Mixed Signals Among Tests for Panel Cointegration

Westerlund, Joakim and Basher, Syed A.

16 May 2007

Mixed Signals Among Tests for Panel Cointegration*

Joakim Westerlund† Syed A. Basher‡

May 16, 2007

Abstract

In this paper, we study the effect that different serial correlation adjustment methods can have on panel cointegration testing. As an example, we consider the very popular tests developed by Pedroni (1999, 2004). Results based on both simulated and real data suggest that different adjustment methods can lead to significant variations in test outcome, and thus also in the conclusions.

JEL Classification: C14; C15; C32; C33.

Keywords: Panel Data; Cointegration Testing; Parametric and Semiparametric Methods.

1 Introduction

By now there exists a burgeoning literature on nonstationary panels, suggesting numerous tests for panel cointegration.[1] There are basically two classes of such tests. The first class is based on the seminal work of Engle and Granger (1987), and develops residual-based tests for use in the panel data context. Among the many contributions within this class, Kao (1999) and Pedroni (1999, 2004) belong to the most well cited ones. The second class builds on the work of Johansen (1988, 1991), and develops likelihood-ratio tests for panel data. The two most notable contributions within this class include Larsson, Lyhagen and Lötgren (2001) and Groen and Kleibergen (2003). As in the time series literature, most of these panel tests take no cointegration as the null hypothesis.

Gutierrez (2003) and Banerjee, Marcellino and Osbat (2004) study small-sample performance of many of these panel tests using Monte Carlo simulations, and find that no one test

*Westerlund gratefully acknowledges financial support from the Jan Wallander and Tom Hedelius Foundation, research grant number W2006-0068:1. The authors would also like to thank numerous people for making their data available to us. The usual disclaimer applies.

†Corresponding author: Department of Economics, Lund University, P. O. Box 7082, S-220 07 Lund, Sweden. Telephone: +46 46 222 4970; Fax: +46 46 222 4118; E-mail address: joakim.westerlund@nek.lu.se.

‡Department of Economics, York University, Canada. E-mail: basher@econ.yorku.ca.

[1] See Breitung and Pesaran (2007) for a recent survey of the literature.

can be said to dominate the others. In terms of applied work, however, the class of residual-based tests has proven to be the most popular one.

Besides the classification as residual or likelihood-based, most existing panel tests can also be classified according to the adjustment method undertaken to eliminate the dependence on nuisance parameters reflecting the serial correlation properties of the data, which would otherwise impair inference.[2] Parametric tests, such as those of Kao (1999), Larsson, Lyhagen and Lötgren (2001) and Groen and Kleibergen (2003), allow for quite general dependencies by assuming that the data admits to an autoregressive representation. By approximating this autoregression using the actual data, it is possible, at least in principle, to construct tests that do not not depend on any nuisance parameters. However, by approximating the data generating process in this way, the researcher faces the complicating factor of having to chose the appropriate order for the autoregression, which is typically unknown. If the order is chosen too small, the problem of nuisance parameter dependency will remain, whereas, if it is chosen too large, the small-sample properties of the test will deteriorate. Only if it is chosen correctly can the test be expected to perform well.

The tests by Pedroni (1999, 2004) can not only be constructed in this fully parametric way but also semiparametrically, in which case the nuisance parameters are estimated directly by using kernel methods. But this necessitates the researcher choosing the correct kernel to use and, perhaps even more importantly, the appropriate bandwidth parameter, which reflects the number of autocovariances to estimate for the kernel. As with the parametric method, this choice makes the testing problematic in practice as the correct bandwidth window to use in any given application is never known. Moreover, because these tests have the property that they can be constructed using both parametric and semiparametric adjustments, their constriction is even more uncertain in comparison to most other tests.

In addition to the problem that the true or optimal lag and bandwidth orders are never available in practice, adjustments of this kind can have a significant impact on test performance in small samples. Indeed, most empirical work tend to suggest that test results can be quite sensitive to different choices of lag lengths and bandwidths. As a result, test results are often reported for more than one value.

Of course, the problem of having to pick the best lag length or bandwidth is not unique to panel data. However, if one admits to the possibility of an heterogeneous data generating process, then this choice must be made not just once but as many times as there are individuals in the panel. This means that the sensitivity of the choice of lag length or bandwidth becomes

[2] Although this paper focuses on the problem of correcting for serial correlation, readers should be aware that there is generally the additional difficulty of correcting for the fact that the regressors may be endogenous.

even greater as the cross-sectional dimension of the panel increases, especially considering that the testing is often performed in situations when the time series dimension of the data is substantially shorter than in the conventional time series case.

Fortunately, there are ways to eliminate this kind of problems. Indeed, in a recent study, Westerlund (2005) proposes two simple residual-based panel data tests for the null of no cointegration, which can be said to be fully nonparametric as they do not require any correction for the temporal dependencies of the data. The key here is that no adjustment is required even though the data generating process is identical to the one used by Pedroni (1999, 2004). Thus, with these tests, there is no lag or bandwidth parameter that needs to be chosen by the researcher, which of course reduces the uncertainty and ambiguity of the test outcome.

In view of this new development, a natural question arises of how robust parametric and semiparametric tests can be expected to be, and what this has for implications in applied work. In this paper, we try to shed some light on the issue using both simulated and actual data. Simulation studies are usually conducted for a single choice of adjustment method. We therefore begin our analysis with a small simulation exercise where we consider different methods. Our main finding is that the Pedroni (1999, 2004) tests are much more prone to erroneous conclusions than the Westerlund (2005) tests.

In the empirical section of the paper, we compare the results of conducting the various tests using data sets from nine published studies that use the tests of Pedroni (1999, 2004). Consistent with our simulation results, we find that different choices of adjustments can lead to completely different conclusions when using these tests. By contrast, the results based on the Westerlund (2005) tests are completely unambiguous in this respect, and are usually in agreement with the results reported in the published studies.

The remainder of this paper is organized as follows. Section 2 provides a brief discussion of the difficulty of having to adjust for serial correlation when testing for panel cointegration, while Sections 3 and 4 report the simulation and empirical, respectively. Section 5 concludes.

2 Adjusting for serial correlation when testing for cointegration in panel data

The purpose of this section is to provide a birds-eye account of the difficulties of having to adjust for the presence of serial correlation when testing for panel cointegration using the Pedroni (1999, 2004) tests. The interested reader is referred to the original papers for a detailed description of the construction of these tests.

The tests of Pedroni (1999, 2004) and Westerlund (2005) are all based on examining the stationarity of the error term in the following regression

$$y_{it} = d_t'\delta_i + x_{it}'\beta_i + e_{it}, \quad (1)$$

where $t = 1, ..., T$ and $i = 1, ..., N$ indexes the time series and cross-sectional dimensions, respectively, d_t is a vector of deterministic components, with a constant and time trend as typical elements, β_i are slope parameters, and x_{it} is a vector of integrated regressors.[3] Now, consider the problem of testing the hypothesis of no cointegration based on the regression in (1). The error e_{it} is stationary when y_{it} and x_{it} are cointegrated and it has a unit root when they are not. Thus, testing the null hypothesis of no cointegration for cross-sectional unit number i is equivalent to testing whether e_{it} possesses a unit root by using the following autoregression

$$e_{it} = \rho_i e_{it-1} + u_{it}. \quad (2)$$

In this paper, however, we are not interested in testing if the no cointegration null holds for a single unit but rather if it holds for the panel as a whole. In other words, we want to test the null that $\rho_i = 1$ for all i. This can be done in two ways depending on how ρ_i is estimated. The group mean approach involves estimating ρ_i separately for each unit i before combining them into a panel statistic, while the panel approach involves first restricting $\rho_i = \rho$ for all i and then using the resulting pooled estimate of ρ as a statistic.

Thus, one way in which the tests differ is in the treatment of ρ_i, which is important in the sense that it has implications for the way a rejection is interpreted.[4] From a practitioners point of view, however, this is not the most important difference. Indeed, as in all testing situations of this type, correcting for serial correlation is a key aspect when testing for cointegration in panel data. This can essentially be done in two ways.

To appreciate this, assume that

$$u_{it} = \sum_{j=1}^{P_i} \phi_{ij} u_{it-j} + \varepsilon_{it}, \quad (3)$$

where ε_{it} is assumed to be mean zero and serially uncorrelated. This equation says that u_{it} follows an autoregressive process of some unknown order P_i, which means that the data can be serially correlated in a very general way. The problem is how to account for this correlation

[3] It might be noted that none of the tests considered here allow for cross-sectional dependence, except if it is in the form of a simple common time effect. However, this is by no means restrictive for what we are trying to do in this study. The issue of serial correlation correction is there whether the data is cross-sectionally dependent or not.

[4] In particular, a rejection by the group mean approach is usually interpreted as that $\rho_i < 1$ for at least one i, whereas, in the panel approach, it is interpreted as $\rho < 1$ for all i. Thus, a rejection of the null has different meanings depending on whether ρ_i is estimated separately or not.

when constructing the cointegration test. The most natural way is to simply substitute for u_{it} in (2), which yields the following augmented test regression

$$e_{it} = \rho_i e_{it-1} + \sum_{j=1}^{P_i} \phi_{ij} \Delta e_{it-j} + \varepsilon_{it}. \tag{4}$$

Thus, since the error ε_{it} is serially uncorrelated by constriction, one way to robustify the cointegartion test is to simply replace (2) with (4). Because this approach involves explicitly modelling the serial correlation, the resulting test is often called parametric. Another way to account for the correlation is to estimate directly the long-run variance of u_{it}, which is given by

$$\text{lrvar}(u_{it}) = \sum_{j=-\infty}^{\infty} \text{cov}(u_{it}, u_{it-j}).$$

This quantity can be estimated semiparametrically as suggested by Newey and West (1994), using the following weighted sum of sample autocovariances

$$\sum_{j=-(K_i-1)}^{K_i-1} \left(1 - \frac{|j|}{K_i}\right) \frac{1}{T} \sum_{t=|j|+1}^{T} u_{it} u_{it-j}, \tag{5}$$

where K_i is a bandwidth truncation parameter that determines the number of autocovariances to use.[5] In other words, choosing K_i in (5) is essentially the same as choosing P_i in (4). Once $\text{lrvar}(u_{it})$ has been estimated, the corresponding semiparametric test can be obtained by basically using this estimate instead of the usual estimate of the contemporaneous variance $\text{var}(u_{it})$.

The tests of Pedroni (1999, 2004) all require correcting for serial correlation, either parametrically as in (4) or semiparametrically as in (5). This makes them uncertain in the sense that their performance depends to a large degree on how well the researcher chooses K_i and P_i. In particular, the problem is that there is no unique way of choosing these parameters, and different choices can lead to very different test outcomes. In fact, as we will demonstrate in the next section, depending on the choice of K_i and P_i, the conclusion of the test can in many cases be completely reversed.

3 Simulation evidence

In this section, we conduct a small simulation exercise to evaluate the performance of the Pedroni (1999, 2004) tests when considering different choices of K_i and P_i. For this purpose, we make $1,000$ replications using (1) through (3) to generate the data.

[5] The advantage with this particular weighting scheme, usually referred to as the Bartlett kernel, is that it ensures nonnegative variance estimates. Although the choice of kernel also affects the performance of the test, in practice the Bartlett kernel is almost always used, and we therefore only consider this choice.

For simplicity, we assume that there is a single regressor such that $\Delta x_{it} \sim N(0,1)$, that (3) is generated with one lag only, and that $\varepsilon_{it} \sim N(0,1)$.[6] All initial values are set to zero. For the deterministic component, we have two configurations, one with an intercept and one with an intercept and trend. In both cases, all the parameters of (1) are set equal to one.

The tests can be classified as group mean or panel, and as parametric or semiparametric. Coefficient type tests will be denoted Z_ρ, while t-ratio type tests will be denoted Z_t. A tilde signifies that the test is of the group mean type, while a star signifies that the test is parametric. For convenience, we employ a uniform bandwidth and lag length truncation window, and evaluate the tests for all truncations between one and 10, which, given the variety of the choices that appear in applied work, seems like a very reasonable range of values.[7] The same truncation is used for all units of the panel. To economize on space, we only report the results for the size of a nominal 5% level test.

The results are reported in Table 1. These are summarized by means of intervals comprising the sizes that were generated by varying K_i and P_i between one and 10. The information contained in the table can be described as follows.

On the one hand, we see that the while the size of the semiparametric tests appear to be quite stable across bandwidths, these tests are usually also very oversized. The only exception is when serial correlation coefficient ϕ is positive, in which case the distortions go in the other direction, thus making the test more conservative.

On the other hand, the size of the parametric tests varies widely depending on the choice of lag length, and can produce almost any result. Indeed, since most intervals contains both endpoints, by simply choosing the appropriate lag length, it should in principle be possible to always obtain the results one would like to have. Nevertheless, in contrast to the semiparametric tests, we see that the parametric tests can actually be decently sized. However, this requires the knowledge of the true lag length, which of course is never known in practice.

These findings present us with an intricate dilemma. On the one hand, when using the semiparametric tests, we obtain results that are more stable across different truncation windows than those obtained using the parametric tests. On the other hand, since the semiparametric tests tend to be very distorted, there is a large risk of obtaining spurious results, especially in the presence of serial correlation. As expected, we see that the distortions have a clear tendency of accumulating, and to become even more serious as N increases. Also, although the distortions

[6]For simplicity, to be able to identify the effect of serial correlation, we assume that the regressor is strictly exogenous. See Pedroni (1999, 2004) and Westerlund (2005) for some results when the regressor is permitted to be endogenous.

[7]Theoretically, an asymptotically valid test requires the lag or bandwidth truncation window to increase with T. However, this is not always done in practice, especially when it comes to lag length selection. It is therefore of interest to see how the tests perform when combined with a fixed truncation.

seem to be exacerbated by the presence of the trend, we see that the deterministic specification does not alter the conclusions.

Because of this difficulty in interpreting the results of the Pedroni (1999, 2004) tests, it is useful to consider as an alternative the tests of Westerlund (2005), here denoted V_g and V_p, which are interesting in the sense that they do not require any adjustment to account for the serial correlation of the data.[8] Thus, with these tests, there is no dependence on the choice of lag length or bandwidth. However, the underlying assumptions are exactly as in Pedroni (1999, 2004), which means that the test outcomes can be easily compared and interpreted. The V_g test is of the group mean type, and is thus comparable with the corresponding group mean tests of Pedroni (1999, 2004), while V_p is of the panel type. Based on the results reported in Table 1, it would appear as that the risk of obtaining misleading results is much lower when using the Westerlund (2005) tests.

4 Empirical evidence

In this section, we reevaluate the results obtained from nine recent empirical studies that are based on the tests of Pedroni (1999, 2004). The purpose is to show how different choices of K_i and P_i can give rise to quite different conclusions.

In doing so, it is important to point out that these studies have been selected based on the availability of the data and not based on their results. Thus, we hypothesize that our critique should apply in general, and not only to the studies considered here. We would also like to stress that the goal here is to examine the robustness of the tests when using actual data, and not to suggest that the authors of the empirical studies have been in any way strategic in their choice of lag length or bandwidth. The results are summarized in Table 2.

The data have been obtained directly from the authors of the studies, and have been processed in accordance with their instructions. The interested reader is referred to the individual papers for further details. Also, although most data sets could be used directly, in a few cases we found that there were some observations missing. In these cases, for simplicity and convenience of comparison, the data sets were excluded.

As explained briefly in the previous section, in the original published papers, the tests and adjustment methods employed to test for cointegration vary, and are rarely explained in detail. In particular, with exception of Pedroni (2004), there is no mentioning about the choice of lag

[8] The idea here is that instead of using a coefficient or t-ratio type test, which makes it necessary to adjust for serial correlation, one looks at the ratio of two sample quantities that share the same nuisance parameter reflecting the serial correlation of the data. Because this parameter is the same for both numerator and denominator it cancels out in the limit, thus making the resulting test asymptotically independent of the serial correlation.

length and bandwidth. Replicating the original published results is therefore very difficult, and thus not attempted. Instead, we impose a unifying approach and use the same battery of tests on all the data sets.

Table 3 summarizes the findings from this empirical exercise. The first column shows the deterministic component of each regression, whereas the third column indicates the type of model that is being estimated. A brief explanation of these models can be found in Table 2. The remaining columns contain the p-values from the various tests, which are again summarized by means of the intervals that were generated by considering all lag lengths and bandwidths between one and 10.

Table 3 points to several interesting results that are worthy of further discussion. Firstly, in accordance with our simulations, the results obtained by using the parametric tests of Pedroni (1999, 2004) vary significantly in all cases considered, and do not appear to be particularly robust with respect to the choice of lag length. The semiparametric tests are more stable and lead to more unambiguous test results. Of course, since the null is almost always rejected, the question is whether these results reflect the actual data generating process or the oversize effect documented in the simulations.

Secondly, given this ambiguity, it is interesting to consider the results obtained by applying the tests of Westerlund (2005). In particular, it is interesting to see whether these tests lead to the same conclusions as those drawn in the original published papers. The results reported in Table 3 suggest that in a majority of cases, V_g and V_p result in a rejection of the no cointegration null, which is consistent with the results obtained in these studies. On the other hand, for the Bahmani-Oskooee, Miteza and Nasir (2002), Harb (2004) and Jenkins and Snaith (2005) data sets, we see that the null hypothesis is usually not rejected, which do not agree with the results provided in these studies.

5 Conclusions

In applied work with nonstationary panel data, researchers often face dilemma about which type of serial correlation adjustment to use in cointegration testing. This can basically be achieved in two ways, each with its own set of truncation problems. In the case of parametric testing, the problem is how to choose the appropriate lag order to use in the autoregressive test regression, while, in the case of semiparametric testing, the problem is how to choose the best bandwidth window.

This issue is very important because different choices of adjustment methods can often lead to significant variations in test outcome, and thus also in the conclusions based upon them.

In particular, since the analysis of most studies critically rely on showing cointegration, such variations could potentially undermine the analysis of the whole study.

In this paper, we systematically analyze this dilemma. As an example, we consider the very popular tests developed by Pedroni (1999, 2004), which all require some kind of adjustment to account for the serial correlation of the data. The properties of these tests are analyzed using both simulated and empirical data. Our main finding is that the choice of adjustment method matters, and that the results can be deceptive unless this choice is exactly right. As a solution to this problem, we suggest using the tests of Westerlund (2005), which do not require this kind of adjustment.

Table 1: Size at the 5% level for different lag lengths and bandwidths.

						$\widetilde{Z}_t$		$\widetilde{Z}_\rho$		$\widetilde{Z}_t^*$		Z_t		Z_ρ		Z_v		Z_t^*	
Determ	ϕ	N	T	V_g	V_p	min	max	min	max	min	max	min	max	min	max	min	max	min	max
Const	0	10	50	4.0	3.8	30.5	37.4	12.3	20.6	1.0	26.6	16.4	19.7	13.8	13.8	9.3	11.7	1.9	29.1
		10	100	3.8	4.7	15.3	24.2	10.6	17.8	1.6	14.0	10.9	11.9	11.3	11.3	12.3	13.6	3.0	18.7
		20	50	5.4	6.2	46.9	58.8	17.2	30.6	0.9	40.7	19.2	23.8	15.5	15.9	8.2	12.0	0.8	33.3
		20	100	4.9	6.6	22.1	36.0	12.3	26.5	1.0	18.8	11.7	13.3	10.4	10.4	10.3	10.8	1.1	19.3
	0.3	10	50	2.2	2.5	3.8	6.3	0.2	0.7	1.1	5.2	2.0	2.9	0.4	0.4	0.4	1.4	1.9	8.3
		10	100	2.4	3.4	1.4	3.8	0.2	1.3	0.8	6.8	1.3	1.5	0.5	0.5	0.4	3.5	2.2	8.9
		20	50	2.2	3.7	2.5	6.3	0.0	0.4	0.7	5.9	1.2	1.7	0.2	0.2	0.2	0.7	0.2	6.0
		20	100	3.1	4.4	0.9	4.3	0.2	1.0	0.0	4.6	0.3	0.7	0.1	0.1	0.1	1.5	1.0	6.7
	-0.3	10	50	6.8	6.8	82.0	94.0	63.6	90.9	1.1	89.8	70.0	73.4	74.0	74.4	46.7	59.3	2.0	82.8
		10	100	5.6	6.3	60.7	84.4	54.5	86.2	1.7	82.1	58.5	62.1	69.4	69.6	37.5	47.9	3.0	72.7
		20	50	10.6	10.6	96.8	99.6	87.4	99.1	0.9	98.9	82.7	89.9	89.9	89.9	62.4	74.4	0.9	93.1
		20	100	7.1	7.9	84.4	97.5	78.8	97.8	1.0	97.5	76.2	80.8	85.8	86.0	51.1	65.5	1.3	86.5
Trend	0	10	50	4.9	4.1	45.1	55.8	7.9	18.9	0.7	40.1	29.6	33.2	13.3	13.7	4.0	8.2	1.5	48.3
		10	100	3.3	4.2	27.6	41.9	11.7	23.3	1.3	22.1	18.6	20.3	13.3	13.4	6.7	12.0	3.8	34.4
		20	50	6.0	7.2	67.3	81.0	8.9	30.7	0.4	57.4	32.8	45.7	16.9	17.4	3.1	9.9	0.5	58.6
		20	100	5.7	6.0	37.9	60.6	14.0	34.4	0.7	30.7	18.6	24.3	14.7	14.7	6.7	11.7	1.6	34.6
	0.3	10	50	1.4	1.3	1.3	5.2	0.1	0.3	0.7	7.1	0.6	1.3	0.0	0.0	0.0	0.6	1.5	9.8
		10	100	1.3	1.8	1.2	3.2	0.0	0.6	0.2	6.4	0.3	0.7	0.0	0.0	0.1	0.8	1.5	13.4
		20	50	0.6	1.6	1.0	6.8	0.0	0.0	0.0	7.9	0.2	0.6	0.0	0.0	0.0	0.3	0.3	8.2
		20	100	2.0	2.6	0.3	2.1	0.0	0.1	0.0	5.6	0.1	0.2	0.0	0.0	0.0	0.5	0.2	8.4
	-0.3	10	50	14.1	12.8	97.8	99.7	78.9	97.4	0.6	99.3	91.6	95.9	91.4	91.5	49.7	68.2	1.8	98.9
		10	100	8.3	8.3	89.9	98.7	79.1	98.0	1.5	97.6	91.3	94.2	93.2	93.2	52.5	66.9	3.9	97.2
		20	50	23.5	20.5	99.9	100.0	96.6	100.0	0.4	100.0	97.9	99.7	99.4	99.4	72.5	88.6	0.6	99.9
		20	100	11.8	10.1	99.1	100.0	94.4	100.0	0.6	100.0	98.9	99.7	99.3	99.3	69.0	84.8	1.6	99.7

Notes: ϕ refers to the autoregressive parameter and measures the serial correlation in the regression errors. The intervals for the Pedroni (1999, 2004) tests in the seven rightmost panels have been generated by allowing the lag lengths and bandwidths to vary between one and 10. A tilde indicates that the test is of group mean type, while no tilde indicates that the test is of panel type. All tests are semiparametric except for those that are star superscripted, which are parametric. The Westerlund (2005) V_g and V_p tests do not require any choice of lag length or bandwidth, and the results are therefore reported by a single value.

Table 2: Summary of the empirical studies that use the tests of Pedroni (1999, 2004)

Study	Sample	Objective and model specification	Findings
Bahmani-Oskooee, Miteza and Nasir (2002)	Annual data for 49 countries, 1973 to 1990.	A bivariate model for the relation between black market and official exchange rates was considered. The deterministic component includes a constant or a constant and trend.	All models were found to be cointegrated.
Camarero and Tamarit (2002)	Annual data for 10 European countries, 1973 to 1992.	Five models were employed to analyze the determinants of the bilateral real exchange rate between Spanish peseta and nine European trading partners. All models were fitted with a constant only.	All five models are found to be cointegrated, but the evidence was strongest for the one with the real interest rate differential and productivity as determinants.
Christopoulos, Loizides and Tsionas (2005)	Annual data for 10 European countries, 1961 to 1999.	A bivariate model for the relation between government size and the unemployment rate. Five specifications were tested to identify the direction of the causality between the variables. All specifications included a constant.	The null of no cointegration can only be rejected when the unemployment rate was considered as the dependent variable.
Edmond (2001)	Annual data for 22 industrial countries, 1971 to 1990.	Uses a trivariate model to test whether investments in R&D at home and abroad has spillover effects on total factor productivity. Two different specifications with a constant were considered.	The null hypothesis of no cointegration can be rejected in both specifications.
Harb (2004)	Annual data for six Gulf nations, 1979 to 2000.	Money demand is modelled as a function of both real GDP and real private consumption. The model is fitted with a constant or with a constant and trend.	Cointegration is found when the trend is not included.
Jenkins and Snaith (2005)	Monthly data for 11 countries, 1981M1 to 1995M6.	Use 25 subindices to test the weak version of the purchasing power parity hypothesis. The regression is fitted with a constant only.	The evidence of cointegration is weaker for nontraded goods.
Lee (2005)	Annual data for 18 developing countries, 1975 to 2001.	Test the direction of causality between energy consumption and GDP. The deterministic component includes a constant or a constant and common time effects.	Cointegration is found in both specifications. The causality seem to be running from energy to GDP.
Pedroni (2004)	Monthly data for 20 countries, 1973M6 to 1994M12.	Reevaluates the weak version of the purchasing power parity hypothesis. The regression is fitted with a constant or a constant and common time effects.	The null hypothesis of no cointegration can be rejected in both specifications.
Sarantis and Stewart (2001)	Annual data for 20 OECD countries, 1955 to 1994.	Examine the determinants of aggregate private savings using a modified version of the life-cycle model of Modigliani. The regression id fitted with a constant only.	Strong support in favor of cointegration is found.

Table 3: Panel cointegration test results for different lag lengths and bandwidths.

				$\widetilde{Z}_t$		$\widetilde{Z}_\rho$		$\widetilde{Z}_t^*$		Z_t		Z_ρ		Z_v		Z_t^*	
Determ	Model	V_g	V_p	min	max	min	max	min	max	min	max	min	max	min	max	min	max
								Bahmani-Oskooee, Miteza and Nasir (2002)									
Const		0.11	0.22	0.00	0.00	0.00	0.03	0.00	0.83	0.00	0.00	0.00	0.00	0.00	0.19	0.00	0.00
Trend		0.02	0.41	0.00	0.00	0.01	0.19	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.18	0.00	0.00
								Camarero and Tamarit (2002)									
Const	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.97	0.00	1.00
	2	0.00	0.01	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.35	1.00	0.00	1.00
	3	0.00	0.01	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00	0.00	1.00
	4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99
	5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00
								Christopoulos, Loizides and Tsionas (2005)									
Const	u	0.03	0.07	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.25	1.00	0.00	0.98
	g	0.00	0.03	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.02	1.00	0.00	0.97
	y	0.00	0.02	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.99
	$\dot{p}$	1.00	0.66	0.00	0.00	0.00	0.06	0.00	1.00	0.00	0.79	0.00	0.00	0.03	0.52	0.00	0.94
	pop	0.00	0.01	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00
								Edmond (2001)									
Const	(i)	0.00	0.01	0.09	0.81	0.83	1.00	0.00	0.96	0.15	0.71	0.41	0.45	0.14	1.00	0.00	0.92
	(iii)	0.00	0.01	0.06	0.72	0.78	1.00	0.00	0.94	0.17	0.84	0.40	0.45	0.09	1.00	0.00	0.88
								Harb (2004)									
Const	GDP	0.11	0.17	0.00	0.00	0.51	0.93	0.00	1.00	0.06	0.48	0.17	0.19	0.77	0.99	0.00	0.97
Trend		0.77	0.34	0.00	0.03	0.83	0.99	0.00	1.00	0.23	0.76	0.52	0.55	0.96	1.00	0.00	1.00
Const	Cons	0.11	0.19	0.00	0.01	0.52	0.97	0.00	0.99	0.20	0.64	0.21	0.23	0.78	0.99	0.00	0.99
Trend		0.80	0.57	0.08	0.15	0.90	1.00	0.00	1.00	0.64	0.97	0.65	0.69	0.99	1.00	0.00	1.00

Continued overleaf

Table 3: Continued.

				$\widetilde{Z}_t$		$\widetilde{Z}_\rho$		$\widetilde{Z}_t^*$		Z_t		Z_ρ		Z_v		Z_t^*	
Determ	Model	V_g	V_p	min	max	min	max	min	max	min	max	min	max	min	max	min	max
								Jenkins and Snaith (2005)									
Const	1000	0.12	0.06	0.00	0.00	0.00	0.00	0.00	0.60	0.00	0.00	0.00	0.00	0.60	0.87	0.00	0.27
	1110	0.24	0.11	0.00	0.00	0.00	0.00	0.00	0.74	0.00	0.00	0.00	0.00	0.67	0.90	0.00	0.40
	1111	0.07	0.04	0.00	0.00	0.00	0.00	0.00	0.48	0.00	0.00	0.00	0.00	0.55	0.85	0.00	0.27
	1112	0.27	0.12	0.00	0.00	0.00	0.00	0.00	0.74	0.00	0.00	0.00	0.00	0.68	0.90	0.00	0.38
	1114	0.28	0.13	0.00	0.00	0.00	0.00	0.00	0.86	0.00	0.00	0.00	0.00	0.69	0.90	0.00	0.46
	1116	0.27	0.13	0.00	0.00	0.00	0.00	0.00	0.77	0.00	0.00	0.00	0.00	0.69	0.90	0.00	0.56
	1150	0.19	0.09	0.00	0.00	0.00	0.00	0.00	0.74	0.00	0.00	0.00	0.00	0.65	0.89	0.00	0.46
	1160	0.13	0.06	0.00	0.00	0.00	0.00	0.00	0.70	0.00	0.00	0.00	0.00	0.61	0.87	0.00	0.32
	1200	0.18	0.09	0.00	0.00	0.00	0.00	0.00	0.83	0.00	0.00	0.00	0.00	0.65	0.89	0.00	0.44
	1210	0.19	0.10	0.00	0.00	0.00	0.00	0.00	0.85	0.00	0.00	0.00	0.00	0.65	0.89	0.00	0.48
	1220	0.21	0.11	0.00	0.00	0.00	0.00	0.00	0.85	0.00	0.00	0.00	0.00	0.66	0.90	0.00	0.44
	1300	0.22	0.10	0.00	0.00	0.00	0.00	0.00	0.87	0.00	0.00	0.00	0.00	0.67	0.89	0.00	0.54
	1330	0.98	0.82	0.29	0.69	0.19	0.78	0.14	1.00	0.00	0.00	0.00	0.00	0.88	0.96	0.00	0.76
	1600	0.33	0.15	0.00	0.00	0.00	0.00	0.00	0.80	0.00	0.00	0.00	0.00	0.71	0.91	0.00	0.44
	1630	0.07	0.03	0.00	0.00	0.00	0.00	0.00	0.60	0.00	0.00	0.00	0.00	0.52	0.83	0.00	0.29
	1640	0.67	0.37	0.00	0.00	0.00	0.00	0.00	0.95	0.00	0.00	0.00	0.00	0.79	0.94	0.00	0.69
	1700	0.14	0.07	0.00	0.00	0.00	0.00	0.00	0.72	0.00	0.00	0.00	0.00	0.61	0.87	0.00	0.38
	1730	0.04	0.03	0.00	0.00	0.00	0.00	0.00	0.50	0.00	0.00	0.00	0.00	0.51	0.83	0.00	0.24
	1800	0.03	0.02	0.00	0.00	0.00	0.00	0.00	0.41	0.00	0.00	0.00	0.00	0.45	0.80	0.00	0.15
								Lee (2005)									
Const		0.04	0.02	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.96	0.00	1.00
Const, CTE		0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.77	0.00	0.97
								Pedroni (2004)									
Const		0.00	0.00	0.71	0.87	0.83	0.93	0.54	0.98	0.27	0.99	0.03	0.03	0.00	0.01	0.00	0.74
Const, CTE		0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.57	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.50
								Sarantis and Stewart (2001)									
Const		0.02	0.09	0.23	0.96	0.97	1.00	0.29	1.00	0.64	1.00	0.01	0.01	0.99	1.00	0.16	1.00

Notes: The values reported in the table are the asymptotic normal p-values. The intervals have been generated by allowing the lag lengths and bandwidths to vary between one and 10. The abbreviation CTE refers to the common time effects specification. See Table 2 for a summary of the various studies. See Table 1 for an explanation of the various tests. For further information regarding the various models, we make reference to the original studies to which they belong.

References

Bahmani-Oskooee, M., Miteza, I., Nasir, A. B. M. (2002). The long-run relation between black market and official exchange rates: evidence from panel cointegration. *Economics Letters* 76, 397-404.

Banerjee, A., Marcellino, M., Osbat, C. (2004). Some cautions on the use of panel methods for integrated series of macroeconomic data. *Econometric Journal* 7, 322-340.

Breitung, J., Pesaran, M. H. (2007). Unit roots and cointegration in panels. Forthcoming in Matyas, L., Sevestre, P. (Eds.), *The Econometrics of Panel Data*, Klüver Academic Press.

Camarero, M., Tamarit, C. (2002). A panel cointegration approach to the estimation of the peseta real exchange rate. *Journal of Macroeconomics* 24, 371-393.

Christopoulos, D. K., Loizides, J., Tsionas, E. G. (2005). The Abrams curve of government size and unemployment: evidence from panel data. *Applied Economics* 37, 11931199.

Edmond, C. (2001). Some panel cointegration models of International R&D spillovers. *Journal of Macroeconomics* 23, 241-260.

Engle, R., Granger, C. (1987). Cointegration and error correction: representation, estimation and testing. *Econometrica* 55, 251-276.

Groen, J. J. J., Kleibergen, F. (2003). Likelihood-based cointegration analysis in panels of vector error correction models. *Journal of Business and Economic Statistics* 21, 295-318.

Gutierrez, L. (2003). On the power of panel cointegration tests: a Monte Carlo comparison. *Economics Letters* 80, 105-111.

Harb, N. (2004). Money demand function: a heterogenous panel application. *Applied Economics Letters* 11, 551-555.

Jenkins, M. A., Snaith, S. M. (2005). Tests of purchasing power parity via cointegration analysis of heterogenous panels with consumer price indices. *Journal of Macroeconomics* 27, 345-362.

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

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

Kao, C. (1999). Spurious regression and residual-based tests for cointegration in panel data. *Journal of Econometrics* 90, 1-44.

Larsson, R., Lyhagen, J., Lötgren, M. (2001). Likelihood-based cointegration test in heterogeneous panels. *Econometrics Journal* 4, 109-142.

Lee, C-C. (2005). Energy consumption and GDP in developing countries: a cointegrated panel analysis. *Energy Economics* 27, 415 427.

Pedroni, P. (1999). Critical values for cointegration tests in heterogeneous panels with multiple regressors. *Oxford Bulletin of Economics and Statistics* 61, 653670.

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, 597-625.

Sarantis, N., Stewart, C. (2001). Saving behavior in OECD countries: evidence from panel cointegration tests. *The Manchester School Supplement*, 22-41.

Westerlund, J. (2005). New simple tests for panel cointegration. *Econometric Reviews* 24, 297-316.