

Joakim Westerlund

Testing for Error Correction in Panel Data

RM/06/056

JEL code : C12, C32



Maastricht research school of Economics  
of **TE**chnology and **OR**ganizations

Universiteit Maastricht  
Faculty of Economics and Business Administration  
P.O. Box 616  
NL - 6200 MD Maastricht

phone : ++31 43 388 3830  
fax : ++31 43 388 4873



# Testing for Error Correction in Panel Data\*

Joakim Westerlund<sup>†</sup>

November 10, 2006

## Abstract

This paper proposes new error correction based cointegration tests for panel data. The limiting distributions of the tests are derived and critical values are provided. Our simulation results suggest that the tests have good small-sample properties with small size distortions and high power relative to other popular residual-based panel cointegration tests. In our empirical application, we present evidence suggesting that international health care expenditures and GDP are cointegrated once the possibility of an invalid common factor restriction has been accounted for.

**JEL Classification:** C12; C32; C33; O30.

**Keywords:** Panel Cointegration Test; Common Factor Restriction; Cross-Section Dependence; International Health Care Expenditures.

## 1 Introduction

The use of panel cointegration techniques to test for the presence of long-run relationships among integrated variables with both a time series dimension  $T$  and a cross-sectional dimension  $N$  has received much attention recently. The literature concerned with the development of such tests has thus far taken two broad directions. The first consists of taking cointegration as the null hypothesis. This is the basis of the panel cointegration tests proposed by McCoskey and Kao (1998) and Westerlund (2005a).

---

\*Previous versions of this paper were presented at the 13<sup>th</sup> International Conference on Panel Data in Cambridge and at a seminar at Lund University. The author would like to thank conference and seminar participants, and in particular Anindya Banerjee, David Edgerton, Rolf Larsson, Johan Lyhagen, Peter Pedroni, Jean-Pierre Urbain, and two anonymous referees for many valuable comments and suggestions. The author would also like to thank the Maastricht Research School of Economics of Technology and Organizations for its hospitality during a visit at the Department of Quantitative Economics at the University of Maastricht, where a part of this paper was written. Thank you also to the Jan Wallander and Tom Hedelius Foundation for financial support under research grant number W2006-0068:1. The usual disclaimer applies.

<sup>†</sup>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.

The second approach is to take no cointegration as the null hypothesis. Tests within this category are almost exclusively based on the methodology of Engle and Granger (1987) whereby the residuals of a static least squares regression is subjected to a unit root test. The most influential theoretical contributions within this category are those of Pedroni (1999, 2004), in which the author generalizes the work of Phillips and Ouliaris (1990) by developing several tests that are appropriate for various cases of heterogeneous dynamics, endogenous regressors, and individual specific constants and trends. Tests are developed both for the case with a common autoregressive root under the alternative hypothesis as well as for the case that permits heterogeneity of the autoregressive roots.

Of course, this new development has not gone unnoticed in the empirical literature, where the new tests have attracted an enormous amount of interest. Although there are many reasons for this, the single most cited rationale for using these tests is the increased power that may be brought to bear on the cointegration hypothesis through the accounting of both the time series and cross-sectional dimensions. In spite of this, however, many studies such as Ho (2002) fail to reject the null hypothesis, even in cases when cointegration is strongly suggested by theory.

One plausible explanation for this failure to reject the null centers on the fact that residual-based tests of this sort require the long-run cointegrating vector for the variables in their levels being equal to the short-run adjustment process for the variables in their differences. Kremers *et al.* (1992) refer to this as a common factor restriction and show that its failure can cause a significant loss of power for residual-based cointegration tests.

In this paper, we propose four new panel tests of the null hypothesis of no cointegration that are based on structural rather than residual dynamics, and therefore do not impose any common factor restriction. The proposed tests are panel extensions of those proposed in the time series context by Banerjee *et al.* (1998). As such, they are designed to test the null by inferring whether the error correction term in a conditional error correction model is equal to zero. If the null hypothesis of no error correction is rejected, then the null hypothesis of no cointegration is also rejected.

Each test is able to accommodate individual specific short-run dynamics, including serially correlated error terms, non-strictly exogenous regressors, individual specific intercept and trend terms, and individual specific slope parameters. Bootstrap tests are also proposed to handle applications with cross-sectional dependence.

Two tests are designed to test the alternative hypothesis that the panel is cointegrated as a whole, while the other two test the alternative that there is at least one individual that is cointegrated. All four tests are shown to be very straightforward and easy to implement. The asymptotic results reveal that the tests have limiting normal distributions, and that they are consistent.

Simulation evidence is also provided to evaluate and compare the small-sample performance of the tests relative to the performance of the popular residual-based tests by Pedroni (2004). The results suggest that the new tests maintain good size accuracy, and that they are more powerful than the residual-based tests provided that the conditions laid out in the paper hold. In our empirical application, we provide evidence suggesting that international health care expenditures and GDP are cointegrated once the long- and short-run adjustment processes are allowed to differ.

The plan of the paper is as follows. Sections 2 and 3 specify the model and outlines the new cointegration tests. Section 4 then analyzes the asymptotic properties of the tests under the assumption of cross-sectional independence, while Section 5 presents bootstrap tests that relaxes this assumption. Section 6 is devoted to the Monte Carlo study, while Section 7 contains the empirical application. Concluding remarks are given in Section 8. Proofs of important results are relegated to the appendix.

## 2 The error correction setup

In this section, we present the basic error correction setup, in which the new cointegration tests will be developed. We begin by specifying the model of interest, and then we go on to discuss how it relates to the model used in the residual-based approach.

### 2.1 Model and assumptions

We consider the following data generating process

$$y_{it} = \phi_{1i} + \phi_{2i}t + z_{it}, \quad (1)$$

$$x_{it} = x_{it-1} + v_{it}, \quad (2)$$

where  $t = 1, \dots, T$  and  $i = 1, \dots, N$  indexes the time series and cross-sectional units, respectively. For simplicity, the  $K$  dimensional vector  $x_{it}$  is modelled as a pure random walk while the scalar  $y_{it}$  consists of both a deterministic term  $\phi_{1i} + \phi_{2i}t$  and a stochastic term  $z_{it}$ , which is modeled as

$$\alpha_i(L)\Delta z_{it} = \alpha_i(z_{it-1} - \beta_i'x_{it-1}) + \gamma_i(L)'v_{it} + e_{it}, \quad (3)$$

where  $\alpha_i(L) = 1 - \sum_{j=1}^{p_i} \alpha_{ij}L^j$  and  $\gamma_i(L) = \sum_{j=0}^{p_i} \gamma_{ij}L^j$  are scalar and  $K$  dimensional polynomials in the lag operator  $L$ . Note that (3) is basically the conditional model for  $z_{it}$  given  $x_{it}$  in a standard vector error correction setup, with (2) being the associated marginal model for  $x_{it}$ . By substituting (1) into (3), we get the following conditional error correction model for  $y_{it}$

$$\alpha_i(L)\Delta y_{it} = \delta_{1i} + \delta_{2i}t + \alpha_i(y_{it-1} - \beta_i'x_{it-1}) + \gamma_i(L)'v_{it} + e_{it}, \quad (4)$$

where  $\delta_{1i} = \alpha_i(1)\phi_{2i} - \alpha_i\phi_{1i} + \alpha_i\phi_{2i}$  and  $\delta_{2i} = -\alpha_i\phi_{2i}$  now represent the deterministic components. Typical deterministic elements include a constant and a linear time trend. To allow for this, we distinguish between three cases. In Case 1,  $\phi_{1i}$  and  $\phi_{2i}$  are both restricted to zero so  $y_{it}$  has no deterministic terms, while in Case 2, then  $\phi_{1i}$  is unrestricted but  $\phi_{2i}$  is zero so  $y_{it}$  is generated with a constant. Finally, in Case 3, then there are no restrictions on either  $\phi_{1i}$  or  $\phi_{2i}$ , which suggests that  $y_{it}$  is generated with both a constant and trend.

In all three cases, note that the error correction model in (4) can only be stable if the variables it comprises are all stationary.<sup>1</sup> Thus, as  $y_{it-1} - \beta'_i x_{it-1}$  must be stationary, the vector  $\beta_i$  defines a long-run equilibrium relationship between  $x_{it}$  and  $y_{it}$ , provided of course that the errors  $v_{it}$  and  $e_{it}$  are also stationary. Any deviation from this equilibrium relationship lead to a correction by a proportion  $-2 < \alpha_i \leq 0$ , which is henceforth referred to as the error correction parameter. If  $\alpha_i < 0$ , then there is error correction, which implies that  $y_{it}$  and  $x_{it}$  are cointegrated, whereas if  $\alpha_i = 0$ , then the error correction will be absent and there is no cointegration. This suggests that the null hypothesis of no cointegration for cross-sectional unit  $i$  can be implemented as a test of  $H_0 : \alpha_i = 0$  versus  $H_1 : \alpha_i < 0$ . In what follows, we shall propose four new panel statistics that are based on this idea.

Two of the statistics are based on pooling the information regarding the error correction along the cross-sectional dimension of the panel. These are referred to as panel statistics. The second pair do not exploit this information and are referred to as group mean statistics. The relevance of this distinction lies in the formulation of the alternative hypothesis. For the panel statistics, the null and alternative hypotheses are formulated as  $H_0 : \alpha_i = 0$  for all  $i$  versus  $H_1^p : \alpha_i = \alpha < 0$  for all  $i$ , which indicates that a rejection should be taken as evidence of cointegration for the panel as a whole. By contrast, for the group mean statistics,  $H_0$  is tested versus  $H_1^g : \alpha_i < 0$  for at least some  $i$ , suggesting that a rejection should be taken as evidence of cointegration for at least one of the cross-sectional units.

We now lay out the assumptions needed for the development of our new panel statistics.

**Assumption 1.** (Error process.) The vector  $w_{it} = (e_{it}, v'_{it})'$  satisfies the following set of conditions:

- (a) The vector  $w_{it}$  is independent and identically distributed (i.i.d.) across both  $i$  and  $t$  with  $E(w_{it}) = 0$  and  $\text{var}(e_{it}) = \sigma_i^2 < \infty$ ;
- (b)  $\text{var}(v_{it}) = \Omega_i$  is positive definite;
- (c)  $E(e_{kt}v_{ij}) = 0$  for all  $k, t, i$  and  $j$ .

---

<sup>1</sup>Formally, for the single equation error correction model in (4) to be stable, for some complex number  $z$ , we require that the roots of the equation  $\alpha_i(z) = 0$  lie outside the unit circle, see for example Zivot (2000).

Assumption 1 establishes the basic conditions needed for developing the new cointegration tests. Many of these are quite restrictive but are made here in order to make the analysis of the tests more transparent, and will be relaxed later on.

For example, Assumption 1 (a) states that the individuals are independent over the cross-sectional dimension. Although not strictly necessary, this condition is convenient as it will allow us to apply standard central limit theory in a relatively simple manner. Some possibilities for how to relax this condition will be discussed in Section 5. Similarly, independence through time is convenient because it facilitates a straightforward asymptotic analysis by application of the conventional methods for integrated processes. In particular, Assumption 1 (a) ensures that the following invariance principle holds individually for each cross-section

$$T^{-1/2} \sum_{t=1}^{[Tr]} w_{it} \Rightarrow B_i \text{ as } T \rightarrow \infty,$$

where the symbol  $\Rightarrow$  signifies weak convergence and  $B_i = (B_{1i}, B'_{2i})'$  is a vector Brownian motion of dimension  $K + 1$  with block-diagonal covariance matrix, with elements  $\sigma_i^2$  and  $\Omega_i$ . Also, for notational convenience, the Brownian motion  $B_i(r)$  is written as  $B_i$  with the argument  $r \in [0, 1]$  suppressed.

Assumptions 1 (b) and (c) are concerned with the covariance matrix of  $B_i$ , equally the long-run covariance matrix of  $w_{it}$ . Specifically, Assumption 1 (c) states that the  $K \times K$  matrix  $\Omega_i$  is positive definite, which is tantamount to requiring that  $x_{it}$  is not cointegrated in case we have multiple regressors. This assumption is very standard and will be maintained throughout.<sup>2</sup>

Assumption 1 (c) requires that the vector of regressors is independent of the regression error  $e_{it}$ , which is satisfied if  $x_{it}$  is strictly exogenous. Although this might seem overly restrictive at first, our model is actually quite general when it comes to the short-run dynamics of the system. In fact, the only really necessary requirement is that the regressors contained in  $x_{it}$  are weakly exogenous with respect to  $\alpha_i$  and  $\beta_i$ , the parameters of interest, which is implicit in our model since  $x_{it}$  is not error correcting. Apart from this, however, our model is flexible. In particular, as we will show later, the orthogonality condition in Assumption 1 (c) can be easily relaxed to accommodate regressors that are weakly but not necessarily strictly exogenous.<sup>3</sup>

---

<sup>2</sup>Note the data generating process used here can be seen as a restricted version of the one used by Larsson *et al.* (2001). The restriction being that the cointegration rank is at most one. Thus, if one suspects that there is cointegration in  $x_{it}$ , then the rank test of Larsson *et al.* (2001) may be used to determine the exact number of cointegration relationships.

<sup>3</sup>Note the distinction here between strict exogeneity, weak exogeneity and endogeneity. In particular, because the error correction model in (4) has been conditioned on all current and past values of  $v_{it}$ , we have by construction that  $E(e_{it}v_{it-j}) = 0$  for all  $j \geq 0$ . In addition, the marginal model for  $x_{it}$  is assumed not to be error correcting. Strict exogeneity corresponds to requiring that  $E(e_{it}v_{it-j}) = 0$  for all  $j$ , including all  $j < 0$ . Weak exogeneity, on the other

Moreover, the fact that  $\alpha_i(L)$  and  $\gamma_i(L)$  are permitted to vary between the individuals of the panel indicates that we are in effect allowing for a completely heterogeneous serial correlation structure. In addition, although the regressors are specified as pure random walks, our results can easily be generalized to allow for more general dynamics. In fact, it is straightforward to show that  $v_{it}$  can be endowed with a general autoregressive representation without affecting the results reported in this paper.

## 2.2 A comparison with the residual-based approach

Although the data generating process adopted in this paper is quite general, since the regressors are not permitted to be endogenous, it is more restrictive than the one used by Pedroni (2004) for developing his residual-based panel cointegration tests. Thus, in view of this, it seems fair to question the relevance of introducing the new tests. The reason is that, while more general when it comes to the endogeneity of the regressors, residual-based tests are also more restrictive because they impose a possibly invalid common factor restriction. Accordingly, to better understand the trade-off between these two competing approaches, it is important to discuss the implications of the weak exogeneity and common factor restrictions.

We begin with weak exogeneity, which essentially says that (4) is the model of interest when testing for cointegration, and that the marginal model in (2) can be ignored. The concept of weak exogeneity has been studied to some length by for example Johanssen (1992) and Urbain (1992), and the reader is referred to these sources for full details. In the present context, Johansen (1992) has shown that  $x_{it}$  is weakly exogenous with respect to  $\alpha_i$  and  $\beta_i$  if the marginal model for  $x_{it}$  is not error correcting. If this is the case, then  $\alpha_i$  and  $\beta_i$  can be efficiently estimated from the single equation conditional error correction model. In particular, weak exogeneity ensures that a test for no cointegration can be implemented as a test for no error correction in (4) only.

On the other hand, if the weak exogeneity assumption does not hold, then the conditional model in (4) does not contain all of the necessary information to conduct the cointegration test. In particular, although irrelevant under the null of no cointegration, it is not difficult to see that such a test might have difficulties finding cointegration under the alternative if it is mainly  $x_{it}$  that is error correcting. The intuition lies in noting that  $\alpha_i$  can be written as

$$\alpha_i = \alpha_{iy} - \gamma'_{i0} \alpha_{ix},$$

where  $\alpha_{iy}$  and  $\alpha_{ix}$  are the error correction parameters of the equations for  $y_{it}$  and  $x_{it}$  in the underlying unconditional error correction model, respectively,

---

hand, only requires that the marginal model is not error correcting. Specifically,  $e_{it}$  and  $v_{it}$  do not have to be uncorrelated at all lags and leads. Finally, endogeneity does not preclude  $x_{it}$  from being error correcting, nor does it prevent  $e_{it}$  and  $v_{it}$  from being correlated.



see for example Zivot (2000). Now, because the signs of  $\gamma_{i0}$  and  $\alpha_{ix}$  are not constrained in any way, we can just as well assume that  $\gamma_{i0}$  is positive, and concentrate on  $\alpha_{ix}$ . In this case, it is clear that a left-sided cointegration test based on the conditional model will have reduced power if  $\alpha_{ix}$  is negative, and it will have no power if  $\gamma'_{i0}\alpha_{ix}$  is greater than  $\alpha_{iy}$ . On the other hand, power can also be increased in the sense that  $\alpha_{ix}$  may just as well be positive.

In order to better understand the common factor restriction in the residual-based setup, consider the conditional error correction model in (4). The test regression that forms the basis for the tests of Pedroni (2004) can be derived from this model, thus establishing a relationship between the error correction and residual-based tests. Indeed, by subtracting  $\alpha_i(L)\beta'_i v_{it}$  from both sides of (4), we obtain

$$\Delta(y_{it} - \beta'_i x_{it}) = \delta_{1i} + \delta_{2i}t + \alpha_i(y_{it-1} - \beta'_i x_{it-1}) + \tilde{e}_{it}. \quad (5)$$

where  $\tilde{e}_{it}$  denotes the composite error term

$$\tilde{e}_{it} = (\gamma_i(L) - \alpha_i(L)\beta_i)'v_{it} + e_{it}.$$

The residual-based approach tests the null hypothesis of no cointegration by inferring if the putative equilibrium error  $y_{it} - \beta'_i x_{it}$  in (5) has a unit root or, equivalently, if  $\alpha_i$  is equal to zero. The problem with this approach is that it imposes a possibly invalid common factor restriction as seen by the fact that  $e_{it}$  and  $\tilde{e}_{it}$  are not equal unless  $\gamma_i(L)$  and  $\alpha_i(L)\beta_i$  are equal.

To get an intuition of this, note that the variance of  $\tilde{e}_{it}$  is given by

$$\text{var}(\tilde{e}_{it}) = (\gamma_i(L) - \alpha_i(L)\beta_i)'\Omega_i(\gamma_i(L) - \alpha_i(L)\beta_i) + \sigma_i^2.$$

Suppose now that  $\sigma_i^2$  is close to zero but that the first term on the right-hand side is large. In this case, the error correction model in (4) has nearly perfect fit with  $\alpha_i$  being estimated with high precision. The error correction tests will therefore have good power. By contrast, the estimation of  $\alpha_i$  in (5) will be much more imprecise, producing tests with low power. Thus, we expect the new tests to enjoy higher power whenever  $\gamma_i(L)$  and  $\alpha_i(L)\beta_i$  differ, and the signal-to-noise ratio of  $\Omega_i$  to  $\sigma_i^2$  is large.

Thus, the difference between the assumptions of the two classes of residual-based and error correction tests essentially boils down to a trade-off in power. If weak exogeneity fails, then the error correction tests may have low power, while if the common factor restriction fails, then the residual-based tests may have low power. In this paper, weak exogeneity is a maintained assumption, and hence most of the power analysis will be focused on the common factor issue.<sup>4</sup>

---

<sup>4</sup>Some simulation results of the effects of a failure of weak exogeneity are reported in Section 6.

### 3 Test construction

In constructing the new statistics, it is useful to rewrite (4) as

$$\Delta y_{it} = \delta_i' d_t + \alpha_i (y_{it-1} - \beta_i' x_{it-1}) + \sum_{j=1}^{p_i} \alpha_{ij} \Delta y_{it-j} + \sum_{j=0}^{p_i} \gamma_{ij} \Delta x_{it-j} + e_{it}, \quad (6)$$

where  $d_t = (1, t)'$  now holds the deterministic components with  $\delta_i = (\delta_{1i}, \delta_{2i})'$  being the associated vector of parameters. The problem is how to estimate the error correction parameter  $\alpha_i$ , which, as mentioned earlier, forms the basis for our new tests. One way is to assume that  $\beta_i$  is known, and to estimate  $\alpha_i$  by least squares. However, as shown by Boswijk (1994) and Zivot (2000), tests based on a prespecified  $\beta_i$  are generally not similar and depend on nuisance parameters, even asymptotically.

As an alternative approach, note that (6) can be reparameterized as

$$\Delta y_{it} = \delta_i' d_t + \alpha_i y_{it-1} + \lambda_i' x_{it-1} + \sum_{j=1}^{p_i} \alpha_{ij} \Delta y_{it-j} + \sum_{j=0}^{p_i} \gamma_{ij} \Delta x_{it-j} + e_{it}. \quad (7)$$

In this regression, the parameter  $\alpha_i$  is unaffected by imposing an arbitrary  $\beta_i$ , which suggests that the least squares estimate of  $\alpha_i$  can be used to provide a valid test of  $H_0$  versus  $H_1$ . Indeed, because  $\lambda_i$  is unrestricted, and because the cointegration vector is implicitly estimated under the alternative hypothesis, as seen by writing  $\lambda_i = -\alpha_i \beta_i$ , this means that it is possible to construct a test based on  $\alpha_i$  that is asymptotically similar, and whose distribution is free of nuisance parameters. In this paper, we therefore propose four new tests that are based on the least squares estimate of  $\alpha_i$  in (7) and its  $t$ -ratio. The construction of these statistics is described next.

#### 3.1 The group mean statistics

The construction of the group mean statistics is particularly simple and can be carried out in three steps. The first step is to estimate (7) by least squares for each individual  $i$ , which yields

$$\Delta y_{it} = \hat{\delta}_i' d_t + \hat{\alpha}_i y_{it-1} + \hat{\lambda}_i' x_{it-1} + \sum_{j=1}^{p_i} \hat{\alpha}_{ij} \Delta y_{it-j} + \sum_{j=0}^{p_i} \hat{\gamma}_{ij} \Delta x_{it-j} + \hat{e}_{it}. \quad (8)$$

The lag order  $p_i$  is permitted to vary across individuals, and can be determined preferably using a data dependent rule. For example, we may use the Campbell and Perron (1991) rule, which is a simple sequential test rule based on the significance of the individual lag parameters  $\hat{\alpha}_{ij}$  and  $\hat{\gamma}_{ij}$ . Another possibility is to use an information criterion, such as the Akaike criterion. Alternatively, the number of lags can be set as a fixed function of  $T$ .

The second step involves estimating

$$\alpha_i(1) = 1 - \sum_{j=1}^{p_i} \alpha_{ij}.$$

A natural way to do this is to use a parametric approach and to estimate  $\alpha_i(1)$  using  $\tilde{\alpha}_i(1) = 1 - \sum_{j=1}^{p_i} \hat{\alpha}_{ij}$ . Unfortunately, tests based on  $\tilde{\alpha}_i(1)$  are known to suffer from poor small-sample performance due to the uncertainty inherent in the estimation of the autoregressive parameters, especially when  $p_i$  is large.

As an alternative approach, consider the following kernel estimator

$$\hat{\omega}_{yi}^2 = \frac{1}{T-1} \sum_{j=-M_i}^{M_i} \left(1 - \frac{j}{M_i+1}\right) \sum_{t=j+1}^T \Delta y_{it} \Delta y_{it-j},$$

where  $M_i$  is a bandwidth parameter that determines how many covariances to estimate in the kernel. The relevance of  $\hat{\omega}_{yi}^2$  is easily appreciated by noting that under the null hypothesis, the long-run variance  $\omega_{yi}^2$  of  $\Delta y_{it}$  is given by  $\omega_{ui}^2/\alpha_i(1)^2$ , where  $\omega_{ui}^2$  is the corresponding long-run variance of the composite error term  $u_{it} = \gamma_i(L)'v_{it} + e_{it}$ . This suggests that  $\alpha_i(1)$  can be estimated alternatively using  $\hat{\omega}_{ui}/\hat{\omega}_{yi}$ , where  $\hat{\omega}_{ui}$  may be obtained as above using kernel estimation with  $\Delta y_{it}$  replaced by

$$\hat{u}_{it} = \sum_{j=0}^{p_i} \hat{\gamma}_{ij} \Delta x_{it-j} + \hat{e}_{it},$$

where  $\hat{\gamma}_{ij}$  and  $\hat{e}_{it}$  are obtained from (8). The resulting semiparametric kernel estimator of  $\alpha_i(1)$  will henceforth be denoted  $\hat{\alpha}_i(1)$ .

The third step is to compute the test statistics as follows

$$G_\tau = \frac{1}{N} \sum_{i=1}^N \frac{\hat{\alpha}_i}{SE(\hat{\alpha}_i)} \quad \text{and} \quad G_\alpha = \frac{1}{N} \sum_{i=1}^N \frac{T\hat{\alpha}_i}{\hat{\alpha}_i(1)},$$

where  $SE(\hat{\alpha}_i)$  is the conventional standard error of  $\hat{\alpha}_i$ .

A few remarks are in order. Firstly, although asymptotically not an issue, the normalization of  $G_\alpha$  by  $T$  may cause the test to reject the null too frequently, especially when the number of lags is comparably large. In such cases, one may want to replace  $T$  with the effective number of observations per individual, which is expected to produce better performance in small samples without affecting the asymptotic properties of the test.

Secondly, note the form of the individual quantities  $T\hat{\alpha}_i/\hat{\alpha}_i(1)$  making up the  $G_\alpha$  statistic. This is different from the coefficient statistic proposed by Banerjee *et al.* (1998), which is just  $T\hat{\alpha}_i$ . Apparently, the effect of allowing for serial correlation in  $\Delta y_{it}$  does not vanish asymptotically as claimed by these authors but impacts the test through  $\hat{\alpha}_i(1)$ , which seems very reasonable given

the results presented by Xiao and Phillips (1998) for the coefficient version of the augmented unit root test.

Thirdly, note also that the above estimation procedure does not account for any deterministic terms, and that it needs to be modified to accommodate the constant and trend in Cases 2 and 3, respectively. This requires replacing  $\Delta y_{it}$  in  $\omega_{y_i}^2$  by the fitted residuals from a first-stage regression of  $\Delta y_{it}$  onto  $d_t$ , the vector of deterministic components.

### 3.2 The panel statistics

The panel statistics are complicated by the fact that the both the parameters and dimension of (7) are allowed to differ between the cross-sectional units, and we therefore suggest a three-step procedure to implement these tests.

The first step is the same as for the group mean statistics and involves determining  $p_i$ , the individual lag order. Once  $p_i$  has been determined, we regress  $\Delta y_{it}$  and  $y_{it-1}$  onto  $d_t$ , the lags of  $\Delta y_{it}$  as well as the contemporaneous and lagged values of  $\Delta x_{it}$ . This yields the projection errors

$$\Delta \tilde{y}_{it} = \Delta y_{it} - \hat{\delta}'_i d_t - \hat{\lambda}'_i x_{it-1} - \sum_{j=1}^{p_i} \hat{\alpha}_{ij} \Delta y_{it-j} - \sum_{j=0}^{p_i} \hat{\gamma}_{ij} \Delta x_{it-j},$$

and

$$\tilde{y}_{it-1} = y_{it-1} - \tilde{\delta}'_i d_t - \tilde{\lambda}'_i x_{it-1} - \sum_{j=1}^{p_i} \tilde{\alpha}_{ij} \Delta y_{it-j} - \sum_{j=0}^{p_i} \tilde{\gamma}_{ij} \Delta x_{it-j}.$$

The second step involves using  $\Delta \tilde{y}_{it}$  and  $\tilde{y}_{it-1}$  to estimate the common error correction parameter  $\alpha$  and its standard error. In particular, we compute

$$\hat{\alpha} = \left( \sum_{i=1}^N \sum_{t=2}^T \tilde{y}_{it-1}^2 \right)^{-1} \sum_{i=1}^N \sum_{t=2}^T \frac{1}{\hat{\alpha}_i(1)} \tilde{y}_{it-1} \Delta \tilde{y}_{it}.$$

The standard error of  $\hat{\alpha}$  is given by

$$SE(\hat{\alpha}) = \left( (\hat{S}_N^2)^{-1} \sum_{i=1}^N \sum_{t=2}^T \tilde{y}_{it-1}^2 \right)^{-1/2} \quad \text{where} \quad \hat{S}_N^2 = \frac{1}{N} \sum_{i=1}^N \hat{S}_i^2.$$

Let  $\hat{\sigma}_i$  denote the estimated regression standard error in (8). The quantity  $\hat{S}_i$  is defined as  $\hat{\sigma}_i / \hat{\alpha}_i(1)$ , which is a consistent estimate of the population counterpart  $\sigma_i / \alpha_i(1)$ , the long-run standard deviation of  $\Delta y_{it}$  conditional on all current and past values of  $\Delta x_{it}$ .

The third step is to compute the panel statistics as

$$P_\tau = \frac{\hat{\alpha}}{SE(\hat{\alpha})} \quad \text{and} \quad P_\alpha = T \hat{\alpha}.$$

As a small-sample refinement, similar to the group mean coefficient statistic  $G_\alpha$ ,  $P_\alpha$  may be normalized by the cross-sectional average of the effective number of observations per individual rather than by  $T$ .

## 4 Asymptotic results

In this section, we study the asymptotic properties of our new panel statistics. In so doing, we make use of the sequential limit theory, which is a convenient method for evaluating the limit of a double indexed process.<sup>5</sup> For simplicity, stochastic integrals such as  $\int_0^1 W_i(r)dr$  will be written  $\int_0^1 W_i$  with the measure of integration omitted.

For the asymptotic null distributions of the tests, it is useful to define

$$\begin{aligned} C_i &= \left( \int_0^1 U_i^2, \int_0^1 U_i dV_i \right)', \\ \tilde{C}_i &= \left( \left( \int_0^1 U_i^2 \right)^{-1} \int_0^1 U_i dV_i, \left( \int_0^1 U_i^2 \right)^{-1/2} \int_0^1 U_i dV_i \right)', \end{aligned}$$

where

$$U_i = V_i - \left( \int_0^1 V_i W_i^{d'} \right) \left( \int_0^1 W_i^d W_i^{d'} \right)^{-1} W_i^d.$$

Here,  $V_i$  and  $W_i$  are two scalar and  $K$  dimensional standard Brownian motions that are independent of each other. Furthermore, in order to succinctly express the limiting distributions of our new test statistics when  $d_t$  is nonempty, it is useful to let  $W_i^d = (d', W_i')'$ , where  $d$  is the limiting trend function. In particular,  $d = 0$  in Case 1,  $d = 1$  in Case 2 and  $d = (1, r)'$  in Case 3. Note that the vector  $W_i^d$  enters  $C_i$  and  $\tilde{C}_i$  through the Brownian motion functional  $U_i$ , which is the projection of  $V_i$  onto the space orthogonal to the vector  $W_i^d$ . It is also useful to let  $\Theta$  and  $\tilde{\Theta}$  denote the expected values of  $C_i$  and  $\tilde{C}_i$ , respectively, and to let  $\Sigma$  and  $\tilde{\Sigma}$  denote their variances.

Now, define  $\phi = (-\Theta_2/\Theta_1^2, 1/\Theta_1)'$  and  $\varphi = (-\Theta_2/(2\Theta_1^{3/2}), 1/\sqrt{\Theta_1})'$ . As indicated by the following theorem, when the test statistics are normalized by the appropriate value  $N$ , their asymptotic distributions only depend on the known values of  $\Theta$ ,  $\tilde{\Theta}$ ,  $\Sigma$ ,  $\tilde{\Sigma}$ ,  $\phi$  and  $\varphi$ .

**Theorem 1.** (Asymptotic distribution.) Under Assumption 1 and the null hypothesis  $H_0$ , as  $T \rightarrow \infty$  and then  $N \rightarrow \infty$  sequentially

$$\sqrt{N}G_\alpha - \sqrt{N}\tilde{\Theta}_1 \Rightarrow N(0, \tilde{\Sigma}_{11}),$$

---

<sup>5</sup>Because limit arguments are taken as  $T \rightarrow \infty$  and then  $N \rightarrow \infty$ , this implies that the new tests may be justified in cases where  $T$  is substantially larger than  $N$ .

$$\begin{aligned}
\sqrt{N}G_\tau - \sqrt{N}\tilde{\Theta}_2 &\Rightarrow N(0, \tilde{\Sigma}_{22}), \\
\sqrt{N}P_\alpha - \sqrt{N}\Theta_2\Theta_1^{-1} &\Rightarrow N(0, \phi'\Sigma\phi), \\
P_\tau - \sqrt{N}\Theta_2\Theta_1^{-1/2} &\Rightarrow N(0, \varphi'\Sigma\varphi).
\end{aligned}$$

**Remark 1.** Theorem 1 is proven in the appendix but it is instructive to consider briefly why it holds. The proof for the group mean statistics is particularly simple and proceeds by showing that the intermediate limiting distribution of the normalized statistics passing  $T \rightarrow \infty$  for a fixed  $N$  can be written entirely in terms of the elements of the vector Brownian motion functional  $\tilde{C}_i$ . Therefore, by subsequently passing  $N \rightarrow \infty$ , asymptotic normality follows by direct application of the Lindeberg-Lévy central limit theorem (CLT) to a sum of i.i.d. random variables. The proof for the panel statistics is similar. It proceeds by showing that the intermediate limiting distribution of the normalized statistics can be described in terms of differentiable functions of i.i.d. vector sequences to which the Delta method is applicable. Hence, taking the limit as  $N \rightarrow \infty$ , we obtain a limiting normal distribution for the panel statistics too.

**Remark 2.** Theorem 1 indicates that each of the normalized statistics, when standardized by the appropriate moments, converges to a standard normal distribution. Thus, to be able to make inference based on the normal distribution, we must first obtain the moments for each statistic. This can be done by Monte Carlo simulations. For this purpose, we make 10,000 draws of  $K + 1$  independent, scaled, random walks of length  $T = 1,000$ . By using these random walks as simulated Brownian motions, we construct approximations of the vector Brownian motion functionals  $C_i$  and  $\tilde{C}_i$ . The means and the variances of these simulated functionals are then used to approximate the asymptotic moments. The results obtained from this exercise are reported for up to six regressors in Table 1.

**Remark 3.** In view of Table 1, note that, although the distributions of the statistics are free of nuisance parameters, they do depend upon the deterministic specification of the test regression in (7), and on the number of regressors as reflected by dependence of  $U_i$  on  $W_i^d$ . Thus, the moments will also depend on the deterministic specification and on the number of regressors. Moreover, notice that the distributions are independent of the short-run dynamics of the data generating process as captured by the first differences of the regressors. Thus, the statistics are asymptotically similar with respect to the short-run parameters in (7). In Table 1 we therefore only report simulated moments for the different deterministic cases and for different numbers of regressors. There is no need to tabulate separate moments for different lag orders.

**Remark 4.** As in the conventional unit root testing, the test regression in (7) involves estimating redundant trend parameters under the null. For example, in Case 2, the parameter  $\delta_{1i}$  vanishes under the restriction that  $\alpha_i = 0$  but it

determines the mean of the error correction term  $y_{it-1} - \beta_i' x_{it-1}$  when  $\alpha_i < 0$ . This redundancy turns up in the properties of the error corrections tests, which then depend on  $\delta_{1i}$  even under the null as seen by the dependence of  $W_i^d$  on  $d$ . Under the alternative, the tests are consistent against cointegration with a constant but they are inconsistent against the more general alternative of cointegration with trend as equation (7) in Case 1 does not allow for a trend under the alternative. Thus, the tests permit for the constant term under the alternative by introducing  $\delta_{1i}$  in (7), which is irrelevant under the null.

**Remark 5.** So far, we have maintained Assumption 1 (c), which requires that  $x_{it}$  is strictly exogenous. This is not necessary. In fact, as pointed out by Banerjee *et al.* (1998), Assumption 1 (c) can be readily relaxed by making the error correction model in (4) conditional not only on the lags but also on the leads of the first difference of the regressors. The intuition is simple, and follows by the familiar dynamic least squares regression arguments of Saikkonen (1991). In particular, given that  $x_{it}$  is weakly exogenous with respect to  $\alpha_i$  and  $\beta_i$ ,  $e_{it}$  must be uncorrelated with all current and past values of  $v_{it}$ . But this does not prevent  $e_{it}$  from being correlated with the future values of  $v_{it}$ , which makes the new tests dependent on nuisance parameters reflecting this correlation. Therefore, in order to eliminate this dependency in case Assumption 1 (c) does not hold, we augment (4), and hence equations (6) through (8), not only with the lags but also with the leads of  $v_{it}$ . In what follows, to be able to accommodate different lag and lead orders, we shall use  $q_i$  to denote the number of leads for each cross-section.<sup>6</sup>

It is important that a statistical test is able to fully discriminate between the null and alternative hypotheses in large samples. The next theorem shows that the test statistics are consistent and that they are divergent under the alternative hypothesis.

**Theorem 2.** (Test consistency.) Under Assumption 1 and the alternative hypotheses  $H_1^g$  and  $H_1^p$ , then  $G_\alpha$ ,  $G_\tau$ ,  $P_\alpha$  and  $P_\tau$  diverges towards negative infinity.

**Remark 6.** The proof of Theorem 2 is provided in the appendix. The theorem establishes that the divergence occurs towards negative infinity, suggesting that the tests can be constructed as one-sided using only the left tail of the normal distribution to reject the null hypothesis. Therefore, to test the null hypothesis of no cointegration based on the moments from Table 1, one simply computes the value of the standardized test statistic so that it is in the form specified

---

<sup>6</sup>In contrast to the new tests, the  $t$ -ratio and coefficient type statistics of Pedroni (2004) do not require any correction to account for the absence of strict exogeneity, at least not asymptotically. Thus, in this respect, the residual-based approach is computationally more convenient. As we will show later in Section 6, however, the price in terms of small-sample performance of this greater convenience can be quite large.

in Theorem 1. This value is then compared with the left tail of the normal distribution. Large negative values imply that the null hypothesis should be rejected.

**Remark 7.** The proof of Theorem 2 uses the sequential limit theory. Although this allows for a relatively straightforward and tractable analysis, it cannot be used to obtain the joint rate of divergence under the two alternatives  $H_1^g$  and  $H_1^p$ , which is indicative of the relative power properties of the tests. It is, however, possible to establish the order of the statistics as  $T \rightarrow \infty$  for a fixed  $N$ . In this case, it is shown in the appendix that  $G_\alpha$  and  $P_\alpha$  are  $O_p(T)$  while  $G_\tau$  and  $P_\tau$  are  $O_p(\sqrt{T})$ , which is in agreement with the results obtained by Phillips and Ouliaris (1990) for their residual-based time series tests. Given their faster rate of divergence, it is likely that the coefficient tests  $G_\alpha$  and  $P_\alpha$  have higher power than  $G_\tau$  and  $P_\tau$  in samples where  $T$  is substantially larger than  $N$ .

## 5 Cross-sectional dependence

In this section, the results of the previous sections are generalized to account for dependence between the cross-sectional units.

As pointed out in Section 2, the our error correction tests are based on the assumption of independence, or at least zero correlation, over the cross-sectional units. Although the potential effects of the breakdown of this assumption is by now well understood, the allowance of such dependence has yet to become standard in the panel cointegration literature.

One solution would be to use data that has been demeaned with respect to common time effects. Thus, in this case, the statistics are calculated as before but with  $\tilde{x}_{it} = x_{it} - \frac{1}{N} \sum_{i=1}^N x_{it}$  in place of  $x_{it}$  and  $\tilde{y}_{it} = y_{it} - \frac{1}{N} \sum_{i=1}^N y_{it}$  in place of  $y_{it}$ , which does not alter their asymptotic distributions. However, as demonstrated by Westerlund (2005b), although very convenient, subtracting the cross-sectional average in this way may not work very well, and may in fact result in a deterioration of the small-sample performance of the test.<sup>7</sup> Another disadvantage of this approach is that it do not permit for correlations that differ between pairs of individual time series, which seem like a more realistic assumption in many empirical applications.

One possible response to this is to employ the bootstrap approach, which makes inference possible even under very general forms of cross-sectional dependence. The particular bootstrap opted for in this section resembles that used by Chang (2004) and proceeds as follows.

---

<sup>7</sup>Unreported simulation results suggest that, although the cross-sectional demeaning is able to mitigate some effects of the dependence, the sizes of these tests can be very unreliable with massive distortions in many cases.



The first step is to fit the following least squares regression

$$\Delta y_{it} = \sum_{j=1}^{p_i} \hat{\alpha}_{ij} \Delta y_{it-j} + \sum_{j=0}^{p_i} \hat{\gamma}_{ij} \Delta x_{it-j} + \hat{e}_{it}. \quad (9)$$

By using the residuals from this regression we then form the vector  $\hat{w}_t = (\hat{e}_t', \Delta x_t')'$ , where  $\hat{e}_t = (\hat{e}_{1t}, \dots, \hat{e}_{Nt})'$  and  $\Delta x_t = (\Delta x_{1t}', \dots, \Delta x_{Nt}')'$ . It is this vector that forms the basis for our bootstrap tests, and it is important that it is obtained from equation (9) with the null hypothesis imposed. Otherwise, if the null hypothesis is not imposed, this will render the subsequent bootstrap tests inconsistent. Note that, in Case 3, then (9) should be augmented with a constant to account for the trend in  $y_{it}$ .

We then generate bootstrap samples  $w_t^* = (e_t^*, \Delta x_t^*)'$  by sampling with replacement the centered residual vector

$$\tilde{w}_t = \hat{w}_t - \frac{1}{T-1} \sum_{j=1}^T \hat{w}_j.$$

Note that by resampling  $\tilde{w}_t$  rather than  $\tilde{w}_{it}$ , we can preserve not only the cross-sectional correlation structure of  $e_{it}$  but also any endogenous effects that may run across the individual regressions of the system.

The next step is to generate the bootstrap sample  $\Delta y_{it}^*$ . This is accomplished by first constructing the bootstrap version of the composite error  $u_{it}$  as

$$u_{it}^* = \sum_{j=0}^{p_i} \hat{\gamma}_{ij} \Delta x_{it-j}^* + e_{it}^*,$$

where the least squares estimate  $\hat{\gamma}_{ij}$  is obtained from the regression in (9). Given  $p_i$  initial values, we then generate  $\Delta y_{it}^*$  recursively from  $u_{it}^*$  as

$$\Delta y_{it}^* = \sum_{j=1}^{p_i} \hat{\alpha}_{ij} \Delta y_{it-j}^* + u_{it}^*, \quad (10)$$

where  $\hat{\alpha}_{ij}$  is again obtained from (9). For initial values, we may use the first  $p_i$  observations of  $\Delta y_{it}$ . However, as pointed out by Chang (2004), this does not ensure the stationarity of  $\Delta y_{it}^*$ . As an alternative approach, we may generate a larger number,  $T+n$  say, of  $\Delta y_{it}^*$  and then discard the first  $n$  values. Because this makes the initiation unimportant, we may simply use zeros to start up the recursion.

Finally, we generate  $y_{it}^*$  and  $x_{it}^*$  with the null hypothesis imposed. This is to ensure the spuriousness of the generated sample as claimed under the null, and to make the resulting bootstrap tests valid. Thus, we obtain  $y_{it}^*$  and  $x_{it}^*$  as

$$y_{it}^* = y_{i0}^* + \sum_{j=1}^t \Delta y_{ij}^* \quad \text{and} \quad x_{it}^* = x_{i0}^* + \sum_{j=1}^t \Delta x_{ij}^*,$$

which again requires initiation through  $x_{i0}^*$  and  $y_{i0}^*$ . The value zero will do.

Having obtained the bootstrap sample  $y_{it}^*$  and  $x_{it}^*$ , we then obtain the bootstrapped error correction statistic of interest, which is constructed exactly as its sample counterpart but with  $y_{it}^*$  and  $x_{it}^*$  in place of  $y_{it}$  and  $x_{it}$ . Denote this initial bootstrap statistic by  $t_1^*$ . If we repeat this procedure  $S$  times, say, we obtain  $t_j^*$  for  $j = 1, \dots, S$ , which is the bootstrap distribution of the statistic. For a one-sided 5% nominal level test, we then obtain the lower 5% quantile  $t_C^*$ , say, of the bootstrap distribution. The null hypothesis is rejected if the calculated sample value of the statistic is smaller than  $t_C^*$ . Also, note that in case the regressors are assumed to be weakly but not necessarily strictly exogenous, then the above bootstrap algorithm has to be modified to account for the leads of  $\Delta x_{it}$ .

## 6 Monte Carlo simulations

In this section, we study the small-sample properties of the new tests relative to those of some of the popular residual-based tests recently proposed by Pedroni (2004). For this purpose, a large number of experiments were performed using the following process to generate the data

$$\Delta y_{it} = \alpha(y_{it-1} - \beta_i x_{it-1}) + \sum_{j=-q}^p \gamma \Delta x_{it-j} + e_{it}, \quad (11)$$

$$\Delta x_{it} = v_{it}. \quad (12)$$

For simplicity, we assume that there is a single regressor, and that there is a common error correction parameter  $\alpha$ . Hence,  $\alpha = 0$  under the null hypothesis, while  $\alpha < 0$  under the alternative. The results are organized in two parts depending on whether there is any cross-section dependence present or not.

In the first part with no cross-section dependence, we have three scenarios, which are all based on the following moving average process

$$e_{it} = u_{it} + \phi u_{it-1}.$$

The innovations  $u_{it}$  and  $v_{it}$  are both assumed to be normal variables with zero mean and, unless otherwise stated, unit variance. The first two scenarios that we consider are concerned with the size of the tests, while the third is concerned with the power. The first scenario resembles the data generating process used by Pedroni (2004), and is designed to study the effects of serial correlation. In this case,  $\gamma = 0$  and  $\phi \neq 0$ . By contrast, in the second scenario,  $\phi = 0$  and  $\gamma \neq 0$ . This scenario is taken from Westerlund (2005c), and the purpose is to evaluate the performance of the tests when the regressor is not strictly exogenous. For simplicity,  $p$  and  $q$ , the true number of lags and leads, are both set to one.

Finally, in the third scenario, we investigate the relative power of the tests, in which case  $\phi$ ,  $p$  and  $q$  are all set to zero, so there is no serial correlation and only

the contemporaneous value of  $\Delta x_{it}$  enters (11). In particular, by setting  $\gamma = 1$ , we can use  $\beta_i$  to determine whether the common factor restriction is satisfied or not. If the restriction is satisfied, then  $\beta_i = 1$ , whereas if it is not satisfied,  $\beta_i \sim N(0, 1)$ . The degree of the violation is controlled by the signal-to-noise ratio  $\text{var}(v_{it})/\text{var}(u_{it})$ , which reduces to  $\text{var}(v_{it})$  if we assume that  $\text{var}(v_{it}) = 1$ .

The second part of this section is designed to study the effects of cross-sectional dependence, and is parameterized as follows

$$e_{it} = b(\lambda_i \Delta f_t) + u_{it},$$

where  $b$  is a scale parameter,  $\lambda_i$  is a loading parameter,  $f_t \sim N(0, 1)$  and  $u_{it} \sim N(0, 1)$ . Thus, in this setup,  $f_t$  is a common factor that generates dependence among the cross-sectional units. It enters (11) in first differences, which ensures that the common factor to  $y_{it}$  is stationary. For simplicity, we set  $\gamma$  to zero, so the regressor is strictly exogenous. All experiments are based on generating 2,000 panels with  $N$  individual and  $T + 50$  time series observations, where the first 50 observations for each series is discarded to attenuate the effect of the initial conditions, which are all set to zero.

For comparison, four of the residual-based tests of Pedroni (2004) are also simulated. Two are based on the group mean principle and are denoted  $\tilde{Z}_t$  and  $\tilde{Z}_\rho$ , while the corresponding panel statistics are denoted  $Z_t$  and  $Z_\rho$ . As explained earlier, the Pedroni (2004) tests are based on the test regression in (5) and are thus subject to the common factor critique. Analogous to the new tests,  $\tilde{Z}_t$  and  $Z_t$  are constructed as  $t$ -ratios, while  $\tilde{Z}_\rho$  and  $Z_\rho$  are coefficient statistics. All four tests are semiparametric with respect to the heteroskedasticity and serial correlation properties of the data.

As recommended by Newey and West (1994), all tests are constructed with the bandwidth chosen according to the rule  $4(T/100)^{2/9}$ . For the number of lags and leads, we used two different rules, one depends on the data while the other depends on  $T$ . Specifically, for former, we used the Akaike information criterion, while for the latter, we used  $2(T/100)^{2/9}$ , which gives an overall lag and lead expansion of the usual Newey and West (1994) rate  $4(T/100)^{2/9}$ . Consistent with the results of Ng and Perron (1995), the maximum number of lags and leads for the Akaike criterion is permitted to grow with  $T$  at the same rate. The bootstrap tests are implemented using 200 bootstrap replications.

To keep the amount of table space manageable, we present only the size-adjusted power and the empirical size on the 5% level for Case 1 when there is no deterministic component. The results for the other cases were similar and did not change the conclusions. These are therefore omitted. Computational work was performed in GAUSS.

## 6.1 No cross-sectional dependence

In this case, there is no cross-sectional dependence, and we only consider the tests without the bootstrap proposal. Instead, we focus on the effects of serial

correlation and the absence of strict exogeneity, and on the common factor issue.

Consider first the results on the size of the tests when  $\alpha = 0$ , which are reported in Table 2. The data in the first scenario are generated with moving average errors, which makes  $\phi$  a convenient nuisance parameter to investigate. It has been well documented in simulation studies such as Haug (1996) that highly persistent moving average errors may cause substantial size distortions when testing the null of no cointegration in time series using residual-based tests with semiparametric type serial correlation corrections. Ng and Perron (1997) discuss the root of this problem to some length, and argue that it is the bias of the least squares estimate for the parameters of the spurious regression, and subsequently the autoregressive parameter of the errors, that causes these distortions. This bias is then imported into the semiparametric serial correlation correction, where it is further exacerbated by the difficulty inherent in long-run variance estimation.<sup>8</sup>

In agreement with these results, Table 2 show that the Pedroni (2004) tests tend to be severely oversized when  $\phi$  is nonzero, especially when it is negative. In fact, with these tests, the results indicate that size can be up to 19 times the nominal 5% level. At the other end of the scale, we have the  $G_\tau$  and  $P_\tau$  statistics, which have relatively small distortions. The  $G_\alpha$  and  $P_\alpha$  statistics lie somewhere in between. As expected, all tests perform well when  $\phi = 0$  and there is no persistence. We also see that there is not much difference between the error correction tests based on the two lag and lead rules, although the size is generally best for the tests based on choosing the number of lags and leads as a function of  $T$ .

Given the findings of Ng and Perron (1997), a natural interpretation if these results is that the error correction test are less distorted because they do not rely on getting a good first-stage estimate of the level relationship. This lessens the bias induced not only in the estimation of the autoregressive parameter, but also in the semiparametric serial correlation correction. In other words, it seem likely that the error correction tests are less distorted because the bias is not compounded in the same way as for the residual-based tests.

The results from the second scenario are reported in the two bottom panels of Table 2. As expected, we see that the new tests are generally able to appropriately correct for the fact that the regressor is no longer strictly exogenous. Among these tests, we see that  $G_\tau$  and  $P_\tau$  generally perform best, which is in agreement with the results from the first scenario. Of all the tests considered, the residual-based tests perform worst. In fact, based on the results reported in Table 2, rejection frequencies of up to 10 times the nominal 5% level are not uncommon, even when  $\gamma = 0.4$ , so the strict exogeneity assumption is only moderately violated. Increasing  $\gamma$  from 0.4 to 0.8 almost uniformly result in the

---

<sup>8</sup>Pedroni (2004) also discusses the difficulty in handling the effects of highly persistent moving average errors, and suggests modifying his tests along the lines of Ng and Perron (1997), which is expected to reduce the size distortions even in the panel context.

size of the residual-based tests going to 100%.

Although theoretically somewhat unexpected, these results are nevertheless consistent with the poor performance of the residual-based tests in the first scenario with serial correlation. Apparently, the problem with bias effects extends also to the case with violations of strict exogeneity. In view of this, it seems desirable to employ some form of dynamic or fully modified least squares correction even for the residual-based tests.

The power of the tests is evaluated next. Due to the different size properties of the tests, all powers are adjusted for size. Moreover, to be able to isolate the effects of the common factor restriction, both  $\gamma$  and  $\phi$  are set equal to zero. The results are reported in Table 3 for the case when  $\alpha = -0.01$ , and in Table 4 for the case when  $\alpha = -0.03$ .

It is seen that the new tests are almost uniformly more powerful than the residual-based tests by Pedroni (2004). In fact, the results suggest that the new tests are the most powerful even when the common factor restriction is satisfied, which is also what Banerjee *et al.* (1998) find when simulating the time series versions of these tests. This is somewhat unexpected since the residual-based tests should be more efficient in this case. One possible explanation for this result is that the error correction tests are even more efficient, as they exploit the fact that  $x_{it}$  is weakly exogenous in this setup.

Although somewhat unexpected in the common factor case, the greater power of the error correction tests in the no common factor case is well in line with theory. In particular, in accordance with our discussion in Section 2, Tables 3 and 4 show that the relative power of the error correction tests increases monotonically as the signal-to-noise ratio increases. This effect is further magnified by the fact that the power of the residual-based tests appears to be decreasing in  $\text{var}(v_{it})$ , which is to be expected as large values of  $\text{var}(v_{it})$  will tend to inflate the test regression in (5) with excess volatility.

The panel tests have the highest power, which is not surprising since they are based on the pooled least squares estimator of the  $\alpha$ , and pooling is efficient under the homogenous alternative considered here. Among the panel tests, we see that  $P_\alpha$  has the highest power when the common factor restriction is satisfied, but that  $P_\tau$  becomes more powerful as  $\text{var}(v_{it})$  increases and the restriction is violated. Among the group mean tests, the results suggest that  $G_\tau$  has the highest power.

## 6.2 Cross-section dependence

In the cross-section dependent case, we use the scale parameter  $b$  to control the dependence. If  $b = 0$ , then  $e_{it}$  reduces to  $u_{it}$  and there is no dependence, whereas if  $b = 1$ , then  $e_{it} = \lambda_i \Delta f_t + u_{it}$ , which implies that  $E(e_{it}e_{jt}) = \lambda_i^2$  if  $i \neq j$  so the cross-sectional units are dependent. The results are reported in Table 5, where we use the star superscript to indicate the bootstrapped versions of the error correction statistics. Also, in this subsection, we drop the Akaike type

error correction tests and focus on the tests based setting the number of lags and leads as a function of  $T$ , which greatly reduces the number of computations required.

Consider first the results on the size of the tests. As expected, we see that all tests perform well with only small distortions when  $b = 0$  and there is no cross-sectional dependence. However, as  $b$  departs from zero, we see that there is a strong tendency for the asymptotic tests to overreject the null hypothesis. We also see that the magnitude of these distortions varies quite significantly. At one end of the scale, we have the  $G_\tau$  and  $P_\tau$  tests, which actually appear to be quite robust to the cross-section correlation. At the other end, we have the residual-based tests, where the results suggest that the size can be very unreliable with severe distortions in most cases. As expected, the bootstrapped tests perform very well even under cross-sectional dependence.

Consider next the power of the tests. There are several things that are noteworthy. First, the power of the bootstrapped error correction tests is generally comparable with the power of their asymptotic counterparts. This result is very interesting because it suggests that we can correct the size of the tests with little or no cost in terms of power. Second, the residual-based tests are generally the least powerful, which corroborates the results reported in Tables 3 and 4.

### 6.3 Conclusions and their robustness

The simulation results presented in this section suggest that, under the maintained assumption of weak exogeneity, the new tests perform well with good size and power in most panels. In particular, we find that the error correction tests have both better size accuracy and vastly superior power in comparison to the residual-based tests. We also find that the bootstrapped versions of the new tests are very effective in eliminating the effects of the cross-sectional dependence without sacrificing power.

These findings appear to be very robust, and extend to all sample sizes examined and to the cases with nonzero constant and trend terms. To also examine the effects of a violation of the weak exogeneity assumption, we carried out some simulations using the following process to generate the data

$$\begin{aligned}\Delta y_{it} &= (\alpha_y - \gamma\alpha_x)(y_{it-1} - \beta_i x_{it-1}) + \gamma\Delta x_{it} + e_{it}, \\ \Delta x_{it} &= \alpha_x(y_{it-1} - \beta_i x_{it-1}) + v_{it}.\end{aligned}$$

Note that this is (11) and (12) with  $p$  and  $q$  set to zero,  $\alpha = \alpha_y - \gamma\alpha_x$  and, perhaps most importantly, error correcting behavior for  $\Delta x_{it}$ . As before, we assume that  $\beta_i \sim N(0, 1)$  and  $\gamma = 1$ . We further disregard all effects of cross-section dependence and serial correlation by making both  $e_{it}$  and  $v_{it}$  a draw from the standard normal distribution.

Table 6 summarizes the results regarding the size and power of the tests at the 5% level with 10 cross-sectional units. In agreement with the discussion of

Section 2, we see that the error correction tests continues to perform well with good power when  $\alpha_x$  is positive, but that the power falls significantly when it is negative. In particular, we see that in the special case when both  $\alpha_x$  and  $\alpha_x$  are negative such that  $\alpha_x < \alpha_x$ , then the error correction tests have no power at all. The residual-based tests are also affected but not as dramatically as the error correction tests. Also, as expected, the size of the tests is not affected by the absence of weak exogeneity.

Thus, even though the error correction tests continues to perform well when  $\alpha_x$  is positive, in general we need the weak exogeneity assumption to ensure they work properly. In spite of this, however, given their grater robustness in terms of both size accuracy and power, the overall impression of the Monte Carlo evidence presented is that the new tests compare favorably with the Pedroni (2004) tests. Also, in case weak exogeneity does not hold, it is probably better to analyze the data using the system-based approach of Larsson *et al.* (2001).<sup>9</sup>

## 7 Health care expenditures and GDP

In this section, we apply the new tests to data on health care expenditures (HCE) and GDP to demonstrate the use of these tests and to reassess previous empirical findings.

The relationship between HCE and GDP is the subject of a large portion of the literature in health economics. Many early contributions employed cross-sectional data to obtain estimates of this relationship. Without exception, it has been found that most of the observed variation in HCE can be explained by variation in GDP. However, many of these studies have been criticized for the smallness of their data sets and for the assumption that HCE is homogeneously distributed across countries. More recent research has therefore resorted to panel data, which offers a number of advantages over pure cross-sectional data. For instance, using multiple years of data increases the sample size while simultaneously allowing researchers to control for a wide range of time invariant country characteristics through the inclusion of country specific constants and trends. In addition, with multiple time series observations for each country, researchers are able to exploit the presence of unit roots and cointegration in HCE and GDP.

---

<sup>9</sup>It should be noted, however, that there are simple ways to reduce the problem of endogeneity, even for error correction based tests. For example, as pointed out by Zivot (2000), in situations with  $\alpha_i$  positive, a natural solution would be to simply perform a reverse test. That is, the error correction tests are implemented in a conditional model for  $x_{it}$  rather than for  $y_{it}$ . Zivot (2000) also presents several reasons for why weak exogeneity may not be too much of a problem in practice. One reason is that it can be readily tested as a restriction on the unconditional model, which in the current panel data setting corresponds to the panel vector error correction model studied by Larsson *et al.* (2001). Another reason is that there appears to be strong support for the weak exogeneity assumption in many applications, see Zivot (2000) and the references therein.

This avenue is taken by Hansen and King (1996), who examine a panel spanning the years 1960 to 1987 across 20 OECD countries. They show that, if one examines the time series for each of the countries separately, one can only rarely reject the unit root hypothesis for either HCE or GDP. Moreover, their country specific tests rarely reject the hypothesis of no cointegration. McCoskey and Selden (1998) use the same data set as Hansen and King (1996). Based on the panel unit root proposed by Im *et al.* (2003), the authors are able to reject the presence of a unit root in both HCE and GDP. Once a linear time trend has been accommodated, however, the null hypothesis cannot be rejected.

Hansen and King (1998) question the preference of McCoskey and Selden (1998) for omitting the time trend from their main results, and argue that this may lead to misleading inference. Indeed, using a panel covering 24 OECD countries between 1960 and 1991, Blomqvist and Carter (1997) challenge the findings of McCoskey and Selden (1998). Drawing on a battery of tests, including the panel unit root test of Levin *et al.* (2002), the authors conclude that HCE and GDP appear to be nonstationary and cointegrated. Gerdttham and Löthgren (2000) present confirmatory evidence using a panel of 21 OECD countries between 1960 and 1997. Similarly, using a panel of 10 OECD member countries over the period 1960 to 1993, Roberts (2000) found clear evidence suggesting that HCE and GDP are nonstationary variables. The results of cointegration were, however, not conclusive.

Apparently, although the evidence seems to support the unit root hypothesis for HCE and GDP, it is less conclusive on the cointegration hypothesis. One possible explanation for these differences may be the common factor restriction implicitly imposed when testing the null hypothesis of no cointegration using residual-based procedures as in Hansen and King (1996).<sup>10</sup>

In this section, we verify this conjuncture by using a panel consisting of 20 OECD countries and covering the period 1970 to 2001. For this purpose, data on annual frequency has been acquired through the OECD Health Data 2003 database. Both HCE and GDP are measured in per capita terms at constant 1995 prices and are transformed in logarithms. Moreover, since both variables are clearly trending, we follow the earlier literature and model HCE and GDP with a linear time trend in their levels. An obvious interpretation of such a trend is that it accounts, in part, for the impact of technological change.

The basic model we postulate between HCE and GDP, denoted  $H_{it}$  and  $Y_{it}$ , respectively, is the following simple log-linear relationship

$$\ln(H_{it}) = \mu_i + \tau_i t + \beta_i \ln(Y_{it}) + e_{it}. \quad (13)$$

The first step in our analysis of this relationship is to test whether the variables are nonstationary or not. To this effect, we employ a battery of unit root tests, which can be classified into two groups depending on whether they allow for

---

<sup>10</sup>Since the data sets used in the previous studies are nearly identical, any differences in test results are not likely to be due to differences in the process generating the data.



cross-sectional correlation or not. The first group consists of the  $Z_t$  test of Im *et al.* (2003) and the Harris and Tzavalis (1999)  $\hat{\varphi}_{lsdv}$  test, which are both based on the assumption of no cross-sectional dependence.

However, like many other macroeconomic variables, HCE and GDP usually exhibit strong comovements across countries. The second group of tests allows for such cross-sectional dependence by assuming that the variables admit to a common factor representation. It includes the  $G_{ols}^{++}$ ,  $P_m$  and  $Z$  tests of Phillip and Sul (2003), the  $t_a$  and  $t_b$  tests of Moon and Perron (2004) and the Bai and Ng (2004)  $P_e^c$  test.

All tests are normally distributed under the common null hypothesis of non-stationarity, and all tests except  $\hat{\varphi}_{lsdv}$ ,  $t_a$  and  $t_b$  permit the individual autoregressive roots to differ across the cross-sectional units. The direction of the divergence under the alternative hypothesis determines whether we should use the right or left tail of the normal distribution to reject the null. The  $Z_t$ ,  $\hat{\varphi}_{lsdv}$ ,  $G_{ols}^{++}$ ,  $P_m$ ,  $t_a$  and  $t_b$  statistics diverge to negative infinity and are compared to the left tail, whereas the  $Z$  and  $P_e^c$  statistics diverge to positive infinity, and are thus compared to the right tail.

For the implementation of the tests, we use the Bartlett kernel, and all bandwidths and lag lengths are chosen according to  $4(T/100)^{2/9}$ . The number of lags and leads are chosen by the Akaike criterion. To determine the number of common factors, we use the Bai and Ng (2004)  $IC_1$  criterion with a maximum of five factors. The test results reported in Table 7 indicate that in only three out of the 16 cases do we end up with a rejection of the null at the 1% level of significance. We therefore conclude that the variables appear to be nonstationary.

The second step in our empirical analysis is to test whether HCE and GDP are cointegrated. One way to do this is to use the residual-based approach, and to subject the regression residuals from (13) to a unit root test. The results presented in Table 7 based on using this procedure suggest that we cannot reject the null hypothesis of no cointegration at the 1% level for any of the residual-based tests. In fact, even if we look at the most liberal 10% level, we end up rejecting the null for three out of the four tests. As pointed out earlier, however, the prospect of imposing an invalid common factor restriction may well result in this procedure having very low power in samples as small as ours. In that case, the error correction tests may be able to produce more powerful tests.

Our results confirm this conjuncture. In particular, unreported estimation results show that the standard errors of the individual test regressions for the residual-based tests are much larger in comparison to those of the corresponding error correction test regressions, which is indicative of an invalid common factor restriction. To tests whether this is in fact the case, we performed a series of individual Wald tests, which all resulted in a strong rejection of the common factor restriction.

The implication of these results is that the new tests may be more powerful. The calculated values of the error correction statistics are presented along with

both asymptotic and bootstrapped  $p$ -values in Table 7. When using the asymptotic  $p$ -values, we see that all four tests lead to a clear rejection of the null, even at the most conservative 1% level, which we take as strong evidence in favor of cointegration. There is no difference depending on whether  $\alpha_i$  is restricted to be equal for all  $i$  or not. Based on the bootstrapped  $p$ -values, we end up with one rejection, for  $P_\tau$ , at the 5% level. However, since this rejection is only marginal, and since the homogenous alternative hypothesis considered for this particular test may be overly restrictive, we choose to interpret these results as evidence in favor of cointegration between HCE and GDP.

## 8 Conclusions

In this chapter, we propose four new panel cointegration tests that are designed to test the null hypothesis of no cointegration by testing whether the error correction term in a conditional error correction model is equal to zero. If the null hypothesis of no error correction is rejected, then the null hypothesis of no cointegration is also rejected. Each test is able to accommodate individual specific short-run dynamics, including serially correlated error terms and non-strictly exogenous exogenous regressors, individual specific intercept and trend terms, as well as individual specific slope parameters. A bootstrap procedure is also proposed to handle applications with cross-sectionally dependent data.

Using sequential limit arguments, we show that the new tests have limiting normal distributions, and that they are consistent. These results are verified in small samples using Monte Carlo simulations. In particular, given that the regressors are weakly exogenous with respect to the parameters of interest, we find that the new tests show both better size accuracy and higher power than the residual-based tests recently developed by Pedroni (2004). We further show that this difference in power arise mainly because the residual-based tests ignore potentially valuable information by imposing a possibly invalid common factor restriction. On the other hand, we also find that the power of the new tests can be quite adversely affected if weak exogeneity fails, in which case the system-based approach of Larsson *et al.* (2001) is recommended.

In our empirical application, we provide evidence suggesting that HCE and GDP are cointegrated once the difference between the short- and long-run relationships has been accounted for.

## Appendix: Mathematical proofs

This appendix proves the asymptotic results for the new test statistics. For convenience, we shall prove the results for Case 1 with no deterministic components. For brevity, the notations introduced in the main text are taken as given and are thus not repeated. Moreover, for any variable  $z_{it}$ , we use  $z_{it}^*$  to indicate the error from projecting  $z_{it}$  onto  $x_{it-1}$ , and  $(z_{it})_p$  to indicate the error from projecting  $z_{it}$  onto  $w_{it} = (\Delta y_{it-1}, \dots, \Delta y_{it-p_i}, \Delta x'_{it}, \dots, \Delta x'_{it-p_i})'$ . Thus,  $(z_{it})_p = z_{it} - w'_{it}a_i$ , where  $a_i$  is a vector of projection parameters.

**Lemma A.1.** (Preliminaries for Theorem 1.) Under Assumption 1 and the null hypothesis, as  $T \rightarrow \infty$

$$\begin{aligned} \text{(a)} \quad T^{-1/2}y_{it} &\Rightarrow \frac{1}{\alpha_i(1)}(\gamma_i(1)'B_{2i} + B_{1i}); \\ \text{(b)} \quad T^{-2}\sum_{t=2}^T(y_{it-1}^*)_p^2 &\Rightarrow \frac{1}{\alpha_i(1)^2}\sigma_i^2\int_0^1 U_i^2; \\ \text{(c)} \quad T^{-1}\sum_{t=2}^T(y_{it-1}^*)_p(e_{it}^*)_p &\Rightarrow \frac{1}{\alpha_i(1)}\sigma_i^2\int_0^1 U_i dV_i. \end{aligned}$$

### Proof of Lemma A.1

Consider (a). The model (4) in Case 1 can be written as

$$\alpha_i(L)\Delta y_{it} = \alpha_i(y_{it-1} - \beta'_i x_{it-1}) + u_{it}, \quad (\text{A1})$$

where  $u_{it} = \gamma_i(L)'v_{it} + e_{it}$ . By using the Beveridge-Nelson (BN) decomposition of  $\gamma_i(L)$  as  $\gamma_i(L) = \gamma_i(1) + \gamma_i^*(L)(1-L)$ , we obtain

$$u_{it} = \gamma_i(L)'v_{it} + e_{it} = \gamma_i(1)'v_{it} + \gamma_i^*(L)'\Delta v_{it} + e_{it}.$$

Therefore, as  $T \rightarrow \infty$  we get

$$\begin{aligned} T^{-1/2}\sum_{j=2}^t u_{ij} &= \gamma_i(1)'T^{-1/2}\sum_{j=2}^t v_{ij} + \gamma_i^*(L)'T^{-1/2}v_{it} + T^{-1/2}\sum_{j=2}^t e_{ij} \\ &= \gamma_i(1)'T^{-1/2}\sum_{j=2}^t v_{ij} + T^{-1/2}\sum_{j=2}^t e_{ij} + o_p(1) \\ &\Rightarrow \gamma_i(1)'B_{2i} + B_{1i}. \end{aligned} \quad (\text{A2})$$

Next, we use the BN decomposition  $\alpha_i(L) = \alpha_i(1) + \alpha_i^*(L)(1-L)$  of  $\alpha_i(L)$ , which implies that

$$\alpha_i(L)\Delta y_{it} = \alpha_i(1)\Delta y_{it} + \alpha_i^*(L)\Delta^2 y_{it} = \alpha_i(y_{it-1} - \beta'_i x_{it-1}) + u_{it},$$

which, under the null, can be rewritten as

$$\Delta y_{it} = -\frac{\alpha_i^*(L)}{\alpha_i(1)}\Delta^2 y_{it} + \frac{1}{\alpha_i(1)}u_{it}.$$

This implies (a) as can be seen by writing

$$\begin{aligned} T^{-1/2}\sum_{j=2}^t\Delta y_{ij} &= -\frac{\alpha_i^*(L)}{\alpha_i(1)}T^{-1/2}\Delta y_{it} + \frac{1}{\alpha_i(1)}T^{-1/2}\sum_{j=2}^t u_{ij} \\ &= \frac{1}{\alpha_i(1)}T^{-1/2}\sum_{j=2}^t u_{ij} + o_p(1) \\ &\Rightarrow \frac{1}{\alpha_i(1)}(\gamma_i(1)'B_{2i} + B_{1i}), \end{aligned}$$

where the last result follows from (A2).

Next, consider (b). By using the rules for projections, the sum  $\sum_{t=2}^T(y_{it-1}^*)^2_p$  can be written as

$$\sum_{t=2}^T(y_{it-1}^*)^2_p = \sum_{t=2}^T y_{it-1}^{*2} - \sum_{t=2}^T y_{it-1}^* w_{it}^{*'} \left( \sum_{t=2}^T w_{it}^* w_{it}^{*'} \right)^{-1} \sum_{t=2}^T w_{it}^* y_{it-1}^*, \quad (\text{A3})$$

where

$$\begin{aligned} \sum_{t=2}^T y_{it-1}^* w_{it}^{*'} &= \sum_{t=2}^T y_{it-1} w'_{it} - \sum_{t=2}^T y_{it-1} x'_{it-1} \\ &\quad \cdot \left( \sum_{t=2}^T x_{it-1} x'_{it-1} \right)^{-1} \sum_{t=2}^T x_{it-1} w'_{it}. \end{aligned} \quad (\text{A4})$$

By Lemma 2.1 of Park and Phillips (1989), we have that  $\sum_{t=2}^T y_{it-1} w'_{it} = O_p(T)$ ,  $\sum_{t=2}^T y_{it-1} x'_{it-1} = O_p(T^2)$ ,  $\sum_{t=2}^T x_{it-1} x'_{it-1} = O_p(T^2)$  and  $\sum_{t=2}^T x_{it-1} w'_{it} = O_p(T)$ . Therefore,  $\sum_{t=2}^T y_{it-1}^* w_{it}^{*'} = O_p(T)$ .

Similarly,  $\sum_{t=2}^T w_{it}^* w_{it}^{*'}$  in (A3) may be expanded as

$$\begin{aligned} \sum_{t=2}^T w_{it}^* w_{it}^{*'} &= \sum_{t=2}^T w_{it} w'_{it} - \sum_{t=2}^T w_{it} x'_{it-1} \left( \sum_{t=2}^T x_{it-1} x'_{it-1} \right)^{-1} \sum_{t=2}^T x_{it-1} w'_{it} \\ &= O_p(T) + O_p(T)O_p(T^{-2})O_p(T) \\ &= O_p(T). \end{aligned} \quad (\text{A5})$$

Thus, by using (A3) to (A5), we can show that

$$\sum_{t=2}^T (y_{it-1}^*)^2_p = \sum_{t=2}^T y_{it-1}^{*2} + O_p(T)O_p(T^{-1})O_p(T)$$

$$= \sum_{t=2}^T y_{it-1}^{*2} + O_p(T), \quad (\text{A6})$$

For the remaining term in (A6) we use (a), which implies

$$\begin{aligned} T^{-1/2} y_{it-1}^* &= T^{-1/2} y_{it-1} - T^{-2} \sum_{t=2}^T y_{it-1} x'_{it-1} \\ &\cdot \left( T^{-2} \sum_{t=2}^T x_{it-1} x'_{it-1} \right)^{-1} T^{-1/2} x_{it-1} \\ &\Rightarrow \frac{1}{\alpha_i(1)} (\gamma_i(1)' B_{2i} + B_{1i}) - \frac{1}{\alpha_i(1)} \int_0^1 (\gamma_i(1)' B_{2i} + B_{1i}) B'_{2i} \\ &\cdot \left( \int_0^1 B_{2i} B'_{2i} \right)^{-1} B_{2i}, \end{aligned}$$

where the last expression can be manipulated to obtain

$$\begin{aligned} \frac{1}{\alpha_i(1)} B_{1i} - \frac{1}{\alpha_i(1)} \int_0^1 B_{1i} B'_{2i} \left( \int_0^1 B_{2i} B'_{2i} \right)^{-1} B_{2i} \\ &= \frac{1}{\alpha_i(1)} \sigma_i V_i - \frac{1}{\alpha_i(1)} \sigma_i \int_0^1 V_i W'_i \left( \int_0^1 W_i W'_i \right)^{-1} W_i \\ &= \frac{1}{\alpha_i(1)} \sigma_i U_i. \end{aligned}$$

Putting everything together, we get

$$T^{-2} \sum_{t=2}^T (y_{it-1}^*)_p^2 = T^{-2} \sum_{t=2}^T y_{it-1}^{*2} + o_p(1) \Rightarrow \frac{1}{\alpha_i(1)^2} \sigma_i^2 \int_0^1 U_i^2,$$

which establishes (b).

Finally, consider (c). We have

$$\begin{aligned} \sum_{t=2}^T (y_{it-1}^*)_p (e_{it}^*)_p &= \sum_{t=2}^T y_{it-1}^* e_{it}^* - \sum_{t=2}^T y_{it-1}^* w_{it}^{*'} \\ &\cdot \left( \sum_{t=2}^T w_{it}^* w_{it}^{*'} \right)^{-1} \sum_{t=2}^T w_{it}^* e_{it}^*. \quad (\text{A7}) \end{aligned}$$

By using the same arguments as before, it is clear that

$$\begin{aligned} \sum_{t=2}^T w_{it}^* e_{it}^* &= \sum_{t=2}^T w_{it} e_{it} - \sum_{t=2}^T w_{it} x'_{it-1} \left( \sum_{t=2}^T x_{it-1} x'_{it-1} \right)^{-1} \sum_{t=2}^T x_{it-1} e_{it} \\ &= O_p(\sqrt{T}) + O_p(T) O_p(T^{-2}) O_p(T) \end{aligned}$$

$$= O_p(\sqrt{T}),$$

which, together with (A4), (A5) and (A7), implies that

$$\begin{aligned} \sum_{t=2}^T (y_{it-1}^*)_p (e_{it}^*)_p &= \sum_{t=2}^T y_{it-1}^* e_{it}^* + O_p(T) O_p(T^{-1}) O_p(\sqrt{T}) \\ &= \sum_{t=2}^T y_{it-1}^* e_{it}^* + O_p(\sqrt{T}). \end{aligned}$$

By using (a), the limit of the remaining term becomes

$$\begin{aligned} T^{-1} \sum_{t=2}^T y_{it-1}^* e_{it}^* &= T^{-1} \sum_{t=2}^T y_{it-1} e_{it} - T^{-2} \sum_{t=2}^T y_{it-1} x'_{it-1} \\ &\quad \cdot \left( T^{-2} \sum_{t=2}^T x_{it-1} x'_{it-1} \right)^{-1} T^{-1} \sum_{t=2}^T x_{it-1} e_{it} \\ &\Rightarrow \frac{1}{\alpha_i(1)} \int_0^1 (\gamma_i(1)' B_{2i} + B_{1i}) dB_{1i} \\ &\quad - \frac{1}{\alpha_i(1)} \int_0^1 (\gamma_i(1)' B_{2i} + B_{1i}) B'_{2i} \left( \int_0^1 B_{2i} B'_{2i} \right)^{-1} \\ &\quad \cdot \int_0^1 B_{2i} dB_{1i}, \end{aligned}$$

where the last expression is

$$\begin{aligned} &\frac{1}{\alpha_i(1)} \int_0^1 B_{1i} dB_{1i} - \frac{1}{\alpha_i(1)} \int_0^1 B_{1i} B'_{2i} \left( \int_0^1 B_{2i} B'_{2i} \right)^{-1} \int_0^1 B_{2i} dB_{1i} \\ &= \frac{1}{\alpha_i(1)} \sigma_i^2 \int_0^1 V_i dV_i \\ &\quad - \frac{1}{\alpha_i(1)} \sigma_i^2 \int_0^1 V_i W'_i \left( \int_0^1 W_i W'_i \right)^{-1} \int_0^1 W_i dV_i \\ &= \frac{1}{\alpha_i(1)} \sigma_i^2 \int_0^1 U_i dV_i. \end{aligned}$$

This proves (c). ■

### Proof of Theorem 1

Consider first the group mean statistics. Note that  $\hat{\alpha}_i$  may be written as

$$\hat{\alpha}_i = \left( \sum_{t=2}^T (y_{it-1}^*)_p \right)^{-1} \sum_{t=2}^T (y_{it-1}^*)_p (\Delta y_{it}^*)_p. \quad (\text{A8})$$

Define  $E_{1i} = T^{-2} \sum_{t=2}^T (y_{it-1}^*)^2$  and  $E_{2i} = T^{-1} \sum_{t=2}^T (y_{it-1}^*) (\Delta y_{it}^*)_p$ . By imposing the null, and by using Lemma A.1 (b) and (c), we can show that

$$\begin{aligned} T\hat{\alpha}_i &= E_{1i}^{-1} E_{2i} \\ &\Rightarrow \left( \frac{1}{\alpha_i(1)^2} \sigma_i^2 \int_0^1 U_i^2 \right)^{-1} \frac{1}{\alpha_i(1)} \sigma_i^2 \int_0^1 U_i dV_i \\ &= \alpha_i(1) \left( \int_0^1 U_i^2 \right)^{-1} \int_0^1 U_i dV_i. \end{aligned}$$

Hence, since  $\hat{\alpha}_i(1) \rightarrow_p \alpha_i(1)$  as  $T \rightarrow \infty$ , where  $\rightarrow_p$  signifies convergence in probability, it follows that

$$G_\alpha = \frac{1}{N} \sum_{i=1}^N \frac{T\hat{\alpha}_i}{\hat{\alpha}_i(1)} \Rightarrow \frac{1}{N} \sum_{i=1}^N \left( \int_0^1 U_i^2 \right)^{-1} \int_0^1 U_i dV_i. \quad (\text{A9})$$

To obtain the corresponding intermediate limit as  $T \rightarrow \infty$  for  $G_\tau$ , we need to evaluate  $SE(\hat{\alpha}_i)$ , which may be written as

$$SE(\hat{\alpha}_i) = \hat{\sigma}_i \left( \sum_{t=2}^T (y_{it-1}^*)^2 \right)^{-1/2}.$$

Consider  $\hat{\sigma}_i^2$ , which, by using the results of Lemma A.1, can be written as

$$\begin{aligned} \hat{\sigma}_i^2 &= T^{-1} \sum_{t=2}^T (e_{it}^*)^2 \\ &= T^{-1} \sum_{t=2}^T e_{it}^{*2} - T^{-1} \sum_{t=2}^T e_{it}^* w_{it}^{*'} \left( \sum_{t=2}^T w_{it}^* w_{it}^{*'} \right)^{-1} \sum_{t=2}^T w_{it}^{*'} e_{it}^* \\ &= T^{-1} \sum_{t=2}^T e_{it}^{*2} + T^{-1} O_p(\sqrt{T}) O_p(T^{-1}) O_p(\sqrt{T}) \\ &= T^{-1} \sum_{t=2}^T e_{it}^{*2} + o_p(1), \end{aligned} \quad (\text{A10})$$

where

$$\begin{aligned} \sum_{t=2}^T e_{it}^{*2} &= \sum_{t=2}^T e_{it}^2 - \sum_{t=2}^T e_{it} x'_{it-1} \left( \sum_{t=2}^T x_{it-1} x'_{it-1} \right)^{-1} \sum_{t=2}^T x_{it-1} e_{it} \\ &= \sum_{t=2}^T e_{it}^2 + O_p(T) O_p(T^{-2}) O_p(T) \\ &= \sum_{t=2}^T e_{it}^2 + O_p(1). \end{aligned}$$

We can therefore show that

$$\widehat{\sigma}_i^2 = T^{-1} \sum_{t=2}^T e_{it}^2 + o_p(1) \rightarrow_p \sigma_i^2. \quad (\text{A11})$$

Hence, by using this result, (A11) and Lemma A.1 (b), we get

$$\begin{aligned} T^2 SE(\widehat{\alpha}_i)^2 &= \widehat{\sigma}_i^2 E_{1i}^{-1} \\ &\Rightarrow \sigma_i^2 \left( \frac{1}{\alpha_i(1)^2} \sigma_i^2 \int_0^1 U_i^2 \right)^{-1} \\ &= \alpha_i(1)^2 \left( \int_0^1 U_i^2 \right)^{-1}, \end{aligned}$$

which implies that

$$\begin{aligned} G_\tau &= \frac{1}{N} \sum_{i=1}^N \frac{\widehat{\alpha}_i}{SE(\widehat{\alpha}_i)} \\ &= \frac{1}{N} \sum_{i=1}^N \widehat{\sigma}_i^{-1} E_{1i}^{-1/2} E_{2i} \\ &\Rightarrow \frac{1}{N} \sum_{i=1}^N \left( \int_0^1 U_i^2 \right)^{-1/2} \int_0^1 U_i dV_i. \end{aligned} \quad (\text{A12})$$

Now, note that the limits in (A9) and (A12) are i.i.d. over the cross-section. Hence,  $E(\widetilde{C}_i) = \widetilde{\Theta}$  for all  $i$ , which makes it useful to expand  $G_\alpha$  and  $G_\tau$  as

$$\begin{aligned} \sqrt{N}G_\alpha - \sqrt{N}\widetilde{\Theta}_1 &= \sqrt{N} \left( N^{-1} \sum_{i=1}^N E_{1i}^{-1} E_{2i} - \widetilde{\Theta}_1 \right), \\ \sqrt{N}G_\tau - \sqrt{N}\widetilde{\Theta}_2 &= \sqrt{N} \left( N^{-1} \sum_{i=1}^N \widehat{\sigma}_i^{-1} E_{1i}^{-1/2} E_{2i} - \widetilde{\Theta}_2 \right). \end{aligned}$$

If we assume that  $\text{var}(\widetilde{C}_i) = \widetilde{\Sigma}$  exists and is finite, then  $\sqrt{NT}G_\alpha - \sqrt{N}\widetilde{\Theta}_1 \Rightarrow N(0, \widetilde{\Sigma}_{11})$  and  $\sqrt{N}G_\tau - \sqrt{N}\widetilde{\Theta}_2 \Rightarrow N(0, \widetilde{\Sigma}_{22})$  as  $T \rightarrow \infty$  and then  $N \rightarrow \infty$  by the Lindeberg-Lévy CLT. This establishes the limiting distributions of the group mean statistics.

Consider next the limiting distribution of the panel statistics. We begin with  $P_\alpha$ , which may be written as

$$P_\alpha = \left( \sum_{i=1}^N \sum_{t=2}^T (y_{it-1}^*)^2 \right)^{-1} \sum_{i=1}^N \sum_{t=2}^T \frac{1}{\widehat{\alpha}_i(1)} (y_{it-1}^*)_p (\Delta y_{it}^*)_p = E_1^{-1} E_2^*,$$

where  $E_1 = \sum_{i=1}^N E_{1i}$  and  $E_2^* = \sum_{i=1}^N E_{2i}/\widehat{\alpha}_i(1)$ . In order to get the sequential limit of this statistic, assume that the limit  $S_N^2 = N^{-1} \sum_{i=1}^N S_i^2 \rightarrow S^2$  as  $N \rightarrow \infty$



exist, where  $S_i^2 = \sigma_i^2/\alpha_i(1)^2$  is a nonrandom sequence. We then use the following expansion

$$\begin{aligned}\sqrt{N}P_\alpha - \sqrt{N}\Theta_2\Theta_1^{-1} &= \sqrt{N}(N^{-1}E_2^* - S^2\Theta_2)(N^{-1}E_1)^{-1} \\ &\quad - S^2\Theta_2\sqrt{N}\left((N^{-1}E_1)^{-1} - (S^2\Theta_1)^{-1}\right).\end{aligned}\quad (\text{A13})$$

By Corollary 1 of Phillips and Moon (1999), since  $\widehat{S}_i^2 \rightarrow_p S_i^2$  as  $T \rightarrow \infty$ , we have  $N^{-1}E_1 \rightarrow_p S^2\Theta_1$  and  $N^{-1}E_2^* \rightarrow_p S^2\Theta_2$  as  $T \rightarrow \infty$  prior to  $N$ . In addition, given that  $\text{var}(C_i) = \Sigma$  exist and is finite, then  $\sqrt{N}(N^{-1}E_2 - S^2\Theta_2) \Rightarrow N(0, S^2\Sigma_{22})$  by the Lindberg-Lévy CLT. Thus, it follows that

$$\sqrt{N}(N^{-1}E_2^* - S^2\Theta_2)(N^{-1}E_1)^{-1} \Rightarrow N(0, \Theta_1^{-2}\Sigma_{22}).$$

The second term on the right hand side of (A13) involve a differentiable transformation of the i.i.d. random variable  $E_1$ , to which the Delta method applies. This leads us to the following limit as  $T \rightarrow \infty$  prior to  $N$

$$\sqrt{N}\left((N^{-1}E_1)^{-1} - (S^2\Theta_1)^{-1}\right) \Rightarrow N(0, S^{-4}\Theta_1^{-4}\Sigma_{11}),$$

which implies that

$$S^2\Theta_2\sqrt{N}\left((N^{-1}E_1)^{-1} - (S^2\Theta_1)^{-1}\right) \Rightarrow N(0, \Theta_2^2\Theta_1^{-4}\Sigma_{11}).$$

By substituting these results into (A13), we get

$$\sqrt{N}P_\alpha - \sqrt{N}\Theta_2\Theta_1^{-1} \Rightarrow \Theta_1^{-1}N(0, \Sigma_{22}) - \Theta_2\Theta_1^{-2}N(0, \Sigma_{11}), \quad (\text{A14})$$

which shows that  $\sqrt{N}P_\alpha - \sqrt{N}\Theta_2\Theta_1^{-1}$  is mean zero with variance  $\Theta_1^{-2}\Sigma_{22} + \Theta_2^2\Theta_1^{-4}\Sigma_{11} - 2\Theta_1^{-3}\Theta_2\Sigma_{12}$ . This completes the proof for  $P_\alpha$ .

Next, consider  $P_\tau$ . This statistic may be rewritten as

$$P_\tau = \frac{\widehat{\alpha}}{SE(\widehat{\alpha})} = (\widehat{S}_N^2)^{-1/2}E_1^{-1/2}E_2^*,$$

which implies the following expansion

$$\begin{aligned}P_\tau - \sqrt{N}\Theta_2\Theta_1^{-1/2} &= (\widehat{S}_N^2)^{-1/2}\sqrt{N}(N^{-1}E_2^* - S^2\Theta_2)(N^{-1}E_1)^{-1/2} \\ &\quad - (\widehat{S}_N^2)^{-1/2}S^2\Theta_2\sqrt{N}\left((N^{-1}E_1)^{-1/2} - (S^2\Theta_1)^{-1/2}\right) \\ &\quad - S^2\Theta_2(S^2\Theta_1)^{-1/2}\sqrt{N}\left((\widehat{S}_N^2)^{-1/2} - (S^2)^{-1/2}\right).\end{aligned}\quad (\text{A15})$$

Now, because  $\widehat{S}_N^2 \rightarrow_p S^2$  as  $T \rightarrow \infty$  and then  $N \rightarrow \infty$  by a law of large numbers (LLN), the first term on the right hand side of (A15) converges to

$$(\widehat{S}_N^2)^{-1/2}\sqrt{N}(N^{-1}E_2^* - S^2\Theta_2)(N^{-1}E_1)^{-1/2} \Rightarrow N(0, \Theta_1^{-1}\Sigma_{22}).$$

To evaluate the second term, we use the Delta method, which implies that

$$(\widehat{S}_N^2)^{-1/2} S^2 \Theta_2 \sqrt{N} \left( (N^{-1} E_1)^{-1/2} - (S^2 \Theta_1)^{-1/2} \right) \Rightarrow N \left( 0, \frac{1}{4} \Theta_2^2 \Theta_1^{-3} \Sigma_{11} \right).$$

The third term in (A15) requires more work. We begin by taking a first order Taylor expansion of  $(\widehat{S}_N^2)^{-1/2}$  around  $S^2$ , which gives

$$\begin{aligned} \sqrt{N} \left( (\widehat{S}_N^2)^{-1/2} - (S^2)^{-1/2} \right) &= \sqrt{N} \left( (S^2)^{-1/2} - \frac{1}{2} (S^2)^{-3/2} (\widehat{S}_N^2 - S^2) \right. \\ &\quad \left. - (S^2)^{-1/2} \right) \\ &= -\frac{1}{2} (S^2)^{-3/2} \sqrt{N} (\widehat{S}_N^2 - S^2). \end{aligned} \quad (\text{A16})$$

Consider  $\widehat{S}_N^2$ . Irrespectively of the estimator used, we have  $\widehat{\alpha}_i(1)^2 = \alpha_i(1)^2 + O_p(T^{-1/2})$ . Therefore, since  $\widehat{\sigma}_i^2 = \sigma_i^2 + O_p(T^{-1})$  by (A11), we get

$$\widehat{S}_i^2 = S_i^2 + O_p(T^{-1/2}),$$

which, for some positive and finite constant  $b_i$ , becomes

$$\widehat{S}_i^2 = S_i^2 + T^{-1/2} b_i + O_p(T^{-1}).$$

This implies that

$$\begin{aligned} \widehat{S}_N^2 &= \frac{1}{N} \sum_{i=1}^N \left( S_i^2 + T^{-1/2} b_i + O_p(T^{-1}) \right) \\ &= S_N^2 + T^{-1/2} b_N + O_p(T^{-1}), \end{aligned} \quad (\text{A17})$$

where  $b_N$  is the average  $b_i$ . Moreover, since  $S_N^2 \rightarrow S^2$  as  $N \rightarrow \infty$ , for some positive constants  $g$  and  $a$ , it must be true that

$$S_N^2 = S^2 + N^{-a} g + o_p(N^{-a}). \quad (\text{A18})$$

Equations (A17) and (A18) implies that

$$\begin{aligned} \widehat{S}_N^2 &= S_N^2 + T^{-1/2} b_N + O_p(T^{-1}) \\ &= S^2 + N^{-a} g + T^{-1/2} b_N + o_p(N^{-a}) + O_p(T^{-1}). \end{aligned}$$

Thus, by direct substitution into (A16), we get

$$\begin{aligned} \sqrt{N} \left( (\widehat{S}_N^2)^{-1/2} - (S^2)^{-1/2} \right) &= -\frac{1}{2} (S^2)^{-3/2} \sqrt{N} (\widehat{S}_N^2 - S^2) \\ &= -\frac{1}{2} (S^2)^{-3/2} \sqrt{N} \left( S^2 + N^{-a} g + T^{-1/2} b_N \right. \end{aligned}$$

$$\begin{aligned}
& + o_p(N^{-a}) + O_p(T^{-1}) - S^2 \Big) \\
& = -\frac{1}{2}(S^2)^{-3/2}\sqrt{N} \left( N^{-a}g + T^{-1/2}b_N \right) \\
& + o_p(N^{-a}) + O_p(T^{-1}).
\end{aligned}$$

The term  $\sqrt{N}T^{-1/2}b_N$  vanishes in the first step with  $N$  held fixed, or if we assume that  $N/T \rightarrow 0$  and  $b_N \rightarrow b < \infty$  as  $N \rightarrow \infty$ . Thus, if we further assume that  $a > 1/2$ , so that  $\sqrt{N}N^{-a}g \rightarrow 0$ , then the entire expression vanishes. It follows that the third term in (A15) converges to zero in probability. Together, these results imply that (A15) obeys

$$P_\tau - \sqrt{N}\Theta_2\Theta_1^{-1/2} \Rightarrow \Theta_1^{-1/2}N(0, \Sigma_{22}) - \frac{1}{2}\Theta_2\Theta_1^{-3/2}N(0, \Sigma_{11}). \quad (\text{A19})$$

Therefore,  $P_\tau - \sqrt{N}\Theta_2\Theta_1^{-1/2}$  has zero mean and variance  $\Theta_1^{-1}\Sigma_{22} + \frac{1}{4}\Theta_2^2\Theta_1^{-3}\Sigma_{11} - \Theta_1^{-2}\Theta_2\Sigma_{12}$ , which completes the proof.  $\blacksquare$

### Proof of Theorem 2

Under the alternative hypothesis  $H_1^g$ ,  $y_{it} - \beta_i'x_{it}$  is stationary. Moreover, if we let  $\hat{\beta}_i$  denote the least squares estimator of  $\beta_i$ , then  $T(\hat{\beta}_i - \beta_i) = O_p(1)$ . As pointed out by Boswijk (1994), this implies that  $\hat{\alpha}_i$  is asymptotically equivalent to the least squares estimate of  $\alpha_i$  in

$$\Delta y_{it} = \alpha_i z_{it-1} + \delta_i' w_{it} + e_{it},$$

where  $z_{it}$  is the putative disequilibrium error  $y_{it} - \beta_i'x_{it}$  and  $w_{it}$  is defined in Lemma A.1. As before, all deterministic terms are suppressed.

Because  $z_{it}$  is stationary under the alternative, we have

$$\begin{aligned}
\sum_{t=2}^T (z_{it-1})_p^2 &= \sum_{t=2}^T z_{it-1}^2 - \sum_{t=2}^T z_{it-1} w_{it}' \left( \sum_{t=2}^T w_{it} w_{it}' \right)^{-1} \sum_{t=2}^T w_{it} z_{it-1} \\
&= O_p(T) + O_p(T)O_p(T^{-1})O_p(T) \\
&= O_p(T).
\end{aligned}$$

By using this result, it is straightforward to show that  $SE(\hat{\alpha}_i) = O_p(T^{-1/2})$  and  $\hat{\alpha}_i - \alpha_i = O_p(T^{-1/2})$ . This implies that

$$T^{-1}G_\alpha = \frac{1}{N} \sum_{i=1}^N \frac{\alpha_i}{\hat{\alpha}_i(1)} + O_p(T^{-1/2}) = O_p(1), \quad (\text{A20})$$

which implies that  $G_\alpha = O_p(T)$ . As usual, the limit arguments are taken as  $T \rightarrow \infty$  for a fixed  $N$  in the first step while  $N \rightarrow \infty$  in the second. The corresponding order of  $G_\tau$  is given by

$$T^{-1/2}G_\tau = \frac{1}{N} \sum_{i=1}^N \frac{\alpha_i}{\sqrt{T}SE(\hat{\alpha}_i)} + O_p(T^{-1/2}) = O_p(1). \quad (\text{A21})$$

This shows that  $G_\tau$  is  $O_p(\sqrt{T})$ .

Similarly, under  $H_1^p$ , it is possible to show that  $\hat{\alpha} - \alpha = O_p((TN)^{-1/2})$  in the sequential limit. For  $SE(\hat{\alpha})$ , write

$$\begin{aligned} SE(\hat{\alpha}) &= (\hat{S}_N^2)^{1/2} \left( \sum_{i=1}^N \sum_{t=2}^T (z_{it-1})_p^2 \right)^{-1/2} \\ &= \left( \frac{1}{N} \sum_{i=1}^N \hat{S}_i^2 \right)^{1/2} \left( \sum_{i=1}^N \sum_{t=2}^T (z_{it-1})_p^2 \right)^{-1/2}. \end{aligned} \quad (\text{A22})$$

From Theorem 1, we know that  $\hat{S}_i \rightarrow_p S_i$  as  $T \rightarrow \infty$ , which implies the following as  $N \rightarrow \infty$

$$\frac{1}{N} \sum_{i=1}^N S_i^2 = O_p(1).$$

Similarly, for the second term in (A22), it is possible to show that as  $T \rightarrow \infty$  and then  $N \rightarrow \infty$

$$\frac{1}{TN} \sum_{i=1}^N \sum_{t=2}^T (z_{it-1})_p^2 = O_p(1).$$

The panel statistics  $P_\alpha$  and  $P_\tau$  therefore have the following orders

$$T^{-1}P_\alpha = \alpha + O_p((TN)^{-1/2}) = O_p(1), \quad (\text{A23})$$

and

$$(TN)^{-1/2}P_\tau = \frac{\alpha}{\sqrt{TNSE(\hat{\alpha})}} + O_p((TN)^{-1/2}) = O_p(1). \quad (\text{A24})$$

Therefore, as  $N$  is kept fixed in the first step as  $T \rightarrow \infty$ ,  $P_\alpha = O_p(T)$  and  $P_\tau = O_p(\sqrt{T})$ . It follows that each of the statistics diverges as  $T \rightarrow \infty$  and then  $N \rightarrow \infty$ . Moreover, as  $\alpha_i < 0$  and  $\alpha < 0$  under our two alternatives, the divergence occurs towards negative infinity. ■

Table 1: Simulated asymptotic moments.

Case	$K$	Expected value				Variance			
		$\tilde{\Theta}_2$	$\tilde{\Theta}_1$	$\Theta_2\Theta_1^{-1/2}$	$\Theta_2\Theta_1^{-1}$	$\tilde{\Sigma}_{22}$	$\tilde{\Sigma}_{11}$	$\varphi'\Sigma\varphi$	$\phi'\Sigma\phi$
1	1	-0.9763	-3.8022	-0.5105	-1.0263	1.0823	20.6868	1.3624	8.3827
	2	-1.3816	-5.8239	-0.9370	-2.4988	1.0981	29.9016	1.7657	24.0223
	3	-1.7093	-7.8108	-1.3169	-4.2699	1.0489	39.0109	1.7177	39.8827
	4	-1.9789	-9.8791	-1.6167	-6.1141	1.0576	50.5741	1.6051	53.4518
	5	-2.1985	-11.7239	-1.8815	-8.0317	1.0351	58.9595	1.4935	63.2406
	6	-2.4262	-13.8581	-2.1256	-10.0074	1.0409	69.5967	1.4244	76.6757
2	1	-1.7776	-7.1423	-1.4476	-4.2303	0.8071	29.6336	0.9885	19.7090
	2	-2.0349	-9.1249	-1.7131	-5.8650	0.8481	39.3428	1.0663	31.2637
	3	-2.2332	-10.9667	-1.9206	-7.4599	0.8886	49.4880	1.1168	42.9975
	4	-2.4453	-12.9561	-2.1484	-9.3057	0.9119	58.7035	1.1735	57.4844
	5	-2.6462	-14.9752	-2.3730	-11.3152	0.9083	67.9499	1.1684	69.4374
	6	-2.8358	-17.0673	-2.5765	-13.3180	0.9236	79.1093	1.1589	81.0384
3	1	-2.3664	-12.0116	-2.1124	-8.9326	0.6603	46.2420	0.7649	37.5948
	2	-2.5284	-13.6324	-2.2876	-10.4874	0.7070	53.7428	0.8137	45.6890
	3	-2.7040	-15.5262	-2.4633	-12.1672	0.7586	64.5591	0.8857	57.9985
	4	-2.8639	-17.3648	-2.6275	-13.8889	0.8228	74.7403	0.9985	74.1258
	5	-3.0146	-19.2533	-2.7858	-15.6815	0.8477	84.7990	0.9918	81.3934
	6	-3.1710	-21.2479	-2.9537	-17.6515	0.8599	94.0024	0.9898	91.2392

*Notes:* Case 1 refers to the regression with no deterministic terms, Case 2 refers to the regression with a constant term, and Case 3 refers to the regression with both constant and trend terms. The value  $K$  refers to the number of regressors excluding any deterministic terms.

Table 2: Size at the 5% level.

$\phi$	$\gamma$	$N$	$T$	Group mean statistics						Panel statistics					
				$G_\tau^a$	$G_\tau^b$	$\tilde{Z}_t$	$G_\alpha^a$	$G_\alpha^b$	$\tilde{Z}_\rho$	$P_\tau^a$	$P_\tau^b$	$Z_t$	$P_\alpha^a$	$P_\alpha^b$	$Z_\rho$
0	0	10	100	9.8	6.6	6.8	8.6	7.1	2.2	7.9	9.8	6.0	13.1	14.4	11.1
		10	200	7.7	6.9	6.4	8.7	6.8	2.4	7.2	8.7	5.8	13.8	13.9	9.3
		20	100	10.8	6.8	8.0	9.4	7.2	1.7	7.3	9.3	7.9	11.3	11.7	10.3
		20	200	7.5	5.0	5.6	7.0	5.3	1.1	7.0	6.8	5.6	9.5	11.2	8.0
-0.4	0	10	100	16.3	14.6	79.5	28.9	28.1	72.5	14.3	14.4	47.7	26.5	29.9	54.9
		10	200	12.5	14.2	66.9	24.9	27.8	61.3	11.6	13.4	35.6	25.2	29.3	44.1
		20	100	20.7	21.2	94.8	36.4	42.4	90.5	17.8	19.8	68.9	31.6	34.8	71.6
		20	200	14.1	21.1	88.6	34.4	41.8	81.9	15.6	17.3	50.4	31.4	33.9	56.4
0.4	0	10	100	8.3	3.9	2.2	5.4	3.1	0.4	6.3	6.5	2.6	9.3	9.8	5.8
		10	200	6.7	3.4	2.3	5.3	3.6	1.1	6.3	5.7	2.8	10.2	7.2	6.3
		20	100	12.2	4.2	2.0	5.7	2.2	0.0	6.9	6.0	4.5	7.9	7.5	6.2
		20	200	7.3	3.2	0.9	4.9	1.5	0.1	5.5	4.8	3.9	7.0	6.2	5.7
0	0.4	10	100	10.6	5.6	39.6	10.0	8.0	32.3	10.4	8.3	21.7	15.6	13.3	29.4
		10	200	8.4	6.0	28.7	9.1	7.5	21.9	8.8	7.4	16.5	14.3	12.8	24.5
		20	100	11.7	6.9	54.6	12.0	7.7	42.9	7.3	9.7	27.7	11.0	12.5	33.5
		20	200	7.7	5.7	40.6	8.4	6.3	28.9	7.4	7.9	20.9	11.8	12.0	27.6
0	0.8	10	100	9.2	6.7	94.0	10.9	7.6	93.9	8.5	9.8	74.3	13.4	16.2	79.3
		10	200	8.1	7.7	89.3	8.1	11.2	89.9	8.0	8.3	66.0	13.5	15.8	73.0
		20	100	9.7	5.5	99.8	10.2	7.0	99.4	7.0	10.2	93.6	10.2	12.9	94.1
		20	200	8.8	5.0	98.3	9.5	5.9	98.6	9.1	7.4	82.9	12.6	11.1	85.6

*Notes:* The value  $\phi$  refers to the moving average parameter and  $\gamma$  refers to the parameter of the first difference of the regressor.

<sup>a</sup>The test is based on choosing the number of lags and leads according to the Akaike information criterion.

<sup>b</sup>The test is based on choosing the number of lags and leads as a function of  $T$ .

Table 3: Power at the 5% level when  $\alpha = -0.01$ .

Comfac	var( $v_{it}$ )	$N$	$T$	Group mean statistics						Panel statistics					
				$G_\tau^a$	$G_\tau^b$	$\tilde{Z}_t$	$G_\alpha^a$	$G_\alpha^b$	$\tilde{Z}_\rho$	$P_\tau^a$	$P_\tau^b$	$Z_t$	$P_\alpha^a$	$P_\alpha^b$	$Z_\rho$
Yes	1	10	100	14.2	15.2	14.6	14.2	14.2	10.9	17.0	17.5	10.6	20.2	20.5	10.5
		10	200	45.5	37.0	23.0	38.9	30.3	18.0	50.9	42.8	14.8	54.5	49.6	14.4
		20	100	31.9	27.7	19.0	21.7	23.4	14.9	36.9	35.9	17.6	46.0	40.3	16.5
		20	200	72.1	77.3	48.6	58.4	62.2	34.7	81.8	88.5	39.7	86.0	92.1	39.0
No	1	10	100	14.0	14.7	10.2	9.7	9.6	9.0	19.2	17.9	7.8	17.6	15.8	7.7
		10	200	49.9	39.1	11.7	20.9	15.3	9.1	59.4	47.8	7.9	39.6	32.9	7.6
		20	100	26.7	23.8	9.5	11.4	11.6	8.0	32.2	34.8	9.2	35.7	31.0	9.1
		20	200	71.7	74.9	17.9	28.3	28.2	16.4	87.1	88.6	16.0	75.0	85.2	17.3
No	2	10	100	12.0	11.0	5.3	5.0	6.5	4.9	18.0	15.8	4.7	10.9	10.1	4.9
		10	200	50.0	42.3	6.7	15.6	10.5	6.6	62.6	52.1	4.6	28.9	23.8	4.6
		20	100	24.7	18.1	5.6	7.1	6.6	5.7	34.9	31.2	5.5	28.8	23.2	5.1
		20	200	72.9	77.8	11.9	20.2	18.6	9.4	89.5	93.3	8.9	61.3	76.1	9.3
No	4	10	100	8.8	11.2	4.9	2.6	4.0	4.8	16.6	14.2	4.2	6.2	6.7	4.7
		10	200	59.0	51.2	4.1	7.7	6.1	4.7	74.1	65.9	1.8	15.7	13.1	2.4
		20	100	19.8	17.5	1.4	3.4	3.5	1.0	30.5	32.5	2.5	15.4	15.4	2.5
		20	200	80.6	84.8	4.2	9.1	7.4	3.8	95.5	96.3	2.3	45.7	59.6	3.6

*Notes:* The first column indicates whether the common factor restriction is satisfied or not. The value var( $v_{it}$ ) refers to the signal-to-noise ratio and  $\alpha$  refers to the error correction parameter.

<sup>a</sup>The test is based on choosing the number of lags and leads according to the Akaike information criterion.

<sup>b</sup>The test is based on choosing the number of lags and leads as a function of  $T$ .

Table 4: Power at the 5% level when  $\alpha = -0.03$ .

Comfac	var( $v_{it}$ )	$N$	$T$	Group mean statistics						Panel statistics					
				$G_\tau^a$	$G_\tau^b$	$\tilde{Z}_t$	$G_\alpha^a$	$G_\alpha^b$	$\tilde{Z}_\rho$	$P_\tau^a$	$P_\tau^b$	$Z_t$	$P_\alpha^a$	$P_\alpha^b$	$Z_\rho$
Yes	1	10	100	60.6	65.3	55.5	56.3	60.4	40.4	78.8	76.6	51.2	86.3	82.6	49.6
			200	99.1	98.9	95.1	98.4	98.1	86.5	99.9	99.9	95.2	100.0	100.0	95.1
		20	100	89.5	90.0	71.1	80.7	82.9	55.4	98.3	97.1	77.0	99.4	99.0	75.6
			200	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0
No	1	10	100	67.7	72.2	27.3	28.2	28.6	21.8	87.6	87.1	22.6	74.3	67.9	23.1
			200	100.0	100.0	80.4	95.3	93.0	65.8	100.0	100.0	72.4	100.0	100.0	73.1
		20	100	95.5	93.2	35.5	47.6	48.3	25.9	99.6	99.3	33.9	99.6	98.7	35.0
			200	100.0	100.0	98.4	99.8	100.0	92.4	100.0	100.0	96.4	100.0	100.0	96.9
No	2	10	100	76.8	74.6	18.5	18.8	19.2	13.7	93.5	89.5	16.6	67.0	53.6	17.9
			200	100.0	100.0	67.1	94.2	89.2	54.1	100.0	100.0	59.1	100.0	100.0	59.6
		20	100	97.8	97.2	19.7	32.3	33.6	13.3	99.8	99.5	17.7	98.9	97.0	18.6
			200	100.0	100.0	96.4	99.7	99.9	88.6	100.0	100.0	90.5	100.0	100.0	93.2
No	4	10	100	88.3	88.7	13.0	12.5	13.5	8.9	97.6	96.3	10.3	57.5	47.8	11.5
			200	100.0	100.0	60.1	92.4	87.0	47.8	100.0	100.0	49.7	100.0	100.0	51.7
		20	100	99.1	99.1	10.8	20.2	20.9	8.0	100.0	99.9	8.9	98.8	97.5	10.6
			200	100.0	100.0	90.0	99.8	100.0	79.8	100.0	100.0	81.9	100.0	100.0	85.0

*Notes:* See Table 3 for an explanation of the various features of the table.

<sup>a</sup>The test is based on choosing the number of lags and leads according to the Akaike information criterion.

<sup>b</sup>The test is based on choosing the number of lags and leads as a function of  $T$ .



Table 5: Size and power at the 5% level with cross-sectional dependence.

$\alpha$	$b$	$N$	$T$	Group mean statistics						Panel statistics					
				$G_\tau$	$G_\tau^*$	$\tilde{Z}_t$	$G_\alpha$	$G_\alpha^*$	$\tilde{Z}_\rho$	$P_\tau$	$P_\tau^*$	$Z_t$	$P_\alpha$	$P_\alpha^*$	$Z_\rho$
0	0	10	100	6.0	4.0	5.7	6.7	3.7	1.3	10.0	6.3	6.0	16.7	7.7	9.3
		10	200	5.3	3.0	5.3	5.7	3.3	1.0	6.3	4.7	5.7	12.0	6.0	9.0
		20	100	7.0	4.7	9.3	6.7	5.0	2.3	11.3	7.0	6.7	14.3	7.7	11.7
		20	200	7.0	5.7	9.7	8.0	4.3	1.7	8.7	7.3	9.0	12.7	9.0	12.7
	1	10	100	26.3	8.0	87.0	48.3	12.3	86.3	27.7	7.0	76.7	48.3	10.0	79.3
		10	200	32.0	12.7	84.0	56.3	20.3	83.0	30.0	8.7	71.3	55.0	13.0	78.7
		20	100	40.7	6.0	96.7	65.0	8.3	96.7	43.3	5.0	92.3	60.7	8.7	92.3
		20	200	44.3	9.3	95.7	72.7	22.7	94.0	40.3	4.7	86.7	59.3	7.7	88.3
-0.01	0	10	100	18.3	15.7	11.7	16.3	12.7	13.0	19.0	21.7	6.3	24.7	29.0	8.3
		10	200	47.7	44.0	28.3	42.3	36.7	25.0	52.7	55.0	28.3	62.0	67.7	28.7
		20	100	25.0	26.7	12.7	20.3	17.3	9.7	28.3	33.3	14.3	29.7	40.3	13.7
		20	200	68.3	68.7	45.7	55.0	57.0	28.0	72.7	83.3	35.7	80.0	90.3	34.7
	1	10	100	20.7	27.3	12.0	19.7	29.0	8.3	17.3	25.7	16.0	17.7	34.3	15.0
		10	200	40.7	64.0	26.0	26.0	69.7	22.0	50.7	63.7	40.7	51.7	79.3	34.0
		20	100	24.0	34.3	15.7	21.0	32.7	12.3	24.3	27.7	27.7	29.0	35.3	26.0
		20	200	57.7	77.0	34.7	30.3	79.0	19.7	78.0	81.3	59.0	72.0	88.3	59.7
-0.03	0	10	100	64.7	59.3	56.7	56.7	50.0	47.0	73.7	75.0	48.7	81.7	84.0	48.3
		10	200	99.7	99.3	96.0	99.0	98.0	90.3	100.0	99.3	97.7	100.0	100.0	97.7
		20	100	85.3	87.3	67.7	79.0	73.7	47.0	97.7	96.7	76.3	98.7	99.0	73.3
		20	200	100.0	100.0	99.7	100.0	100.0	99.0	100.0	100.0	100.0	100.0	100.0	100.0
	1	10	100	72.7	80.0	44.3	58.7	73.0	33.7	68.3	82.7	57.0	70.0	89.7	56.0
		10	200	98.0	100.0	79.0	90.0	100.0	64.7	100.0	100.0	91.3	100.0	100.0	91.0
		20	100	86.0	93.0	72.3	78.0	87.0	58.3	92.7	96.0	87.0	96.0	98.0	85.0
		20	200	99.3	100.0	94.7	98.7	100.0	81.0	100.0	100.0	99.3	100.0	100.0	99.3

*Notes:* The value  $\alpha$  refers to the error correction parameter and  $b$  refers to the factor scale parameter. If  $b = 0$ , then there is no common factor, whereas if  $b = 1$ , then there is a common factor. The star superscript is used to denote the bootstrapped test statistics.

Table 6: Size and power at the 5% level with endogenous regressors.

$\alpha_y$	$\alpha_x$	$T$	Group mean statistics				Panel statistics			
			$G_\tau$	$\tilde{Z}_t$	$G_\alpha$	$\tilde{Z}_\rho$	$P_\tau$	$Z_t$	$P_\alpha$	$Z_\rho$
0	0	100	6.7	8.0	6.5	3.0	8.5	6.4	13.6	11.1
		200	5.1	6.7	5.6	2.7	7.4	4.9	12.6	10.6
-0.01	-0.01	100	2.7	3.7	4.8	3.6	2.6	4.6	2.8	5.0
		200	1.2	4.8	2.0	4.7	1.9	8.6	1.9	7.9
	-0.03	100	0.0	0.9	0.0	0.5	0.0	2.0	0.0	2.2
		200	0.0	8.6	0.0	8.8	0.0	16.1	0.0	14.4
-0.03	-0.01	100	31.3	18.1	19.0	13.4	47.7	19.0	39.3	22.2
		200	91.7	68.3	66.7	59.7	97.1	79.3	96.1	79.3
	-0.03	100	1.1	11.0	4.1	8.3	1.6	13.5	3.5	14.9
		200	0.7	51.7	3.3	49.8	2.1	62.9	6.5	61.2
-0.01	0.01	100	37.3	4.7	12.5	4.7	53.6	3.2	25.7	3.9
		200	98.2	9.2	47.3	9.8	99.7	5.5	96.2	5.6
	0.03	100	96.8	6.5	23.1	6.5	97.0	3.5	53.8	4.3
		200	100.0	22.6	85.8	28.0	97.8	7.6	97.5	7.3
-0.03	0.01	100	94.5	22.4	45.1	15.8	98.4	15.8	93.5	19.2
		200	100.0	78.8	99.9	72.0	100.0	61.0	100.0	61.3
	0.03	100	99.9	25.6	78.4	22.0	100.0	11.3	100.0	13.2
		200	100.0	74.1	100.0	78.6	99.9	27.8	99.9	26.9

*Notes:* The values  $\alpha_x$  and  $\alpha_y$  refer to the error correction parameters in the model for  $x_{it}$  and  $y_{it}$ , respectively.

Table 7: Panel unit root and cointegration tests.

Unit root testing						Cointegration testing				
Study	Test	$\ln(H_{it})$	$p\text{-val}^a$	$\ln(Y_{it})$	$p\text{-val}^a$	Study	Test	Value	$p\text{-val}^a$	$p\text{-val}^b$
Im <i>et al.</i> (2003)	$Z_t$	-0.367	0.357	0.189	0.575	This study	$G_\tau$	-9.613	0.000	0.000
Harris and Tzavalis (1999)	$\hat{\varphi}_{lsdv}$	1.704	0.956	2.037	0.979		$G_\alpha$	-10.626	0.000	0.000
Bai and Ng (2004)	$P_e^c$	1.153	0.124	4.482	0.000		$P_\tau$	-4.100	0.000	0.048
Phillips and Sul (2003)	$G_{ols}^{++}$	0.068	0.527	2.176	0.985		$P_\alpha$	-10.119	0.000	0.000
	$Z$	-2.531	0.006	-0.263	0.396	Pedroni (2004)	$\tilde{Z}_t$	-0.401	0.344	-
	$P_m$	2.869	0.002	0.162	0.435		$\tilde{Z}_\rho$	2.001	0.977	-
Moon and Perron (2004)	$t_a$	0.311	0.622	0.173	0.569		$Z_t$	-1.878	0.030	-
	$t_c$	9.590	1.000	13.518	1.000		$Z_\rho$	0.190	0.575	-

*Notes:* All tests are implemented with a constant and trend in the test regression. For semiparametric corrections, the Bartlett kernel is employed. The lags and leads in the error correction test are chosen according to the Akaike criterion. All other bandwidth and lag orders are set according to the rule  $4(T/100)^{2/9}$ . The number of common factors is estimated using the  $IC_1$  criterion of Bai and Ng (2004) with the maximum number of factors set equal to five. All unit root tests take a unit root as the null hypothesis while the cointegration tests take no cointegration as the null.

<sup>a</sup>The  $p$ -values are for a one-sided test based on the normal distribution.

<sup>b</sup>The  $p$ -values are for a one-sided test based on the bootstrapped distribution. We use 500 bootstrap replications.

## References

- Bai, J., and S. Ng (2004). A Panic Attack on Unit Roots and Cointegration. *Econometrica*, Vol. 72, pp. 1127-1177.
- Banerjee, A., J. J. Dolado and R. Mestre (1998). Error-Correction Mechanism Tests for Cointegration in a Single-Equation Framework. *Journal of Time Series Analysis*, Vol. 19, pp. 267-283.
- Blomqvist, A. G., and R. A. L. Carter (1997). Is Health Care Really Luxury? *Journal of Health Economics*, Vol. 16, pp. 207-229.
- Boswijk, P. H. (1994). Testing for an Unstable Root in Conditional and Structural Error Correction Models. *Journal of Econometrics*, Vol. 63, pp. 37-60.
- Campbell, J., and P. Perron (1991). Pitfalls and Opportunities: What Macroeconomists should Know about Unit Roots. In Blanchard, O., and S. Fishers (Eds.), *NBER Macroeconomics Annual*, MIT Press, Cambridge, MA.
- Chang, Y. (2004). Bootstrap Unit Root Tests in Panels with Cross-Sectional Dependency. *Journal of Econometrics*, Vol. 120, pp. 263-293.
- Engle, R., and C. Granger (1987). Cointegration and Error Correction: Representation, Estimation and Testing. *Econometrica*, Vol. 55, pp. 251-276.
- Gerdtham, U. G., and M. Löthgren (2000). On Stationarity and Cointegration of International Health Expenditure and GDP. *Journal of Health Economics*, Vol. 19, pp. 461-475.
- Hansen, P., and A. King (1996). The Determinants of Health Care Expenditures: A Cointegration Approach. *Journal of Health Economics*, Vol. 15, pp. 127-137.
- Hansen, P., and A. King (1998). Health Care Expenditures and GDP: Panel Data Unit Root Test Results - Comment. *Journal of Health Economics*, Vol. 17, pp. 377-381.
- Harris, R. D. F., and E. Tzavalis (1999). Inference for Unit Roots in Dynamic Panels where the Time Dimension is Fixed. *Journal of Econometrics*, Vol. 91, pp. 201-226.
- Haug, A. A. (1996). Tests for Cointegration: A Monte Carlo Comparison. *Journal of Econometrics*, Vol. 71, pp. 891-915.
- Ho, T. (2002). A Panel Cointegration Approach to the Saving-Investment Correlation. *Empirical Economics*, Vol. 27, pp. 91-100.

- Im, K. S., M. Peseran and Y. Shin (2003). Testing for Unit Roots in Heterogeneous Panels. *Journal of Econometrics*, Vol. 115, pp. 53-74.
- Johansen, S. J. (1992). Cointegration in Partial Systems and the Efficiency of Single Equation Analysis. *Journal of Econometrics*, Vol. 52, pp. 389-402.
- Kremers, J., N. Ericsson and J. Dolado (1992). The Power of Cointegration Tests. *Oxford Bulletin of Economics and Statistics*, Vol. 54, pp. 325-348.
- Larsson, R., J. Lyhagen and M. Lötgren (2001). Likelihood-Based Cointegration Test in Heterogeneous Panels. *Econometrics Journal*, Vol. 4, pp. 109-142.
- McCoskey, S., and C. Kao (1998). A Residual-Based Test of the Null of Cointegration in Panel Data. *Econometric Reviews*, Vol. 17, pp. 57-84.
- McCoskey, S., and T. Selden (1998). Health Care Expenditures and GDP: Panel Data Unit Root Test Results. *Journal of Health Economics*, Vol. 17, pp. 369-376.
- Moon, H. R., and B. Perron (2004). Testing for a Unit Root in Panels with Dynamic Factors. *Journal of Econometrics* Vol. 122, pp. 81-126.
- Newey, W., and K. West (1994). Autocovariance Lag Selection in Covariance Matrix Estimation. *Review of Economic Studies*, Vol. 61, pp. 613-653.
- Ng, S., and P. Perron (1995). Unit Root Test in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag. *Journal of the American Statistical Association*, Vol. 90, pp. 268-281.
- Ng, S., and P. Perron (1997). Estimation and Inference in Nearly Unbalanced Nearly Cointegrated Systems. *Journal of Econometrics*, Vol. 79, pp. 5381.
- Park, J. Y., and P. B. C. Phillips (1988). Statistical Inference in Regressions with Integrated Processes: Part 1. *Econometric Theory*, Vol. 4, pp. 468-497.
- Pedroni, P. (1999). Critical Values for Cointegration Tests in Heterogeneous Panels with Multiple Regressors. *Oxford Bulletin of Economics and Statistics*, Vol. 61, pp. 653-670.
- Pedroni, P. (2004). Panel Cointegration: Asymptotic and Finite Sample Properties of Pooled Time Series Tests with an Application to the PPP Hypothesis. *Econometric Theory*, Vol. 3, pp. 579-625.
- Phillips, P. C. B., and H. R. Moon (1999). Linear Regression Limit Theory of Nonstationary Panel Data. *Econometrica*, Vol. 67, pp. 1057-1111.

- Phillips, P. C. B., and S. Ouliaris (1990). Asymptotic Properties of Residual Based Tests for Cointegration. *Econometrica*, Vol. 58, pp. 165-193.
- Phillips, P. C. B. and D. Sul (2003). Dynamic Panel Estimation and Homogeneity Testing Under Cross Section Dependence. *Econometrics Journal* 6, 217-259.
- Roberts, J. (2000). Spurious Regression Problems in the Determinants of Health Care Expenditure: A Comment on Hitiris (1997). *Applied Economics Letters*, Vol. 7, pp. 279-283.
- Saikkonen, P. (1991). Asymptotic Efficient Estimation of Cointegration Regressions. *Econometric Theory*, Vol. 7, pp. 1-21.
- Urbain, J-P. (1992). On Weak Exogeneity in Error Correction Models. *Oxford Bulletin of Economics and Statistics*, Vol. 54, pp. 187-207.
- Zivot, E. (2000). The Power of Single Equation Tests for Cointegration when the Cointegration Vector is Prespecified. *Econometric Theory*, Vol. 16, pp. 407-439.
- Xiao, Z., and P. C. B. Phillips (1998). An ADF Coefficient Test for a Unit Root in ARMA Models of Unknown Order with Empirical Applications to the US Economy. *Econometrics Journal*, Vol. 1, pp. 27-43.
- Westerlund, J. (2005a). A Panel CUSUM Test of the Null of Cointegration. *Oxford Bulletin of Economics and Statistics*, Vol. 62, pp. 231-262.
- Westerlund, J. (2005b). New Simple Tests for Panel Cointegration. *Econometric Reviews*, Vol. 24, pp. 297-316.
- Westerlund, J. (2005c). Data Dependent Endogeneity Correction in Cointegrated Panels. *Oxford Bulletin of Economics and Statistics*, Vol. 67, pp. 691-705.