A Meta Analytic Approach to Testing for Panel Cointegration

Christoph Hanck*

January 15, 2007

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

*Research supported by Ruhr Graduate School in Economics and DFG under Sonderforschungsbereich 475, Universität Dortmund. Contact Information: Tel. (+49) 0231-7553127, christoph.hanck@uni-dortmund.de. I am grateful to James MacKinnon for providing estimation routines and helpful discussion. All errors remain mine.

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 introduces 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 new 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 new tests. Section 3 presents the finite sample study. Section 4 concludes.

2 *P*-Value Combination Tests for Panel Cointegration

The present section develops the new 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 kth variable of unit i of the panel, where $i = 1, \dots, N$ and $k = 1, \dots, K$. The K variables may contain 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 ith 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 distribution function of θ_{i,T_i}. Since the tests considered 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}. We only consider time series tests with the null of no cointegration.

We are interested in testing the following null hypothesis

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

against the alternative

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

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 stationary 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, 1970; Hedges and Olkin, 1985]. Consider the testing problem on the panel as consisting of N testing problems 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*)
Under H_0, θ_{i,T_i} has a continuous distribution function for all $i = 1, \ldots, N$.

Assumption 2 (*Cross-Sectional Uncorrelatedness*)
$x_{ik,t} = x_{ik,t-1} + \xi_{ik,t},\ t = 1, \ldots, T_i,\ i = 1, \ldots, N,\ k = 1, \ldots, K$. Let $\boldsymbol{\xi}_{i,t} \equiv (\xi_{i1,t}, \ldots, \xi_{iK,t})'$. We require $E[\boldsymbol{\xi}_{i,t}\boldsymbol{\xi}'_{j,s}] = \mathbf{0}\ \forall\ s, t = 1, \ldots, T_i$ 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.

Remarks

[1] Both 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.

- Assumption 1 asymptotically ensures 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 tests considered in this paper.

- 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 new tests. Combine the N p-values of the individual time series cointegration tests, p_i, $i = 1, \ldots, N$, 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 $T_i \to \infty$ for all i, the test statistics are

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

asymptotically distributed as

$$\begin{aligned}
&P_{\chi^2} \to_d \chi^2_{2N} \\
&P_{\Phi^{-1}} \to_d \mathcal{N}(0,1) \\
&P_t \overset{\text{approx.}}{\to_d} \mathcal{T}_{5N+4},
\end{aligned}$$

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. Using consistent time series cointegration tests, $p_i \to_p 0$ 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 cointegration when P_{χ^2} exceeds the critical value from a χ^2_{2N} distribution at the desired significance level. For (2b) and (2c) one would reject for large negative values of the panel test statistics $P_{\Phi^{-1}}$ and P_t, respectively.

We now discuss how to obtain the p-values required for computation of the P test statistics. Hanck [2006] shows that using accurate p-values for meta analytic panel tests is crucial to achieve a precise control of the type I error rate. The null distributions of both residual and system-based time series cointegration tests converge to functionals of Brownian motion. Hence, analytic expressions of the distribution functions are hard to obtain, and p-values of the tests cannot simply be obtained by evaluating the corresponding cdf. A remedy frequently adopted in the literature is to derive the critical values (and, consequently, the p-values) by Monte Carlo simulation. However, this approach is unsatisfactory for (at least) the following reason. These simulations are typically only performed for one sample size which is meant to provide an approximation to the asymptotic distribution. This sample size need neither be large enough to be useful as an asymptotic approximation nor does it generally yield accurate critical values for other sample sizes. MacKinnon et al. [1999] show for cases where analytic expressions of the distribution functions are available that this approach may deliver fairly inaccurate critical values. In the time series case, it is now fairly standard practice to report p-values of unit root and cointegration tests using the results of the response surface regressions introduced by MacKinnon [1991]. We follow this approach here.

The null hypothesis (1) formulates no precise econometric characterization of (non-) cointegration. This is to allow for generality in testing the long-run equilibrium properties of the series, enabling the researcher to use whichever time series tests seem suitable to test for time series (non-)cointegration in the different units of the panel. We use p-values of the Augmented Dickey-Fuller (ADF) cointegration tests [Engle and Granger, 1987] as provided by MacKinnon [1996].[3] That is, the p-values are obtained from the t-statistic of $\gamma_i - 1$ 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 cointegrating relationships by estimating the number of significantly non-zero eigenvalues of the matrix $\hat{\Pi}_i$ estimated from the Vector Error Correction Model

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

by the λ_{trace}-test

$$\lambda_{trace,i}\left(h\right) = -T \sum_{k=h+1}^{K} \ln\left(1 - \hat{\pi}_{k,i}\right) \tag{3}$$

and the λ_{max}-test

$$\lambda_{max,i}\left(h|h+1\right) = -T \ln\left(1 - \hat{\pi}_{h+1,i}\right). \tag{4}$$

Here, $\hat{\pi}_{k,i}$ denotes the kth largest eigenvalue of $\hat{\Pi}_i$. In (3), the alternative is a general one, while one tests against $h+1$ cointegration relationships in (4).

Hence, 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

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

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

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

where

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

and

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

Remarks

- When $|\rho| < 1$ the equilibrium error in the first equation is stationary such that $x_{i1,t}$ and $x_{i2,t}$ are cointegrated with $\boldsymbol{\beta}_i = (1 \;\; \alpha_i \;\; \beta)'$.
- 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. $\delta \in \{0, 0.1, \ldots, 1\}$. 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 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 $r = 0$ vs. an unrestricted number of cointegrating relationships.

[4] Uniform 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.

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. Two conclusions are obvious. First, the Engle/Granger-

Table I—Empirical Size of the P Tests

		ADF					λ_{trace}				
T	N	10	20	50	100	150	10	20	50	100	150
(i) P_{χ^2}											
10		.038	.040	.024	.018	.011	.956	.999	1.00	1.00	1.00
30		.035	.031	.021	.014	.009	.184	.260	.467	.702	.845
50		.041	.033	.027	.022	.021	.102	.137	.216	.344	.437
100		.047	.042	.036	.034	.029	.074	.077	.109	.143	.171
250		.046	.044	.046	.045	.036	.052	.056	.063	.076	.086
500		.049	.048	.049	.048	.047	.056	.052	.054	.068	.068
(ii) $P_{\Phi^{-1}}$											
10		.027	.019	.006	.002	.001	.954	.998	1.00	1.00	1.00
30		.034	.022	.016	.009	.005	.180	.265	.468	.711	.848
50		.038	.030	.026	.018	.016	.102	.136	.218	.355	.447
100		.046	.038	.032	.033	.025	.072	.081	.111	.139	.177
250		.043	.045	.044	.041	.032	.051	.061	.061	.079	.086
500		.049	.047	.045	.044	.041	.056	.053	.055	.070	.065
(iii) P_t											
10		.030	.019	.006	.002	.001	.957	.998	1.00	1.00	1.00
30		.035	.023	.016	.010	.005	.183	.264	.473	.716	.852
50		.039	.030	.027	.018	.014	.102	.139	.217	.36	.447
100		.046	.038	.033	.032	.024	.074	.079	.109	.141	.176
250		.046	.045	.044	.041	.031	.053	.061	.063	.082	.086
500		.050	.048	.046	.045	.041	.056	.057	.059	.070	.067

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.

based tests are undersized. This issue is particularly severe in short panels but vanishes with increasing T. Oddly, all tests become *more* undersized as N increases. Hanck [2006] provides an analysis of this behavior. The P_{χ^2} test seems to have slightly better size than the other two. We 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

[5] The 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.

the sample sizes considered. Interestingly, however, there does not seem to be a trend towards lower size with increasing N. For medium- and large-dimensional panels neither approach has a clear advantage over the other.

Second, the Johansen-based tests are grossly oversized in panels of small and medium dimensions. Two reasons may be put forward for this disappointing performance. First, the underlying λ_{trace}-test overrejects in short time series when using asymptotic critical values [see also Cheung and Lai, 1993]. This flaw then inevitably carries over to the panel tests via erroneously small p-values. Second, MacKinnon et al. [1999] emphasize that the p-values estimated for the Johansen [1988] tests, unlike those estimated in MacKinnon [1996] for the Engle and Granger [1987] test, are only valid asymptotically. It may thus not be appropriate to use these for shorter time series. We therefore waive to report the essentially meaningless power for shorter panels.

Table II shows the raw power of the tests at $\rho = 0.9$.[6] The major findings are as follows. First, after having discarded the severely size-distorted panels, both the Engle/Granger- and Johansen-based tests behave consistently in that power for all variants grows with both dimensions. The use of panel data is therefore justified. 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 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.

We now relate our results to those of Gutierrez [2003]. 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 [forthcoming] provide surveys of the literature.

Pedroni [2004]

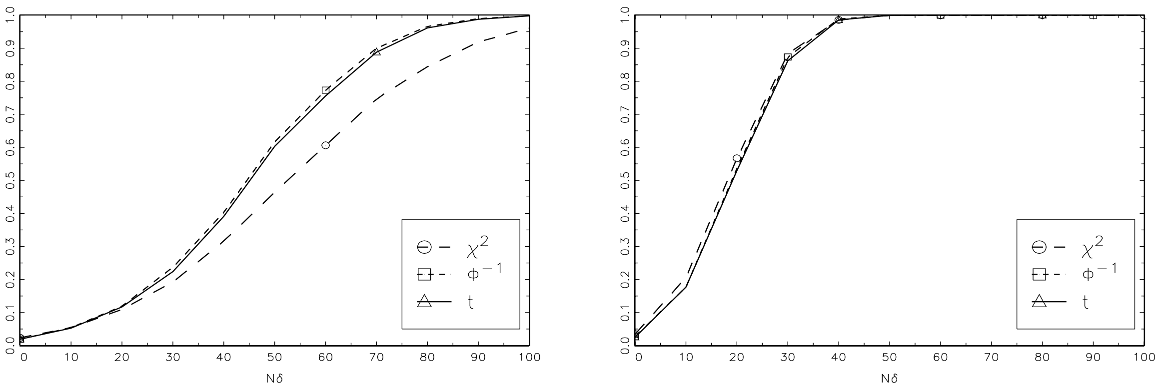
[6] Horowitz and Savin [2000] emphasize that size-adjusted critical values are of little use in empirical work. We therefore do not calculate size-adjusted power.

Table II—Empirical Power of the P Tests

T	N	*ADF* 10	20	50	100	150	λ_{trace} 10	20	50	100	150
(i) P_{χ^2}											
10		.061	.052	.040	.037	.033					
30		.062	.070	.088	.120	.135					
50		.115	.158	.287	.465	.609	.082	.131	.138	.152	.176
100		.403	.659	.955	1.00	1.00	.110	.217	.246	.254	.309
250		.999	1.00	1.00	1.00	1.00	.564	.966	.997	.999	1.00
500		1.00	1.00	1.00	1.00	1.00	.998	1.00	1.00	1.00	1.00
(ii) $P_{\Phi^{-1}}$											
10		.039	.031	.015	.006	.005					
30		.063	.078	.108	.151	.188					
50		.136	.201	.376	.617	.771	.076	.130	.128	.129	.143
100		.426	.700	.968	1.00	1.00	.106	.223	.238	.238	.271
250		.990	1.00	1.00	1.00	1.00	.443	.892	.975	.996	1.00
500		1.00	1.00	1.00	1.00	1.00	.941	1.00	1.00	1.00	1.00
(iii) P_t											
10		.041	.030	.015	.005	.004					
30		.063	.076	.105	.150	.182					
50		.132	.196	.359	.603	.751	.078	.135	.131	.134	.148
100		.423	.685	.961	1.00	1.00	.108	.222	.240	.241	.277
250		.994	1.00	1.00	1.00	1.00	.480	.915	.982	.997	1.00
500		1.00	1.00	1.00	1.00	1.00	.979	1.00	1.00	1.00	1.00

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.

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 "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)



(A) $T = 50,\ N = 100$. (B) $T = N = 100$.

$\rho = 0.9,\ \psi = 0,\ \sigma = 1$ AND $a_1 = 0$

FIGURE I—POWER OF THE P PANEL COINTEGRATION TESTS

"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,

$$DF_\rho^* = \frac{\sqrt{N}T(\hat{\rho}-1) + \frac{3\sqrt{N}\hat{\sigma}_\nu^2}{\hat{\sigma}_{0\nu}^2}}{\sqrt{3+\frac{36\hat{\sigma}_\nu^4}{5\hat{\sigma}_{0\nu}^4}}} \quad \text{and}$$

$$DF_t^* = \frac{t_\rho + \frac{\sqrt{6N}\hat{\sigma}_\nu^2}{2\hat{\sigma}_{0\nu}}}{\sqrt{\frac{\hat{\sigma}_{0\nu}^2}{2\hat{\sigma}_\nu^2}+\frac{3\hat{\sigma}_\nu^2}{10\hat{\sigma}_{0\nu}^2}}}.$$

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 (3). Defining $\overline{\lambda}_{trace} = N^{-1}\sum_{i=1}^{N}\lambda_{trace,i}$, their panel cointegration test statistic is given by

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

Under some conditions, including $\sqrt{N}T^{-1} \rightarrow 0$, 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.

Now, let us compare the results in Figure I with those obtained by Gutierrez [2003].[7] The P tests are somewhat less powerful than the residual-based panel tests $\tilde{Z}_{\hat{\rho}_{NT}-1}$, $Z_{\hat{\rho}_{NT}-1}$, DF_ρ^* and DF_t^* for shorter panels. However, power for longer panels is similar. Furthermore, the P tests always outperform the system-based Υ_{LR} test by Larsson et al. [2001]. Note, though, that these results are not based on size-adjusted critical values as in Gutierrez [2003]. Given that the P tests seem to be undersized (see Table I), their power would be higher if it were reported on the basis of exact rather than nominal critical values.

[7] Figure I corresponds to the middle and lower right panel in Fig. 1 in Gutierrez [2003].

We think that DGP A is restrictive. Apart from 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 comparison of the different panel cointegration tests under these circumstances is available in the literature, we also include some of the tests presented above.

Consider the following modification of DGP A to introduce higher order serial correlation in the equilibrium error z_t. We draw, for each cointegrated series in the panel, the order of the AR-process according to $\tilde{\zeta}_i = [\zeta_i]$, where $\zeta_i \sim \mathcal{U}[1,6],\ i = 1,\ldots,\delta N$ and $[y]$ is the integer part of y. We then generate the AR-coefficients from $\varphi_{i,p} \sim \mathcal{U}[0,0.99],\ i = 1,\ldots,\delta N;\ p_i = 1,\ldots,\tilde{\zeta}_i$, discarding all processes with eigenvalues outside the unit circle.

DGP B

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

Table III gives results on the power of the tests for $\sigma = 1,\ \psi = 0,\ a_1 = 0$ and $\delta = 0.5$. The dimensions are $T, N \in \{10, 30, 50\}$. The second part of the subscript ('DF') indicates that Engle and Granger's [1987] ADF test is chosen as the underlying time series test for the P tests. The number of lagged differences for the ADF regression is chosen according to the automatic procedure suggested by Ng and Perron [2001]. It is not possible to compare the power with the results from Table II because the alternative is now different. But, Table III shows that the first five tests clearly have higher power than the last one. This is intuitive as the P and Pedroni [2004] tests are designed to accommodate cross-sectional heterogeneity. The tests put forward in this paper may therefore be useful in a fairly wide range of practical applications.

Table III—Power with AR(p) Errors

T	N	10	30	50	10	30	50
			$P_{\chi^2 DF}$			$P_{\Phi^{-1}DF}$	
20		.899	.999	1.00	.792	.992	.999
30		.990	1.00	1.00	.961	1.00	1.00
50		.999	1.00	1.00	.997	1.00	1.00
			P_{tDF}			$Z_{\hat{\rho}_{NT}-1}$	
20		.832	.995	.999	.806	.996	1.00
30		.973	1.00	1.00	.899	.999	1.00
50		.998	1.00	1.00	.980	1.00	1.00
			$Z_{\hat{t}_{NT}}$			DF^*_ρ	
20		.968	1.00	1.00	.030	.019	.015
30		.969	1.00	1.00	.101	.109	.130
50		.988	1.00	1.00	.276	.355	.424

Note: $\rho = 0.9,\ \sigma = 1,\ \delta = 0.5,\ \psi = a_1 = 0$.
$M = 5,000$ replications. 5% nominal level.

4 Conclusion

We introduce new tests for panel cointegration. 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 suggested tests is somewhat undersized in either very short and wide panels. However, the empirical size of the tests is very close to the nominal one for panel dimensions often encountered in applied macroeconometric work. In terms of power, their performance is intermediate between other widely used panel cointegration tests.

As most tests in this literature, the ones suggested here rely on the assumption of cross-sectional uncorrelatedness (see Assumption 2). This assumption is likely to be overly strong for many macroeconomic panels and may lead, if violated, to erroneous conclusions [cf. O'Connell, 1998]. We therefore suggest to extend the tests developed here to allow for cross-sectional correlation by, e.g., the bootstrap method. Maddala and Wu [1999] report encouraging results along these lines for their panel unit root test. There is a growing literature on bootstrapping cointegrating regressions [see Li and Maddala, 1997] that can be fruitfully applied to the present problem. Recent useful contributions include Chang and Park [2003] and Chang et al. [2006]. Investigation of this extension is currently under way by the author.

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: forthcoming, 'Unit Roots and Cointegration in Panels'. In: L. Matyas and P. Sevestre (eds.): *The Econometrics of Panel Data*. Dordrecht: Kluwer Academic Publishers, 3rd edition.

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* (forthcoming).

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.: 1970, *Statistical Methods for Research Workers*. London: Oliver and Boyd, 14th edition.

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.: 2006, 'The Error-in-Rejection Probability of Meta Analytic Panel Tests'. Technical report, RGS.

Hedges, L. and I. Olkin: 1985, *Statistical Methods for Meta-Analysis*. San Diego: Academic Press.

Horowitz, J. L. and N. Savin: 2000, 'Empirically Relevant Critical Values for Hypothesis Tests: A Bootstrap Approach'. *Journal of Econometrics* **95**, 375–389.

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.

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.: 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. and S. Ouliaris: 1990, ‘Asymptotic Properties of Residual Based Tests for Cointegration’. *Econometrica* **58**(1), 165–193.

Shin, Y.: 1994, ‘A Residual-Based Test of the Null of Cointegration against the Alternative of No Cointegration’. *Econometric Theory* **10**(1), 91–115.