

General Diagnostic Tests for Cross Section Dependence in Panels

M. Hashem Pesaran*
University of Cambridge & USC

June 2004

Abstract

This paper proposes simple tests of error cross section dependence which are applicable to a variety of panel data models, including stationary and unit root dynamic heterogeneous panels with short T and large N . The proposed tests are based on average of pair-wise correlation coefficients of the OLS residuals from the individual regressions in the panel, and can be used to test for cross section dependence of any fixed order p , as well as the case where no *a priori* ordering of the cross section units is assumed, referred to as $CD(p)$ and CD tests, respectively. Asymptotic distribution of these tests are derived and their power function analyzed under different alternatives. It is shown that these tests are correctly centred for fixed N and T , and are robust to single or multiple breaks in the slope coefficients and/or error variances. The small sample properties of the tests are investigated and compared to the Lagrange multiplier test of Breusch and Pagan using Monte Carlo experiments. It is shown that the tests have the correct size in very small samples and satisfactory power, and as predicted by the theory, quite robust to the presence of unit roots and structural breaks. The use of the CD test is illustrated by applying it to study the degree of dependence in per capita output innovations across countries within a given region and across countries in different regions. The results show significant evidence of cross dependence in output innovations across many countries and regions in the World.

Keywords: Cross section dependence, Spatial dependence, Diagnostic tests, Dynamic heterogenous panels, Empirical growth.

JEL-Classification: C12, C13, C33.

*I am grateful to Mutita Akusuwan for providing me with excellent research assistance and for carrying out some of the computations. I would also like to thank Ron Smith and Takashi Yamagata for helpful discussions and the latter for carrying out the Monte Carlo results reported in Tables 3-6. Financial support from the ESRC (Grant No.RES-000-23-0135) is gratefully acknowledged.

1 Introduction

It is typically assumed that disturbances in panel data models are cross sectionally independent. This is particularly true of panels with large cross section dimension (N). In the case of panels where N is small (say 10 or less) and the time dimension of the panel (T) is sufficiently large the cross correlations of the errors can be modelled (and tested statistically) using the seemingly unrelated regression equation (SURE) framework originally developed by Zellner (1962). Since N is fixed as $T \rightarrow \infty$, traditional time series techniques, including log-likelihood ratio tests, can be applied. A simple example of such a test is the Lagrange multiplier (LM) test of Breusch and Pagan (1980) which is based on the average of the squared pair-wise correlation of the residuals. However, in cases where N is large standard techniques will not be applicable and other approaches must be considered.

In the literature on spatial statistics the extent of cross section dependence is measured with respect to a given “connection or spatial matrix” that characterizes the pattern of spatial dependence accordingly to a pre-specified set of rules. For example, the (i, j) elements of a connection matrix, w_{ij} , could be set equal to 1 if the i^{th} and j^{th} regions are joined, and zero otherwise. See Moran (1948) and further elaborations by Cliff and Ord (1973, 1981). More recent accounts and references can be found in Anselin (1988, 2001), and Haining (2003, Ch. 7). This approach, apart from being dependent on the choice of the spatial matrix, is not appropriate in many economic applications where space is not a natural metric and economic and sociopolitical factors could be more appropriate.¹

In this paper we propose a simple diagnostic test that does not require an *a priori* specification of a connection matrix and is applicable to a variety of panel data models, including stationary dynamic and unit-root heterogeneous panels with short T and large N . The proposed test is based on a simple average of *all* pair-wise correlation coefficients of the Ordinary Least Squares (OLS) residuals from the individual regressions in the panel.

This test, referred to as the *CD* (Cross-section Dependence) test, is shown to be correctly centred for fixed N and T , assuming that the underlying error processes are symmetrically distributed. Its asymptotic distribution under the null hypotheses is established and it is shown to be robust to single or multiple breaks in the slope coefficients and/or error variances, so long as the unconditional means of variables in the panel remain constant over time.

In cases where the cross section units can be ordered *a priori*, as with spatial observations, the CD test might not be sufficiently powerful as it does not exploit the spatial information. To deal with this problem we also propose a generalization of the CD test that captures the spatial patterns when deemed relevant. We shall refer to this test as the p^{th} order spatial dependence test, denoted simply as *CD*(p). The order p measures the extent of local dependence, and specifies the number of contiguous layers of neighbours that the i^{th} cross

¹For empirical applications where economic distance such as trade patterns are used in modelling of spatial correlations see Conley and Topa (2002) and Pesaran, Schuermann, and Weiner (2004).

section unit depends on. $CD(p)$ reduces to CD for $p = N - 1$, when all possible spatial dependencies are taken into account. The $CD(p)$ test can be easily adapted to match different types of spatial connection matrices used in the literature, and potentially provides a simple alternative to a number of tests available in the literature of spatial panels. In addition to its simplicity, the proposed $CD(p)$ test is also applicable to heterogeneous dynamic panels, and like the CD test is robust to structural breaks.

The power of the test naturally depends on the nature of the alternative hypothesis, and depends on the average value of the pair-wise correlations under the alternative. When this average is different from zero, the power of the test rises with $N\sqrt{T}$. In case of spatial panels with local dependence the power of the global CD test rises only with \sqrt{T} , and is not suitable for small T panels, and spatial versions of the test ought to be used. For first-order local dependence it is shown that the $CD(1)$ test has power that rises with \sqrt{NT} .

The small sample properties of the CD and the $CD(1)$ tests are investigated and compared to the LM test by means of a number of Monte Carlo experiments. It is shown that the CD test has the correct size and satisfactory power even under a weak degree of cross section dependence. It also turns out to be remarkably robust to major departures from normal errors, particularly for $T \geq 10$, and, as predicted by the theory, it is not affected by multiple breaks, so long as the unconditional means of the individual processes remain constant over time. In particular, in the case of symmetric errors the test has the correct size for T as small as 5 with N as large as 1000. The same also applies to the $CD(1)$ test investigated under first-order spatial dependence.

The use of the CD test in empirical contexts is illustrated by applying it to study the degree of dependence in per capita output innovations across countries within a given region and across countries in different regions. We also apply the test to an unbalanced panel where the dependence of UK output innovations on those of countries in other regions (including Europe, USA and Canada and Asia and Australia) is investigated. The results show significant evidence of cross dependence in output innovations across many countries and regions in the World, although with varying degrees.

The plan of the paper is as follows. The panel data model is set up in Section 2. The existing tests of cross section dependence are briefly reviewed in Section 3. The new CD test is formulated in Section 4, and its distribution established for strictly exogenous regressors. The robustness of the CD test to structural breaks is addressed in Section 5. The application of the test to heterogeneous dynamic panels with a fixed T is discussed in Section 6. Section 7 considers the problem of testing for the presence of local cross section dependence, using spatial panels as an example of such dependence. The power of the CD test is investigated in Section 8. The modification of the CD test statistic for unbalanced panels is discussed in Section 9. Section 10 reports the results of the Monte Carlo experiments. Section 11 discusses the empirical application, and Section 12 concludes.

2 Panel Data Models

Consider the following panel data model

$$y_{it} = \alpha_i + \beta_i' \mathbf{x}_{it} + u_{it}, \text{ for } i = 1, 2, \dots, N; t = 1, 2, \dots, T, \quad (1)$$

where i indexes the cross section dimension and t the time series dimension, \mathbf{x}_{it} is a $k \times 1$ vector of observed time-varying regressors (individual-specific as well as common regressors). The individual intercepts, α_i , and the slope coefficients, β_i , are defined on a compact set and allowed to vary across i . For each i , $u_{it} \sim IID(0, \sigma_{iu}^2)$, for all t , although they could be cross-sectionally correlated.² The dependence of u_{it} across i could arise in a number of different ways. It could be due to spatial dependence, omitted unobserved common components, or idiosyncratic pair-wise dependence of u_{it} and u_{jt} ($i \neq j$) with no particular pattern of spatial or common components. The regressors could contain lagged values of y_{it} , be either stationary (or integrated of order zero, $I(0)$) or have unit roots (or integrated of order 1, $I(1)$). But in the derivations below we assume $\mathbf{x}_{it} \sim I(0)$, and distinguish between the static and dynamic cases where the regressors are strictly exogenous and when they are weakly exogenous, specifically when $\mathbf{x}_{it} = (y_{i,t-1}, \dots, y_{i,t-p})$.

In panel data literature the primary focus of the analysis is often an average estimate of the slope coefficients, with the intercepts, α_i , being typically treated as nuisance (or incidental) parameters. For example, the fixed effects model imposes the slope homogeneity restrictions³

$$\beta_i = \beta, \text{ for all } i.$$

A less restrictive specification is given by the Swamy's (1970) random coefficient model

$$\beta_i = \beta + \mathbf{v}_i, \mathbf{v}_i \sim iid(\mathbf{0}, \mathbf{\Omega}_v),$$

with the means, β , being the focus of the analysis. The idea is to pool regressions from different cross section units in order to improve the efficiency with which the (mean) slope coefficients, β , are estimated. As we shall see our proposed test of cross section dependence is applicable to fixed and random effects models as well as to the more general heterogeneous slope or random coefficient specifications.

3 Existing Tests of Cross Section Dependence

Currently, there are two alternative approaches to testing for cross section dependence in panels, namely testing for spatial correlation pioneered by Moran (1948) and the Lagrange multiplier approach of Breusch and Pagan (1980, pp. 247-248).

²The assumption that u_{it} 's are serially uncorrelated is not restrictive and can be accommodated by including a sufficient number of lagged values of y_{it} amongst the regressors.

³The standard fixed effects estimator also assumes that $\sigma_i^2 = \sigma^2$.

3.1 Tests of Spatial Dependence

When the cross section units admit a time-invariant ordering as in spatial applications, the cross section dependence (known as spatial autocorrelation in the spatial literature) is tested with respect to a pre-specified connection matrix. The test is typically considered in the context of a single cross section, but can be readily extended to panels. In the context of (1), spatial dependence of the errors can be modelled using the spatial weight matrix, $\mathbf{W} = (w_{ij})$,

$$u_{it} = \lambda \left(\sum_{j=1}^N w_{ij} u_{jt} \right) + \sigma_i \varepsilon_{it} \text{ for } i = 1, 2, \dots, N, \quad (2)$$

where $\varepsilon_{it} \sim IID(0, 1)$ for all i and t . The spatial weights, w_{ij} , are typically assumed to be pre-specified, and the cross section dependence of the errors is investigated by testing the null hypothesis of $\lambda = 0$. In specification of \mathbf{W} , different types of geographical connections are used as set out, for example, in Cliff and Ord (1973, 1981). The statistical theory behind testing for $\lambda = 0$ is reviewed, for example, in Anselin (1988), Anselin and Bera (1998), Haining (2003), with further results and extensions provided in Baltagi, Song and Koh (2003).

In economic applications it is possible to use economic distance such as trade or output patterns for calibration of the spatial matrix. This, for example, allows the spatial weights to be time varying. Other extensions are also possible. But, in the final analysis the spatial dependence (autocorrelation) test could critically depend on the choice of \mathbf{W} . This might be inevitable for pure cross sections, although in a panel context it is possible to develop other types of tests that do not depend on \mathbf{W} .

3.2 Breusch and Pagan's Test of Cross Section Dependence

In the SURE context with N fixed and as $T \rightarrow \infty$, Breusch and Pagan (1980) proposed an Lagrange multiplier (LM) statistic for testing the null of zero cross equation error correlations which is particularly simple to compute and does not require the system estimation of the SURE model. The test is based on the following LM statistic

$$CD_{lm} = T \sum_{i=1}^{N-1} \sum_{j=i+1}^N \hat{\rho}_{ij}^2,$$

where $\hat{\rho}_{ij}$ is the sample estimate of the pair-wise correlation of the residuals. Specifically,

$$\hat{\rho}_{ij} = \hat{\rho}_{ji} = \frac{\sum_{t=1}^T e_{it} e_{jt}}{\left(\sum_{t=1}^T e_{it}^2 \right)^{1/2} \left(\sum_{t=1}^T e_{jt}^2 \right)^{1/2}}, \quad (3)$$

and e_{it} is the Ordinary Least Squares (OLS) estimate of u_{it} defined by

$$e_{it} = y_{it} - \hat{\alpha}_i - \hat{\beta}'_i \mathbf{x}_{it}, \quad (4)$$

with $\hat{\alpha}_i$ and $\hat{\beta}_i$ being the estimates of α_i and β_i computed using the OLS regression of y_{it} on an intercept and \mathbf{x}_{it} for each i , separately. Unlike the spatial dependence test, the LM test is more generally applicable and does not require a particular ordering of the cross section units. However, it is valid for N relatively small and T sufficiently large. In this setting Breusch and Pagan show that under the null hypothesis of no cross section dependence, specified by

$$\text{Cov}(u_{it}, u_{jt}) = 0, \text{ for all } t, i \neq j, \quad (5)$$

CD_{lm} is asymptotically distributed as chi-squared with $N(N-1)/2$ degrees of freedom. As it stands this test is not applicable when $N \rightarrow \infty$. However, noting that under H_0 ,

$$T\hat{\rho}_{ij}^2 \stackrel{a}{\sim} \chi_1^2$$

with $\hat{\rho}_{ij}^2$, $i = 1, 2, \dots, N-1$, $j = i+1, 2, \dots, N$, being asymptotically independent, the following scaled version of CD_{lm} can be considered for testing the hypothesis of cross dependence even for N and T large:

$$CD_{lm} = \sqrt{\frac{1}{N(N-1)}} \sum_{i=1}^{N-1} \sum_{j=i+1}^N (T\hat{\rho}_{ij}^2 - 1). \quad (6)$$

It is now easily seen that under H_0 with $T \rightarrow \infty$ first and then $N \rightarrow \infty$ we would have

$$CD_{lm} \stackrel{a}{\sim} N(0, 1).$$

However, this test is likely to exhibit substantial size distortions for N large and T small, a situation that can frequently arise in empirical applications. This is primarily due to the fact that for a finite T , $E(T\hat{\rho}_{ij}^2 - 1)$ will not be correctly centered at zero, and with N large the incorrect centering of the LM statistic is likely to be accentuated, resulting in size distortions that tend to get worse with N .

4 A New Test of Cross Section Dependence

There is clearly a need for a test of cross section dependence with reasonable small sample properties that does not depend on a particular spatial weight matrix, particularly when N is large and T small. Recognizing the shortcoming of the Breusch and Pagan's LM test when N is large, we propose the following simple alternative which is based on the pair-wise correlation coefficients rather than their squares used in the LM test:

$$CD = \sqrt{\frac{2T}{N(N-1)}} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^N \hat{\rho}_{ij} \right). \quad (7)$$

Unlike the LM test statistic, CD_{lm} , the above statistic has exactly mean zero for fixed values of T and N , under a wide class of panel data models, including heterogeneous dynamic models subject to multiple breaks in their slope coefficients and error variances, so long as the unconditional means of y_{it} and \mathbf{x}_{it} are time-invariant and their innovations are symmetrically distributed.

Initially, we provide a proof in the case of the standard panel data model, (1) subject to the following assumptions:

Assumption 1: For each i , the disturbances, u_{it} , are serially independent with zero means and the variance, σ_i^2 , such that $0 < \sigma_i < \infty$.

Assumption 2: Under the null hypothesis defined by

$$H_0 : u_{it} = \sigma_i \varepsilon_{it}, \text{ with } \varepsilon_{it} \sim IID(0, 1) \text{ for all } i \text{ and } t,$$

the disturbances, ε_{it} , are symmetrically distributed around 0.

Assumption 3: The regressors, \mathbf{x}_{it} , are strictly exogenous such that

$$E(u_{it} | \mathbf{X}_i) = \mathbf{0}, \text{ for all } i \text{ and } t,$$

where $\mathbf{X}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iT})'$, and $\mathbf{X}_i' \mathbf{X}_i$ is a positive definite matrix.

Assumption 4: $T > k + 1$ and the OLS residuals, e_{it} , defined by (4), are not all zero.⁴

Theorem 1 *Under Assumptions 1-4,*

$$E(\hat{\rho}_{ij}) = 0,$$

and

$$E(CD) = 0,$$

for all N and $T > k + 1$, where $\hat{\rho}_{ij}$ and CD are defined by (3) and (7), respectively.⁵

Proof. First note that the pair-wise correlation coefficients can be written as

$$\hat{\rho}_{ij} = \sum_{t=1}^T \xi_{it} \xi_{jt}, \quad (8)$$

where ξ_{it} are the scaled residuals defined by

$$\xi_{it} = \frac{e_{it}}{(\mathbf{e}_i' \mathbf{e}_i)^{1/2}}, \quad (9)$$

⁴The requirement $T > k + 1$ can be relaxed under slope homogeneity assumption, $\beta_i = \beta$ where fixed effects residuals can be used in the construction of the CD statistic instead of e_{it} .

⁵Similar results can also be obtained for fixed or random effects models. It suffices if the OLS residuals used in the computation of $\hat{\rho}_{ij}$ are replaced with associated residuals from fixed or random effects specifications. But the CD test based on the individual-specific OLS residuals are robust to slope and error-variance heterogeneity whilst the fixed or random effects residuals are not.

e_{it} is the OLS residuals from the individual-specific regressions, defined by (4), and $\mathbf{e}_i = (e_{i1}, e_{i2}, \dots, e_{iT})'$. Also,

$$\mathbf{e}_i = \sigma_i \mathbf{M}_i \boldsymbol{\varepsilon}_i, \text{ with } \mathbf{M}_i = \mathbf{I}_T - \mathbf{X}_i (\mathbf{X}_i' \mathbf{X}_i)^{-1} \mathbf{X}_i',$$

where $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{iT})'$. Therefore, conditional on \mathbf{x}_{it} , the scaled residuals, ξ_{it} , are odd functions of the disturbances, ε_{it} , and under Assumption 2 we have

$$E(\xi_{it} | \mathbf{X}_i) = 0, \text{ for all } i \text{ and } t.$$

Hence, unconditionally we also have

$$E(\xi_{it}) = 0, \text{ for all } i \text{ and } t.$$

Using this result in (8) now yields,

$$E(\hat{\rho}_{ij}) = 0,$$

which in turn establishes that (using (7))

$$E(CD) = 0,$$

for any N , and all $T > k + 1$. ■

Consider now the asymptotic distribution of

$$z_N = \sqrt{\frac{2}{N(N-1)}} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^N \hat{\rho}_{ij} \right).$$

under the null hypothesis, H_0 . First note that under Assumptions 1-4, $\hat{\rho}_{ij}$ and $\hat{\rho}_{is}$ are cross sectionally independent for i, j and s , such that $i \neq j \neq s$. In particular,

$$E(\hat{\rho}_{ij} \hat{\rho}_{is}) = \sum_{t=1}^T \sum_{t'=1}^T E(\xi_{it} \xi_{jt} \xi_{it'} \xi_{st'}) = \sum_{t=1}^T \sum_{t'=1}^T E(\xi_{it} \xi_{it'}) E(\xi_{jt}) E(\xi_{st'}) = 0,$$

and $\text{Var}(\hat{\rho}_{ij}) = E(\hat{\rho}_{ij}^2) \leq 1$. Therefore, by standard central limit theorems, for any fixed $T > k + 1$ and as $N \rightarrow \infty$

$$z_N \xrightarrow{N} N(0, v_z^2),$$

where

$$v_z^2 = \lim_{N \rightarrow \infty} \left[\frac{2}{N(N-1)} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^N E(\hat{\rho}_{ij}^2) \right) \right] \leq 1.$$

Hence, for a fixed $T > k + 1$, and as $N \rightarrow \infty$ we have

$$\frac{z_N}{v_z} = \frac{\sum_{i=1}^{N-1} \sum_{j=i+1}^N \hat{\rho}_{ij}}{\sqrt{\sum_{i=1}^{N-1} \sum_{j=i+1}^N E(\hat{\rho}_{ij}^2)}} \xrightarrow{N} N(0, 1).$$

An operational version of this test can be obtained using the non-parametric estimate

$$\hat{E}(\hat{\rho}_{ij}^2) = \sum_{t=1}^T \xi_{it}^2 \xi_{jt}^2 = \frac{\sum_{t=1}^T e_{it}^2 e_{jt}^2}{(\mathbf{e}'_i \mathbf{e}_i)(\mathbf{e}'_j \mathbf{e}_j)}.$$

In terms of the scaled residuals, ξ_{it} , the general form of the CD statistic will be given by

$$CD^* = \frac{\sum_{i=1}^{N-1} \sum_{j=i+1}^N \sum_{t=1}^T \xi_{it} \xi_{jt}}{\left\{ \sum_{i=1}^{N-1} \sum_{j=i+1}^N \sum_{t=1}^T \xi_{it}^2 \xi_{jt}^2 \right\}^{1/2}}.$$

But in most applications one would expect $\hat{E}(\hat{\rho}_{ij}^2)$ to be well approximated by $1/T$, in which case the simple version of the test, CD , given by (7) could be used.

The above result also holds as $T \rightarrow \infty$. To see this note that

$$\sqrt{T} \hat{\rho}_{ij} = \frac{\frac{1}{\sqrt{T}} (\boldsymbol{\varepsilon}'_i \mathbf{M}_i \mathbf{M}_j \boldsymbol{\varepsilon}_j)}{\left(\frac{\boldsymbol{\varepsilon}'_i \mathbf{M}_i \boldsymbol{\varepsilon}_i}{T} \right)^{1/2} \left(\frac{\boldsymbol{\varepsilon}'_j \mathbf{M}_j \boldsymbol{\varepsilon}_j}{T} \right)^{1/2}},$$

Therefore,

$$\sqrt{T} \hat{\rho}_{ij} = \frac{\frac{1}{\sqrt{T}} (\boldsymbol{\varepsilon}'_i \boldsymbol{\varepsilon}_j - \boldsymbol{\varepsilon}'_i \mathbf{A}_i \boldsymbol{\varepsilon}_j - \boldsymbol{\varepsilon}'_i \mathbf{A}_j \boldsymbol{\varepsilon}_j + \boldsymbol{\varepsilon}'_i \mathbf{A}_i \mathbf{A}_j \boldsymbol{\varepsilon}_j)}{\left(\frac{\boldsymbol{\varepsilon}'_i \mathbf{M}_i \boldsymbol{\varepsilon}_i}{T} \right)^{1/2} \left(\frac{\boldsymbol{\varepsilon}'_j \mathbf{M}_j \boldsymbol{\varepsilon}_j}{T} \right)^{1/2}}.$$

where $\mathbf{M}_i = \mathbf{I}_T - \mathbf{X}_i (\mathbf{X}'_i \mathbf{X}_i)^{-1} \mathbf{X}'_i = \mathbf{I}_T - \mathbf{A}_i$. However,⁶

$$\begin{aligned} \frac{\boldsymbol{\varepsilon}'_i \mathbf{M}_i \boldsymbol{\varepsilon}_i}{T} &= 1 + o_p(1), \\ \frac{\boldsymbol{\varepsilon}'_i \mathbf{A}_i \boldsymbol{\varepsilon}_j}{\sqrt{T}} &= \left(\frac{\boldsymbol{\varepsilon}'_i \mathbf{X}_i}{T} \right) \left(\frac{\mathbf{X}'_i \mathbf{X}_i}{T} \right)^{-1} \left(\frac{\mathbf{X}'_i \boldsymbol{\varepsilon}_j}{\sqrt{T}} \right) = o_p(1), \\ \frac{\boldsymbol{\varepsilon}'_i \mathbf{A}_i \mathbf{A}_j \boldsymbol{\varepsilon}_j}{\sqrt{T}} &= \left(\frac{\boldsymbol{\varepsilon}'_i \mathbf{X}_i}{T} \right) \left(\frac{\mathbf{X}'_i \mathbf{X}_i}{T} \right)^{-1} \left(\frac{\mathbf{X}'_i \mathbf{X}_j}{T} \right) \left(\frac{\mathbf{X}'_j \boldsymbol{\varepsilon}_j}{\sqrt{T}} \right) = o_p(1). \end{aligned} \quad (10)$$

Therefore,

$$\begin{aligned} \sqrt{T} \hat{\rho}_{ij} &= T^{-1/2} \boldsymbol{\varepsilon}'_i \boldsymbol{\varepsilon}_j + o_p(1), \\ &= \frac{1}{\sqrt{T}} \sum_{t=1}^T \varepsilon_{it} \varepsilon_{jt} + o_p(1), \end{aligned}$$

But under H_0 , ε_{it} and ε_{jt} are independently distributed and serially uncorrelated with mean 0 and a unit variance. Therefore, for each $i \neq j$, as $T \rightarrow \infty$

$$\sqrt{T} \hat{\rho}_{ij} \xrightarrow{T} N(0, 1).$$

⁶The following derivations assume that the regressors are stationary. But, it is easily seen that the same results follow for $I(1)$ regressors so long as the Assumptions 1-4 hold.

Using this result in (7) and letting $N \rightarrow \infty$, we now have (under H_0)

$$CD = \sqrt{\frac{2T}{N(N-1)}} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^N \hat{\rho}_{ij} \right) \Rightarrow N(0, 1).$$

It is therefore clear that the CD test is valid for N and T tending to infinity in any order. It is also clear that since the mean of CD is exactly equal to zero for all fixed $T > k + 1$ and N , the test is likely to have good small sample properties (for both N and T small), a conjecture which seems to be supported by extensive Monte Carlo experiments to be reported in Section 10.

5 Robustness of the CD Test to Structural Breaks

One of the key features of the proposed CD test is its robustness to single and even multiple structural breaks in the slope coefficients and the error variances of the individual regressions. Consider, for example, the following generalization of (1)

$$y_{it} = \begin{cases} \mu_{iy} + \beta'_{i1}(\mathbf{x}_{it} - \boldsymbol{\mu}_{ix}) + \sigma_{i1}\varepsilon_{it}, & t = 1, 2, \dots, T_{i1}, \\ \mu_{iy} + \beta'_{i2}(\mathbf{x}_{it} - \boldsymbol{\mu}_{ix}) + \sigma_{i2}\varepsilon_{it}, & t = T_{i1} + 1, \dots, T \end{cases} \quad \text{for } i = 1, 2, \dots, N, \quad (11)$$

where the slope coefficients, β_{ij} , and the error variances, σ_{ij}^2 , for $j = 1, 2$ are subject to a single break at time $t = T_{i1}$, and $\varepsilon_{it} \sim iid(0, 1)$. Note that the break point need not be the same across i . This specification also allows for induced changes in the regression intercepts

$$\alpha_{ij} = \mu_{iy} - \beta'_{ij}\boldsymbol{\mu}_{ix}, \quad j = 1, 2, \quad (12)$$

but restricts the unconditional means of y_t and \mathbf{x}_t , namely $\boldsymbol{\mu}_y$ and $\boldsymbol{\mu}_x$, to be time-invariant. Under this set up and assuming that ε_{it} and the innovations in \mathbf{x}_{it} are symmetrically distributed around zero, it can be shown that the OLS residuals, e_{it} , defined by (4), which ignores the possibility of structural breaks continues to be an odd function of ε_{it} , $t = 1, \dots, T$, and ensures that $\hat{\rho}_{ij}$ based on the OLS residuals that ignore the break(s) will have zero mean exactly for any fixed $T > k + 1$.⁷ Therefore, the CD test will have the correct size under structural breaks. In fact, the analysis readily extends to situations where the slopes and the errors variances have been subject to multiple breaks. In contrast, the LM test of cross section dependence need not have the correct size, even asymptotically, when the underlying regressions are subject to breaks.

6 Cross Section Dependence in Heterogeneous Dynamic Panels

Consider now the more demanding problem of testing cross section dependence in dynamic heterogeneous models, and as an example consider the following

⁷See Appendix A for a proof.

first-order dynamic panel data model

$$y_{it} = \alpha_i + \beta_i y_{i,t-1} + \sigma_i \varepsilon_{it}, \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T,$$

where $\alpha_i = \mu_i(1 - \beta_i)$, and the process initialized with $y_{i0} = \mu_i + \delta_i \varepsilon_{i0}$. Also, as before, we assume that for each i the errors, ε_{it} , $t = 0, 1, \dots, T$ are serially uncorrelated with a zero mean and a unit variance but could be cross-sectionally correlated.

The above specification is quite general and allows the underlying AR(1) processes to be stationary for some individuals and have a unit root for some other individuals in the panel. In the stationary case, if the process has started a long time in the past we would have $\delta_i = \sigma_i (1 - \beta_i^2)^{-1/2}$. In the unit root case where $\beta_i = 1$, δ_i could still differ across i depending on the number of periods that the i^{th} unit root process has been in operation before the initial observation, y_{i0} .

Given the complicated nature of the dynamics and the mix of stationary and unit processes that could prevail in a given panel, testing for cross section dependence might seem a hopeless undertaking for a fixed T and as $N \rightarrow \infty$. As it is well known the OLS estimates of α_i and β_i for the individual series, as well as the fixed and random effects panel estimates used under slope homogeneity ($\beta_i = \beta$) are biased when T is small.⁸ The bias could be substantial for values of β_i near unity. Nevertheless, as it turns out the CD test advanced in this paper is still valid for all values of β_i including unity. The main reason lies in the fact that despite the small sample bias of the parameter estimates, the OLS or fixed effects residuals have exactly mean zero even for a fixed T , so long as ε_{it} $t = 0, 1, \dots, T$ are symmetrically distributed. To see this we first write the individual AR(1) processes in matrix notations as

$$\mathbf{B}_i(\mathbf{y}_i^* - \mu_i \boldsymbol{\tau}_{T+1}) = \mathbf{D}_i \boldsymbol{\varepsilon}_i^*, \quad (13)$$

where $\mathbf{y}_i^* = (y_{i0}, y_{i1}, \dots, y_{iT})'$, $\boldsymbol{\varepsilon}_i^* = (\varepsilon_{i0}, \varepsilon_{i1}, \dots, \varepsilon_{iT})'$, $\boldsymbol{\tau}_{T+1}$ is a $(T+1) \times 1$ vector of ones, \mathbf{D}_i is a $(T+1) \times (T+1)$ diagonal matrix with its first element equal to δ_i and the remaining elements equal to σ_i , and

$$\mathbf{B}_i = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -\beta_i & 1 & 0 & & 0 & 0 \\ 0 & -\beta_i & 1 & & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & -\beta_i & 1 \end{pmatrix}.$$

The OLS estimates of individual intercepts and slopes can now be written as

$$\hat{\beta}_i = \frac{\boldsymbol{\varepsilon}_i^{*'} \mathbf{H}_i' \mathbf{G}_1' \mathbf{M}_\tau \mathbf{G}_0 \mathbf{H}_i \boldsymbol{\varepsilon}_i^*}{\boldsymbol{\varepsilon}_i^{*'} \mathbf{H}_i' \mathbf{G}_1' \mathbf{M}_\tau \mathbf{G}_1 \mathbf{H}_i \boldsymbol{\varepsilon}_i^*},$$

⁸See, for example, Nickell (1981).

$$\hat{\alpha}_i = \mu_i \left(1 - \hat{\beta}_i\right) + \left(\frac{\boldsymbol{\tau}'_T \mathbf{G}_0 \mathbf{H}_i \boldsymbol{\varepsilon}_i^*}{T}\right) - \left(\frac{\boldsymbol{\tau}'_T \mathbf{G}_1 \mathbf{H}_i \boldsymbol{\varepsilon}_i^*}{T}\right) \hat{\beta}_i.$$

where $\mathbf{H}_i = \mathbf{B}_i^{-1} \mathbf{D}_i$, $\mathbf{G}_0 = (\mathbf{0}_{T \times 1}, \mathbf{I}_T)$, $\mathbf{G}_1 = (\mathbf{I}_T, \mathbf{0}_{T \times 1})$, and $\mathbf{0}_{T \times 1}$ is a $T \times 1$ vector of zeros. Using these results we now have the following expression for the OLS residuals, $e_{it} = y_{it} - \hat{\alpha}_i - \hat{\beta}_i y_{i,t-1}$,

$$e_{it} = -\left(\hat{\beta}_i - \beta_i\right) (y_{i,t-1} - \mu_i) + \sigma_i \varepsilon_{it} - \left(\frac{\boldsymbol{\tau}'_T \mathbf{G}_0 \mathbf{H}_i \boldsymbol{\varepsilon}_i^*}{T}\right) + \left(\frac{\boldsymbol{\tau}'_T \mathbf{G}_1 \mathbf{H}_i \boldsymbol{\varepsilon}_i^*}{T}\right) \hat{\beta}_i.$$

Using (13) we also note that $y_{i,t-1} - \mu_i = \mathbf{s}'_{t-1} \mathbf{H}_i \boldsymbol{\varepsilon}_i^*$, where \mathbf{s}_{t-1} is a $(T+1) \times 1$ selection vector with zero elements except for its t^{th} element which is unity. Therefore, e_{it} , and hence $\xi_{it} = (\mathbf{e}'_i \mathbf{e}_i)^{-1/2} e_{it}$ will be an odd function of $\boldsymbol{\varepsilon}_i^*$, and we have $E(\xi_{it}) = 0$, $t = 1, 2, \dots, T$, under the assumption that $\boldsymbol{\varepsilon}_i^*$ has a symmetric distribution. Thus, under the null hypothesis that ε_{it} and ε_{jt} are cross sectionally independent we have $E(\hat{\rho}_{ij}) = 0$, and the CD test advanced here continues to hold for dynamic heterogeneous panel data models. Furthermore, the test will be robust to structural breaks so long as the unconditional mean of the process remains unchanged, namely if $E(y_{it}) = \mu_i$, for all t . A proof in the case of a single break, but in the context a general AR(p) process, is provided in Appendix B. The robustness of the CD test in the presence of multiple breaks will be illustrated by means of Monte Carlo experiments in Section 10.

7 CD Tests for Local Cross Section Dependence

The power of the CD test can be substantially enhanced when the dependence under the alternative hypothesis is local. A prominent example of such dependence is the various spatial patterns defined on a lattice in the spatial literature.⁹ Local dependence is usually defined with respect to a weight matrix, $\mathbf{W} = (w_{ij})$, applied to a particular ordering of the cross section units. A well known example is the so-called “rook” formation under which $w_{ij} = 1$ if the i^{th} cross section units is topologically adjacent to the j^{th} cross section unit, and $w_{ij} = 0$, otherwise.¹⁰ This is an example of a first-order neighbourhood dependence, and can be readily extended to higher orders. It is often convenient to order the cross section units by their topological position, so that the p^{th} order neighbours of the i^{th} cross section unit can be defined as the $i + p$ and the $i - p$ cross section units.

To test for the presence of local dependence we shall condition on the order of local dependence. Under the alternative hypothesis of a p^{th} order local dependence we propose the following generalization of the *CD* statistic defined by

⁹See, for example, Cliff and Ord (1973).

¹⁰Another possibility that could be more relevant for the analysis of economic and financial panels would be to set $w_{ij} = 1$, if the “economic distance” between the i^{th} and the j^{th} cross section units is less than a threshold, \bar{d} , and $w_{ij} = 0$, otherwise.

(7):

$$\begin{aligned} CD(p) &= \sqrt{\frac{2T}{p(2N-p-1)}} \left(\sum_{s=1}^p \sum_{i=s+1}^N \hat{\rho}_{i,i-s} \right), \\ &= \sqrt{\frac{2T}{p(2N-p-1)}} \left(\sum_{s=1}^p \sum_{i=1}^{N-s} \hat{\rho}_{i,i+s} \right), \end{aligned} \quad (14)$$

where $p = 1, 2, \dots, N-1$. We shall refer to this test as the local CD test of order p , or $CD(p)$ for short. Clearly, $CD(N-1)$ reduces to the CD statistic. This is not surprising as local dependence makes sense only for values of $p < N-1$.

Like CD , under Assumptions 1-4 the $CD(p)$ statistic has mean zero exactly for any N and all $T > k+1$. It is also easily seen that for a given p , and for N and T sufficiently large, $CD(p) \stackrel{d}{\sim} N(0, 1)$. Note, however, that $CD(p)$ test can be implemented only if the cross section units can be given an ordering which remains immutable over time. In the absence of such ordering, $CD = CD(N-1)$ can be used which is invariant to re-ordering of the cross section units.

8 Power of the CD Tests

Naturally the power of the CD depends on the form of the alternative hypothesis. Here we consider two possible specifications, namely a global cross dependence alternative defined via a multi-factor model, and a local cross dependence alternative based on a first-order spatial model.

8.1 Power under Global Cross Dependence

A general alternative specification that allows for global cross section dependence in panels is the unobserved multi-factor model defined by

$$y_{it} = \alpha_i + \beta_i \mathbf{x}_{it} + u_{it},$$

where

$$u_{it} = \sigma_i (\boldsymbol{\gamma}'_i \mathbf{f}_t + \varepsilon_{it}), \quad \varepsilon_{it} \sim IID(0, 1).$$

The $m \times 1$ vector of unobserved common factors could be any general stationary process and without loss of generality can be orthonormalized to have mean zero and the variance covariance matrix \mathbf{I}_m . Alternative assumptions regarding the dependence of \mathbf{f}_t on ε_{it} and the regressors, \mathbf{x}_{it} could also be entertained. But to simplify the analysis we assume that for each i

$$\begin{cases} T^{-1} \sum_{t=1}^T \varepsilon_{it} \mathbf{f}_t = O_p(T^{-1/2}), \quad T^{-1} \sum_{t=1}^T \mathbf{x}_{it} \mathbf{f}_t' = O_p(T^{-1/2}), \\ T^{-1} \sum_{t=1}^T \mathbf{f}_t \mathbf{f}_t' = \mathbf{I}_m + O_p(T^{-1/2}). \end{cases} \quad (15)$$

We also make the following standard assumptions about the regressors¹¹

$$\left(\frac{\mathbf{X}'_i \mathbf{X}_j}{T}\right) = \boldsymbol{\Sigma}_{ij} + O_p(T^{-1/2}), \quad \left(\frac{\mathbf{X}'_i \boldsymbol{\varepsilon}_i}{T}\right) = O_p(T^{-1/2}), \quad (16)$$

where $\boldsymbol{\Sigma}_{ii}$ is a positive definite matrix.

The extent of pair-wise cross section dependence in the panel is measured by the factor loadings $\boldsymbol{\gamma}_i$ and $\boldsymbol{\gamma}_j$. The null hypothesis of cross section independence is now defined by

$$H_0 : \boldsymbol{\gamma}_i = \mathbf{0},$$

as compared to the alternatives that

$$H_g : \boldsymbol{\gamma}_i \neq \mathbf{0}, \text{ for some } i = 1, \dots, N, \quad (17)$$

Under the alternative hypothesis the true value of ρ_{ij} is given by

$$\rho_{ij} = \frac{\boldsymbol{\gamma}'_i \boldsymbol{\gamma}_j}{(1 + \boldsymbol{\gamma}'_i \boldsymbol{\gamma}_i)^{1/2} (1 + \boldsymbol{\gamma}'_j \boldsymbol{\gamma}_j)^{1/2}} \neq 0. \quad (18)$$

Consider now the CD test statistic defined by (7) and note that under H_a , the vector of the OLS residuals is given by

$$\mathbf{e}_i = \sigma_i (\mathbf{M}_i \boldsymbol{\varepsilon}_i + \mathbf{M}_i \mathbf{F} \boldsymbol{\gamma}_i),$$

where $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_T)'$, and as before $\mathbf{M}_i = \mathbf{I}_T - \mathbf{X}_i (\mathbf{X}'_i \mathbf{X}_i)^{-1} \mathbf{X}'_i$. Hence,

$$\hat{\rho}_{ij} = \frac{\boldsymbol{\varepsilon}'_i \mathbf{M}_i \mathbf{M}_j \boldsymbol{\varepsilon}_j + \boldsymbol{\gamma}'_i \mathbf{F}' \mathbf{M}_i \mathbf{M}_j \mathbf{F} \boldsymbol{\gamma}_j + \boldsymbol{\varepsilon}'_i \mathbf{M}_i \mathbf{M}_j \mathbf{F} \boldsymbol{\gamma}_j + \boldsymbol{\varepsilon}'_j \mathbf{M}_j \mathbf{M}_i \mathbf{F} \boldsymbol{\gamma}_i}{(\boldsymbol{\varepsilon}'_i \mathbf{M}_i \boldsymbol{\varepsilon}_i + 2\boldsymbol{\varepsilon}'_i \mathbf{M}_i \mathbf{F} \boldsymbol{\gamma}_i + \boldsymbol{\gamma}'_i \mathbf{F}' \mathbf{M}_i \mathbf{F} \boldsymbol{\gamma}_i)^{1/2} (\boldsymbol{\varepsilon}'_j \mathbf{M}_j \boldsymbol{\varepsilon}_j + 2\boldsymbol{\varepsilon}'_j \mathbf{M}_j \mathbf{F} \boldsymbol{\gamma}_j + \boldsymbol{\gamma}'_j \mathbf{F}' \mathbf{M}_j \mathbf{F} \boldsymbol{\gamma}_j)^{1/2}}. \quad (19)$$

The finite sample property of $\hat{\rho}_{ij}$ under H_g is quite complicated and depends on the magnitude of the factor loadings and the cross correlation patterns of the regressors and the unobserved factors. It does not, however, depend on the error variances, σ_i^2 . However, it is possible to derive the asymptotic power function of the CD test under the “local” alternatives:¹²

$$H_{gNT} : \boldsymbol{\gamma}_i = \frac{1}{T^{1/4} N^{1/2}} \boldsymbol{\delta}_i, \quad (20)$$

where

$$\psi = p \lim_{N \rightarrow \infty} \left(\frac{2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \boldsymbol{\delta}'_i \boldsymbol{\delta}_j}{N(N-1)} \right) \neq 0. \quad (21)$$

Under (20), the numerator of (19) scaled by T^{-1} can be written as

$$\begin{aligned} \frac{T^{-1} \mathbf{e}'_i \mathbf{e}_j}{\sigma_i \sigma_j} &= \frac{\boldsymbol{\varepsilon}'_i \mathbf{M}_i \mathbf{M}_j \boldsymbol{\varepsilon}_j}{T} + N^{-1} T^{-1/2} \boldsymbol{\delta}'_i \left(\frac{\mathbf{F}' \mathbf{F}}{T} \right) \boldsymbol{\delta}_j \\ &\quad + N^{-1/2} T^{-1/4} \left(\frac{\boldsymbol{\varepsilon}'_i \mathbf{M}_i \mathbf{M}_j \mathbf{F} \boldsymbol{\delta}_j + \boldsymbol{\varepsilon}'_j \mathbf{M}_j \mathbf{M}_i \mathbf{F} \boldsymbol{\delta}_i}{T} \right), \end{aligned}$$

¹¹These assumptions allow for the inclusion of lagged dependent variables amongst the regressors and can be relaxed further to take account of non-stationary $I(1)$ regressors.

¹²Here “local” is used in the mathematical statistical sense originally developed by Pitman.

and using (10), (15), (16) it is easily seen that the denominator of (19) is equal to $1 + o_p(1)$ (as T and $N \rightarrow \infty$). Hence, using these results in (7), and after some algebra we obtain

$$\begin{aligned} CD &= \sqrt{\frac{2}{TN(N-1)}} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \sum_{t=1}^T \varepsilon_{it} \varepsilon_{jt} + \\ &\quad \frac{1}{N} \sqrt{\frac{2}{N(N-1)}} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \delta'_i \left(\frac{\mathbf{F}'\mathbf{F}}{T} \right) \delta_j + \\ &\quad N^{-1/2} T^{-1/4} \sqrt{\frac{2}{N(N-1)}} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \left(\frac{\varepsilon'_i \mathbf{M}_i \mathbf{M}_j \mathbf{F} \delta_j + \varepsilon'_j \mathbf{M}_j \mathbf{M}_i \mathbf{F} \delta_i}{\sqrt{T}} \right) + o_p(1). \end{aligned}$$

The first term represents sums of *iid*(0, 1) random variables and tend to $N(0, 1)$ as both N and T tend to infinity. The second term tends to ψ , the non-centrality parameter defined by (21). Under the assumption that ε_{it} and δ_j are distributed independently for all i, j , and t , the third term tends to zero in probability as N and T tend to infinity. Hence, under the Pitman local alternatives, $H_{g,NT}$, and the assumptions (15) and (16) we have

$$CD | H_{g,NT} \xrightarrow{(N,T)} N(\psi, 1).$$

This result does not depend on the order by which N and T tend to infinity. It also establishes the consistency of the CD test at the rate of $N\sqrt{T}$, under the multi-factor alternatives, so long as $\psi \neq 0$. This condition is clearly satisfied under the homogeneous alternative, $\delta_j = \delta \neq \mathbf{0}$, which yields $\psi = \delta' \delta > 0$. Under heterogeneous alternatives it is possible for ψ to be zero, even if some or all of the individual factor loadings are non-zero under fixed alternatives. This can arise if γ_i are independently distributed with mean zeros and finite second-order moments. In such a case the CD test will no longer be consistent. This is clearly a limitation (not shared by the LM test), but its practical significance seems rather doubtful. As it is shown in Pesaran (2002), under $E(\gamma_i) = \mathbf{0}$, the cross section averages, \bar{y}_t and $\bar{\mathbf{x}}_t$, defined by

$$\bar{y}_t = N^{-1} \sum_{i=1}^N y_{it}, \text{ and } \bar{\mathbf{x}}_t = N^{-1} \sum_{i=1}^N \mathbf{x}_{it},$$

will be perfectly correlated for sufficiently large N , an outcome rarely encountered in the analysis of stationary aggregate time series.

8.2 Power Under Local Cross Dependence

Similar power analysis can also be carried under local cross section dependence alternatives discussed in Section 7. Here we focus on the following first-order neighborhood alternative

$$u_{it} = \lambda(b_i u_{i-1,t} + a_i u_{i+1,t}) + \sigma_i \varepsilon_{it}, \text{ for } i = 2, \dots, N-1, \quad (22)$$

with end points set at $u_{1t} = a_1 u_{2t} + \varepsilon_{1t}$ and $u_{Nt} = b_N u_{N-1,t} + \varepsilon_{Nt}$. Higher order spatial models can also be similarly analysed. Writing all the N errors in matrix notations for each period we have

$$\mathbf{u}_{\circ t} = (\mathbf{I}_N - \lambda \mathbf{W})^{-1} \boldsymbol{\Sigma}^{1/2} \boldsymbol{\varepsilon}_{ot}, \quad (23)$$

where $\mathbf{u}_{\circ t} = (u_{1t}, \dots, u_{Nt})'$, $\boldsymbol{\varepsilon}_{ot} = (\varepsilon_{1t}, \dots, \varepsilon_{Nt})'$, $\boldsymbol{\Sigma}$ is an $N \times N$ diagonal matrix with σ_i^2 on its i^{th} diagonal, and \mathbf{W} is the spatial weight matrix¹³

$$\mathbf{W} = \begin{pmatrix} 0 & a_1 & 0 & 0 & \cdots & \cdots & 0 \\ b_2 & 0 & a_2 & 0 & & & \vdots \\ 0 & b_3 & 0 & a_3 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & & & \ddots & b_{N-1} & 0 & a_{N-1} \\ 0 & \cdots & \cdots & \cdots & 0 & b_N & 0 \end{pmatrix} \quad (N \times N),$$

where $0 < a_i, b_i < \infty$. The correlation matrix for this model is given by

$$\mathbf{R} = (\rho_{ij}) = \boldsymbol{\Sigma}^{-1/2} (\mathbf{I}_N - \lambda \mathbf{W})^{-1} \boldsymbol{\Sigma} (\mathbf{I}_N - \lambda \mathbf{W}')^{-1} \boldsymbol{\Sigma}^{-1/2},$$

which simplifies to

$$\mathbf{R} = [\mathbf{I}_N - \lambda (\mathbf{G} + \mathbf{G}') + \lambda^2 \mathbf{G}' \mathbf{G}]^{-1},$$

where

$$\mathbf{G} = \boldsymbol{\Sigma}^{-1/2} \mathbf{W} \boldsymbol{\Sigma}^{1/2}.$$

For this alternative the CD test has power only with respect to T . To see this consider the alternatives

$$H_{\ell T} : \lambda = \frac{\delta}{\sqrt{T}},$$

and note that \mathbf{R} can be written as

$$\mathbf{R} = \mathbf{I}_N + \frac{\delta}{\sqrt{T}} (\mathbf{G} + \mathbf{G}') + O(T^{-1}). \quad (24)$$

Hence, using (7), for a fixed N and as $T \rightarrow \infty$, we have

$$p \lim_{T \rightarrow \infty} \left(\frac{CD}{\sqrt{T}} | H_{\ell T} \right) = \left(\frac{2(N-1)}{N} \right)^{1/2} \varphi_{N-1},$$

¹³In the spatial literature it is typically assumed that $a_i = b_i = 0.5$ and \mathbf{W} is known as the ‘‘rook’’ formation.

where (noting that $0 < a_i, b_i, \sigma_i < \infty$)

$$\varphi_{N-1} = \frac{\sum_{i=2}^N \left[a_{i-1} \left(\frac{\sigma_i}{\sigma_{i-1}} \right) + b_i \left(\frac{\sigma_{i-1}}{\sigma_i} \right) \right]}{N-1} > 0.$$

Therefore, the CD test has power against spatial dependence only as $T \rightarrow \infty$, and its power is unaffected along the cross section dimension.

This problem is circumvented by making use of the spatial $CD(p)$ test statistic defined by (14), for a fixed order p , as $N \rightarrow \infty$. For example, for $p = 1$ we have

$$CD(1) = \sqrt{\frac{T}{N-1}} \left(\sum_{i=1}^{N-1} \hat{\rho}_{i,i+1} \right), \quad (25)$$

and using (24) under the Pitman alternatives

$$H_{\ell TN} : \lambda = \frac{\delta}{\sqrt{TN}},$$

it is easily seen that

$$p \lim_{T, N \rightarrow \infty} \left(\frac{CD(1)}{\sqrt{NT}} \mid H_{\ell TN} \right) = \delta\varphi, \quad (26)$$

where

$$\varphi = \lim_{N \rightarrow \infty} \left\{ \frac{\sum_{i=2}^N \left[a_{i-1} \left(\frac{\sigma_i}{\sigma_{i-1}} \right) + b_i \left(\frac{\sigma_{i-1}}{\sigma_i} \right) \right]}{N-1} \right\} > 0. \quad (27)$$

It is also easily seen that

$$CD(1) \mid H_{\ell TN} \xrightarrow{(N,T)} N(\delta\varphi, 1). \quad (28)$$

Once again the outcome does not depend on the order by which N and T are allowed to tend to infinity. To establish this result we first note that for each i and t , under $H_{\ell TN}$ and using (23) we have¹⁴

$$u_{it} = \sigma_i \varepsilon_{it} + \frac{\delta}{\sqrt{NT}} (b_i \sigma_{i-1} \varepsilon_{i-1,t} + a_i \sigma_{i+1} \varepsilon_{i+1,t}) + O_p \left(\frac{1}{NT} \right). \quad (29)$$

Now following similar lines of reasoning as before, it is possible to show that

$$CD(1) = \sqrt{\frac{1}{T(N-1)}} \left(\sum_{i=1}^{N-1} \sum_{t=1}^T \left(\frac{u_{it}}{\sigma_i} \right) \left(\frac{u_{i+1,t}}{\sigma_{i+1}} \right) \right) + O_p \left(\frac{1}{\sqrt{NT}} \right). \quad (30)$$

Also using (29) we have

$$\begin{aligned} \left(\frac{u_{it}}{\sigma_i} \right) \left(\frac{u_{i+1,t}}{\sigma_{i+1}} \right) &= \varepsilon_{it} \varepsilon_{i+1,t} + \frac{\delta}{\sqrt{NT}} \left(\frac{b_{i+1} \sigma_i}{\sigma_{i+1}} \varepsilon_{it} + \frac{a_i \sigma_{i+1}}{\sigma_i} \varepsilon_{i+1,t} \right) \\ &\quad + \frac{\delta}{\sqrt{NT}} \left(\frac{b_i \sigma_{i-1}}{\sigma_i} \varepsilon_{i+1,t} \varepsilon_{i-1,t} + \frac{a_{i+1} \sigma_{i+2}}{\sigma_{i+1}} \varepsilon_{it} \varepsilon_{i+2,t} \right) + O_p \left(\frac{1}{NT} \right). \end{aligned}$$

¹⁴Note that under $H_{\ell TN}$, $\mathbf{u}_{\circ t} = \left(\mathbf{I}_N + \frac{\delta}{\sqrt{NT}} \mathbf{W} \right) \boldsymbol{\Sigma}^{1/2} \boldsymbol{\varepsilon}_{\circ t} + O_p \left(\frac{1}{NT} \right)$.

Substituting this result in (30) and recalling that ε_{it} are $iid(0, 1)$, we have the desired result in (28).

To summarize, the CD test, defined by (7), is suitable for unordered data under global alternatives such as the multi-factor residual models, whilst the $CD(p)$ test defined by (14), with p fixed as $N \rightarrow \infty$, is appropriate for the analysis of spatial dependence where the cross section units can be ordered. Under suitably defined Pitman local alternatives the power functions of CD and $CD(1)$ tests are shown to be asymptotically normally distributed, with the power function of the $CD(1)$ test being symmetric with respect to positive and negative departures from the null hypothesis, $\lambda = 0$.¹⁵ It is clearly desirable also to derive the power function of the $CD(p)$ test under p^{th} order spatial alternatives, but this is beyond the scope of the present paper.

9 CD Tests in the Case of Unbalanced Panels

The $CD(p)$ tests can be readily extended to unbalanced panels, a situation that frequently arises in practice. Denote by \mathcal{T}_i , the set of dates over which time series observations on y_{it} and \mathbf{x}_{it} are available for the i^{th} individual, and the number of the elements in the set by $\#\mathcal{T}_i$. For each i compute the OLS residuals based on full set of time series observations for that individual. As before, denote these residuals by e_{it} , for $t \in \mathcal{T}_i$, and compute the pair-wise correlations of e_{it} and e_{jt} using the *common* set of data points in $\mathcal{T}_i \cap \mathcal{T}_j$. Since, the estimated residuals need not sum to zero over the common sample period ρ_{ij} could be estimated by

$$\hat{\rho}_{ij} = \frac{\sum_{t \in \mathcal{T}_i \cap \mathcal{T}_j} (e_{it} - \bar{e}_i)(e_{jt} - \bar{e}_j)}{\left[\sum_{t \in \mathcal{T}_i \cap \mathcal{T}_j} (e_{it} - \bar{e}_i)^2 \right]^{1/2} \left[\sum_{t \in \mathcal{T}_i \cap \mathcal{T}_j} (e_{jt} - \bar{e}_j)^2 \right]^{1/2}},$$

where

$$\bar{e}_i = \frac{\sum_{t \in \mathcal{T}_i \cap \mathcal{T}_j} e_{it}}{\#(\mathcal{T}_i \cap \mathcal{T}_j)}.$$

The (global) CD statistic for the unbalanced panel, for example, is given by

$$CD = \sqrt{\frac{2}{N(N-1)}} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^N \sqrt{T_{ij}} \hat{\rho}_{ij} \right), \quad (31)$$

where $T_{ij} = \#(\mathcal{T}_i \cap \mathcal{T}_j)$. Under the null hypothesis $CD \sim N(0, 1)$ for $T_i > k+1$, $T_{ij} > 3$, and sufficiently large N . Similar adjustments can also be made to the $CD(p)$ test defined by (14).

¹⁵The asymptotic power function of the CD test is also symmetric under homogeneous alternatives, $\gamma_i = \gamma$.

10 Small Sample Properties of LM and CD Tests

In comparing the small sample properties of the LM and the CD tests we carried out a large number of Monte Carlo experiments, both for models with strictly exogenous regressors and for stationary and unit root dynamic panels with and without parameter heterogeneity and structural breaks. In all cases the CD test has size very close to its nominal value and exhibit increasing power with N and T , although as predicted by the theory the power increases much faster with N than with T . In what follows we shall focus on the dynamic experiments as they might be thought more likely to be subject to the familiar small T bias. Results from the other experiments are available from the author on request.

10.1 Monte Carlo Setup

The dynamic experiments were based on the following data generating process:

$$\begin{aligned} y_{it} &= \mu_i(1 - \beta_i) + \beta_i y_{i,t-1} + u_{it} \\ u_{it} &= \gamma_i f_t + \varepsilon_{it}, \quad i = 1, 2, \dots, N; t = 1, 2, \dots, T. \end{aligned} \quad (32)$$

The idiosyncratic errors, ε_{it} , were generated under two different schemes. (i) Normal errors: $\varepsilon_{it} \sim iidN(0, 1)$, and (ii) Chi-squared errors, $\varepsilon_{it} \sim iid \chi_{(1)}^2$. We also tried the Student t distribution with 4 degrees of freedom, but the results were indistinguishable from the ones based on normal errors. The use of $\chi_{(1)}^2$ errors is intended to check the robustness of the CD test to extreme departures from symmetry of error distributions required for the test to be valid in small samples. For the slope coefficients, β_i , we experimented with homogeneous as well as heterogeneous values. The results turned out to be very similar. Here we focus on the heterogeneous slope experiments where $\beta_i \sim iidU(0, 1)$. The fixed effects, μ_i , were drawn as $\varepsilon_{i0} + \eta_i$, with $\eta_i \sim iidN(0, 1)$, thus allowing for the possibility of correlations between fixed effects and the initial values, y_{i0} .

Under the null hypothesis we have $\gamma_i = 0$. Under the alternative hypothesis we experimented with different degrees of cross section error correlations as set out below

$$\begin{aligned} f_t &\sim iidN(0, 1) \\ \text{Low} &: \gamma_i \sim iidU(0.1, 0.3), \\ \text{Moderate} &: \gamma_i \sim iidU(0.2, 0.4), \\ \text{Sizeable} &: \gamma_i \sim iidU(0.3, 0.5). \end{aligned}$$

It turned out the CD test has ample power even for the low cross-dependence scenario, and therefore we shall be reporting power results only under $\gamma_i \sim iidU(0.1, 0.3)$. The average pair-wise correlation for this scenario is

$$q = E \left(\frac{\gamma_i \gamma_j}{\sqrt{(1 + \gamma_i^2)(1 + \gamma_j^2)}} \right) = \left[E \left(\frac{\gamma_i}{\sqrt{1 + \gamma_i^2}} \right) \right]^2 = 0.038.$$

We also considered the alternatives $\gamma_i \sim iidU(-0.2, 0.6)$, where γ_i is allowed to cover negative as well as positive values. The value of q for this case is 0.0324.

The parameters γ_i, β_i and η_i are drawn independently of ε_{it} once for each experiment, and then fixed throughout the replications. ε_{it} and the time-specific common effect, f_t , are newly drawn for each replication, independently of each other. All experiments are carried out for N and $T = 5, 10, 20, 30, 50, 100$. The LM and the CD statistics, defined by (6) and (7), are computed using the OLS residuals from the individual regressions of y_{it} on an intercept and $y_{i,t-1}$. The nominal size of the tests is set at the 5 per cent significance level. The number of replications is 1,000.

10.2 Results

Table 1a reports the size of the LM and the CD tests for the normal errors, with the associated powers under the two alternatives, $\gamma_i \sim iidU(0.1, 0.3)$ and $\gamma_i \sim iidU(-0.2, 0.6)$, given in Table 1b. Perhaps not surprisingly, the LM test is not at all effective at controlling the size of the test, except for small values of N relative to T . It tends to over-reject particularly for $N \geq T$, with the extent of over-rejection being rather severe for values of T in the range of 20 – 30. By contrast the CD test has the correct size for all values of N and T , (even as small as 5), and as predicted by theory its power tends to rise with N and T , quite rapidly. The power of the CD test tends to be slightly higher under the positive alternatives $\gamma_i \sim iidU(0.1, 0.3)$, as compared to the factor loadings that cover both positive and negative values. We also carried out a number of experiments with N relatively large (100, 200, and 300) and obtained sizes very close to 5% for the CD test.¹⁶ Furthermore, we obtained very similar results using AR(2) specifications, which are available from the author on request.

The results for the $\chi^2(1)$ errors are summarized in Table 2, and show the CD test to be remarkably robust to departures from symmetric error distributions, unless T is very small, around 5 or less in the current set of experiments.

As pointed out in Sections 5 and 6, another important feature of the CD is its robustness to structural change. To evaluate the small sample properties of the test in the presence of breaks in the slope parameters and/or in the error variances we employed two different DGPs: one with a single break, another with breaks in every period! Under the former we generated y_{it} as

$$y_{it} - \mu_i = \beta_{it}(y_{it-1} - \mu_i) + u_{it}, \quad u_{it} = \gamma_i f_t + \sigma_{it} \varepsilon_{it},$$

with $\mu_i \sim iidN(1, 1)$, $\beta_{it} = \beta_t = 0.6$ for $t = -50, \dots, T/2$, $\beta_t = 0.8$ for $t = T/2, \dots, T$; $\sigma_{it} = \sigma_t = \sqrt{1.5}$ for $t = -50, \dots, T/2$, $\sigma_t = 1$ for $t = T/2, \dots, T$, and $\varepsilon_{it} \sim iidN(0, 1)$. This postulates a single break in slopes and the errors variances right in the middle of the sample. Under the second DGP the slopes and error variances were generated to vary randomly every time and across individuals, namely we set $\beta_{it} \sim iidU(0, 1)$ for $t = 1, \dots, T$, $i = 1, \dots, N$; $\sigma_{it}^2 \sim$

¹⁶The size and power of the CD test for $T = 5$ and $N = 1000$ turned out to be 0.055 and 0.990, respectively. The corresponding values for the LM test were 1.00 and 1.00!

$iid\chi^2(2)/2$ for $t = -50, \dots, 0, 1, 2, \dots, T$, $i = 1, \dots, N$, and discarded the first 49 observations. Size and power of LM and CD tests for these two DGPs are summarized in Tables 3 and 4. As can be seen the CD test continues to perform remarkably well. It has the correct size under both DGP's and for all sample sizes. The LM test continues to over-reject, now slightly more as compared to the no-break case. The power of the CD test is very little affected by a single break, although there is a noticeable improvement in power of the CD test under the multiple break scenario.

Table 5 provides the results for the pure unit root case where β_{it} is set to unity for all i and t , and $\sigma_{it}^2 \sim iid\chi^2(2)/2$, as before. As can be seen the test sizes are hardly affected by the presence of the unit root, which is in line with our theoretical results. The power of the CD test is marginally better, however. In contrast, the LM test continues to show substantial over-rejections particularly for $N \geq T$.

Finally, we considered the size and power of the first-order spatial CD test, based on the CD(1) statistic given by (25). For this purpose we used the same DGP as in Table 1, under the null hypothesis of cross section independence, but for the analysis of the power of the CD(1) test used the spatial alternative given by

$$u_{it} = \lambda(0.5u_{i-1,t} + 0.5u_{i+1,t}) + \sigma_i\varepsilon_{it}, \text{ for } i = 2, \dots, N - 1,$$

with the end points set at $u_{1t} = u_{2t} + \varepsilon_{1t}$ and $u_{Nt} = u_{N-1,t} + \varepsilon_{Nt}$. Size and power of the CD(1) test, along with the LM test for comparison, are summarized in Tables 6a and 6b, respectively. The CD(1) test tends to have the correct size, except for $T = 5$, where the test is slightly over-sized. The power of the CD(1) test is computed under $\lambda = \pm 0.1$. These are the smallest values considered for λ by Baltagi, Song and Koh (2003) in their experiments. Under both of these alternatives, the CD(1) test exhibit reasonable power and rises with \sqrt{TN} , as predicted by the theory. The power is very similar for positive and negative values of λ .

The LM test continues to be substantially over-sized unless $N < T$. For such sample sizes, however, the CD(1) test that exploits the spatial information exhibits much better power than the LM test. The power of the CD(1) test also seems to be quite close to the ones reported by Baltagi, Song and Koh (2003) in the context of a panel data model with homogeneous slopes and non-stochastic regressors; although a formal comparison is beyond the scope of the present paper.

Overall, the Monte Carlo experiments confirm our theoretical findings, and show that the proposed test has good small sample properties, even for small T and large N panels.

11 Cross Section Dependence in Per Capita Output Innovations Within and Across Regions

One of the maintained assumptions of the empirical growth literature has been the cross section independence of innovations in per capita output, although it has been recognized by many researchers in the field that such an assumption might not be justified.¹⁷ The CD test is eminently suited to shedding some empirical light on this issue. We shall make use of per capita output data (Purchasing Power Parity adjusted Real GDP per capita in constant prices) from the latest Penn World Tables, Version 6.1.¹⁸ We computed CD statistics both for countries within a given region, as well as across countries from different regions, over the periods 1971-2000 and 1981-2000. For each country, i , we first computed OLS residuals from regressions of its log per capita output (y_{it}) on an intercept, a linear trend, $y_{i,t-1}$ and $y_{i,t-2}$. Using these residuals we then computed the various CD statistics summarized in Table A.

Table A
Cross Section Dependence of Output Innovations
Within Country Groupings

Country Groups	CD Statistics	
	1971-2000	1981-2000
Europe	19.81	13.80
North and Latin America	6.30	0.94
Asia and Australia	6.75	4.71
Middle East and North Africa (MENA)	1.55	-0.81
Rest of the World	5.05	6.05
The World	16.67	7.54

Notes The CD statistic is computed for various groups of countries based on the OLS residuals from the AR(2) regressions in y_{it} (log per capita output) with a linear trend. For the membership of different country groupings see Table 7.

The hypothesis that output innovations are cross sectionally independent is rejected for all regions except for the MENA (Middle East and North Africa) countries for both periods. The results also show a significant weakening of the degree of cross section dependence in the case of output innovations in North and Latin America over the more recent period of 1981-2000. The highest degree of cross section dependence is found to be across the countries in Europe.

¹⁷See, for example, Barro (1997) and Lee, Pesaran and Smith (1997). The latter recognizes the potential importance of cross section error dependence in panel output regressions but does not provide evidence of its statistical significance.

¹⁸The PWT code for the series is RGDPL, and is constructed in international dollars, with 1996 as the reference year. For further details see Heston, Summers and Aten (2002).

The CD test can also be applied to output innovations across regions. For example, correlations of the innovations for each country in Europe can be computed with the output innovations of the countries in other regions. Table B summarizes the results for a balanced panel of countries over the period 1971-2000. Perhaps not surprisingly, the within region dependence tends to be higher than cross region dependence, although there are important exceptions particularly the MENA region. There are nevertheless significant evidence of across as well as within region dependence.

Table B
Cross Dependence of Output Innovations Across Country Groups

Country Groups	Europe	America	Asia and Australia	MENA	Rest of World
Europe	19.81	6.68	2.12	1.72	6.05
America	6.68	6.30	2.57	3.51	5.78
Asia and Australia	2.12	2.57	6.75	1.09	-0.40
MENA	1.72	3.51	1.09	1.55	1.95
Rest of World	6.05	5.78	-0.40	1.95	5.05

The CD statistics are computed over the period 1971-2000. The diagonal elements refer to the within country CD statistics and are reproduced here from Table A for convenience. The off diagonal elements are the CD statistics computed based on pair-wise correlation of the innovations across regions.

Finally, we computed the CD statistic for testing the dependence of UK's output innovations on output innovations of countries in a number of different regions over the 1950-2000 period. All countries with at least 10 years of data over the 1950-2000 period were included in this exercise, thus providing an unbalanced panel with T_{ij} ranging from 10 to 40. Using (31), we obtained the following CD statistics.

Country/Region	Test statistic
Europe	6.51
USA and Canada	4.56
Latin America	-0.14
Asia and Australia	2.96
Rest of the World	1.92

These results show that UK output innovations are significantly related to output innovations of countries in Europe, USA and Canada, and Asia and Australia, in that order. The hypothesis that UK output innovations are uncorrelated with the countries in Latin America is not rejected, whilst it is rejected vis-a-vis the countries in the rest of the world at the 10% level, but not at the 5% level. The results seem plausible and reflects the closer links that UK economy has been having with the countries in Europe than with the USA and Canada.

12 Concluding Remarks

In this paper we have developed simple, yet general tests of cross section dependence of errors in linear panel data models which is applicable in a variety of contexts, including dynamic heterogeneous panels with (possibly) multiple breaks and unit roots, as well as for spatial panels. The tests are based on simple averages of pair-wise correlation coefficients of OLS residuals from individual regressions and is shown to be valid under fairly general conditions even when T is small and N large. The tests can be applied to balanced and unbalanced panels and is shown to have a standard normal distribution assuming that the errors are symmetrically distributed. The extensive Monte Carlo evidence reported in the paper shows that the proposed CD and its spatial version, $CD(p)$, tests have good small sample properties. The Monte Carlo evidence also documents the pitfalls of using the LM test for large N panels, unless T is much larger than N .

The tests could be particularly useful in the case of panels with small T and large N where hitherto there is little or no empirical evidence provided in support of the routinely made assumption that the errors are cross sectionally independent. As an illustration, using output series from the Penn World Table, we have investigated the extent to which output innovations are cross sectionally independent within and across various regions. The results clearly show significant evidence of cross section dependence in output innovations, that ought to be taken into account in cross country growth analysis. Appropriate estimation and inference techniques for the analysis of panel data models subject to cross section dependence is an important subject in its own right and has been discussed, for example, by Ahn, Lee and Schmidt, (2001) and Pesaran (2002). Panel unit root tests are also being developed that allow for cross section dependence.

The approach developed in this paper might also prove useful for the analysis of cross section dependence in non-linear panel data models such Probit and Logit specifications. But this is an area for further research and falls outside the scope of the present paper. It would also be of interest to compare the small sample properties of the spatial versions of the CD test proposed in this paper to the spatial correlation tests developed by Anselin (1988), and recently extended by Baltagi, Song and Koh (2003) to panels with random effects.

Table 1(a)
Size of LM and CD Tests
(Heterogenous AR(1) with Normal errors)

T/N	5	10	20	30	50	100
Size						
LM test						
5	0.094	0.289	0.831	1.000	1.000	1.000
10	0.065	0.151	0.371	0.666	0.982	1.000
20	0.043	0.079	0.136	0.217	0.481	0.966
30	0.053	0.065	0.108	0.152	0.255	0.667
50	0.043	0.054	0.063	0.087	0.124	0.285
100	0.056	0.052	0.055	0.083	0.089	0.142
CD test						
5	0.082	0.048	0.070	0.057	0.059	0.059
10	0.052	0.064	0.049	0.053	0.061	0.052
20	0.054	0.055	0.063	0.056	0.066	0.055
30	0.042	0.055	0.053	0.041	0.052	0.048
50	0.047	0.064	0.044	0.047	0.056	0.053
100	0.064	0.072	0.053	0.057	0.045	0.050

Notes: All experiments are based on 1,000 replications. The data generating process (DGP) is specified as $y_{it} = \mu_i(1 - \beta_i) + \beta_i y_{i,t-1} + u_{it}$, $u_{it} = \gamma_i f_t + \varepsilon_{it}$, where $\beta_i \sim iidU(0, 1)$, $\mu_i \sim \varepsilon_{i0} + \eta_i$, $\eta_i \sim iidN(1, 2)$, $\beta_i \sim iidU(0, 1)$, and $\varepsilon_{it} \sim iidN(0, 1)$. The cross dependence (CD) or correlation test statistic is defined as $CD = \sqrt{\frac{2T}{N(N-1)}} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \hat{\rho}_{ij}$, where $\hat{\rho}_{ij}$ is the pair-wise simple correlation coefficient between e_{it} and e_{jt} for all i and $i \neq j$. The Lagrange Multiplier (LM) test is defined as $CD_{LM} = T \sum_{i=1}^{N-1} \sum_{j=i+1}^N \hat{\rho}_{ij}^2$. Under the null hypothesis of zero cross dependence ($\gamma_i = 0$), the CD test is carried out at the 5% 2-sided nominal significant level; the null is rejected if $|CD| \geq 1.96$. For the LM test, the null is rejected when the test statistic is larger than the 1-sided 5% critical of the Chi-squared distribution with $N(N-1)/2$ degrees of freedom.

Table 1(b)
Power of LM and CD Tests
(Heterogenous AR(1) with Normal errors)

T/N	5	10	20	30	50	100
Power						
$\gamma_i \sim iidU(0.1, 0.3)$						
LM test						
5	0.096	0.287	0.861	1.000	1.000	1.000
10	0.075	0.153	0.411	0.720	0.987	1.000
20	0.064	0.108	0.179	0.386	0.709	0.989
30	0.065	0.109	0.171	0.348	0.598	0.952
50	0.065	0.113	0.199	0.457	0.698	0.974
100	0.089	0.237	0.372	0.775	0.937	1.000
CD test						
5	0.115	0.117	0.218	0.292	0.407	0.694
10	0.090	0.177	0.295	0.494	0.697	0.928
20	0.113	0.248	0.464	0.761	0.932	0.995
30	0.128	0.312	0.584	0.889	0.977	1.000
50	0.140	0.431	0.793	0.975	1.000	1.000
100	0.218	0.694	0.957	1.000	1.000	1.000
Power						
$\gamma_i \sim iidU(-0.2, 0.6)$						
LM test						
5	0.085	0.304	0.880	1.000	1.000	1.000
10	0.087	0.160	0.433	0.700	0.982	1.000
20	0.089	0.164	0.366	0.568	0.883	0.998
30	0.109	0.225	0.487	0.636	0.926	0.997
50	0.184	0.341	0.738	0.882	0.989	1.000
100	0.371	0.694	0.986	0.999	1.000	1.000
CD test						
5	0.094	0.138	0.152	0.251	0.502	0.556
10	0.098	0.177	0.214	0.409	0.769	0.813
20	0.091	0.253	0.321	0.617	0.957	0.969
30	0.115	0.321	0.392	0.802	0.995	0.998
50	0.143	0.468	0.513	0.941	1.000	1.000
100	0.228	0.693	0.837	0.998	1.000	1.000

Notes: The data generating process is the same as that specified at foot of Table 1a, except for γ_i which is specified in this Table.

Table 2
Size and Power of LM and CD Tests
(Heterogenous AR(1) with $\chi^2(1)$ errors)

T/N	5	10	20	30	50	100
Size						
LM test						
5	0.100	0.280	0.837	1.000	1.000	1.000
10	0.064	0.147	0.350	0.607	0.979	1.000
20	0.070	0.112	0.151	0.225	0.443	0.940
30	0.070	0.083	0.116	0.163	0.315	0.626
50	0.076	0.088	0.119	0.125	0.175	0.330
100	0.072	0.079	0.100	0.112	0.118	0.168
CD test						
5	0.078	0.072	0.084	0.072	0.067	0.080
10	0.059	0.063	0.047	0.038	0.037	0.051
20	0.048	0.051	0.047	0.053	0.041	0.046
30	0.042	0.049	0.055	0.041	0.059	0.041
50	0.047	0.060	0.053	0.050	0.043	0.060
100	0.069	0.056	0.053	0.059	0.047	0.053
Power						
LM test						
5	0.125	0.305	0.867	1.000	1.000	1.000
10	0.097	0.187	0.448	0.776	0.991	1.000
20	0.093	0.149	0.303	0.524	0.774	0.998
30	0.113	0.158	0.258	0.525	0.740	0.984
50	0.118	0.197	0.299	0.594	0.799	0.990
100	0.129	0.280	0.452	0.820	0.956	0.999
CD test						
5	0.162	0.208	0.463	0.536	0.710	0.888
10	0.123	0.309	0.483	0.714	0.891	0.983
20	0.131	0.330	0.625	0.880	0.978	0.999
30	0.162	0.405	0.717	0.951	0.999	1.000
50	0.193	0.520	0.823	0.987	1.000	1.000
100	0.222	0.718	0.972	0.999	1.000	1.000

Notes: The data generating process is the same as that described at the foot of Table 1a, except for $\varepsilon_{it} \sim iid\chi^2(1)$. Power is calculated under $\gamma_i \sim iidU(0.1, 0.3)$.

Table 3
Size and Power of LM and CD Tests
(DGP Subject to a Single Break)

T/N	5	10	20	30	50	100
Size						
LM test						
5	0.109	0.279	0.865	1.000	1.000	1.000
10	0.070	0.152	0.452	0.781	1.000	1.000
20	0.061	0.100	0.227	0.399	0.834	1.000
30	0.069	0.099	0.180	0.325	0.660	0.995
50	0.064	0.088	0.161	0.229	0.556	0.986
100	0.051	0.081	0.149	0.219	0.498	0.968
CD test						
5	0.068	0.065	0.056	0.068	0.067	0.061
10	0.058	0.069	0.066	0.062	0.057	0.046
20	0.048	0.049	0.048	0.058	0.062	0.047
30	0.059	0.064	0.046	0.059	0.040	0.048
50	0.066	0.057	0.061	0.053	0.052	0.043
100	0.054	0.056	0.050	0.047	0.051	0.059
Power						
LM test						
5	0.119	0.285	0.889	1.000	1.000	1.000
10	0.080	0.150	0.475	0.801	0.999	1.000
20	0.066	0.101	0.250	0.500	0.878	1.000
30	0.078	0.109	0.253	0.430	0.804	0.999
50	0.078	0.151	0.264	0.482	0.817	0.999
100	0.098	0.190	0.434	0.656	0.959	1.000
CD test						
5	0.101	0.118	0.198	0.242	0.342	0.600
10	0.106	0.132	0.275	0.399	0.592	0.881
20	0.088	0.191	0.408	0.615	0.841	0.990
30	0.101	0.231	0.515	0.749	0.956	1.000
50	0.137	0.317	0.707	0.899	0.997	1.000
100	0.198	0.472	0.918	0.995	1.000	1.000

Notes: The data generating process used is $y_{it} - \mu_i = \beta_{it}(y_{it-1} - \mu_i) + u_{it}$, $u_{it} = \gamma_i f_t + \sigma_{it} \varepsilon_{it}$, where $\mu_i \sim iidN(1, 1)$, $\beta_{it} = \beta_t = 0.6$ for $t = -50, \dots, T/2$, $\beta_t = 0.8$ for $t = T/2, \dots, T$; $\sigma_{it} = \sigma_t = \sqrt{1.5}$ for $t = -50, \dots, T/2$, $\sigma_t = 1$ for $t = T/2, \dots, T$, and $\varepsilon_{it} \sim iidN(0, 1)$. First 49 observations are discarded. Power is calculated under $\gamma_i \sim iidU(0.1, 0.3)$. See also the notes to Table 1a.

Table 4
Size and Power of LM and CD Tests
(Multiple Structural Break: $\beta_{it} \sim U(0, 1), \sigma_{it} \sim \sqrt{\chi^2(2)/2}$)

T/N	5	10	20	30	50	100
Size						
LM test						
5	0.083	0.231	0.855	1.000	1.000	1.000
10	0.071	0.078	0.270	0.630	0.968	1.000
20	0.060	0.071	0.104	0.124	0.395	0.921
30	0.053	0.052	0.100	0.083	0.149	0.683
50	0.054	0.047	0.064	0.066	0.083	0.252
100	0.044	0.052	0.060	0.057	0.067	0.070
CD test						
5	0.078	0.057	0.064	0.062	0.059	0.066
10	0.067	0.044	0.044	0.064	0.060	0.060
20	0.049	0.063	0.057	0.050	0.062	0.062
30	0.056	0.052	0.056	0.055	0.055	0.049
50	0.047	0.052	0.042	0.033	0.056	0.064
100	0.046	0.043	0.049	0.039	0.052	0.047
Power						
LM test						
5	0.117	0.229	0.855	1.000	1.000	1.000
10	0.063	0.095	0.320	0.677	0.981	1.000
20	0.054	0.098	0.154	0.257	0.616	0.985
30	0.070	0.090	0.171	0.197	0.454	0.955
50	0.072	0.111	0.148	0.250	0.528	0.928
100	0.087	0.146	0.341	0.514	0.820	0.997
CD test						
5	0.121	0.105	0.235	0.337	0.498	0.750
10	0.112	0.169	0.356	0.489	0.692	0.925
20	0.079	0.231	0.445	0.685	0.917	0.991
30	0.122	0.281	0.634	0.812	0.984	1.000
50	0.156	0.421	0.795	0.943	0.999	1.000
100	0.198	0.584	0.934	0.999	1.000	1.000

Notes: The data generating process is the same as that specified at the foot of Table 1a except $\beta_{it} = 0.5$ for $t = -50, \dots, 0$ and all i , $\beta_{it} \sim iidU(0, 1)$ for $t = 1, \dots, T$, $i = 1, \dots, N$; $\sigma_{it}^2 \sim iid\chi^2(2)/2$ for $t = -50, \dots, T$, $i = 1, \dots, N$. Power is calculated under $\gamma_i \sim iidU(0.1, 0.3)$.

Table 5
Size and Power of LM and CD Tests
(The Unit Root Case: $\beta_{it} = 1$, and $\sigma_{it} \sim \sqrt{\chi^2(2)/2}$)

T/N	5	10	20	30	50	100
Size						
LM test						
5	0.064	0.197	0.766	1.000	1.000	1.000
10	0.045	0.091	0.256	0.560	0.943	1.000
20	0.043	0.080	0.118	0.163	0.466	0.952
30	0.071	0.073	0.071	0.118	0.224	0.768
50	0.062	0.056	0.066	0.081	0.149	0.374
100	0.042	0.050	0.055	0.059	0.076	0.109
CD test						
5	0.060	0.063	0.069	0.074	0.069	0.052
10	0.056	0.051	0.050	0.056	0.069	0.048
20	0.059	0.054	0.048	0.067	0.049	0.049
30	0.070	0.057	0.048	0.041	0.043	0.041
50	0.073	0.049	0.049	0.056	0.058	0.059
100	0.057	0.043	0.037	0.053	0.042	0.046
Power						
LM test						
5	0.059	0.200	0.787	1.000	1.000	1.000
10	0.057	0.087	0.272	0.577	0.965	1.000
20	0.052	0.104	0.162	0.309	0.712	0.989
30	0.082	0.110	0.175	0.254	0.580	0.982
50	0.095	0.108	0.196	0.344	0.694	0.975
100	0.086	0.171	0.385	0.619	0.883	0.998
CD test						
5	0.080	0.126	0.268	0.353	0.503	0.797
10	0.098	0.191	0.344	0.512	0.743	0.943
20	0.117	0.250	0.506	0.739	0.925	0.999
30	0.152	0.332	0.675	0.819	0.978	0.999
50	0.159	0.441	0.807	0.959	0.999	1.000
100	0.238	0.653	0.973	0.997	1.000	1.000

Notes: The data generating process is the same as that specified at the foot of Table 4, except that $\beta_{it} = \beta = 1$ for all i and t . Power is calculated under $\gamma_i \sim iidU(0.1, 0.3)$.

Table 6a
Size of LM and the Spatial First-Order, CD(1), Tests
(Heterogenous AR(1) with spatially correlated errors)

T/N	5	10	20	30	50	100
SIZE						
LM test						
5	0.094	0.310	0.846	1.000	1.000	1.000
10	0.068	0.105	0.256	0.547	0.964	1.000
20	0.062	0.072	0.095	0.167	0.366	0.915
30	0.044	0.059	0.058	0.099	0.176	0.527
50	0.048	0.058	0.063	0.065	0.071	0.211
100	0.058	0.050	0.066	0.058	0.058	0.079
CD(1) test						
5	0.073	0.095	0.077	0.071	0.075	0.079
10	0.075	0.069	0.071	0.055	0.080	0.063
20	0.057	0.064	0.055	0.051	0.062	0.072
30	0.048	0.062	0.060	0.050	0.052	0.062
50	0.045	0.040	0.072	0.053	0.059	0.049
100	0.037	0.046	0.043	0.052	0.050	0.052

Notes: The $CD(1)$ test is based on $CD(1) = \sqrt{\frac{T}{N-1}} \sum_{i=2}^N \hat{\rho}_{i,i-1}$, which is distributed as $N(0, 1)$. The data generating process is the same as the one used under the null hypothesis of cross section independence described at the foot of Table 1a.

Table 6b
Power of LM and the Spatial First-Order, CD(1), Tests
(Heterogenous AR(1) with spatially correlated errors)

T/N	5	10	20	30	50	100
Power: $\lambda = 0.1$						
LM test						
5	0.096	0.312	0.850	1.000	1.000	1.000
10	0.084	0.127	0.281	0.570	0.964	1.000
20	0.095	0.101	0.138	0.201	0.419	0.938
30	0.093	0.094	0.101	0.152	0.258	0.614
50	0.157	0.135	0.137	0.125	0.171	0.374
100	0.326	0.239	0.256	0.218	0.246	0.332
CD(1) test						
5	0.108	0.132	0.140	0.160	0.221	0.362
10	0.132	0.164	0.244	0.323	0.478	0.757
20	0.171	0.294	0.475	0.603	0.838	0.984
30	0.233	0.381	0.653	0.817	0.966	0.998
50	0.386	0.621	0.876	0.972	1.000	1.000
100	0.675	0.908	0.996	1.000	1.000	1.000
Power: $\lambda = -0.1$						
LM test						
5	0.099	0.318	0.846	1.000	1.000	1.000
10	0.078	0.120	0.279	0.582	0.970	1.000
20	0.107	0.096	0.131	0.208	0.413	0.936
30	0.102	0.093	0.122	0.148	0.255	0.620
50	0.159	0.123	0.139	0.135	0.168	0.381
100	0.340	0.248	0.249	0.245	0.249	0.342
CD(1) test						
5	0.091	0.133	0.146	0.178	0.260	0.384
10	0.116	0.178	0.244	0.335	0.487	0.742
20	0.174	0.276	0.475	0.634	0.825	0.977
30	0.279	0.438	0.666	0.818	0.957	0.999
50	0.413	0.630	0.876	0.967	0.998	1.000
100	0.695	0.911	0.999	1.000	1.000	1.000

Notes: Power is calculated under $u_{it} = \lambda(0.5u_{i-1,t} + 0.5u_{i+1,t}) + \sigma_i \varepsilon_{it}$ for $i = 2, \dots, N - 1$, with the end points $u_{1t} = u_{2t} + \varepsilon_{1t}$ and $u_{Nt} = u_{N-1,t} + \varepsilon_{Nt}$. See also the notes to Tables 1a and 6a.

Table 7
Details of Country Groups

Country Groups	Member Countries
Sample period: 1971-2000	
Europe	Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and United Kingdom
North America	United States and Canada
Latin America	Argentina, Bolivia, Brazil, Barbados, Chile, Colombia, Costa Rica, Dominican Republic, Ecuador, Guatemala, Honduras, Jamaica, Mexico, Nicaragua, Panama, Peru, Paraguay, El Salvador, Trinidad & Tobago, Uruguay and Venezuela.
Asia and Australia	Australia, Bangladesh, China, Hong Kong, Indonesia, India, Japan, Korea, Sri Lanka, Malaysia, Nepal, New Zealand, Pakistan, Philippines and Thailand.
Middle East and North Africa (MENA)	Algeria, Egypt, Iran, Israel, Jordan, Morocco, Syria, Tunisia and Turkey
Rest of the World	Burundi, Benin, Burkina Faso, Chad, Cote d'Ivoire, Cameroon, Congo, Comoros, Cape Verde, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Equatorial Guinea, Ethiopia, Hungary, Iceland, Kenya, Lesotho, Madagascar, Mali, Mauritius, Mozambique, Malawi, Niger, Nigeria, Romania, Rwanda, Senegal, Seychelles, South Africa, Tanzania, Togo, Uganda, Zambia and Zimbabwe.
Sample period: 1981-2000	
Countries to be added to the lists given above:	
Latin America	Antigua, Belize, Grenada, St.Kitts&Nevis, St.Lucia and St.Vincent & Gredine.
Rest of the World	Poland.

A Properties of Residuals in Regression Models Subject to Structural Breaks

Abstracting from the cross section index i , the regression model (11) with a single break can be written as

$$y_t = \begin{cases} \mu_y + \beta'_1(\mathbf{x}_t - \boldsymbol{\mu}_x) + \sigma_1 \varepsilon_t, & t = 1, 2, \dots, T_1, \\ \mu_y + \beta'_2(\mathbf{x}_t - \boldsymbol{\mu}_x) + \sigma_2 \varepsilon_t, & t = T_1 + 1, \dots, T \end{cases}$$

where the $k \times 1$ slope coefficients, β_j , and the error variances, σ_j^2 , for $j = 1, 2$, are subject to a single break at time $t = T_1$, and $\varepsilon_t \sim iid(0, 1)$. The implied intercepts are given by

$$\alpha_j = \mu_y - \beta'_j \boldsymbol{\mu}_x, \quad j = 1, 2,$$

The unconditional means of y_t and \mathbf{x}_t , namely $\boldsymbol{\mu}_y$ and $\boldsymbol{\mu}_x$, are not subject to change. We also assume that \mathbf{x}_t follows the covariance stationary process:

$$\mathbf{x}_t = \boldsymbol{\mu}_x + \sum_{s=0}^{\infty} \boldsymbol{\Psi}_s \boldsymbol{\nu}_{t-s}, \quad \boldsymbol{\nu}_t \sim iid(0, \mathbf{I}_k), \quad \sum_{s=0}^{\infty} \|\boldsymbol{\Psi}_s\| < \infty. \quad (33)$$

where $\boldsymbol{\nu}_t$ and $\varepsilon_{t'}$, are independently distributed for all t and t' . We shall also assume that the innovations in the \mathbf{x}_t process, $\boldsymbol{\nu}_t$, are symmetrically distributed around zero.

Suppose now that the breaks are ignored and the residuals, e_t , are computed by running the ordinary least squares regression of y_t on \mathbf{x}_t over the whole sample, $t = 1, 2, \dots, T$. We have

$$e_t = y_t - \hat{\alpha} - \hat{\boldsymbol{\beta}}' \mathbf{x}_t, \quad (34)$$

where

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \left[\sum_{t=1}^T (\mathbf{x}_t - \bar{\mathbf{x}})(\mathbf{x}_t - \bar{\mathbf{x}})' \right]^{-1} \left[\sum_{t=1}^T (\mathbf{x}_t - \bar{\mathbf{x}})(y_t - \bar{y}) \right], \\ \hat{\alpha} &= \bar{y} - \hat{\boldsymbol{\beta}}' \bar{\mathbf{x}}, \quad \bar{\mathbf{x}} = \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t, \quad \text{and} \quad \bar{y} = \frac{1}{T} \sum_{t=1}^T y_t. \end{aligned}$$

In what follows we establish that for all $t = 1, 2, \dots, T$, the OLS residuals e_t are odd functions of the disturbances, ε_t and $E(e_t) = 0$.

We first note that

$$\begin{aligned} \bar{y} &= \mu_y + \lambda_1 \beta'_1 (\bar{\mathbf{x}}_1 - \boldsymbol{\mu}_x) + (1 - \lambda_1) \beta'_2 (\bar{\mathbf{x}}_2 - \boldsymbol{\mu}_x) \\ &\quad + \sigma_1 \lambda_1 \bar{\varepsilon}_1 + \sigma_2 (1 - \lambda_1) \bar{\varepsilon}_2, \end{aligned}$$

where $\lambda_1 = T_1/T$, $\bar{\mathbf{x}}_1 = \frac{1}{T_1} \sum_{t=1}^{T_1} \mathbf{x}_t$, $\bar{\mathbf{x}}_2 = \frac{1}{T-T_1} \sum_{t=T_1+1}^T \mathbf{x}_t$, etc. Hence, for $t \leq T_1$

$$y_t - \bar{y} = \mu_y - \bar{y} + \beta'_1 (\mathbf{x}_t - \boldsymbol{\mu}_x) + \sigma_1 \varepsilon_t,$$

and for $t > T_1$

$$y_t - \bar{y} = \mu_y - \bar{y} + \beta_2'(\mathbf{x}_t - \boldsymbol{\mu}_x) + \sigma_2 \varepsilon_t.$$

Also since

$$E(\mathbf{x}_t) = \boldsymbol{\mu}_x,$$

then for all t ,

$$E(y_t - \bar{y}) = 0, \text{ for } t = 1, 2, \dots, T.$$

Consider now the residuals defined by (34) and note that

$$e_t = (y_t - \bar{y}) - \widehat{\boldsymbol{\beta}}'(\mathbf{x}_t - \bar{\mathbf{x}}). \quad (35)$$

Hence it is sufficient to show that the second term has zero expectations for all t . Under the data generating mechanism

$$\begin{aligned} \mathbf{Q} \widehat{\boldsymbol{\beta}} &= \mathbf{Q}_1 \boldsymbol{\beta}_1 + \mathbf{Q}_2 \boldsymbol{\beta}_2 + \sigma_1 \sum_{t=1}^{T_1} (\mathbf{x}_t - \bar{\mathbf{x}}) \varepsilon_t \\ &\quad + \sigma_2 \sum_{t=T_1+1}^T (\mathbf{x}_t - \bar{\mathbf{x}}) \varepsilon_t, \end{aligned}$$

where

$$\begin{aligned} \mathbf{Q} &= \sum_{t=1}^T (\mathbf{x}_t - \bar{\mathbf{x}})(\mathbf{x}_t - \bar{\mathbf{x}})' \\ \mathbf{Q}_1 &= \sum_{t=1}^{T_1} (\mathbf{x}_t - \bar{\mathbf{x}})(\mathbf{x}_t - \bar{\mathbf{x}})', \text{ and } \mathbf{Q}_2 = \mathbf{Q} - \mathbf{Q}_1. \end{aligned}$$

Therefore

$$\begin{aligned} \widehat{\boldsymbol{\beta}}'(\mathbf{x}_t - \bar{\mathbf{x}}) &= \boldsymbol{\beta}_1' \mathbf{Q}_1 \mathbf{Q}^{-1} (\mathbf{x}_t - \bar{\mathbf{x}}) + \boldsymbol{\beta}_2' \mathbf{Q}_2 \mathbf{Q}^{-1} (\mathbf{x}_t - \bar{\mathbf{x}}) \\ &\quad + \sigma_1 \sum_{t=1}^{T_1} \varepsilon_t (\mathbf{x}_t - \bar{\mathbf{x}})' \mathbf{Q}^{-1} (\mathbf{x}_t - \bar{\mathbf{x}}) \\ &\quad + \sigma_2 \sum_{t=T_1+1}^T \varepsilon_t (\mathbf{x}_t - \bar{\mathbf{x}})' \mathbf{Q}^{-1} (\mathbf{x}_t - \bar{\mathbf{x}}). \end{aligned}$$

Since \mathbf{x}_t and $\varepsilon_{t'}$ are independently distributed for all t and t' , the last two terms have zero unconditional expectations. Also using (33) it is easily seen that

$$\mathbf{x}_t - \bar{\mathbf{x}} = \sum_{s=0}^{\infty} \boldsymbol{\Psi}_s (\boldsymbol{\nu}_{t-s} - \bar{\boldsymbol{\nu}}_{-s}), \quad \bar{\boldsymbol{\nu}}_{-s} = \frac{1}{T} \sum_{t=1}^T \boldsymbol{\nu}_{t-s},$$

which establishes that $\mathbf{x}_t - \bar{\mathbf{x}}$ is an odd function of $\{\boldsymbol{\nu}_t\}$, the innovations in \mathbf{x}_t . Since \mathbf{Q} and \mathbf{Q}_1 are even functions of $\{\boldsymbol{\nu}_t\}$, it follows that $\mathbf{Q}_j \mathbf{Q}^{-1}(\mathbf{x}_t - \bar{\mathbf{x}})$,

for $j = 1, 2$ are also odd functions of $\{\nu_t\}$ and in view of the symmetry of $\{\nu_t\}$ will have zero mean unconditionally. Thus, under our assumptions, e_t and $\xi_t = (\sum e_t^2)^{-1/2} e_t$ are odd functions of $\{\varepsilon_t, \nu_t\}$, and therefore have zero expectations for all t , despite the breaks in the slopes and the error variances. This result continues under multiple breaks and/or even if Ψ_i are subject to one or more breaks. The key assumptions are symmetry of the innovations, ε_t and ν_t , and the time-invariance of the unconditional means of y_t and \mathbf{x}_t .

B Residuals from AR(p) Models Subject to Breaks

Consider the AR(p) model defined over the period $t = 1, 2, \dots, T$; and assumed to have been subject to a single structural break at time T_1 :

$$y_t = \begin{cases} \alpha_1 + \beta_{11}y_{t-1} + \beta_{12}y_{t-2} + \dots + \beta_{1p}y_{t-p} + \sigma_1\varepsilon_t, & \text{for } t \leq T_1, \\ \alpha_2 + \beta_{21}y_{t-1} + \beta_{22}y_{t-2} + \dots + \beta_{2p}y_{t-p} + \sigma_2\varepsilon_t, & \text{for } t > T_1, \end{cases} \quad (36)$$

where $\varepsilon_t \sim iid(0, 1)$ for all t ,

$$\alpha_j = \mu_j(1 - \tau_p' \beta_j), \quad j = 1, 2, \quad (37)$$

$\beta_j = (\beta_{j1}, \beta_{j2}, \dots, \beta_{jp})'$ and τ_p is a $p \times 1$ unit vector.

Suppose that the structural break is ignored and residuals are computed by estimating the AR(p) model in y_t using the OLS regression y_t on an intercept and $\mathbf{x}_t = (y_{t-1}, y_{t-2}, \dots, y_{t-p})'$ making use of the available observations $F_T = (y_{1-p}, y_{2-p}, \dots, y_0, y_1, \dots, y_T)$. In this case the fitted residuals are given by

$$e_t = y_t - \hat{\alpha} - \hat{\beta}' \mathbf{x}_t, \quad t = 1, 2, \dots, T \quad (38)$$

where $\mathbf{x}_t = (y_{t-1}, y_{t-2}, \dots, y_{t-p})'$, $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p)'$

$$\hat{\beta} = (\mathbf{X}'\mathbf{M}\mathbf{X})^{-1} \mathbf{X}'\mathbf{M}\mathbf{y}, \quad (39)$$

$$\hat{\alpha} = \frac{\tau_T' \mathbf{y} - \tau_T' \mathbf{X} \hat{\beta}}{T}, \quad (40)$$

$\mathbf{y} = (y_1, y_2, \dots, y_T)$, $\mathbf{X} = (\mathbf{y}_0, \mathbf{y}_{-1}, \dots, \mathbf{y}_{-p+1})$, $\mathbf{y}_{-j+1} = (y_{-j+1}, y_{-j+2}, \dots, y_{T-j})'$, τ_T is a $T \times 1$ vector of ones, and $\mathbf{M} = \mathbf{I}_T - \tau_T(\tau_T' \tau_T)^{-1} \tau_T'$. In what follows we shall establish that $E(e_t) = 0$ for $t = 1, 2, \dots, T$ so long as $\mu_1 = \mu_2$, ε_t is symmetrically distributed, and $E(e_t)$ exists. We shall provide a proof for the stationary case with a single break, although the result holds much more generally both in the presence of multiple breaks and if there are unit roots in the pre- and/or post-break processes.

In the case where the pre-break regime is stationary, the distribution of the initial values, $\mathbf{x}_p = (y_{1-p}, y_{2-p}, \dots, y_0)'$, can be written as

$$\mathbf{x}_p - \mu_1 \tau_p \sim (\mathbf{0}, \sigma_1^2 \mathbf{V}_p), \quad (41)$$

where \mathbf{V}_p is a positive definite matrix.

Using (41) and (36) for $t = 1, 2, \dots, T$, in matrix notations we have

$$\mathbf{B} \mathbf{y}^* = \mathbf{d} + \mathbf{D} \boldsymbol{\varepsilon}^*, \quad (42)$$

where $\mathbf{y}^* = (\mathbf{x}'_p, \mathbf{y}')'$, $\boldsymbol{\varepsilon}^* = (\varepsilon_{1-p}, \varepsilon_{2-p}, \dots, \varepsilon_0, \varepsilon_1, \varepsilon_2, \dots, \varepsilon_T)'$

$$\mathbf{D} = \sigma_1 \begin{pmatrix} \boldsymbol{\psi}_p & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{T_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & (\sigma_2/\sigma_1) \mathbf{I}_{T_2} \end{pmatrix}, \quad \mathbf{d} = \begin{pmatrix} \mu_1 \boldsymbol{\tau}_p \\ \mu_1 (1 - \beta_1^*) \boldsymbol{\tau}_{T_1} \\ \mu_2 (1 - \beta_2^*) \boldsymbol{\tau}_{T_2} \end{pmatrix}, \quad (43)$$

$$\mathbf{B} = \begin{pmatrix} \mathbf{I}_p & \mathbf{0} & \mathbf{0} \\ \mathbf{B}_{21} & \mathbf{B}_{22} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_{32} & \mathbf{B}_{33} \end{pmatrix}. \quad (44)$$

The sub-matrices, \mathbf{B}_{ij} , depend only on the slope coefficients, β_1 and β_2 and are as defined in Appendix B of Pesaran and Timmermann (2004). \mathbf{I}_{T_1} and \mathbf{I}_{T_2} are identity matrices of order T_1 and T_2 , respectively, $T_2 = T - T_1$, $\boldsymbol{\varepsilon}^* \sim (\mathbf{0}, \mathbf{I}_{T+p})$, and $\boldsymbol{\psi}_p$ is a lower triangular Cholesky factor of \mathbf{V}_p , namely $\mathbf{V}_p = \boldsymbol{\psi}_p \boldsymbol{\psi}'_p$.

Using (42) it is easily seen that

$$\mathbf{y}_{-j+1} = \mathbf{G}_j(\mathbf{c} + \mathbf{H}\boldsymbol{\varepsilon}^*), \quad \text{for } j = 0, 1, \dots, p, \quad (45)$$

where \mathbf{G}_j are $T \times (T+p)$ selection matrices defined by $\mathbf{G}_j = (\mathbf{0}_{T \times p-j}, \mathbf{I}_T, \mathbf{0}_{T \times j})$, $\mathbf{H} = \mathbf{B}^{-1} \mathbf{D}$, and $\mathbf{c} = \mathbf{B}^{-1} \mathbf{d}$. In particular,

$$\mathbf{y} = \mathbf{G}_0(\mathbf{c} + \mathbf{H}\boldsymbol{\varepsilon}^*),$$

and

$$\mathbf{X} = \left[\mathbf{G}_1(\mathbf{c} + \mathbf{H}\boldsymbol{\varepsilon}^*), \mathbf{G}_2(\mathbf{c} + \mathbf{H}\boldsymbol{\varepsilon}^*), \dots, \mathbf{G}_p(\mathbf{c} + \mathbf{H}\boldsymbol{\varepsilon}^*) \right].$$

However, as shown in Pesaran and Timmermann (2004, Appendix B), under $\mu_1 = \mu_2 = \mu$, $\mathbf{G}_j \mathbf{c} = \boldsymbol{\mu} \boldsymbol{\tau}_T$, and the (i, j) element of $\mathbf{X}' \mathbf{M} \mathbf{X}$ will be given by $\boldsymbol{\varepsilon}^{*'} \mathbf{H}' \mathbf{G}'_i \mathbf{M}_\tau \mathbf{G}_j \mathbf{H} \boldsymbol{\varepsilon}^*$, for $i, j = 1, 2, \dots, p$, and the j^{th} element of $\mathbf{X}' \mathbf{M} \mathbf{y}$ by $\boldsymbol{\varepsilon}^{*'} \mathbf{H}' \mathbf{G}'_j \mathbf{M}_\tau \mathbf{G}_0 \mathbf{H} \boldsymbol{\varepsilon}^*$, for $j = 1, 2, \dots, p$. Hence, under $\mu_1 = \mu_2$, $\hat{\beta}$ will be an even function of $\boldsymbol{\varepsilon}$. Similarly, using (40) and recalling that $\mathbf{G}_j \mathbf{c} = \boldsymbol{\mu} \boldsymbol{\tau}_T$, we have

$$\hat{\alpha} = \mu \left(1 - \sum_{j=1}^p \hat{\beta}_j \right) + \left(\frac{\boldsymbol{\tau}'_T \mathbf{G}_0 \mathbf{H} \boldsymbol{\varepsilon}^*}{T} \right) - \sum_{j=1}^p \left(\frac{\boldsymbol{\tau}'_T \mathbf{G}_j \mathbf{H} \boldsymbol{\varepsilon}^*}{T} \right) \hat{\beta}_j. \quad (46)$$

Using this result in (38) and noting that for $t \leq T_1$

$$y_t - \mu = \sum_{j=1}^p \beta_{1j} (y_{t-j} - \mu) + \sigma_1 \varepsilon_t,$$

we have

$$e_t = - \sum_{j=1}^p (\hat{\beta}_j - \beta_{1j}) (y_{t-j} - \mu) + \sigma_1 \varepsilon_t - \left(\frac{\boldsymbol{\tau}'_T \mathbf{G}_0 \mathbf{H} \boldsymbol{\varepsilon}^*}{T} \right) + \sum_{j=1}^p \left(\frac{\boldsymbol{\tau}'_T \mathbf{G}_j \mathbf{H} \boldsymbol{\varepsilon}^*}{T} \right) \hat{\beta}_j.$$

Similarly, for $t > T_1$:

$$e_t = - \sum_{j=1}^p (\hat{\beta}_j - \beta_{2j}) (y_{t-j} - \mu) + \sigma_2 \varepsilon_t - \left(\frac{\boldsymbol{\tau}'_T \mathbf{G}_0 \mathbf{H} \boldsymbol{\varepsilon}^*}{T} \right) + \sum_{j=1}^p \left(\frac{\boldsymbol{\tau}'_T \mathbf{G}_j \mathbf{H} \boldsymbol{\varepsilon}^*}{T} \right) \hat{\beta}_j.$$

It is now easily seen that in both regimes e_t is an odd function of the standardized errors, ε_t , $t = -p + 1, -p + 2, \dots, T$, and under the distributional symmetry of the errors, we have

$$E(\xi_t) = E \left[\left(\sum_{t=1}^T e_t^2 \right)^{-1/2} e_t \right] = 0, \text{ for } t = 1, 2, \dots, T.$$

Note that for ξ_t to be well defined we need $T > p + 1$, and $E(\xi_t)$ exists for all $T > p + 1$. In contrast, the condition for the existence of the moments of e_t is much more complicated and demanding. For example, $E(e_t)$ exists if $E(\hat{\beta}_i)$ exists. A sufficient condition for the latter is known in the literature only for the simple case of $p = 1$. In this case

$$\hat{\beta}_1 = \frac{\boldsymbol{\varepsilon}^{*'} \mathbf{H}' \mathbf{G}'_1 \mathbf{M}_\tau \mathbf{G}_0 \mathbf{H} \boldsymbol{\varepsilon}^*}{\boldsymbol{\varepsilon}^{*'} \mathbf{H}' \mathbf{G}'_1 \mathbf{M}_\tau \mathbf{G}_1 \mathbf{H} \boldsymbol{\varepsilon}^*},$$

and $E(\hat{\beta}_1)$ exists if $T > 3$. (See Pesaran and Timmermann (2004)).

References

- [1] Ahn, S.G., Lee, Y.H., and Schmidt, P. (2001), "GMM Estimation of Linear Panel Data Models with Time-varying Individual Effects", *Journal of Econometrics*, 102, 219-255.
- [2] Anselin, L. (1988), *Spatial Econometrics: Methods and Models*, Dordrecht: Kluwer Academic Publishers.
- [3] Anselin, L. (2001), "Spatial Econometrics", in B. Baltagi (ed.), *A Companion to Theoretical Econometrics*, Blackwell, Oxford.
- [4] Anselin, L. and Bera, A.K. (1998), "Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics", in Ullah, A, and Giles, D.E.A., (Eds.), *Handbook of Applied Economic Statistics*, Marcel Dekker, New York.
- [5] Baltagi, B.H., S.H. Song and Koh, W. (2003), "Testing Panel Data Regression Models with Spatial Error Correlation," , *Journal of Econometrics*, 117, 123-150.
- [6] Barro, R.J. (1997), *Determinants of Economic Growth: A Cross-Country Empirical Study*, Cambridge, MA: MIT Press.
- [7] Breusch, T.S., and Pagan, A.R. (1980), "The Lagrange Multiplier Test and its Application to Model Specifications in Econometrics", *Review of Economic Studies*, 47, 239-53.
- [8] Cliff, A. and Ord, J.K. (1973), *Spatial Autocorrection*, London: Pion.
- [9] Cliff, A and Ord, J.K. (1981), *Spatial Processes: Models and Applications*, London: Pion.
- [10] Conley, T.G. and Topa, G. (2002), "Socio-economic Distance and Spatial Patterns in Unemployment", *Journal of Applied Econometrics*, 17, 303 - 327.
- [11] Haining, R.P. (2003), *Spatial data Analysis: Theory and Practice*, Cambridge University Press, Cambridge.
- [12] Heston, A., Summers, R. and Aten, B. (2002): Penn World Table Version 6.1, Center for International Comparisons at the University of Pennsylvania (CICUP), October 2002.
- [13] Lee, K., Pesaran, M. H., and Smith, R. (1997), "Growth and Convergence in Multi-country Empirical Stochastic Solow Model", *Journal of Applied Econometrics*, 12, 357-92.
- [14] Moran, P.A.P. (1948), "The Interpretation of Statistical Maps", *Biometrika*, 35, 255-60.
- [15] Nickell, S. (1981), "Biases in Dynamic Models with Fixed Effects", *Econometrica*, 49, 1399-1416.

- [16] Pesaran, M.H. (2002), "Estimation and Inference in Large Heterogeneous Panels with Cross Section Dependence", University of Cambridge DAE Working Paper No.0305 and CESifo Working Paper Series No 869, October 2002.
- [17] Pesaran, M.H., Schuermann, T., and Weiner, S.M. (2004), "Modeling Regional Interdependencies using a Global Error-Correcting Macroeconomic Model", *Journal of Business Economics and Statistics* (with Discussions and a Rejoinder), 22, 129-181.
- [18] Pesaran, M.H. and Timmermann, A. (2004), "Small Sample Properties of Forecasts from Autoregressive Models under Structural Breaks", Forthcoming in *Journal of Econometrics*.
- [19] Swamy, P.A.V.B. (1970), "Efficient Inference in Random Coefficient Regression Model", *Econometrica*, 38, 311-23.
- [20] Zellner, A. (1962), "An Efficient Method for Estimating Seemingly Unrelated Regressions and Tests of Aggregation Bias", *Journal of American Statistical Association*, 58, 977-992.