	0.034
100		0.036	0.034	0.024	0.018	0.014	0.052	0.050	0.045	0.046	0.049
250		0.042	0.043	0.029	0.028	0.028	0.064	0.069	0.073	0.080	0.088
500		0.047	0.045	0.039	0.035	0.031	0.061	0.078	0.085	0.094	0.105

Note: $\rho_i = 1$, $\psi = 0$, $\sigma = 1$ and $a_1 = 0$. $M = 5,000$ replications.

5% nominal level. $\zeta_i \sim \mathcal{U}[1, 2]$, $i = 1, \dots, \delta N$ and $\varphi_{i,p} \sim \mathcal{U}[0.1, 0.35]$

heterogeneity. In the case of Pedroni's tests it is however not clear to what extent this result is practically useful in view of the large size distortions found above.

DGPs A and B closely mirror the framework of residual-based cointegration tests. This potentially affords the residual-based tests (i.e. the ADF -based P tests as well as Kao's [1999] and Pedroni's [2004]) a relative advantage against the system-based tests (i.e. the λ_{trace} -based P tests and Larsson et al.'s [2001]).¹³ To verify this claim, we consider the following Vector Error Correction-type DGP. This DGP is more closely related to the latter system-based tests, as, for instance, it is not built around an equilibrium error $z_{i,t}$ whose stationarity properties the residual-based tests attempt to detect. Also, it does not require a left-hand side variable.

¹³Indeed, Tables III-VIII report that the λ_{trace} -based P tests and the Υ_{LR} test are generally less powerful than the residual-based tests under DGPs A and B.

TABLE VI—SIZE OF OTHER COINTEGRATION TESTS, AR(P) ERRORS

T	N	10	20	50	100	150	10	20	50	100	150
		$(i) Z_{\hat{\rho}_{NT-1}}$					$(ii) Z_{\hat{t}_{NT}}$				
10		0.070	0.143	0.364	0.701	0.876	0.692	0.906	0.999	1.000	1.000
30		0.343	0.502	0.811	0.978	0.999	0.463	0.668	0.940	0.998	1.000
50		0.347	0.519	0.802	0.966	0.995	0.393	0.609	0.900	0.994	1.000
100		0.392	0.538	0.809	0.963	0.995	0.411	0.609	0.892	0.991	0.999
250		0.380	0.506	0.762	0.938	0.984	0.458	0.650	0.906	0.992	1.000
500		0.400	0.521	0.767	0.942	0.987	0.515	0.704	0.939	0.997	1.000
		$(iii) \tilde{Z}_{\hat{\rho}_{NT-1}}$					$(iv) DF_t^*$				
10		0.000	0.000	0.000	0.000	0.000	0.010	0.008	0.015	0.026	0.039
30		0.128	0.194	0.329	0.548	0.708	0.094	0.077	0.083	0.110	0.137
50		0.256	0.403	0.686	0.912	0.976	0.127	0.130	0.121	0.146	0.167
100		0.384	0.583	0.869	0.989	0.999	0.171	0.153	0.143	0.170	0.182
250		0.471	0.689	0.939	0.995	1.000	0.179	0.163	0.163	0.170	0.190
500		0.514	0.716	0.955	0.999	1.000	0.206	0.166	0.158	0.180	0.200
		$(v) DF_\rho^*$					$(vi) \Upsilon_{LR}$				
10		0.100	0.124	0.229	0.401	0.540	0.580	0.792	0.976	0.999	1.000
30		0.111	0.106	0.135	0.172	0.227	0.053	0.048	0.043	0.038	0.040
50		0.108	0.119	0.123	0.163	0.192	0.057	0.054	0.045	0.051	0.048
100		0.119	0.117	0.123	0.142	0.166	0.065	0.063	0.059	0.066	0.069
250		0.122	0.118	0.118	0.131	0.149	0.080	0.082	0.086	0.100	0.117
500		0.129	0.117	0.117	0.140	0.153	0.076	0.096	0.102	0.125	0.144

Note: $\rho_i = 1$, $\psi = 0$, $\sigma = 1$ and $a_1 = 0$. $M = 5,000$ replications.
 5% nominal level. $\zeta_i \sim \mathcal{U}[1, 2]$, $i = 1, \dots, \delta N$ and $\varphi_{i,p} \sim \mathcal{U}[0.1, 0.35]$

DGP C Let $X_{1,t} = (x_{1,1t}, \dots, x_{N,1t})^\top$ and $X_{2,t} = (x_{1,2t}, \dots, x_{N,2t})^\top$. $X_{1,t}$ and $X_{2,t}$ are generated as

$$\begin{pmatrix} \Delta X_{1,t} \\ \Delta X_{2,t} \end{pmatrix} = \alpha \beta^\top \begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix} + \Gamma \begin{pmatrix} \Delta X_{1,t-1} \\ \Delta X_{2,t-1} \end{pmatrix} + \varepsilon_t,$$

where

$$\alpha = -\tilde{\alpha} \cdot \begin{pmatrix} \mathbf{I}_N & \mathbf{0}_N \\ \mathbf{0}_N & \mathbf{I}_N \end{pmatrix} \quad \text{and} \quad \beta^\top = \begin{pmatrix} \mathbf{I}_N & -\mathbf{I}_N \\ \mathbf{0}_N & \mathbf{0}_N \end{pmatrix}.$$

with \mathbf{I}_A ($\mathbf{0}_A$) the A -dimensional identity (zero) matrix, $\Gamma = 0.15\mathbf{I}_{2N}$ and $\varepsilon_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{2N})$.

For $\tilde{\alpha} = 0$, $x_{i,1t}$ and $x_{i,2t}$, $i = 1, \dots, N$, are independent nonstationary processes. We then study size of the tests. For $\tilde{\alpha} > 0$, this formulation ensures cointegrating relationships $x_{i,1t} - x_{i,2t}$, such that we analyze power of the tests. By increasing $\tilde{\alpha}$, the degree of reversion to the cointegration relationship increases, such that the tests should become more powerful. For the degree of mean reversion, we choose $\tilde{\alpha} \in \{0, 0.05, 0.1, 0.15, 0.2\}$.

TABLE VII—SIZE-ADJUSTED POWER OF THE P TESTS WITH AR(P) ERRORS

T	N	ADF					λ_{trace}				
		10	20	50	100	150	10	20	50	100	150
<i>(i) P_{χ^2}</i>											
10		0.108	0.120	0.096	0.079	0.064	0.546	0.788	0.982	1.000	1.000
30		0.058	0.053	0.048	0.033	0.034	0.041	0.033	0.038	0.031	0.018
50		0.077	0.081	0.086	0.116	0.124	0.044	0.040	0.058	0.042	0.020
100		0.149	0.215	0.378	0.600	0.735	0.079	0.069	0.170	0.168	0.106
250		0.663	0.898	0.998	1.000	1.000	0.300	0.344	0.861	0.942	0.905
500		0.975	1.000	1.000	1.000	1.000	0.707	0.858	1.000	1.000	1.000
<i>(ii) $P_{\Phi^{-1}}$</i>											
10		0.017	0.005	0.001	0.000	0.000	0.468	0.699	0.955	0.998	1.000
30		0.039	0.026	0.016	0.008	0.006	0.036	0.028	0.033	0.020	0.009
50		0.069	0.063	0.069	0.085	0.093	0.041	0.036	0.041	0.029	0.012
100		0.155	0.218	0.406	0.640	0.780	0.066	0.056	0.140	0.129	0.076
250		0.589	0.833	0.995	1.000	1.000	0.211	0.260	0.721	0.863	0.781
500		0.927	0.997	1.000	1.000	1.000	0.534	0.717	0.996	1.000	1.000
<i>(iii) P_t</i>											
10		0.045	0.022	0.004	0.000	0.000	0.502	0.728	0.962	0.998	1.000
30		0.045	0.034	0.023	0.011	0.008	0.036	0.028	0.032	0.022	0.009
50		0.073	0.071	0.077	0.091	0.103	0.040	0.036	0.044	0.029	0.013
100		0.157	0.222	0.406	0.635	0.776	0.068	0.059	0.147	0.139	0.079
250		0.610	0.847	0.995	1.000	1.000	0.246	0.287	0.769	0.890	0.825
500		0.951	0.999	1.000	1.000	1.000	0.619	0.781	0.999	1.000	1.000

Note: Half of the series has $\rho_i \sim \mathcal{U}[0.9, 1]$, $\rho_i = 1$ else. $\psi = 0$, $\sigma = 1$ and $a_1 = 0$.

$M = 5,000$. 5% nominal level. $\zeta_i \sim \mathcal{U}[1, 2]$, $i = 1, \dots, \delta N$ and $\varphi_{i,p} \sim \mathcal{U}[0.1, 0.35]$

Results are summarized in Table IX.¹⁴ As regards size, a pattern comparable to the results of DGP A emerges. For example, the system-based $P_{\Phi^{-1},\lambda}$ and Υ_{LR} tests are oversized for very small T . We also again observe a mild “adding up” of size distortions for the $P_{\Phi^{-1},DF}$ and the DF_t^* test. A notable exception are the tests by Pedroni, that seem to be quite severely undersized under DGP C. Concerning power, we find that the above conjecture that this error-correction DGP is more favorable to the system-based tests is corroborated. In particular, unlike in Tables III and IV, the system-based $P_{\Phi^{-1},\lambda}$ and Υ_{LR} tests are, together with the DF_t^* test, the most powerful ones under DGP C.

To summarize, the P tests discussed in this paper enjoy a rather consistently attractive

¹⁴Results for the other tests considered above are available upon request. Broadly speaking, the other P tests behave very much like the $P_{\Phi^{-1}}$ test when using the same time series test, with the P_{χ^2} again being slightly less powerful. Furthermore, the other tests by Kao and Pedroni are somewhat less powerful than the ones reported here.

TABLE VIII—SIZE-ADJUSTED POWER OF OTHER COINTEGRATION TESTS, AR(P) ERRORS

T	N	10	20	50	100	150	10	20	50	100	150
		$(i) Z_{\hat{\rho}_{NT-1}}$					$(ii) Z_{\hat{t}_{NT}}$				
10		0.089	0.156	0.438	0.785	0.935	0.732	0.925	1.000	1.000	1.000
30		0.461	0.669	0.944	0.998	1.000	0.560	0.797	0.986	1.000	1.000
50		0.556	0.752	0.969	0.999	1.000	0.574	0.802	0.985	0.999	1.000
100		0.692	0.863	0.992	1.000	1.000	0.667	0.868	0.995	1.000	1.000
250		0.799	0.924	0.999	1.000	1.000	0.804	0.937	0.999	1.000	1.000
500		0.872	0.960	1.000	1.000	1.000	0.873	0.967	1.000	1.000	1.000
		$(iii) \tilde{Z}_{\hat{\rho}_{NT-1}}$					$(iv) DF_t^*$				
10		0.000	0.000	0.000	0.000	0.000	0.009	0.006	0.008	0.015	0.027
30		0.208	0.314	0.607	0.875	0.968	0.140	0.145	0.202	0.316	0.434
50		0.481	0.718	0.969	0.999	1.000	0.255	0.276	0.395	0.559	0.701
100		0.838	0.975	1.000	1.000	1.000	0.385	0.448	0.620	0.812	0.915
250		0.996	1.000	1.000	1.000	1.000	0.530	0.601	0.802	0.939	0.983
500		1.000	1.000	1.000	1.000	1.000	0.596	0.688	0.874	0.970	0.995
		$(v) DF_\rho^*$					$(vi) \Upsilon_{LR}$				
10		0.096	0.134	0.230	0.417	0.582	0.569	0.794	0.982	1.000	1.000
30		0.149	0.172	0.248	0.364	0.484	0.048	0.038	0.041	0.034	0.018
50		0.200	0.232	0.340	0.501	0.619	0.052	0.049	0.064	0.047	0.022
100		0.274	0.341	0.499	0.688	0.811	0.087	0.075	0.186	0.182	0.115
250		0.393	0.467	0.671	0.851	0.938	0.310	0.355	0.853	0.941	0.899
500		0.465	0.550	0.762	0.912	0.973	0.701	0.854	1.000	1.000	1.000

Note: Half of the series has $\rho_i \sim \mathcal{U}[0.9, 1]$, $\rho_i = 1$ else. $\psi = 0$, $\sigma = 1$ and $a_1 = 0$.
 $M = 5,000$. 5% nominal level. $\zeta_i \sim \mathcal{U}[1, 2]$, $i = 1, \dots, \delta N$ and $\varphi_{i,p} \sim \mathcal{U}[0.1, 0.35]$

performance in terms finite sample size and power across a range of realistic and challenging scenarios. The P tests may therefore be useful in a fairly wide range of practical applications.

4 Conclusion

We study new tests for panel cointegration, labeled P tests. As in Maddala and Wu [1999] and Choi [2001], we use a meta analytic p -value combination approach to develop tests for nonstationary panel data. The new tests are flexible, intuitively appealing and easy to implement. The tests employ highly accurate p -values obtained from response surface regressions [MacKinnon, 1996; MacKinnon et al., 1999]. A finite sample study reveals that the Engle and Granger [1987]-based variant of the controls size even for rather short panels. The empirical size of the tests is very close to the nominal one for

TABLE IX—REJECTION RATES FOR DGP C

$\tilde{\alpha}$	T	N	$P_{\Phi^{-1},DF}$			$P_{\Phi^{-1},\lambda}$			$\tilde{Z}_{\rho_{NT}-1}$			DF_t^*			τ_{LR}						
			10	20	50	100	10	20	50	100	10	20	50	100	10	20	50	100			
<i>(i) Size</i>																					
10	.037	.029	.017	.010	.303	.431	.740	.928	.007	.004	.001	.001	.001	.035	.043	.094	.224	.357	.505	.820	.964
30	.039	.033	.029	.019	.059	.063	.062	.069	.010	.005	.001	.000	.000	.072	.055	.041	.038	.071	.079	.082	.090
50	.037	.036	.027	.024	.053	.052	.063	.056	.006	.003	.000	.000	.000	.082	.062	.044	.032	.067	.066	.074	.076
100	.047	.045	.037	.035	.052	.050	.052	.053	.008	.002	.000	.000	.000	.108	.066	.041	.031	.066	.062	.065	.070
250	.044	.043	.042	.044	.054	.049	.048	.049	.004	.001	.000	.000	.000	.103	.074	.046	.031	.066	.059	.062	.062
<i>(ii) Power</i>																					
10	.046	.047	.055	.056	.049	.063	.056	.078	.044	.042	.053	.039	.039	.048	.057	.060	.062	.044	.064	.056	.078
30	.055	.061	.070	.077	.133	.205	.385	.603	.053	.052	.060	.079	.079	.125	.185	.387	.611	.142	.196	.355	.562
50	.084	.087	.141	.202	.369	.604	.904	.994	.089	.096	.183	.302	.302	.285	.537	.908	.997	.347	.582	.880	.990
100	.245	.410	.782	.970	.955	.999	1.00	1.00	.346	.615	.956	1.00	1.00	.907	.999	1.00	1.00	.932	.998	1.00	1.00
250	.999	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
10	.052	.054	.059	.065	.062	.079	.089	.117	.038	.040	.042	.031	.031	.056	.074	.080	.102	.061	.076	.081	.116
30	.081	.115	.203	.326	.393	.611	.934	.999	.091	.129	.240	.445	.445	.422	.705	.987	1.00	.392	.578	.905	.997
50	.247	.386	.769	.968	.893	.996	1.00	1.00	.337	.572	.948	.999	.999	.907	.999	1.00	1.00	.864	.991	1.00	1.00
100	.938	.998	1.00	1.00	1.00	1.00	1.00	1.00	.995	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
250	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
10	.056	.049	.072	.080	.063	.086	.112	.181	.042	.042	.039	.039	.039	.062	.083	.113	.174	.060	.900	.103	.166
30	.165	.292	.587	.867	.695	.920	1.00	1.00	.215	.408	.786	.978	.978	.780	.985	1.00	1.00	.674	1.00	.999	1.00
50	.631	.893	1.00	1.00	.997	1.00	1.00	1.00	.838	.988	1.00	1.00	1.00	.999	1.00	1.00	1.00	.994	1.00	1.00	1.00
100	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
250	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
10	.057	.074	.094	.125	.078	.103	.151	.234	.041	.047	.056	.054	.054	.066	.105	.157	.233	.073	.100	.135	.207
30	.346	.588	.933	.999	.892	.993	1.00	1.00	.504	.814	.995	1.00	1.00	.972	1.00	1.00	1.00	.873	.987	1.00	1.00
50	.929	.997	1.00	1.00	1.00	1.00	1.00	1.00	.995	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
100	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
250	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Note: $M = 5,000$ replications. 5% nominal level. Power results ($\tilde{\alpha} > 0$) are size adjusted. ADF is the underlying time series test for $P_{\Phi^{-1},DF}$, and λ_{trace} for $P_{\Phi^{-1},\lambda}$.

1
2
3
4
5
6 panel dimensions often encountered in applied macroeconometric work. The λ_{trace} -based
7 P tests exhibit an upward size distortion for small T , which however vanishes quickly. In
8 that regard, we find the P tests to compare favorably with existing tests. In terms of
9 power, their performance is often as good as and, in certain cases, better than the more
10 powerful among other widely used panel cointegration tests.
11
12

13
14
15 As most tests in this literature, the ones suggested here rely on the assumption of cross-
16 sectional uncorrelatedness (see Assumption 2). This assumption is likely to be overly
17 strong for many macroeconomic panels and may lead, if violated, to erroneous conclusions
18 [cf. O'Connell, 1998]. We therefore suggest to extend the tests developed here to allow for
19 cross-sectional correlation by, e.g., the bootstrap method. Maddala and Wu [1999] report
20 encouraging results along these lines for their panel unit root test. There is a growing
21 literature on bootstrapping cointegrating regressions [see Li and Maddala, 1997] that can
22 be fruitfully applied to the present problem. Recent useful contributions include Chang
23 and Park [2003] and Chang et al. [2006]. Investigation of this extension is currently under
24 way by the author.
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

References

- Baltagi, B. H. and C. Kao: 2000, 'Nonstationary Panels, Cointegration in Panels and Dynamic Panels: A Survey'. In: B. H. Baltagi (ed.): *Advances in Econometrics*, Vol. 15. Amsterdam: Elsevier, Chapt. 2, pp. 7–51.
- Banerjee, A.: 1999, 'Panel Data Unit Roots and Cointegration: An Overview'. *Oxford Bulletin of Economics and Statistics* **61**(S1), 607–629.
- Banerjee, A., M. Marcellino, and C. Osbat: 2005, 'Testing for PPP: Should We Use Panel Methods?'. *Empirical Economics* **30**(1), 77–91.
- Bickel, P. J. and K. A. Doksum: 2001, *Mathematical Statistics*, Vol. 1. Upper Saddle River, New Jersey: Prentice Hall, 2nd edition.
- Breitung, J. and M. H. 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*. Dordrecht: Kluwer Academic Publishers, 3rd edition, Chapt. 9, pp. 279–322.
- Chang, Y. and J. Y. Park: 2003, 'A Sieve Bootstrap for the Test of a Unit Root'. *Journal of Time Series Analysis* **24**(4), 379–400.
- Chang, Y., J. Y. Park, and K. Song: 2006, 'Bootstrapping Cointegrating Regressions'. *Journal of Econometrics* **133**, 703–739.
- Cheung, Y.-W. and K. S. Lai: 1993, 'Finite Sample Sizes of Johansen's Likelihood Ratio Tests for Cointegration'. *Oxford Bulletin of Economics and Statistics* **55**(3), 313–328.
- Choi, I.: 2001, 'Unit Root Tests for Panel Data'. *Journal of International Money and Finance* **20**(2), 249–272.
- Engle, R. and C. Granger: 1987, 'Co-Integration and Error Correction: Representation, Estimation, and Testing'. *Econometrica* **55**(2), 251–76.
- Fisher, R.: 1932, *Statistical Methods for Research Workers*. London: Oliver and Boyd.
- Gonzalo, J.: 1994, 'Five Alternative Methods of Estimating Long-Run Equilibrium Relationships'. *Journal of Econometrics* **60**(1–2), 203–233.
- Gutierrez, L.: 2003, 'On the Power of Panel Cointegration Tests: A Monte Carlo Comparison'. *Economics Letters* **80**(1), 105–111.
- Hanck, C.: 2008, 'The Error-in-Rejection Probability of Meta Analytic Panel Tests'. *Economics Letters* **100**(3), to appear.
- Hedges, L. V. and I. Olkin: 1985, *Statistical Methods for Meta-Analysis*. San Diego: Academic Press.
- Hsiao, C.: 2003, *Analysis of Panel Data*. New York: Cambridge University Press, second edition.
- Johansen, S.: 1988, 'Statistical Analysis of Cointegration Vectors'. *Journal of Economic Dynamics and Control* **12**(2–3), 231–254.
- Kao, C.: 1999, 'Spurious Regression and Residual Based Tests for Cointegration in Panel Data'. *Journal of Econometrics* **90**(1), 1–44.
- Larsson, R., J. Lyhagen, and M. Löthgren: 2001, 'Likelihood-Based Cointegration Tests in Heterogeneous Panels'. *The Econometrics Journal* **4**(1), 109–142.
- Li, H. and G. Maddala: 1997, 'Bootstrapping Cointegrating Regressions'. *Journal of Econometrics* **80**(2), 297–318.
- MacKinnon, J. G.: 1991, 'Critical Values for Cointegration Tests'. In: R. F. Engle and C. W. Granger (eds.): *Long-run Economic Relationships: Readings in Cointegration*. Oxford: Oxford University Press, Chapt. 13, pp. 267–276.
- MacKinnon, J. G.: 1996, 'Numerical Distribution Functions for Unit Root and Cointegration Tests'. *Journal of Applied Econometrics* **11**(6), 601–618.
- MacKinnon, J. G., A. A. Haug, and L. Michelis: 1999, 'Numerical Distribution Functions of Likelihood Ratio Tests for Cointegration'. *Journal of Applied Econometrics* **14**(5), 563–577.
- Maddala, G. and S. Wu: 1999, 'A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test'. *Oxford Bulletin of Economics and Statistics* **61**(S1), 631–652.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- McCoskey, S. and C. Kao: 1998, 'A Residual Based Test of the Null of Cointegration in Panel Data'. *Econometric Reviews* **17**(1), 57–84.
- Ng, S. and P. Perron: 2001, 'Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power'. *Econometrica* **69**(6), 1519–1554.
- O'Connell, P. G.: 1998, 'The Overvaluation of Purchasing Power Parity'. *Journal of International Economics* **44**(1), 1–19.
- Pedroni, P.: 2004, 'Panel Cointegration: Asymptotic and Finite Sample Properties of Pooled Time Series Tests with an Application to the PPP Hypothesis'. *Econometric Theory* **20**(3), 597–625.
- Phillips, P. C. B. and S. Ouliaris: 1990, 'Asymptotic Properties of Residual Based Tests for Cointegration'. *Econometrica* **58**(1), 165–193.
- Said, S. E. and D. A. Dickey: 1984, 'Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order'. *Biometrika* **71**(3), 599–607.
- Shin, Y.: 1994, 'A Residual-Based Test of the Null of Cointegration against the Alternative of No Cointegration'. *Econometric Theory* **10**(1), 91–115.