

Some cautions on the use of panel methods for integrated series of macroeconomic data

ANINDYA BANERJEE[†], MASSIMILIANO MARCELLINO[‡]
AND CHIARA OSBAT[§]

[†]*Department of Economics, European University Institute, 50133 Florence, Italy*
E-mail: anindya.banerjee@iue.it

[‡]*IEP—Bocconi University, IGIER and CEPR, 20136 Milan, Italy*

[§]*Department of Economics, European University Institute, 50133 Florence, Italy; European Central Bank*

Received: June 2003

Summary Existing panel cointegration tests rule out cross-unit cointegrating relationships, while economic theory and empirical observation argue strongly in favour of their presence. Using an extensive set of simulation experiments, we show that both univariate and multivariate panel cointegration tests can be substantially oversized in the presence of cross-unit cointegration. We also propose a test for cross-unit cointegration that performs well in practice and can be used to decide upon the usefulness of panel methods.

Keywords: *Panel data, Cointegration, Unit roots.*

1. INTRODUCTION

In recent years, the use of panel data techniques to test macroeconomic hypotheses on integrated data has become increasingly common. Examples of empirical applications of panel cointegration include testing for cross-country convergence in per capita GDP, testing for purchasing power parity (PPP) and investigating the Balassa–Samuelson effect.¹ The argument presented in favour of panel data methods is that they have greater power than standard unit root and cointegration tests, by virtue of the reduction in noise caused by the averaging or pooling across the units of the panel.

The emphasis in the theoretical literature has primarily taken the form of considering the asymptotic properties of panel data estimators and test statistics as T , the length of the time series,

¹On growth and convergence, see e.g. Evans and Karras (1996) and Lee et al. (1997). Panel studies of PPP include Frankel and Rose (1996), Papell (1997) and Chiu (2004). Lyhagen (2000) cautions against the use of panel unit root tests in testing PPP on the grounds that the distributions of these tests are influenced by a common stochastic trend which is not accounted for in their construction. For studies of the Balassa–Samuelson hypothesis, see Drine and Rault (2003) and Egert (2002). Other relevant papers include Johansen (2002) and Song and Wu (1997) who study hysteresis in unemployment in Norwegian counties and US states respectively, Pesaran and Smith (1995) who analyse UK labour demand functions for 38 industries over 30 years, and Dembla (2000) who estimates production functions in the Indian manufacturing sector.

and N , the number of units in the panel, go to infinity, possibly at rates of divergence that are controlled appropriately, as discussed in McCoskey and Kao (1998), Kao (1999) and Pedroni (1999, 2000, 2001), *inter alia*. It has been shown that an added advantage of some of these test statistics is that under the null hypothesis they have Gaussian distributions asymptotically. Inference in large samples is therefore, in some sense, made ‘standard’.

The analyses mentioned above assume a unique cointegrating vector in each unit, either homogeneous (Kao 1999) or heterogeneous (Pedroni 1999) across the units of the panel. In the context of the analysis of multivariate data sets, however, it is natural to focus on methods which relax the assumption of a unique cointegrating vector. Groen and Kleibergen (2003), Larsson and Lyhagen (2000a, 2000b) and Larsson et al. (2001) have therefore developed techniques, analogous to the Johansen (1995) maximum-likelihood method, that allow for multiple cointegrating vectors in each unit. The maximum-likelihood-based panel cointegration methods allow for cross-unit dependence through the effects of the dynamics of the short run and, in some cases, through contemporaneous error correlations *via* generalized least squares (GLS) methods such as those used by O’Connell (1998) in a single-equation framework. Yet, no account is taken of the possibility of long-run cross-unit dependence induced by the existence of cross-unit cointegrating relationships.

In this paper we show, through detailed Monte Carlo simulations, that the consequences of using panel cointegration methods when the restriction of no cross-unit cointegration is violated are dramatic. Since in many empirical applications this restriction is likely to be violated because of economic links across regions and units,² we argue that many of the conclusions in the empirical literature may be based upon misleading inference.

On the other hand, we also confirm that there are important gains in efficiency when the use of the panel approach is justified. Thus, we suggest testing for the validity of the no cross-unit cointegration hypothesis prior to applying panel cointegration methods. Specifically, we propose the extraction of the common trends from each unit using the Johansen maximum-likelihood method, and then testing for cointegration among these trends to rule out the existence of cross-unit cointegration. Several simulation experiments indicate that this procedure performs well in practice.

The paper is organised as follows. Section 2 briefly reviews the panel multivariate cointegration tests. In Section 3, we evaluate their finite sample properties by means of an extensive simulation study.³ Panel univariate cointegration tests are considered in Section 4. Section 5 summarizes the main findings of the paper and provides suggestions for further theoretical developments.

2. PANEL MULTIVARIATE COINTEGRATION TESTS

The most general panel cointegration model, as presented in Larsson and Lyhagen (2000a) (henceforth LL), corresponds to a restricted cointegrated vector autoregression (VAR). The model under consideration is

$$\Delta X_t = \alpha\beta' X_{t-1} + \sum_{k=1}^{m-1} \Gamma_k \Delta X_{t-k} + \varepsilon_t, \quad (1)$$

²See Banerjee et al. (2004) for evidence of cointegration of real exchange rates.

³Our routines were written and compiled in Ox (Doornik 2001). The usual disclaimer applies.

where the vector X_t of dimension $Np \times 1$ is given by stacking the N vectors

$$X_{i,t} = (x_{i,1,t}, x_{i,2,t}, \dots, x_{i,p,t})', \quad i = 1, 2, \dots, N, \quad t = 1, 2, \dots, T.$$

The number of units composing the panel is denoted by N , p is the dimension of each unit and T is the time dimension. Hence $x_{i,p,t}$ denotes a univariate observation of the process. Analogously,

$$\varepsilon_{i,t} = (\varepsilon_{i,1,t}, \varepsilon_{i,2,t}, \dots, \varepsilon_{i,p,t})', \quad i = 1, 2, \dots, N, \quad t = 1, 2, \dots, T,$$

and the $Np \times 1$ dimensional vector ε_t is given by stacking the N vectors $\varepsilon_{i,t}$ (each of dimension $p \times 1$). The error process ε_t has a multivariate normal distribution with zero mean and covariance matrix of dimension $Np \times Np$ given by

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1N} \\ \Sigma_{21} & \Sigma_{22} & \cdots & \Sigma_{2N} \\ \vdots & & \ddots & \vdots \\ \Sigma_{N1} & \Sigma_{N2} & \cdots & \Sigma_{NN} \end{pmatrix},$$

where each of the individual Σ_{ij} matrices (denoting the covariance matrix of $\varepsilon_{i,t}$ with $\varepsilon_{j,t}$) is of dimension $p \times p$. The matrices of fixed coefficients Γ_k have dimension $Np \times Np$ while the matrices α and β have dimension $Np \times \sum_{i=1}^N r_i$, where r_i is the rank in each unit.⁴ α and β have the following structure:

$$\alpha = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1N} \\ \alpha_{21} & \alpha_{22} & & \\ \vdots & & \ddots & \vdots \\ \alpha_{N1} & \alpha_{N2} & \cdots & \alpha_{NN} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_{11} & 0 & \cdots & 0 \\ 0 & \beta_{22} & & \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \beta_{NN} \end{pmatrix}.$$

The model therefore allows for interaction among the units through the long-run adjustment coefficients α , the short-run coefficients Γ_k , and the off-diagonal elements Σ_{ij} of Σ , but the restriction $\beta_{ij} = 0 \forall i \neq j$ rules out cointegrating relationships across the units. A further restriction is that r_i is assumed to be the same for each unit, i.e. there is a common maximum rank. These two restrictions are the most important issues to be investigated, since they may be inconsistent with both theory and data in most empirical applications.⁵

The estimation algorithm for the individual cointegrating relations is a series of reduced rank regressions in which each β_{ii} is estimated by concentrating out the remaining $N - 1$ matrices in the β matrix. Hence, β_{11} to β_{NN} are estimated at each iteration, and the procedure is repeated until convergence.⁶

The panel trace test for cointegrating rank of $r_i = q$ versus $r_i = p$, for each i and for $q = 0, \dots, p - 1$, is derived in LL. Since the rank is tested under the block-diagonality restriction

⁴The notation is very similar to the standard one used in the literature on cointegrated VAR models; see e.g. Johansen (1995, p. 39).

⁵Note that in practice, as N grows for given T , the estimation of this model becomes infeasible, since only some degrees of freedom are saved, with respect to an unrestricted VAR, by restricting the parameters in the off-diagonal terms of β to zero.

⁶For starting values, LL propose using the β_{ii} estimated from a standard cointegration analysis on each unit separately. We instead use the initial values suggested in Johansen (1995, p. 110).

on β , the asymptotic distribution of the panel trace test statistic (henceforth LL test statistic) is a convolution of the distribution of the standard Johansen trace test statistic (henceforth Johansen test statistic) and an independent χ^2 random variable with $N(N-1)(p-q)q$ degrees of freedom. This implies that the panel trace test for $r_i = 0$ is the same as the Johansen test, while for $r_i > 0$ there is an additional component in the distribution of the panel trace statistic, which accounts for the additional zero restrictions imposed on β . Note that testing for $r_i = q$ versus $r_i = p$ in this framework corresponds to testing for rank Nq versus Np in a Johansen context, whereas testing for rank in the Johansen framework would allow for all the intermediate possibilities. For example, if $N = p = 2$, in the framework of LL one would test for 0 versus 4 followed by 2 versus 4, while in the Johansen framework one would test for 0 versus 4, 1 versus 4 and so on.

3. SIMULATION STUDY

In this section, we evaluate by simulation the finite sample properties of the LL test statistic. We first consider the case where the hypothesis of no cross-unit cointegration is satisfied. We next evaluate the consequences of the violation of this assumption, and lastly analyse the role of the panel dimension.

3.1. Results with block-diagonal β matrix

Table 1 lists the data generation processes (DGPs) used in the Monte Carlo experiments. The DGPs considered in this section are characterized by $N = 2$, a block-diagonal β matrix, the same cointegrating rank in all units and, for simplicity, homogeneous cointegration vectors.⁷

More specifically, DGP 1 is the simplest null, with rank zero in both units. DGP 2A has rank 1 in both units, with the block-diagonal loading matrix α constructed such that each of the two equilibrium correction terms enters only one equation in its own unit. In DGP 2B, with rank 1 in both units and non-block-diagonal α , the first equilibrium correction term enters one equation in each of the two units, whereas the second equilibrium correction term enters only one equation of the second unit. DGP 2C is a variant of DGP 2A with lower stationary eigenvalues (equal to 0.8, versus 0.9 in DGP 2A).⁸

In Table 2 we report the size and power of the LL test statistic and of the Johansen test statistic performed unit by unit, jointly, and in full systems. Thus the column headed LL gives the rejection frequencies of $H_0: r_i = r_{LL} = 0$, $i = 1, 2$, at the 5% significance level when the LL test statistic is calculated for DGP 1 and LL critical values are used. For DGP 2A to DGP 2C the rejection frequencies of $H_0: r_i = r_{LL} = 1$ at the 5% significance level is given, again with LL critical values.

⁷We do not present here the results for DGPs with deterministic components such as constants or trends or break dummies, complications that are very likely to occur in practice. In this latter case, provided these deterministic variables enter into the system unrestrictedly, we may consider our analysis as a proxy for working with detrended or demeaned data.

⁸Other experiments with different configurations for α yielded qualitatively similar results, while with smaller stationary roots the statistics have better size and higher power.

Table 1. Description of data generating processes.

DGP	r_{LL}	r_J	α'	β'	Roots
<i>Block-diagonal β</i>					
1	0	0	0	0	(1,1,1,1)
2A	1	2	$\begin{pmatrix} -0.1 & 0 & 0 & 0 \\ 0 & 0 & -0.1 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}$	(1,1,0.9,0.9)
2B	1	2	$\begin{pmatrix} -0.1 & 0 & 0.1 & 0 \\ 0 & 0 & -0.1 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}$	(1,1,0.9,0.9)
2C	1	2	$\begin{pmatrix} -0.2 & 0 & 0 & 0 \\ 0 & 0 & -0.2 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}$	(1,1,0.8,0.8)
<i>Non-block-diagonal β</i>					
3A	0	1	(-0.1 0 0 0)	(1 0 -1 0)	(1,1,1,0.9)
3B	0	1	(-0.1 -0.1 0 0)	(1 0 -1 0)	(1,1,1,0.9)
3C	0	1	(-0.1 0 0 -0.1)	(1 0 -1 0)	(1,1,1,0.9)
3D	0	1	(-0.2 0 0 0)	(1 0 -1 0)	(1,1,1,0.8)
4A	0	2	$\begin{pmatrix} 0 & -0.1 & 0 & 0 \\ -0.1 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 0 & -1 \\ 1 & 0 & -1 & 0 \end{pmatrix}$	(1,1,0.9,0.9)
4B	0	2	$\begin{pmatrix} 0 & -0.1 & 0 & 0 \\ -0.1 & -0.1 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 0 & -1 \\ 1 & 0 & -1 & 0 \end{pmatrix}$	(1,1,0.9,0.9)
4C	0	2	$\begin{pmatrix} 0 & -0.1 & 0 & 0 \\ -0.1 & 0 & -0.1 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 0 & -1 \\ 1 & 0 & -1 & 0 \end{pmatrix}$	(1,1,0.9,0.9)
4D	0	2	$\begin{pmatrix} 0 & -0.2 & 0 & 0 \\ -0.2 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 0 & -1 \\ 1 & 0 & -1 & 0 \end{pmatrix}$	(1,1,0.8,0.8)
5A	1	3	$\begin{pmatrix} -0.1 & 0 & 0 & 0 \\ 0 & 0 & -0.1 & 0 \\ 0 & 0 & 0 & 0.1 \end{pmatrix}$	$\begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 1 & 0 & -1 \end{pmatrix}$	(1,0.9,0.9,0.9)
5B	1	3	$\begin{pmatrix} -0.1 & 0 & 0 & -0.1 \\ 0 & 0 & -0.1 & 0 \\ 0 & 0 & 0 & 0.1 \end{pmatrix}$	$\begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 1 & 0 & -1 \end{pmatrix}$	(1,0.9,0.9,0.9)
5C	1	3	$\begin{pmatrix} -0.1 & 0 & 0 & -0.1 \\ 0 & 0 & -0.1 & 0 \\ 0 & 0 & -0.1 & 0.1 \end{pmatrix}$	$\begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 1 & 0 & -1 \end{pmatrix}$	(1,0.9,0.9,0.9)
5D	1	3	$\begin{pmatrix} -0.2 & 0 & 0 & 0 \\ 0 & 0 & -0.2 & 0 \\ 0 & 0 & 0 & 0.2 \end{pmatrix}$	$\begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 1 & 0 & -1 \end{pmatrix}$	(1,0.8,0.8,0.8)

r_{LL} : Cointegrating rank for each unit in sense of Larsson and Lyhagen (2000a). r_J : Cointegrating rank for full system in sense of Johansen (1995).

Table 2. Size and power of multivariate tests with block diagonal β matrix.

DGP	T	LL	J unit-by-unit			J system	GG
			Unit 1	Unit 2	Joint		
<i>Size experiments</i>							
$H_0 : r_{LL} = 0 (r_J = 0)$							
1	100	0.064	0.048	0.058	0.056	0.065	0.051
	200	0.059	0.051	0.051	0.056	0.060	0.045
	400	0.057	0.052	0.051	0.052	0.057	0.045
$H_0 : r_{LL} = 1 (r_J = 2)$							
2A	100	0.098	0.048	0.049	0.815	0.022	0.021
	200	0.087	0.055	0.052	0.142	0.051	0.024
	400	0.069	0.051	0.052	0.051	0.056	0.028
2B	100	0.118	0.049	0.035	0.887	0.024	0.038
	200	0.093	0.055	0.046	0.384	0.045	0.033
	400	0.069	0.051	0.048	0.055	0.055	0.031
2C	100	0.098	0.054	0.054	0.136	0.057	0.027
	200	0.072	0.054	0.050	0.054	0.058	0.029
	400	0.060	0.050	0.053	0.050	0.053	0.031
<i>Power experiments $H_0 : r_{LL} = 0 (r_J = 0)$</i>							
2A	100	0.513	0.576	0.575	0.229	0.136	–
	200	0.978	0.982	0.985	0.914	0.635	–
	400	1.000	1.000	1.000	1.000	0.999	–
2B	100	0.699	0.576	0.367	0.151	0.165	–
	200	0.999	0.982	0.792	0.667	0.522	–
	400	1.000	1.000	0.998	0.995	0.974	–
2C	100	0.983	0.985	0.983	0.917	0.658	–
	200	1.000	1.000	1.000	1.000	0.999	–
	400	1.000	1.000	1.000	1.000	1.000	–

Note: The first column denotes the DGP as defined in Table 1. The second gives the length of the time series. The third reports rejection frequencies of the null hypothesis at 5% using the LL test statistic and Larsson and Lyhagen's (2000a) asymptotic critical values. The fourth and fifth columns report the rejection frequencies of the null hypothesis at 5% of unit-by-unit Johansen tests using Johansen (1995) asymptotic critical values. The sixth column reports the rejection frequencies of the null hypothesis using the joint test, with the significance level of the unit-by-unit Johansen tests set at $1 - (0.95)^{1/2}$. The seventh reports rejection frequencies at 5% for the full system using Johansen (1995) asymptotic critical values. The last column reports the rejection frequencies of cointegration among the common trends derived from the unit-by-unit cointegration analysis when the null hypothesis is accepted in each unit.

The columns headed 'unit 1' and 'unit 2' provide the corresponding rejection frequencies, for each unit individually, of the Johansen test statistic at the 5% significance level using Johansen (1995) critical values. The rejection frequencies in the column headed 'joint' are calculated by setting the significance level of the unit-by-unit tests at $1 - (0.95)^{1/2}$, so that the significance of the joint test (i.e. the probability of rejecting $r_i = 0$ in both units in DGP 1, and of rejecting $r_i = 1$ in both units in DGP 2A to 2C) is 5%. The column headed 'J system' gives the rejection frequencies

(of $r_J = 0$ for DGP 1 and $r_J = 2$ for DGPs 2A to 2C) when the full four-dimensional system is estimated without restriction (apart from those necessary for identification). Note that r_{LL} denotes the cointegrating rank in the LL sense, and r_J is the cointegrating rank for the full system in the sense of Johansen. The final column of Table 2 gives the rejection frequencies of cointegration among the common trends derived from the unit-by-unit cointegration analysis, along the lines of Gonzalo and Granger (1995), whenever the null hypothesis is accepted in each unit. In all data generation processes considered in Table 2, cross-unit cointegration is ruled out so that the figures in this column should be close to 0.05.⁹

The results on size, as reported in the first panel of Table 2, are encouraging. Except for the joint test, no substantial size distortions are evident. The distortions of the joint test disappear as the sample size increases to 400 but, particularly for DGP 2A and 2B, they are present and important at empirically relevant sample sizes of 100 (or 200, not reported here). The distortions increase with more feedback (compare DGP 2B with 2A) and are lower with lower stationary eigenvalues (compare DGP 2C with 2A).

The Gonzalo and Granger test is slightly undersized whenever cointegrating vectors are estimated, as in DGP 2A to 2C, even at large sample sizes of 400. This is likely to be a consequence of using critical values that are not strictly applicable with constructed, as opposed to raw, series. While it is possible in principle to adjust the critical values for size, we do not pursue this further as such corrections would be *ad hoc*, the size distortions are not great, and the general points can be made in the absence of such size corrections.

The results of the power experiments are reported in the second panel of Table 2. For DGP 2A to 2C the rejection frequencies of $r_{LL} = 0$ are given in the column headed LL, while the remaining columns provide the same information for the other tests. All the tests have power approaching 100% as sample size increases. Broadly speaking however, the LL test statistic has the best power properties for a majority of cases at all sample sizes. The test based on estimating the full Johansen system is the least powerful (although increasing rapidly to 1) for all but one of the cases. This is to be expected since, when the restriction of block-diagonality of β is satisfied, estimating the full system leads to a loss in efficiency and power. Estimating the system unit by unit is partly beneficial (by cutting down on the number of parameters to be estimated), although cross-unit links via the α matrix are not taken into account. The joint test therefore occupies the middle ground in terms of power performance.

In terms of bias of the estimates of the cointegrating vectors, even for $T = 100$, the LL procedure and the unit-by-unit Johansen procedure perform reasonably well.¹⁰ In terms of efficiency, the latter is slightly better than the former, in the sense that it yields lower standard errors for the estimated coefficients. The full-system cointegrated VAR ranks third both in terms of bias and efficiency.

3.2. Results with full β matrix

From results reported in the previous section, the panel trace test is seen to have the correct size and good power properties compared to other tests for cointegration. It also offers some efficiency

⁹It is worth noting that Gonzalo and Granger suggest that the asymptotic distribution of the cointegration test is not affected by having estimated the common trends in the first stage. Hence, the critical values used are taken from Johansen (1995).

¹⁰Detailed results are available in Table 4 of Appendix A of the working paper version of this paper (<http://www.iue.it/Personal/Banerjee/Welcome.html>).

gains when estimating the cointegrating coefficients. Yet, as we stress in Section 1, the framework proposed in LL (and in the more restrictive models in the panel unit root and cointegration literature) may often be inappropriate in applications with macroeconomic data, especially when there are cross-unit cointegrating relationships. This section evaluates the performance of the LL test statistic when such cross-unit cointegrating relationships exist, and considers some *ad hoc* ‘diagnostic tests’ that can be used to detect their presence.

When the variables in different units are related by cointegrating relationships, the hypothesis of a block-diagonal β is violated. Several structures for β are now possible, and we focus on three of these in our simulation experiments. In DGP 3 there exists only one cross-unit cointegrating relationship, so that $r_{LL} = 0$ and $r_J = 1$. DGP 3 comes in four ‘varieties’: DGPs 3A, 3B and 3C differ in their specification of the loading matrix α , while in 3D the stationary eigenvalues of the companion matrix are equal to 0.8, compare to 0.9 for the first three variants. DGP 4 considers the case where $r_{LL} = 0$ and $r_J = 2$, namely, two cross-unit cointegrating relationships and no within-unit cointegration. In DGP 5 we also allow for within-unit cointegration, and keep one cross-unit relationship, so that $r_{LL} = 1$ and $r_J = 3$. Several subcases of DGP 4 and 5 are analysed, which again differ in the structure of the α matrix (indicated by A,B,C) and the magnitude of the roots (D), as detailed in Table 1.

The rejection frequencies of the LL test statistic are reported in the third and fourth columns of Table 3. For DGP 3, where $r_{LL} = 0$, the probability of rejecting $r_{LL} = 0$ quickly increases towards 1, while that of rejecting $r_{LL} = 1$ is contained in the range 0.22–0.29 when $T = 400$. The probability of rejecting $r_{LL} = 1$ increases to one for DGP 4, and the same is true for DGP 5.

These results therefore indicate massive overrejection of the null hypothesis by the LL test statistic in the presence of cross-unit cointegration. Intuitively, cointegration across the units is wrongly attributed to cointegration within each unit. It is interesting to evaluate whether this is also a problem in the unit-by-unit analysis using the Johansen trace test. Indeed, the numbers reported in columns ‘unit 1’ and ‘unit 2’ of Table 3, show that there are cases when the size of the test is severely biased upwards. In comparison with LL however, even the highest rejection probabilities of $r_{LL} = 0$ are much smaller under this approach, rising to approximately 0.37 for DGPs 3 and 4 for $T = 400$. The distortion is related to the cointegrating relationships affecting several variables of the system (DGPs of types B and C). As a consequence, the joint test based on combining the unit-by-unit results also presents size distortions in these cases (column ‘joint’ of Table 3).¹¹ For DGP 5 the rejection of $r_J = 0$ soon rises to one while $r_J = 1$ is rejected, as expected, approximately 5% of the time in both units.

In light of these results, it seems particularly important to use either a full-system approach, or at a minimum to evaluate whether there is cross-unit cointegration. Since the former approach is not feasible with a larger number of units, consideration of the latter becomes even more important. The last three columns of Table 3 report the size and power of the Gonzalo and Granger test. The notation r_C is used to indicate that the rank of the system consisting of the *extracted* common trends is being estimated. The power of the test (columns $r_C = 0$ for all DGPs and also $r_C = 1$ for DGP 4) is in general rather low for $T = 100$, but quickly increases with T and is often close to one for $T = 400$. As in the previous section, the size (columns $r_C = 1$ for DGPs 3 and 5, and column $r_C = 2$ for DGP 4) is slightly lower than the nominal value even asymptotically ($T = 400$).

¹¹Note that the same random numbers are used for all experiments.

Table 3. Rejection frequencies of multivariate tests with non-block-diagonal β matrix.

DGP	T	J unit-by-unit										
		LL		Unit 1				Unit 2		GG		Joint
		$r_{LL} = 0$	$r_{LL} = 1$	$r_J = 0$	$r_J = 1$	$r_J = 0$	$r_J = 1$	$r_C = 0$	$r_C = 1$	$r_C = 2$		
3A	100	0.231	0.105	0.054	0.004	0.054	0.006	0.055	0.210	0.020	-	
	200	0.621	0.235	0.051	0.004	0.053	0.005	0.056	0.605	0.036	-	
	400	0.993	0.281	0.053	0.003	0.051	0.004	0.050	0.993	0.041	-	
3B	100	0.583	0.221	0.263	0.026	0.054	0.006	0.219	0.502	0.037	-	
	200	0.970	0.288	0.324	0.026	0.053	0.005	0.276	0.960	0.039	-	
	400	1.000	0.275	0.369	0.028	0.051	0.005	0.320	1.000	0.038	-	
3C	100	0.580	0.218	0.054	0.004	0.265	0.023	0.221	0.500	0.035	-	
	200	0.969	0.285	0.051	0.004	0.341	0.030	0.288	0.960	0.042	-	
	400	1.000	0.264	0.053	0.003	0.382	0.029	0.333	1.000	0.031	-	
3D	100	0.628	0.236	0.056	0.005	0.054	0.006	0.067	0.613	0.038	-	
	200	0.995	0.285	0.060	0.005	0.053	0.005	0.059	0.994	0.040	-	
	400	1.000	0.271	0.059	0.003	0.051	0.005	0.052	1.000	0.038	-	
4A	100	0.512	0.350	0.063	0.008	0.054	0.006	0.058	0.495	0.115	0.016	
	200	0.977	0.937	0.063	0.006	0.053	0.005	0.058	0.976	0.632	0.038	
	400	1.000	1.000	0.058	0.006	0.051	0.005	0.045	1.000	1.000	0.033	
4B	100	0.696	0.416	0.164	0.025	0.054	0.006	0.126	0.663	0.121	0.014	
	200	0.998	0.861	0.222	0.034	0.053	0.005	0.171	0.997	0.470	0.026	
	400	1.000	1.000	0.264	0.034	0.051	0.005	0.197	1.000	0.966	0.020	
4C	100	0.843	0.638	0.142	0.023	0.265	0.023	0.241	0.798	0.240	0.020	
	200	0.997	0.993	0.180	0.026	0.341	0.302	0.306	0.999	0.837	0.026	
	400	1.000	1.000	0.203	0.026	0.382	0.029	0.358	1.000	1.000	0.021	
4D	100	0.979	0.939	0.066	0.006	0.054	0.006	0.057	0.978	0.630	0.036	
	200	1.000	1.000	0.070	0.005	0.053	0.005	0.054	1.000	0.999	0.032	
	400	1.000	1.000	0.070	0.007	0.051	0.005	0.047	1.000	1.000	0.026	
5A	100	0.771	0.364	0.584	0.049	0.595	0.061	0.803	0.520	0.039	-	
	200	1.000	0.842	0.984	0.056	0.986	0.062	0.131	0.956	0.023	-	
	400	1.000	1.000	1.000	0.051	1.000	0.058	0.041	1.000	0.018	-	
5B	100	0.954	0.424	0.584	0.049	0.715	0.068	0.751	0.714	0.024	-	
	200	1.000	0.896	0.984	0.056	0.993	0.083	0.121	0.905	0.013	-	
	400	1.000	1.000	1.000	0.051	1.000	0.103	0.068	1.000	0.009	-	
5C	100	0.993	0.687	0.584	0.049	0.911	0.172	0.681	0.698	0.016	-	
	200	1.000	0.997	0.984	0.056	0.999	0.221	0.204	0.928	0.006	-	
	400	1.000	1.000	1.000	0.051	1.000	0.270	0.199	1.000	0.003	-	
5D	100	1.000	0.849	0.985	0.053	0.991	0.071	0.120	0.956	0.023	-	
	200	1.000	1.000	1.000	0.054	1.000	0.063	0.046	1.000	0.020	-	
	400	1.000	1.000	1.000	0.049	1.000	0.061	0.039	1.000	0.017	-	

Note: The first column denotes the DGP as defined in Table 1. The second gives the length of the time series. The columns labelled ‘LL’ report the rejection frequencies of $r_{LL} = 0$ or 1 at 5% using Larsson and Lyhagen’s (2000a) asymptotic critical values, when their test statistic is calculated. The columns labelled ‘J unit-by-unit’ report the rejection frequencies of unit-by-unit Johansen tests at 5% of rank 0 or 1, using Johansen (1995) asymptotic critical values. The column labelled ‘Joint’ reports the rejection frequencies of the null hypothesis using the joint test, where the significance level of the unit by unit Johansen tests is set at $1 - (0.95)^{1/2}$. The columns labelled ‘GG’ report the rejection frequencies of cointegration among common trends derived from unit-by-unit analysis when the null hypothesis is accepted in each unit.

Overall, these results suggest that whenever there exists cross-unit cointegration the LL test statistic and, to a lesser extent, the unit-by-unit analysis suffer from over-sizing. Thus it is important to emphasize that *under this setting, investigators are bound to take the wrong decision*: if they do not reject the null hypothesis of no cointegration, they will proceed under the assumption that $r = 0$, possibly estimating a model in differences, thus disregarding important long-run information. On the other hand, if they reject this null hypothesis, they will conclude that there is cointegration within each unit, while in fact the cointegrating relationships link variables belonging to different units.

However, since the presence of cross-unit cointegration is detected by the Gonzalo and Granger test, at least for large enough sample sizes, it is important to make use of this information when considering cointegration in panels.

Further simulations conducted on DGPs with different cointegrating ranks in each unit indicate that the LL test statistic tends to overreject the null hypothesis of no cointegration in each unit. In other words, cointegration in one unit biases the test towards the rejection of no cointegration in the other unit. Unit-by-unit analysis is accurate when T is large enough, and can provide a useful diagnostic test for different cointegrating ranks across units.

3.3. Results for higher panel dimensions

Since the case of $N = 2$ might be thought of as being unduly restrictive, in this section we summarize the results of simulations performed for $N = 4, 8$. Conventional panel studies often consider cases where the number of units is even larger, and a limitation of the highly parameterized maximum-likelihood methods is that they cannot be implemented unless T is very large with respect to N or the dependence across the units is severely restricted. Imposing these restrictions leads to the framework considered by Larsson and Lyhagen (2000a), or by Kao (1999) and Pedroni (1999, 2004) *inter alia* (the papers by Kao and Pedroni make the further assumption of one cointegrating vector).

The trade-off between higher dimensionality and *a priori* restrictions is an issue that merits further investigation. Nevertheless, a panel of 4 or 8 units is a reasonable size when considering, for example, data from the G7 countries, or from the largest economies of the European Union, or economic groupings such as NAFTA. The key economic indicators of the main sectors of an economy, of geographically differentiated regions within a country, or of the different segments of the labour market could also be investigated within this context.

In order to evaluate the size and power of the competing approaches, we make use of DGPs which replicate the structures used for $N = 2$, possibly with more than two cross-unit cointegrating vectors or different degrees of cointegration within each unit. A detailed account of these results (available upon request) is not presented here for space constraints. The following three important features are however worth noting:

(1) As the N panel dimension is increased, the size of the LL test statistic becomes distorted if asymptotic critical values are used. For example, when data are generated with $r_i = 1$ for each i , $N = 4$, $T = 200$, and the largest stationary roots are 0.8, the LL test statistic rejects the null with a frequency of 19% at a significance level of 5%. This distortion becomes very severe for $N = 8$, with the rejection frequency increasing to 90%. The distortion also increases with the magnitude of the largest stationary roots, if the other features of the DGP are left unchanged. Thus, with $r_i = 1$ for each i , $N = 4$, $T = 200$, and largest stationary roots of 0.9, the corresponding rejection frequency is 32%.

The issue of size distortions is evident in the discussion of the simulations presented in LL, leading these authors to suggest the use of ‘Bartlett corrections’ for their test statistic. The form of the correction is given by $\tilde{C}_T = \frac{E(C_T)}{E(C_\infty)} C_\infty$, where C_∞ is the asymptotic critical value and $E(C_T)$ and $E(C_\infty)$ are, respectively, the expectations of the finite sample and asymptotic distributions of the LL test statistic. Both $E(C_T)$ and $E(C_\infty)$ can be approximated by simulation. Our results suggest that the use of \tilde{C}_T is indeed effective in correcting the size of the tests. In other words, \tilde{C}_T is close in magnitude to the corresponding empirical quantile of the distribution of the LL test statistic. The power of the test when using \tilde{C}_T is satisfactory, and is close to one for $T = 200$.

(2) We also investigate the performance of the LL test statistic in the presence of cross-unit cointegration and different cointegrating ranks in each unit. It may be noted that in common with the results for $N = 2$, the presence of even a few cross-unit cointegrating relationships substantially biases the LL test statistic towards rejection of no cointegration in each unit. For example, with $T = 200$, $r_i = 0$ for each i , and one cointegrating relationship among all units, the probability of rejecting the null $r_i = 0$ for each i is 56% at a 5% significance level when $N = 8$. We conclude that the importance of cross-unit cointegrating relationships does not decrease when the number of units increases.

For the case of different ranks across the units, the overrejection of no cointegration in each unit reported above is confirmed. The degree of overrejection increases with the number of units which have cointegration. For example, if $N = 4$, $T = 200$, the rejection frequencies of $r_i = 0$ for each i are 18%, 44% and 69% for, respectively, $(r_1 = 1, r_2 = r_3 = r_4 = 0)$, $(r_1 = r_2 = 1, r_3 = r_4 = 0)$ and $(r_1 = r_2 = r_3 = 1, r_4 = 0)$.

(3) As far as the various versions of the Johansen test are concerned (unit-by-unit, joint, and system), these retain their good size and power properties when the units do not cointegrate with each other. However, as for $N = 2$, unit-by-unit analysis (and therefore the joint test) can lead to overrejection of the null of no cointegration in the presence of cross-unit cointegration. The size of the GG test remains slightly lower than the nominal value, at about 4%. Its power is relatively low when there is one cross-unit relationship, but quickly increases with the amount of cross-unit cointegration and the number of observations.

In summary, it is satisfying to conclude that the overall performance for higher dimensions of the various tests considered remains qualitatively unaltered from that reported in detail in the previous subsections for $N = 2$.

4. PANEL UNIVARIATE COINTEGRATION TESTS

Pedroni (1999, 2004) proposed seven panel cointegration tests. Unlike the LL test statistic discussed above, Pedroni’s tests are based on estimating a static cointegrating regression in a panel single-equation framework, with fixed effects and heterogeneous time trends. The presence of cointegration is investigated by conducting panel unit root tests on the estimated residuals, taking into account the effect on the distributions of the unit root tests of using estimated residuals instead of ‘raw data’. The cointegrating regression is given by

$$y_{i,t} = \alpha_i + \delta_i t + Z_{i,t} \beta_i + e_{i,t}, \quad t = 1, 2, \dots, T; \quad i = 1, 2, \dots, N, \quad (2)$$

where α_i and δ_i are scalars denoting fixed effects and unit-specific linear trend parameters, respectively, $y_{i,t}$ are univariate observations of the dependent variable, the $p - 1$ regressors are

$Z_{i,t} = (z_{1,i,t}, z_{2,i,t}, z_{3,i,t}, \dots, z_{p-1,i,t})$ and the cointegration coefficients are $\beta_i = (\beta_{1,i}, \beta_{2,i}, \beta_{3,i}, \dots, \beta_{p-1,i})'$. Under the null hypothesis of no cointegration, the error process $e_{i,t}$ is $I(1)$.

To discuss the distribution of the tests, it is convenient to denote the $(p \times 1)$ -dimensional vector of observations $(y_{i,t}, Z_{i,t})'$ as $x_{i,t}$ to make the notation consistent with the more general cointegration framework above.

Under H_0 (no cointegration), $x_{i,t}$ is generated as

$$x_{i,t} = x_{i,t-1} + \xi_{i,t}. \tag{3}$$

The error term $\xi_{i,t} = (\xi_{i,t}^y, \xi_{i,t}^Z)'$ is a $(p \times 1)$ -dimensional stationary ARMA process which satisfies $\frac{1}{T^{1/2}} \sum_{t=1}^{[Tr]} \xi_{i,t} \implies B_i(\Omega_i)^{12}$ for each i as $T \rightarrow \infty$ where \implies indicates weak convergence and $B_i(\Omega_i)$ is a vector Brownian motion with $(p \times p)$ -dimensional asymptotic covariance matrix given by Ω_i . A further assumption is that $E(\xi_{i,t}, \xi_{j,s}') = 0 \forall i \neq j$ and $\forall s, t$.

Compared to the panel multivariate maximum-likelihood cointegration framework presented in the previous sections, two important restrictions may be noted. First, the cointegrating rank in each unit, regardless of the dimension p , is either 0 or 1, with the unique cointegrating vector, if it exists, given by $(1, -\beta_i')$. Secondly, the specification for $\xi_{i,t}$ rules out all forms of cross-sectional dependence across the units (both in the short and in the long run) although a wide range of temporal dependence within each unit is permitted. These restrictions allow for the estimation of cointegrating relationships in large panels (large N), while this is not feasible within the maximum-likelihood set-up due to degrees-of-freedom restrictions.

Pedroni proposes the use of seven panel cointegration statistics, described in the Appendix, four based on pooling along the within-dimension (denoted ‘panel tests’) and three based on pooling along the between-dimension (denoted ‘group tests’). Three of the four tests in the first category use nonparametric corrections, while the fourth is a parametric ADF test. In the second category, two use nonparametric corrections and the third is again an ADF test.

The test statistics are described in detail in Pedroni (1999, 2004). In order to provide some intuition into the construction of these statistics, we briefly describe here an example of how one of the tests, the parametric panel t -test, is derived. The starting point is a panel ADF regression on the residuals estimated via a static fixed effects estimator from Equation (2)

$$\Delta \hat{e}_{i,t} = \rho \hat{e}_{i,t-1} + \sum_{k=1}^{K_i} \gamma_{ik} \Delta \hat{e}_{i,t-k} + u_{i,t}, \tag{4}$$

where ρ is an autoregressive parameter constrained to be the same across all units, while the coefficients of the lagged differences, γ_{ik} , are allowed to vary across units. The error process $u_{i,t}$ is white noise. The first step in the calculation of the test statistic is analogous to the first step of the Johansen procedure, i.e. the short-run parameters γ_{ik} are filtered out and two sets of residuals are defined:

$$\hat{u}_{1,i,t} = \Delta \hat{e}_{i,t} - \sum_{k=1}^{K_i} \hat{\gamma}_{1ik} \Delta \hat{e}_{i,t-k}, \tag{5}$$

$$\hat{u}_{2,i,t} = \hat{e}_{i,t-1} - \sum_{k=1}^{K_i} \hat{\gamma}_{2ik} \Delta \hat{e}_{i,t-k}, \tag{6}$$

¹² $r \in [0, 1]$ and $[Tr]$ denotes the integer part of Tr .

where $\hat{\gamma}_{1ik}$ ($k = 1, \dots, K_i$) and $\hat{\gamma}_{2ik}$ ($k = 1, \dots, K_i$) are the coefficient estimates in a regression of $\Delta \hat{e}_{i,t}$ on $\Delta \hat{e}_{i,t-k}$ ($k = 1, \dots, K_i$) and $\hat{e}_{i,t-1}$ on $\Delta \hat{e}_{i,t-k}$ ($k = 1, \dots, K_i$), respectively. The lag-length K_i for each unit is either fixed or chosen according to an information criterion.

The next step is to estimate the following regression:

$$\hat{u}_{1,i,t} = \theta_i \hat{u}_{2,i,t} + u_{i,t}, \quad (7)$$

and normalize $\hat{u}_{1,i,t}$ and $\hat{u}_{2,i,t}$ using $\hat{\sigma}_{u,i}$, the estimated variance of $u_{i,t}$, to control for heterogeneity across units, obtaining

$$\tilde{u}_{1,i,t} = \hat{u}_{1,i,t} / \hat{\sigma}_{u,i}, \quad (8)$$

$$\tilde{u}_{2,i,t} = \hat{u}_{2,i,t} / \hat{\sigma}_{u,i}. \quad (9)$$

The normalized residuals as constructed in Equations (8) and (9) are then used to estimate the following pooled regression:

$$\tilde{u}_{1,i,t} = \theta \tilde{u}_{2,i,t} + \tilde{u}_{i,t} \quad (10)$$

and the t -statistic for $\theta = 0$ is calculated.

Alternatively, the t -statistic can be constructed via a nonparametric correction or a test based on the estimator of θ can be used.¹³

The ‘group’ versions of the Pedroni t -test are based not on pooling the normalized residuals as in (10), but on averaging the test statistics estimated unit by unit, based on Equation (7).

Denoting by ρ_i the autoregressive coefficient of the residuals in the i th unit, the panel tests impose a common coefficient under the alternative hypothesis:

$$H_0 : \rho_i = 0, \quad H_1 : \rho_i = \rho, \quad -1 < \rho < 0 \quad \forall i; \quad (11)$$

while the group tests allow for heterogeneous coefficients under the alternative hypothesis:

$$H_0 : \rho_i = 0, \quad H_1 : -1 < \rho_i < 0 \quad \forall i. \quad (12)$$

The Pedroni test statistics, after proper standardization, have standard normal asymptotic distributions under the null hypothesis. Pedroni (1999) tabulates the required moments for the standardizations by simulation.

Pedroni’s panel univariate cointegration tests are usually implemented when N is large and the assumption of maximal rank of 1 is not controversial, a leading example being tests of the weak form of the purchasing power parity (PPP) hypothesis. Yet, the assumption of cross-unit long-run independence is more contentious, and its violation has important consequences, as we demonstrate below by means of simulation experiments.¹⁴

In order to make our results in this section directly comparable with those for LL presented in Section 3, the structure of the DGPs used in these simulations simply involves replicating the DGPs used in the previous sections for $N = 2$ an appropriate number of times (twice for $N = 4$, four times for $N = 8$ etc.). Table 4 provides the results for the size of the Pedroni test statistics—i.e. when the cointegrating rank is zero in each unit and the units are independent in the long run. It also reports the power of the tests for DGP 2A (cross-unit independence and cointegrating rank 1

¹³Details on the other variants of the Pedroni tests are given in Pedroni’s papers and are not repeated here.

¹⁴Panel unit root tests for the strong-form of PPP and their properties have been analysed by us previously in Banerjee et al. (2004).

Table 4. Size and power of univariate Pedroni tests.

Test	T	Size—DGP 1				Power—DGP 2A			
		$N = 2$	$N = 8$	$N = 32$	$N = 128$	$N = 2$	$N = 8$	$N = 32$	$N = 128$
Panel ν	20	0.057	0.035	0.016	0.011	0.071	0.030	0.029	0.029
	50	0.121	0.080	0.055	0.045	0.233	0.172	0.608	0.993
	100	0.141	0.096	0.068	0.051	0.573	0.801	1.000	1.000
	200	0.155	0.108	0.080	0.061	0.991	1.000	1.000	1.000
Panel ρ	20	0.047	0.028	0.014	0.007	0.052	0.027	0.021	0.015
	50	0.088	0.066	0.044	0.034	0.187	0.215	0.630	0.994
	100	0.104	0.072	0.055	0.048	0.546	0.931	1.000	1.000
	200	0.115	0.085	0.064	0.055	0.993	1.000	1.000	1.000
Nonparametric panel t	20	0.111	0.140	0.240	0.548	0.120	0.132	0.298	0.691
	50	0.073	0.083	0.111	0.176	0.134	0.212	0.688	0.998
	100	0.060	0.064	0.078	0.099	0.340	0.841	1.000	1.000
	200	0.059	0.062	0.684	0.083	0.924	1.000	1.000	1.000
Parametric panel t	20	0.206	0.325	0.666	0.990	0.219	0.332	0.748	0.997
	50	0.097	0.127	0.215	0.457	0.173	0.312	0.854	1.000
	100	0.071	0.083	0.117	0.190	0.377	0.881	1.000	1.000
	200	0.064	0.070	0.086	0.118	0.936	1.000	1.000	1.000
Group ρ	20	0.008	0.002	0.000	0.000	0.009	0.000	0.000	0.000
	50	0.036	0.022	0.008	0.001	0.080	0.069	0.164	0.443
	100	0.049	0.038	0.023	0.011	0.333	0.712	1.000	1.000
	200	0.062	0.048	0.041	0.031	0.951	1.000	1.000	1.000
Nonparametric group t	20	0.106	0.132	0.208	0.410	0.113	0.129	0.239	0.504
	50	0.067	0.073	0.086	0.123	0.120	0.172	0.523	0.962
	100	0.057	0.054	0.066	0.076	0.307	0.746	1.000	1.000
	200	0.051	0.053	0.060	0.069	0.902	1.000	1.000	1.000
Parametric group t	20	0.215	0.390	0.757	0.998	0.226	0.381	0.808	0.999
	50	0.092	0.131	0.229	0.489	0.163	0.276	0.772	0.999
	100	0.125	0.076	0.112	0.198	0.345	0.807	1.000	1.000
	200	0.057	0.064	0.080	0.114	0.915	1.000	1.000	1.000

Note: Entries are rejection frequencies of $r = 0$ at 5 using $N(0,1)$ critical values. The test statistics are described in the Appendix. The DGPs are described in Table 1.

in each unit). Table 5 considers the case of cross-unit cointegration (DGPs 4A and 4B). Detailed results for the other DGPs considered in the LL context are available upon request.

A comparison of the ‘size’ columns for $N = 2$ in Table 4 with the block corresponding to DGP 1 (column labelled LL) in Table 2 shows that at sample sizes of 100 or 200, the univariate panel tests denoted *panel ν -test* and *panel ρ -test* are more oversized than the LL test statistic. The *nonparametric panel t -test*, and its parametric and nonparametric group equivalents, given by the *parametric group t -test* and the *nonparametric group t -test*, respectively, have size performance

Table 5. Rejection frequencies of univariate Pedroni tests with non-block-diagonal β matrix.

Test	T	DGP 4A				DGP 4B			
		$N = 2$	$N = 8$	$N = 32$	$N = 128$	$N = 2$	$N = 8$	$N = 32$	$N = 128$
Panel ν	20	0.069	0.043	0.009	0.009	0.068	0.043	0.036	0.012
	50	0.201	0.115	0.076	0.087	0.192	0.109	0.065	0.069
	100	0.288	0.141	0.096	0.128	0.278	0.128	0.081	0.098
	200	0.325	0.153	0.107	0.158	0.314	0.141	0.095	0.122
Panel ρ	20	0.055	0.030	0.008	0.064	0.049	0.025	0.019	0.003
	50	0.126	0.111	0.058	0.058	0.114	0.062	0.035	0.027
	100	0.186	0.098	0.093	0.102	0.173	0.083	0.050	0.040
	200	0.230	0.143	0.123	0.156	0.213	0.109	0.071	0.068
Nonparametric panel t	20	0.119	0.141	0.191	0.532	0.113	0.126	0.251	0.474
	50	0.103	0.097	0.122	0.210	0.097	0.079	0.084	0.129
	100	0.114	0.098	0.117	0.170	0.103	0.083	0.075	0.084
	200	0.134	0.120	0.125	0.181	0.124	0.091	0.075	0.091
Parametric panel t	20	0.214	0.324	0.618	0.988	0.209	0.306	0.673	0.982
	50	0.134	0.146	0.222	0.474	0.126	0.122	0.165	0.357
	100	0.128	0.123	0.156	0.275	0.118	0.100	0.107	0.157
	200	0.145	0.133	0.145	0.234	0.134	0.103	0.094	0.128
Group ρ	20	0.009	0.001	0.000	0.000	0.006	0.001	0.000	0.000
	50	0.057	0.028	0.009	0.003	0.053	0.022	0.004	0.001
	100	0.091	0.061	0.043	0.044	0.081	0.049	0.024	0.018
	200	0.120	0.096	0.087	0.159	0.109	0.078	0.057	0.076
Nonparametric group t	20	0.113	0.128	0.157	0.401	0.112	0.121	0.216	0.362
	50	0.096	0.084	0.100	0.162	0.087	0.073	0.067	0.100
	100	0.097	0.088	0.100	0.162	0.087	0.074	0.063	0.085
	200	0.112	0.104	0.108	0.213	0.100	0.085	0.074	0.120
Parametric group t	20	0.223	0.372	0.698	0.996	0.216	0.362	0.753	0.994
	50	0.125	0.139	0.232	0.532	0.116	0.127	0.184	0.418
	100	0.110	0.112	0.056	0.313	0.099	0.096	0.107	0.192
	200	0.122	0.113	0.133	0.282	0.110	0.097	0.093	0.175

Note: Entries are rejection frequencies of $r = 0$ at 5% using $N(0,1)$ critical values. The test statistics are described in the Appendix. The DGPs are described in Table 1.

comparable to the LL test. Looking at Table 4 in more detail, we also see that the size performance of the *panel ν -test* and the *panel ρ -test* is more satisfactory as the N dimension of the panel increases. In general, size distortions are a decreasing function of sample size for fixed N , except for the *parametric group ρ -test* which is very undersized at small sample sizes and the *parametric group t -test* which displays some nonmonotonicities in rejection rates as a function of sample size. When T is small (20 or 50) there is a tendency for the size of the tests to become more and more distorted with an increase in N , while this is not the case for larger sample sizes.

Taking all these features into account, the most reliable comparisons of the univariate and multivariate frameworks may be made by considering the *nonparametric panel t-test*, the *parametric panel t-test* and the *nonparametric group t-test* for values of T equal to 100 or 200, for all values of N . When N is in excess of 8 or 16, the *panel v-test* and the *panel ρ -test* may also be considered.

As far as power is concerned, comparing the relevant block of Table 2 (column LL, DGP2A, in panel labelled Power experiments) with the 'power' columns of Table 4 shows the LL test statistic to have more power for $N = 2$, but the single-equation tests gain in power rapidly with an increase in N (for which construction of the LL test statistic eventually becomes infeasible).

Table 5 may be compared directly with the relevant blocks of Table 3, i.e. DGP 4A and 4B. Recall that for DGP 4A and 4B, the cointegrating rank within each unit is 0, while there are two cross-unit cointegrating relationships (for each pair of units, i.e. four variables for $N = 2$). The two DGPs differ in their specification of the α matrix which allows for short-run linkages. It may be seen in Table 5 that the rejection rates of no cointegration (within each unit) can be quite large, especially for small T and large N . When N is small the distortions are however lower than those for LL resulting from Table 3, where the rejection rates quickly increase to 1. Allowing for different cointegrating rank in each unit also has severe distortionary effects, with the rejection rates for the Pedroni test statistics reflecting accurately the conclusions reached earlier for the LL test.¹⁵

5. CONCLUSIONS

Panel multivariate cointegration methods (such as Larsson and Lyhagen 2000a) and the panel analogue of the univariate Engle-Granger test (as developed by Kao (1999) and Pedroni (1999) *inter alia*) assume a block-diagonal β matrix, i.e. no cross-unit cointegration, an assumption that is often at odds with economic theory and empirical results.

Our simulation results indicate that when the hypotheses underlying the multivariate panel framework are satisfied, the LL test has good size and power properties, and often yields gains in efficiency relative to full-system analysis for estimation of the cointegrating parameters. However, the consequences of violations of the assumptions of block-diagonality of β and equal rank across the units can be very serious, with both univariate and multivariate tests displaying size distortions. When N is small the presence of cross-unit cointegration is less harmful for the single-equation tests than for the LL test statistic.

By contrast, unit-by-unit analysis is accurate when T is large enough, and can provide a useful diagnostic test for different cointegrating ranks across units, while cross-unit cointegration can be detected by the Gonzalo and Granger test, at least for large enough sample sizes.

The suggestion for empirical analysis is therefore to use full-system estimation whenever possible. If this is not feasible, the first step should be a unit-by-unit cointegration analysis. The second step would then be to test for the presence of cross-unit cointegration by means of the Gonzalo and Granger test. If the null hypothesis of no cross-unit cointegration is accepted, and the unit-by-unit analysis does not indicate the presence of different ranks across units, the third step is then to apply the LL or Pedroni tests (depending on the size of N), since these methods

¹⁵Details given in Tables 19 and 20 in the working paper version of this paper.

can yield efficiency gains in terms of higher power and lower standard errors for the estimated cointegrating coefficients.

The way forward in the longer term is either to develop tests which do not impose such severe restrictions, or to find more reasonable and testable ways of incorporating restrictions within the maximum-likelihood framework. The paper by Bai and Ng (2004) is a first step in this direction, and is also the focus of our continuing research in this area.

ACKNOWLEDGMENTS

We thank seminar participants at the first meeting of the ESF-EMM network in Arona, especially David Hendry, Søren Johansen and Katarina Juselius, for their comments. We are very grateful to Johan Lyhagen and Peter Pedroni for kindly making their codes available to us. The financial generosity of the European University Institute for partially funding this research under research project 'Modelling and Forecasting Economic Time Series in the Presence of High Orders of Integration and Structural Breaks in Data' is acknowledged. The paper has also benefited from comments by two anonymous referees and our Associate Editor Siem Jan Koopman.

REFERENCES

- Bai, J. and S. Ng (2004). A PANIC attack on unit roots and cointegration, *Econometrica* (forthcoming).
- Banerjee, A., M. Marcellino and C. Osbat (2004). Testing for PPP: Should we use panel methods? *Empirical Economics* (forthcoming).
- Chiu, R.-L. (2004). Testing the purchasing power parity in panel data, *International Review of Economics and Finance* 11, 349–62.
- Dembla, P. (2000). Production in the Indian manufacturing sector: a panel data analysis, *Indian Economic Review* 35, 41–65.
- Doornik, J. A. (2001). *Object-Oriented Matrix Programming Using Ox*. 4th ed., London: Timberlake Consultants Press. Available at www.nuff.ox.ac.uk/Users/Doornik.
- Drine, I. and C. Rault (2003). Do panel data permit the rescue of the Balassa–Samuelson hypothesis for Latin American countries? *Applied Economics* 35, 351–59.
- Egert, B. (2002). Investigating the Balassa–Samuelson hypothesis in the transition. *Economics of Transition* 10, 273–309.
- Evans, P. and G. Karras (1996). Convergence revisited, *Journal of Monetary Economics* 37, 249–65.
- Frankel, J. A. and A. K. Rose (1996). A panel project on purchasing power parity: Mean reversion within and between countries, *Journal of International Economics* 40, 209–24.
- Gonzalo, J. and C. W. J. Granger (1995). Estimation of common long-memory components in cointegrated systems, *Journal of Business and Economic Statistics* 13, 27–36.
- Groen, J. J. and F. R. Kleibergen (2003). Likelihood-based cointegration analysis in panels of vector error correction models, *Journal of Business and Economic Statistics* 21, 295–318.
- Johansen, K. (2002). Hysteresis in unemployment: evidence from Norwegian counties, Discussion paper, Department of Economics, Norwegian University of Science and Technology.
- Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford and New York: Oxford University Press.
- Kao, C. (1999). Spurious regression and residual-based tests for cointegration in panel data, *Journal of Econometrics* 90, 1–44.

- Larsson, R. and J. Lyhagen (2000a). Likelihood-based inference in multivariate panel cointegration models, Discussion Paper 1313, Econometric Society World Congress 2000 Contributed Papers.
- Larsson, R. and J. Lyhagen (2000b). Testing for common cointegrating rank in dynamic panels, Discussion Paper Working Paper Series in Economics and Finance 378, Stockholm School of Economics.
- Larsson, R., J. Lyhagen and M. Löthgren (2001). Likelihood-based cointegration tests in heterogeneous panels, *Econometrics Journal* 4, 41.
- Lee, K. E., M. H. Pesaran and R. Smith (1997). Growth and convergence in a multicountry empirical stochastic Solow model, *Journal of Applied Econometrics* 12, 357–92.
- Lyhagen, J. (2000). Why not use standard panel unit root test for testing PPP, Discussion Paper 413, SSE/EFI, Working Paper Series in Economics and Finance.
- McCoskey, S. and C. Kao (1998). A residual-based test for the null of cointegration in panel data, *Econometric Reviews* 17, 57–84.
- O'Connell, P. (1998). The overvaluation of purchasing power parity, *Journal of International Economics* 44, 1–19.
- Papell, D. H. (1997). Searching for stationarity: purchasing power parity under the current float, *Journal of International Economics* 43, 313–32.
- Pedroni, P. (1999). Critical values for cointegration tests in heterogeneous panels with multiple regressors, *Oxford Bulletin of Economics and Statistics*, 61, 653–70.
- Pedroni, P. (2000). Fully modified OLS for heterogeneous cointegrated panels. In B. Baltagi and C. D. Kao (eds.), *Advances in Econometrics. Nonstationary Panels, Panel Cointegration and Dynamic Panels*, pp. 93–130. New York: Elsevier Science.
- Pedroni, P. (2001). Purchasing power parity tests in cointegrated panels, *The Review of Economics and Statistics* 83, 727–31.
- Pedroni, P. (2004). Panel cointegration; asymptotic and finite sample properties of pooled time series tests with an application to the PPP hypothesis, *Econometric Theory* (forthcoming).
- Pesaran, M. H. and R. Smith (1995). Estimating long-run relationships from dynamic heterogeneous panels, *Journal of Econometrics* 68, 79–113.
- Song, F. M. and Y. Wu (1997). Hysteresis in unemployment: Evidence from 48 US states, *Economic Inquiry* 35, 235–43.

APPENDIX

Panel Cointegration Statistics (Pedroni 1999)

1. Panel ν -statistic

$$Z_{\hat{\nu}_{N,T}} \equiv \left(\sum_{i=1}^N \sum_{t=2}^T \hat{L}_{11,i}^{-2} \hat{e}_{i,t-1}^2 \right)^{-1}.$$

2. Panel ρ -statistic

$$Z_{\hat{\rho}_{N,T}} \equiv \left(\sum_{i=1}^N \sum_{t=2}^T \hat{L}_{11,i}^{-2} \hat{e}_{i,t-1}^2 \right)^{-1} \left(\sum_{i=1}^N \sum_{t=2}^T \hat{L}_{11,i}^{-2} (\hat{e}_{i,t-1} \Delta \hat{e}_{i,t} - \hat{\lambda}_i) \right).$$

3. Panel t -statistic (nonparametric)

$$Z_{t_{N,T}} \equiv \left(\hat{\sigma}_{N,T}^2 \sum_{i=1}^N \sum_{t=2}^T \hat{L}_{11,i}^{-2} \hat{e}_{i,t-1}^2 \right)^{-1/2} \left(\sum_{i=1}^N \sum_{t=2}^T \hat{L}_{11,i}^{-2} (\hat{e}_{i,t-1} \Delta \hat{e}_{i,t} - \hat{\lambda}_i) \right).$$

4. Panel *t*-statistic (parametric)

$$Z_{iN,T}^* \equiv \left(\widehat{s}_{N,T}^{*2} \sum_{i=1}^N \sum_{t=3}^T \widehat{L}_{11,i}^{-2} \widehat{u}_{2,i,t}^2 \right)^{-1/2} \left(\sum_{i=1}^N \sum_{t=3}^T \widehat{L}_{11,i}^{-2} \widehat{u}_{2,i,t} \widehat{u}_{1,i,t} \right).$$

5. Group ρ -statistic

$$\widetilde{Z}_{\widehat{\rho}_{N,T-1}} \equiv \sum_{i=1}^N \left(\sum_{t=2}^T \widehat{e}_{i,t-1}^2 \right)^{-1} \sum_{t=2}^T (\widehat{e}_{i,t-1} \Delta \widehat{e}_{i,t} - \widehat{\lambda}_i).$$

6. Group *t*-statistic (nonparametric)

$$\widetilde{Z}_{tN,T} \equiv \sum_{i=1}^N \left(\widehat{\sigma}_i^2 \sum_{t=2}^T \widehat{e}_{i,t-1}^2 \right)^{-1/2} \sum_{t=2}^T (\widehat{e}_{i,t-1} \Delta \widehat{e}_{i,t} - \widehat{\lambda}_i).$$

7. Group *t*-statistic (parametric)

$$\widetilde{Z}_{iN,T}^* \equiv \sum_{i=1}^N \left(\sum_{t=3}^T \widehat{s}_i^{*2} \widehat{u}_{2,i,t}^2 \right)^{-1/2} \sum_{t=3}^T \widehat{u}_{2,i,t} \widehat{u}_{1,i,t}.$$

$\widehat{e}_{i,t}$ is the OLS residual from Equation (2) in the text. The residuals $\widehat{\mu}_{i,t}$, and $\widehat{\eta}_{i,t}$ are defined based on the following regressions, respectively:

$$\begin{aligned} \widehat{e}_{i,t} &= \widehat{\gamma}_i \widehat{e}_{i,t-1} + \widehat{\mu}_{i,t}, \\ \Delta y_{i,t} &= \widehat{\delta}_i + \sum_{m=1}^{p-1} \widehat{\beta}_{m,i} \Delta z_{m,i,t} + \widehat{\eta}_{i,t}, \end{aligned}$$

where $\widehat{}$ indicates sample estimates. This leads to the following definitions:

$$\begin{aligned} \widehat{\lambda}_i &= \frac{1}{T} \sum_{s=1}^{k_i} \left(1 - \frac{s}{k+1} \right) \sum_{t=s+1}^T \widehat{\mu}_{i,t} \widehat{\mu}_{i,t-s}, \\ \widehat{s}_i^2 &\equiv \frac{1}{T} \sum_{t=1}^T \widehat{\mu}_{i,t}^2, \widehat{\sigma}_i^2 = \widehat{s}_i^2 + 2\widehat{\lambda}_i, \\ \widetilde{\sigma}_{N,T}^2 &\equiv \frac{1}{N} \sum_{i=1}^N \widehat{L}_{11,i}^{-2} \widehat{\sigma}_i^2, \\ \widehat{s}_i^{*2} &\equiv \frac{1}{T} \sum_{t=1}^T \widehat{u}_{1,t}^2, \\ \widetilde{s}_{N,T}^{*2} &\equiv \frac{1}{N} \sum_{i=1}^N \widehat{s}_i^{*2}, \\ \widehat{L}_{11,i}^2 &= \frac{1}{T} \sum_{t=1}^T \widehat{\eta}_{i,t}^2 + \frac{2}{T} \sum_{s=1}^{k_i} \left(1 - \frac{s}{k+1} \right) \sum_{t=s+1}^T \widehat{\eta}_{i,t} \widehat{\eta}_{i,t-s}. \end{aligned}$$

Finally, $\widehat{u}_{i,t}$, $\widehat{u}_{1,i,t}$ and $\widehat{u}_{2,i,t}$ are as defined in Equations (7), (8) and (9) respectively.