Mixed Signals Among Panel Cointegration Tests

Christoph Hanck*

November 13, 2006

Abstract

Time series cointegration tests, even in the presence of large sample sizes, often yield conflicting conclusions ("mixed signals") as measured by, inter alia, a low correlation of empirical p-values [see Gregory et al., 2004, *Journal of Applied Econometrics*]. Using their methodology, we present evidence suggesting that the problem of mixed signals persists for popular panel cointegration tests. As expected, there is weaker correlation between residual and system-based tests than between tests of the same group.

Keywords: Panel cointegration tests, Monte Carlo comparison

*Research supported by DFG under Sonderforschungsbereich 475. Universität Dortmund/Ruhr Graduate School in Economics, Contact Information: Tel. (+49) 0231-7553127, christoph.hanck@uni-dortmund.de, `www.rgs-econ.org`.

1 Introduction

An extensive battery of tests is available to investigate the unit root and cointegration properties of economic time series. Typically, however, an applied researcher has little practical guidance as to which test to use, as most tests test very similar hypotheses. It would therefore be reassuring if rejection or acceptance of a particular economic hypothesis did not depend on which of the tests is used. For instance, in the context of hypothesis testing with stationary variables it is well-known that the classical likelihood ratio, Lagrange multiplier and Wald tests are asymptotically numerically equivalent under quite general conditions [Davidson and MacKinnon, 1993, Ch. 13].

As analytical characterizations of the correlations of the various test statistics for cointegration are difficult to obtain, Gregory et al. [2004] analyze this question by means of Monte Carlo methods. They generate replications of two independent random walks and test the null of no cointegration using the popular residual-based tests by Engle and Granger [1987] and Phillips and Ouliaris [1990] as well as the system-based λ_{trace} and λ_{max} tests [Johansen, 1988]. They then calculate p-values from the empirical distribution of the test statistics by taking rank order of the latter and dividing by M, the number of replications. Disturbingly, for most pairs of tests, virtually any combination of p-values can arise. That is, while the combinations should ideally cluster around the 45°-line, it frequently occurs that a particularly high test statistic of, say, the λ_{trace}-test is associated with a low test statistic of, say, the Engle and Granger [1987] Augmented Dickey-Fuller (ADF)-test. The main conclusion is that using different tests is likely to yield conflicting conclusions in applications.

In recent years, the cointegration methodology has been extended to panel data. Pedroni [2004] and Kao [1999] generalize residual-based tests, Larsson et al. [2001] extend the Johansen [1988] tests, while McCoskey and Kao [1998] propose a test for the null of panel cointegration in the spirit of Shin [1994]. Hanck [2005] extends the p-value combination panel unit root tests of Maddala and Wu [1999] and Choi [2001] to the panel cointegration case.

Under cross-sectional independence all the above-mentioned panel tests provide a means to better exploit the variation in the data. Furthermore, Phillips and Moon [1999] show that panel data can help mitigate the spurious regression phenomenon. The contribution of this paper is to investigate whether the availability of panel data is also useful for obtaining more consistent decisions among the competing tests. To shed light on this question we adopt the methodology suggested by Gregory et al. [2004] and extend it to the panel data setting.

The remainder of the paper is organized as follows. Section 2 briefly reviews the panel cointegration tests compared in this paper. Section 3 describes the simulation setup of the comparative study and reports the results. Section 4 concludes.

2 Panel Cointegration Tests

We give the key statistics of the various tests that are considered. For more details, refer to the original contributions. Furthermore, Banerjee [1999] or Baltagi and Kao [2000] provide surveys of the literature. We focus on tests with the null of no panel cointegration.

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 "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 $\tilde{e}_{i,t} = (\Delta\hat{e}_{i,t}, \hat{e}_{i,t-1})'$ and T is sample size. The $\hat{e}_{i,t}$ are obtained from heterogenous Engle/Granger-type first stage OLS multivariate time series regressions of one of the variables $\mathbf{x}_{ik}$ 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,

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

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

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 the set of N λ_{trace} test statistics [Johansen, 1988] for each unit in the panel. (See also (2) below.) 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 $\sqrt{N}T^{-1} \to 0$, 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.

Hanck [2005]

The main idea of the 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. The test statistics are obtained by combining the p-values of the N tests into panel test statistics as follows:

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

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

where Φ^{-1} denotes the inverse of the cumulative distribution function (cdf) of the standard normal distribution. When considered together we refer to Eqs. (1a) and (1b) as P tests from now on. Assuming continuous distribution functions of the time series test statistics under H_0, as $T_i \to \infty$ for all i, the test statistics are asymptotically distributed as

$$P_{\chi^2} \to_d \chi^2_{2N}$$
$$P_{\Phi^{-1}} \to_d \mathcal{N}(0,1),$$

where χ^2_{2N} is a χ^2 random variable with $2N$ degrees of freedom. The decision rule 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. On the other hand, for (1b) one would reject for large negative values of $P_{\Phi^{-1}}$.

We obtain the p-values from the ADF cointegration tests [Engle and Granger, 1987] as provided by MacKinnon [1996]. That is, the p-values are 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 multivariate OLS time series regressions of one of the variables $\mathbf{x}_{ik}$ 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

$$\lambda_{trace,i}\left(h\right) = -T \sum_{k=h+1}^{K} \ln\left(1 - \hat{\pi}_{k,i}\right) \quad (2)$$

and the λ_{max}-test

$$\lambda_{max,i}\left(h|h+1\right) = -T \ln\left(1 - \hat{\pi}_{h+1,i}\right). \quad (3)$$

Here, $\hat{\pi}_{k,i}$ denotes the kth largest eigenvalue of $\hat{\Pi}_i$. In (2), the alternative is a general one, while one tests against $h+1$ cointegration relationships in (3).

3 Do Panel Cointegration Tests Produce "Mixed Signals"?

We now use the panel cointegration tests outlined in the previous section to investigate the extent to which different widely used panel cointegration tests yield the same decision for a given (artificial) sample. Gregory et al. [2004] observe mixed signals, i.e. a relatively high test statistic for one test and a relatively low test statistic for another, for time series cointegration tests.[1] This effect is particularly strong when comparing residual- and system-based tests.

It might be conjectured that the availability of panel data, leading to standard (normal) null distributions of the test statistics, could help alleviate this problem. To shed light on this question, we adopt the methodology of Gregory et al. [2004].[2] More precisely, we generate many replications of two integrated time series for each of the N units in the panel. For each replication, we store the different panel cointegration test statistics. The

[1] Berndt and Savin [1977] study the related problem of conflicting decisions among the classical hypothesis tests in linear regression models. A crucial difference is that the numerical relationship between the criteria is well understood for these simpler models. Furthermore, in this context the situation is resolved asymptotically.

[2] Gregory et al. [2004] complement their simulation study with an extensive analysis of all applications of the cointegration methodology published in the *Journal of Applied Econometrics* in recent years. While such an approach has obvious appeal it is not yet promising in the panel data context due to the small number of empirical applications. We therefore exclusively rely on artificial data.

extent to which the different tests yield identical decisions is measured by two related criteria. First, we compute empirical p-values of the tests by taking rank order of the test statistics and dividing by M. We then compute the correlation of the empirical p-values for each pair of tests. If both have the same null and the same alternative, the correlation should therefore ideally be close to one, i.e. a strong rejection of one test should also be a strong one of the other. Second, we record all the instances of each pair of tests rejecting jointly. The critical values are either taken from the asymptotic distribution of the tests or the empirical distribution arising from the replications under the null. Thus, when testing a sample generated under the null at the 5% level, all pairs of tests should ideally jointly reject in close to 5% of the replications.

We compare the tests of Kao [1999], Pedroni [2004] and Larsson et al. [2001] presented in the previous section. We further include the two P tests. For each, we use both Engle and Granger's [1987] ADF test with one lagged difference ($P_{\chi^2 DF}$ and $P_{\Phi^{-1}DF}$) as well as Johansen's [1988] λ_{trace} test for $r = 0$ versus $r \leqslant K = 2$ cointegrating relationships ($P_{\chi^2 J}$ and $P_{\Phi^{-1}J}$). Following Gregory et al. [2004], we choose relatively large time series dimensions to limit size distortions. More specifically, $T \in \{250, 500, 1000, 2000\}$ and $N \in \{10, 20, 50, 100, 150\}$. 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 we only consider the bivariate case:

DGP

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

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_{i1t} and x_{i2t} 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_{i2t} is weakly exogenous when $a_1 = 0$.

We consider the parameter space $\beta = 2,\ a_2 = -1,\ \sigma \in \{0.5, 1\},\ \psi \in \{-0.5, 0, 0.5\}$ and $a_1 \in \{0, 1\}$. This implies that, for instance, the Pedroni [2004] and P tests cannot exhibit their comparative advantage of being able to detect cross-sectional heterogeneity in the slope coefficients. Similarly, a bivariate system necessarily has at most one cointegrating relationship. Thus, the Larsson et al. [2001] test have no opportunity to detect multiple cointegration. But, the Kao [1999] tests require a common β for all i. Hence, in order to be able to validly compare all tests under both the null and the alternative we use this simple DGP. We carry out the experiments under both the null and the alternative.[3] For the latter we set $\rho = 0.98$. The fraction of cointegrated series in the panel is either zero or one, $\delta \in \{0, 1\}$. For a given cross-sectional dimension we draw the unit specific intercepts as $\alpha_i \sim \mathcal{U}[0, 10]$ and keep them fixed for all T. The number of replications for each experiment is $M = 10,000$.

Here, we report the (representative) results for $a_1 = 1,\ \sigma = 1,\ \psi = 0$.[4] Table I shows the correlation of the empirical p-values for $N = 50$. Panels (a) and (b) consider $T = 250$ and $T = 2000$, respectively. Within each of the panels there is a fairly high correlation among the different residual-based tests (rows 2-8) and, especially, among the different system-based tests (rows 1, 9-10). The pattern is not uniform, though. For the residual-based tests, the correlation ranges from roughly 30% ($P_{\chi^2 DF}$ and $Z_{\hat{t}_{NT}}$) to almost 95% (DF^*_ρ and $Z_{\hat{\rho}_{NT}-1}$). For a graphical illustration, see the scatter plot of the empirical p-values for these cases in Figure I. Panel (a), depicting the correlation of $P_{\chi^2 DF}$ and $Z_{\hat{t}_{NT}}$, shows that, even within the group of residual-based tests, virtually any combination of empirical p-values can arise. On the other hand, Panel (b) reveals that for some cases the p-values cluster around the 45°-line, indicating a close correspondence.

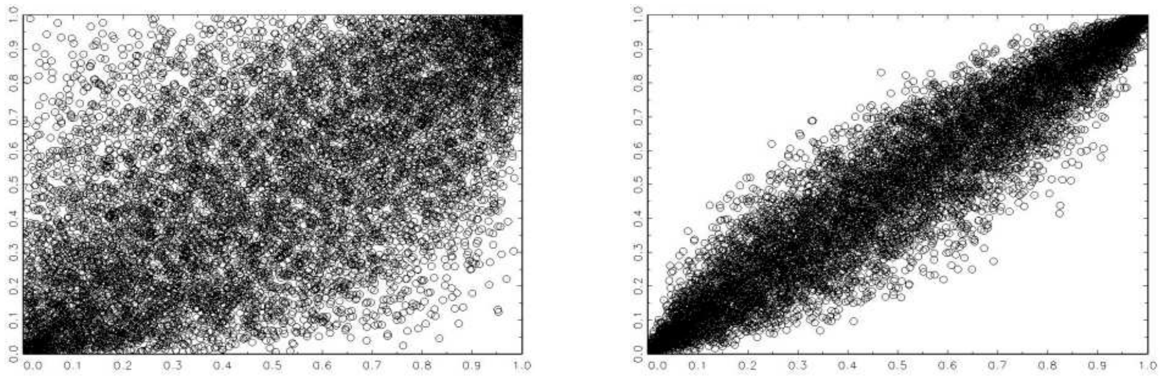
Furthermore, different tests by the same author do not seem to be any more related than tests by different authors. Across the two groups the correlation typically is substantially lower, with several entries even being negative (see, e.g., the first column). Finally, compare Panels (a) and (b). Increasing the time series dimension barely affects the correlation of the empirical p-values. (Similar results obtain for increasing N.)

We provide some further insights in Table II. Using 5% size-adjusted critical values we report the fraction of each pair of test rejecting jointly.[5] The case considered in Panel

[3]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.

[4]The full set of results of the finite sample study is available upon request.

[5]Horowitz and Savin [2000] correctly point out that size-adjusted critical values are usually of little



(A) $P_{\chi^2 DF}$ AND $Z_{\hat{t}_{NT}}$ (B) DF_ρ^* AND $Z_{\hat{\rho}_{NT}-1}$

FIGURE I—CORRELATION OF EMPIRICAL p-VALUES

(a) corresponds to Panel (a) of Table I. The entries under (b) give results under the alternative of panel cointegration. As expected from Table I, no pair of tests achieves a fraction of joint rejections of 5%. Reassuringly, the combinations having a high correlation of empirical p-values also have a relatively high fraction of joint rejections. However, in spite of fairly high correlation (take $P_{\chi^2 J}$ and $P_{\Phi^{-1} J}$ with more than 90%) we still observe pairs of tests jointly rejecting for a rather small fraction of samples (1.6% for this example). That is, conflicting testing decisions are not uncommon. As all tests reject more frequently under the alternative, the fraction of joint rejections of course increases (see Panel b). Nevertheless, there is still a large amount of disagreement especially across groups of tests.

Comparing the results with Gregory et al. [2004], we state that the consensus in test decisions among panel data cointegration tests generally does not seem to be higher than among time series cointegration tests. Thus, it seems all but unlikely that a researcher will find conflicting evidence when applying some pairs of panel cointegration tests to a given dataset. The issue is not resolved asymptotically. A possible explanation of this phenomenon could be that the complexities inherent to panel data—such as treatment of cross-sectional heterogeneity—lead to different implicit alternatives of the tests. Consequently, we observe a rather low correlation of empirical p-values and fractions of joint rejections when the data is generated under the null.

use for applied work. Here, however, we use them to avoid spurious results that could arise if, say, two tests were both heavily oversized and would therefore also frequently reject jointly.

Table I—Correlation of the Empirical p-values under the Null

	Υ_{LR}	DF_t^*	$Z_{\hat{\rho}_{NT}-1}$	$\tilde{Z}_{\hat{\rho}_{NT}-1}$	DF_ρ^*	$Z_{\hat{t}_{NT}}$	$P_{\chi^2 DF}$	$P_{\Phi^{-1}DF}$	$P_{\chi^2 J}$	$P_{\Phi^{-1}J}$
					(a)	$T = 250$				
Υ_{LR}	1.00									
DF_t^*	-.055	1.00								
$Z_{\hat{\rho}_{NT}-1}$	.115	.445	1.00							
$\tilde{Z}_{\hat{\rho}_{NT}-1}$	.264	.312	.698	1.00						
DF_ρ^*	.098	.514	.944	.658	1.00					
$Z_{\hat{t}_{NT}}$	-.087	.935	.486	.341	.492	1.00				
$P_{\chi^2 DF}$	.235	.314	.583	.927	.599	.304	1.00			
$P_{\Phi^{-1}DF}$	.213	.466	.764	.919	.806	.439	.898	1.00		
$P_{\chi^2 J}$	.984	-.059	.116	.268	.099	-.089	.245	.213	1.00	
$P_{\Phi^{-1}J}$	.961	-.045	.106	.242	.090	-.078	.205	.198	.898	1.00
					(b)	$T = 2000$				
Υ_{LR}	1.00									
DF_t^*	-.096	1.00								
$Z_{\hat{\rho}_{NT}-1}$	.131	.466	1.00							
$\tilde{Z}_{\hat{\rho}_{NT}-1}$	.346	.320	.652	1.00						
DF_ρ^*	.094	.552	.949	.614	1.00					
$Z_{\hat{t}_{NT}}$	-.112	.938	.505	.359	.530	1.00				
$P_{\chi^2 DF}$	.265	.330	.545	.929	.561	.330	1.00			
$P_{\Phi^{-1}DF}$	.242	.487	.736	.915	.782	.467	.896	1.00		
$P_{\chi^2 J}$	.984	-.097	.135	.351	.099	-.111	.279	.245	1.00	
$P_{\Phi^{-1}J}$	.964	-.090	.119	.318	.083	-.107	.229	.222	.903	1.00

Note: (a) $N = 50$ (b) $\rho = 1,\ \psi = 0,\ \sigma = 1,\ \delta = 1$ and $a_1 = 1$.
(c) $M = 10,000$ replications.

4 Conclusion

We perform a simulation study to investigate whether several widely used panel cointegration tests yield the same acceptance or rejection decisions. Broadly in accordance with the evidence presented by Gregory et al. [2004] for time series tests, the panel versions also exhibit a low correlation of empirical p-values under the null. The persistence of the phenomenon even at $T = 2000$ indicates that this problem does not seem to be resolved asymptotically. When analyzing the relative frequency of joint rejections, we constrain the tests to have the desired size by using size-adjusted critical values. Low fractions of joint rejections (relative to the size of the tests) show that the tests do not reject for the same samples. This phenomenon is less prevalent under the alternative.

The practical upshot is that researchers are likely to be confronted with conflicting test decisions when using different tests in applied work. Given that there rarely is a com-

Table II—Fraction of joint rejections under H_0 and H_1

	Υ_{LR}	DF_t^*	$Z_{\hat{\rho}_{NT}-1}$	$\tilde{Z}_{\hat{\rho}_{NT}-1}$	DF_ρ^*	$Z_{\hat{t}_{NT}}$	$P_{\chi^2 DF}$	$P_{\Phi^{-1} DF}$	$P_{\chi^2 J}$	$P_{\Phi^{-1} J}$
Υ_{LR}	.050									
DF_t^*	.002	.050			*(a)*	$\rho = 1$				
$Z_{\hat{\rho}_{NT}-1}$	.005	.011	.050							
$\tilde{Z}_{\hat{\rho}_{NT}-1}$	.007	.008	.021	.050						
DF_ρ^*	.004	.015	.037	.020	.050					
$Z_{\hat{t}_{NT}}$	.002	.036	.012	.011	.014	.050				
$P_{\chi^2 DF}$	.006	.008	.016	.036	.017	.008	.050			
$P_{\Phi^{-1} DF}$	.006	.012	.025	.036	.028	.013	.032	.050		
$P_{\chi^2 J}$	.043	.002	.004	.007	.004	.002	.007	.006	.050	
$P_{\Phi^{-1} J}$	.040	.002	.004	.007	.004	.002	.005	.006	.016	.050
Υ_{LR}	.191									
DF_t^*	.191	.999			*(b)*	$\rho = .98$				
$Z_{\hat{\rho}_{NT}-1}$	.190	.975	.975							
$\tilde{Z}_{\hat{\rho}_{NT}-1}$	.183	.872	.869	.873						
DF_ρ^*	.176	.857	.858	.809	.858					
$Z_{\hat{t}_{NT}}$	.191	.999	.975	.873	.858	1.00				
$P_{\chi^2 DF}$	.135	.516	.516	.516	.504	.516	.516			
$P_{\Phi^{-1} DF}$	.169	.750	.750	.749	.733	.750	.513	.750		
$P_{\chi^2 J}$	.164	.174	.173	.166	.161	.174	.124	.155	.174	
$P_{\Phi^{-1} J}$	.173	.227	.225	.215	.208	.227	.153	.197	.028	.227

Note: (a) $N = 50,\ T = 250,\ \psi = 0,\ \sigma = 1,\ \delta = 1$ and $a_1 = 1$.
(b) $M = 10,000$ replications. (c) Size-adjusted 5% critical values.

pelling theoretical reason to prefer one test over another in practice, this issue is rather troublesome. More research clarifying the theoretical relationship between the different tests would be welcome.

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.

Berndt, E. R. and N. E. Savin: 1977, 'Conflict Among Criteria for Testing Hypothesis in the Multivariate Linear Regression Model'. *Econometrica* **45**(5), 1263–1277.

Choi, I.: 2001, 'Unit Root Tests for Panel Data'. *Journal of International Money and Finance* **20**(2), 249–272.

Davidson, R. and J. G. MacKinnon: 1993, *Estimation and Inference in Econometrics*. New York: Oxford University Press.

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.

Gregory, A. W., A. A. Haug, and N. Lomuto: 2004, 'Mixed Signals Among Tests for Cointegration'. *Journal of Applied Econometrics* **19**(1), 89–98.

Hanck, C.: 2005, 'A Meta Analytic Approach to Testing for Panel Cointegration'. 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.

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.

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.

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 H. R. Moon: 1999, 'Linear Regression Limit Theory for Nonstationary Panel Data'. *Econometrica* **67**(5), 1057–1111.

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.