Testing Weak Cross-Sectional Dependence in Large Panels

M. Hashem Pesaran*

University of Southern California, and Trinity College, Cambridge

January 11, 2013

Abstract

This paper considers testing the hypothesis that errors in a panel data model are weakly cross sectionally dependent, using the exponent of cross-sectional dependence α, introduced recently in Bailey, Kapetanios and Pesaran (2012). It is shown that the implicit null of the CD test depends on the relative expansion rates of N and T. When $T = O(N^\epsilon)$, for some $0 < \epsilon \leq 1$, then the implicit null of the CD test is given by $0 \leq \alpha < (2 - \epsilon)/4$, which gives $0 \leq \alpha < 1/4$, when N and T tend to infinity at the same rate such that $T/N \rightarrow \kappa$, with κ being a finite positive constant. It is argued that in the case of large N panels, the null of weak dependence is more appropriate than the null of independence which could be quite restrictive for large panels. Using Monte Carlo experiments, it is shown that the CD test has the correct size for values of α in the range $[0, 1/4]$, for all combinations of N and T, and irrespective of whether the panel contains lagged values of the dependent variables, so long as there are no major asymmetries in the error distribution.

Keywords: Exponent of cross-sectional dependence, Diagnostic tests, Panel data models, Dynamic heterogenous panels.

JEL-Classification: C12, C13, C33.

1 Introduction

This paper is concerned with tests of error dependence in the case of large linear regression panels where N (the cross section dimension) is large. 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). In such panels where 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 coefficients 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-sectional dependence is measured with respect to a given "connection or spatial matrix" that characterizes the pattern of spatial dependence according 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 ith 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 the absence of ordering, tests of cross-sectional independence in the case of large N panels have been considered in Frees (1995), Pesaran (2004), Pesaran, Ullah and Yamagata (2008), Sarafidis, Yamagata, Robertson (2009), and Baltagi, Feng and Kao (2011). Recent surveys are provided by Moscone and Tosetti (2009), and Sarafidis and Wansbeek (2012). The null hypothesis of these tests is the cross-sectional independence of the errors in the panel regressions, and the tests are based on pair-wise correlation coefficients of the residuals, $\hat{\rho}_{ij}$, for the (i,j) units, computed assuming homogeneous or heterogeneous slopes.

The original LM test of Breusch and Pagan (1980), and its modified version for large N panels by Pesaran, Ullah and Yamagata (2008), are based on $\hat{\rho}_{ij}^2$, and test the hypothesis that *all* pair-wise error covariances, $Cov(u_{it}, u_{jt})$, are equal to zero for $i \neq j$. In contrast, we show that the implicit null of the CD test, proposed in Pesaran (2004), which is based on $\hat{\rho}_{ij}$, is weak cross-sectional dependence discussed in Chudik, Pesaran and Tosetti (2011), and further developed in Bailey, Kapetanios and Pesaran (2012, BKP). More specifically, we show that the implicit null of the CD test depends on the relative expansion rates of N and T. In general, if $T = O(N^\epsilon)$ for some ϵ in the range $(0,1]$, then the implicit null of the CD test is given by $0 \leq \alpha < (2-\epsilon)/4$, where α is the exponent of cross-sectional dependence defined by $\bar{\rho}_N = [2/(N(N-1))] \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \rho_{ij} = O(N^{2\alpha-2})$, with ρ_{ij} denoting the population correlation coefficient of u_{it} and u_{jt}. α measures the degree of cross-sectional dependence amongst the errors and attains its highest value of unity if the number of non-zero pair-wise error correlations tend to infinity at the same rate as N^2. This is related to the rate at which the largest eigenvalue of the correlation matrix of the errors, $\lambda_{\max}(\mathbf{R})$, expands with N, where $\mathbf{R} = (\rho_{ij})$. $\alpha = 1$ corresponds to the strong cross-sectional dependence case where $\lambda_{\max}(\mathbf{R})$ rises at the same rate as N. (see, for example, Chamberlain, 1983). The value of $\alpha = 0$ corresponds to the opposite extreme where $\lambda_{\max}(\mathbf{R})$ is fixed in N. Intermediate values of α

---

[1]For 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).

in the range $(0, 1)$ correspond to cases where $\lambda_{\max}(\mathbf{R})$ changes with N but at a slower rate. There is also a correspondence between the rate of expansion of $\lambda_{\max}(\mathbf{R})$ and the number of non-zero factor loadings in a factor representation of the errors which is discussed in the literature, although the focus has been on the extreme values of $\alpha = 0$ and $\alpha = 1$. (see, for example, Chudik, Pesaran and Tosetti (2011)).

BKP show that α is identified and can be estimated consistently if $1/2 < \alpha \leq 1$. This paper complements BKP by showing that the null hypothesis that α lies in the range $[0, 1/2)$ can be tested using the CD statistic if ϵ is close to zero (T almost fixed as $N \to \infty$), but in the case where $\epsilon = 1$ (N and $T \to \infty$ at the same rate) then the implicit null of the CD test is given by $\alpha < 1/4$.

The null of weak cross-sectional dependence also seems more appropriate than the null of cross-sectional independence in the case of large panel data models where only pervasive cross dependence is of concern. For example, in portfolio analysis full diversification of idiosyncratic errors is achieved if the errors are weakly correlated, and cross-sectional error independence is not required. In estimation of panels only strong cross-sectional error dependence can pose real problems, and in most applications weak cross-sectional error dependence does not pose serious estimation and inferential problems.

The small sample properties of the CD test for different values of α and sample sizes are investigated by means of a number of Monte Carlo experiments. It is shown that the CD test has the correct size for values of α in the range $[0, 1/4]$, for all combinations of N and T, and irrespective of whether the panel contains lagged values of the dependent variables, so long as there are not major asymmetries in the error distributions. This is in contrast to the LM based tests (such as the one proposed by Pesaran, Ullah and Yamagata, 2008) that require the regressors to be strictly exogenous. In line with the theoretical results, the CD test tends to over-reject if T is large relative to N and α is within the interval $(1/4, 1/2]$. The CD test also has satisfactory power for all values of $\alpha > 1/2$ and rises with $N\sqrt{T}$ so long as $\alpha > 1/2$.

The rest of the paper is organized as follows. The panel data model and the LM tests of error cross-sectional independence are introduced in Section 2. The concept of weak cross-sectional dependence is introduced and discussed in Section 3. The use of CD statistic for testing weak cross-sectional dependence is discussed in Section 4, where the asymptotic distribution of the test is rigorously established under the null of independence. The distribution of CD statistic under the more general null of weak dependence is considered in Section 5, and the conditions under which it tends to $N(0, 1)$ are derived. The application of the test to heterogeneous dynamic panels is discussed in Section 6. Small sample evidence on the performance of the test is provided in Section 7. Section 8 concludes.

2 Panel Data Models and the LM Type Tests of Cross-Sectional Error Independence

Consider the following panel data model

$$y_{it} = \boldsymbol{\beta}_i' \mathbf{x}_{it} + u_{it}, \text{ for } i = 1, 2, ..., N; \ t = 1, 2, ..., T, \tag{1}$$

where i indexes the cross section dimension and t the time series dimension, $\mathbf{x}_{it}$ is a $(k+1) \times 1$ vector of observed time-varying regressors (individual-specific as well as common regressors). An individual-specific intercept can be included by setting the first element of $\mathbf{x}_{it}$ to unity. The coefficients, $\boldsymbol{\beta}_i$, are defined on a compact set and allowed to vary across i. For each i, $u_{it} \sim$

$IID(0, \sigma_i^2)$, 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 $\mathbf{x}_{it} = (1, y_{i,t-1}, ..., y_{i,t-p})$. The testing procedure is applicable to fixed and random effects models as well as to the more general heterogeneous slope or random coefficient specifications.

2.1 LM Type Tests

In the SURE context with N fixed and $T \to \infty$, Breusch and Pagan (1980) proposed a 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,

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

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

$$e_{it} = y_{it} - \hat{\boldsymbol{\beta}}_i' \mathbf{x}_{it}, \tag{3}$$

with $\hat{\boldsymbol{\beta}}_i$ being the OLS estimator of $\boldsymbol{\beta}_i$ computed using the regression of y_{it} on $\mathbf{x}_{it}$ for each i, separately. The LM test 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-sectional dependence, specified by

$$Cov(u_{it}, u_{jt}) = 0, \text{ for all } t, i \neq j, \tag{4}$$

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 $N \to \infty$. However, noting that under H_0, $T\hat{\rho}_{ij}^2 \overset{a}{\sim} \chi_1^2$ with $\hat{\rho}_{ij}^2$, $i = 1, 2, .., N-1$, $j = i+1, 2, ..., N$, 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:

$$CD_{lm} = \sqrt{\frac{1}{N(N-1)}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} (T\hat{\rho}_{ij}^2 - 1). \tag{5}$$

---

[2] The 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.

It is now easily seen that under H_0 with $T \to \infty$ first followed by $N \to \infty$ we would have $CD_{lm} \overset{a}{\sim}$ $N(0,1)$. 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. A bias corrected version of CD_{lm} is proposed in Pesaran, Ullah and Yamagata (2008) under the assumptions that the regressors are strongly exogenous and the errors are normally distributed. In what follows we propose a test of weak cross-sectional dependence, which we argue to be more appropriate for large panels, where mere incidence of isolated dependencies are of little consequence for estimation or inference.

3 Weak Error Cross-Sectional Dependence

As noted in the introduction when N is large it is often more appropriate to consider the extent of error cross-sectional dependence rather than the extreme null hypothesis of error independence that underlies the LM type tests. This is in line, for example, with the assumption of approximate factor models discussed in Chamberlain (1983) in the context of capital asset pricing models. To this end we consider the following factor model for the errors

$$u_{it} = \omega_i \left(\boldsymbol{\gamma}_i' \mathbf{f}_t + \varepsilon_{it} \right), \tag{6}$$

where $\mathbf{f}_t = (f_{1t}, f_{2t}, ..., f_{mt})'$ is the $m \times 1$ vector of unobserved common factors (m being fixed) with $E(\mathbf{f}_t) = \mathbf{0}$, and $Cov(\mathbf{f}_t) = \mathbf{I}_m$, $\boldsymbol{\gamma}_i = (\gamma_{i1}, \gamma_{i2}, ..., \gamma_{im})'$ is the associated vector of factor loadings, ε_{it} are idiosyncratic errors that are cross-sectionally and serially independent with mean zero and a unit variance, namely $\varepsilon_{it} \sim IID(0,1)$, and ω_i is a scaler that controls the variance of u_{it}. The degree of cross-sectional dependence of the errors, u_{it}, is governed by the rate at which the average pairwise error correlation coefficient, $\bar{\rho}_N = [2/N(N-1)] \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \rho_{ij}$, tends to zero in N, where $\rho_{ij} = Corr(u_{it}, u_{jt})$. In the case of the above factor model we have, $Var(u_{it}) = \sigma_i^2 = \omega_i^2(1 + \boldsymbol{\gamma}_i'\boldsymbol{\gamma}_i)$, $\rho_{ij} = \boldsymbol{\delta}_i'\boldsymbol{\delta}_j$, for $i \neq j$, where

$$\boldsymbol{\delta}_i = \frac{\boldsymbol{\gamma}_i}{\sqrt{1 + \boldsymbol{\gamma}_i'\boldsymbol{\gamma}_i}}. \tag{7}$$

Then it is easily seen that

$$\bar{\rho}_N = \left(\frac{N}{N-1} \right) \left(\bar{\boldsymbol{\delta}}_N' \bar{\boldsymbol{\delta}}_N - \frac{\sum_{i=1}^{N} \boldsymbol{\delta}_i'\boldsymbol{\delta}_i}{N^2} \right), \tag{8}$$

where $\bar{\boldsymbol{\delta}}_N = N^{-1} \sum_{i=1}^{N} \boldsymbol{\delta}_i$.

Consider now the effects of the jth factor, f_{jt}, on the ith error, u_{it}, as measured by γ_{ij}, and suppose that these factor loadings take non-zero values for M_j out of the N cross-section units under consideration. Then following BKP, the degree of cross-sectional dependence due to the jth factor can be measured by $\alpha_j = \ln(M_j)/\ln(N)$, and the overall degree of cross-sectional dependence of the errors by $\alpha = \max_j(\alpha_j)$. BKP refer to α as the exponent of cross-sectional dependence. α can take any value in the range 0 to 1, with 1 indicating the highest degree of cross-sectional dependence. Considering that $\boldsymbol{\gamma}_i'\boldsymbol{\gamma}_i = O(m)$ where m is fixed as $N \to \infty$, the exponent of cross-sectional dependence of the errors can be equivalently defined in terms of the scaled factor loadings,

$\boldsymbol{\delta}_i = (\delta_{i1}, \delta_{i2}, ..., \delta_{im})'$. Without loss of generality, suppose that only the first M_j elements of δ_{ij} over i are non-zero, and note that[3]

$$\bar{\delta}_{j,N} = \frac{1}{N} \left(\sum_{i=1}^{M_j} \delta_{ij} + \sum_{i=M_j+1}^{N} \delta_{ij} \right) = \frac{M_j}{N} \left(M_j^{-1} \sum_{i=1}^{M_j} \delta_{ij} \right) = N^{\alpha_j - 1} \mu_j = O(N^{\alpha_j - 1}),$$

where $\mu_j = \left(M_j^{-1} \sum_{i=1}^{M_j} \delta_{ij} \right) \neq 0$, for a finite M_j and as $M_j \to \infty$. Similarly, $N^{-2} \sum_{i=1}^{N} \delta_{ij}^2 = O(N^{\alpha_j - 2})$, and using (8) we have

$$\bar{\rho}_N = O(N^{2\alpha - 2}).$$

In what follows we develop a test of the null hypothesis that $\alpha < 1/2$. The case where $\alpha > 1/2$ is covered in BKP. The values of α in the range $[0, 1/2)$ correspond to different degrees of weak cross-sectional dependence, as compared to values of α in the range $(1/2, 1]$ that relate to different degrees of strong cross-sectional dependence.

4   A Test of Weak Cross-Sectional Dependence

Given that α is defined by the contraction rate of $\bar{\rho}_N$, we base the test of weak cross-sectional error dependence on its sample estimate, given by

$$\widehat{\bar{\rho}}_N = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\rho}_{ij}, \tag{9}$$

where $\hat{\rho}_{ij}$ is already defined by (2). The CD test of Pesaran (2004) is in fact a scaled version of $\widehat{\bar{\rho}}_N$ which can be written as

$$CD = \left[\frac{TN(N-1)}{2} \right]^{1/2} \widehat{\bar{\rho}}_N. \tag{10}$$

In what follows we consider the distribution of the CD statistic under three different null hypotheses. To establish comparability and some of the basic results we begin with CD statistic under hypothesis of cross-sectional independence defined by

$$H_0 : \boldsymbol{\gamma}_i = \mathbf{0}, \text{ for all } i. \tag{11}$$

We then consider the asymptotic distribution of the CD statistic as N and $T \to \infty$, such that $T = O(N^\epsilon)$, for $0 < \epsilon \leq 1$, and show that the implicit null of the CD test is given by

$$H_0^w : \alpha < (2 - \epsilon)/4. \tag{12}$$

As argued earlier, such a null is much less restrictive for large N panels than the pair-wise error independence assumption that underlies the LM type tests which are based on $\hat{\rho}_{ij}^2$.

Initially, we derive the asymptotic distribution of the CD test in the case of the standard panel data model, (1) subject to the following assumptions:

---

[3]The main results in the paper remain valid even if $\sum_{i=M_j+1}^{N} \delta_{ij} = O(1)$. But for expositional simplicity we maintain the assumption that $\sum_{i=M_j+1}^{N} \delta_{ij} = 0$.

Assumption 1: The factor model, (6), holds. The idiosyncratic errors, ε_{it}, are serially and cross-sectionally independent with mean zero and a unit variance, and are symmetrically distributed around 0 for all i and t. $\mathbf{f}_t \sim IID(\mathbf{0}, \mathbf{I}_m)$, $\mathbf{f}_t$ and $\varepsilon_{i,t'}$ are distributed independently for all t and t', and $0 < \omega_i^2 < K < \infty$. The factor loadings, $\boldsymbol{\gamma}_i$, are independently distributed across i.

Assumption 2a: The regressors, $\mathbf{x}_{it}$, are strictly exogenous such that

$$E\left(\varepsilon_{it} \mid \mathbf{X}_j\right) = 0, \text{ for all } i, \ j \text{ and } t, \tag{13}$$

where $\mathbf{X}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{iT})'$. Also $T^{-1}\mathbf{X}_i'\mathbf{X}_i$ is a positive definite matrix for any fixed T, $T^{-1}\mathbf{X}_i'\mathbf{X}_i \to_p \boldsymbol{\Sigma}_{ii}$, as $T \to \infty$, with $\boldsymbol{\Sigma}_{ii}$ being a positive definite matrix, and $T^{-1}\mathbf{X}_i'\mathbf{X}_j \to_p \boldsymbol{\Sigma}_{ij} = O(1)$.

Assumption 2b: The following standardized errors

$$\zeta_{it,T} = \varepsilon_{it}/(T^{-1}\boldsymbol{\varepsilon}_i'\mathbf{M}_i\boldsymbol{\varepsilon}_i)^{1/2}, \text{ for } i = 1, 2, ..., N; t = 1, 2, ..., T \tag{14}$$

where $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, ..., \varepsilon_{iT})'$, are cross-sectionally and serially independent and have fourth-order moments, namely

$$E\left(\zeta_{it,T}^4\right) < K < \infty. \tag{15}$$

It is also assumed that the regressors are suitably bounded such that

$$h_{it,T} = -\mathbf{x}_{it}'\left(\frac{\mathbf{X}_i'\mathbf{X}_i}{T}\right)^{-1}\left(\frac{\mathbf{X}_i'\boldsymbol{\zeta}_{iT}}{\sqrt{T}}\right), \tag{16}$$

have fourth-order moments, namely $E\left(h_{it,T}^4\right) < K < \infty$, where $\boldsymbol{\zeta}_{iT} = (\zeta_{i1,T}, \zeta_{i2,T}, ..., \zeta_{iT,T})$, and $\mathbf{M}_i = \mathbf{I}_T - \mathbf{X}_i\left(\mathbf{X}_i'\mathbf{X}_i\right)^{-1}\mathbf{X}_i'$.

Assumption 3: $T > k + 4$ and the OLS residuals, e_{it}, defined by (3), are not all zero.

Assumption 4: The factor loadings, $\boldsymbol{\gamma}_i$, defined by (6) satisfy the α-summability condition

$$\sum_{i=1}^{N} \boldsymbol{\gamma}_i = O(N^\alpha). \tag{17}$$

Remark: *Assumption 1 is standard in the literature on multi-factor models, except for the assumption that the idiosyncratic errors are symmetrically distributed. This assumption is needed to ensure that the standardized errors, $\zeta_{it,T}$, also have mean zero. The first part of Assumption 2b is met if $\sup_i E(\varepsilon_{it}^6) < K < \infty$, and $T > k + 4$. This result can be established using Lemmas in Lieberman (1994). For future use it is also worth noting that*

$$E\left(\zeta_{it,T}^2\right) = E\left(\frac{\varepsilon_{it}^2}{T^{-1}\boldsymbol{\varepsilon}_i'\mathbf{M}_i\boldsymbol{\varepsilon}_i}\right) = 1 + O\left(\frac{1}{T}\right). \tag{18}$$

In fact when ε_{it} are normally distributed we have the following exact result

$$E\left(\zeta_{it,T}^2\right) = \frac{T}{T - k - 3},$$

which follows since $(T - k - 1)\varepsilon_{it}^2/\varepsilon_i'M_i\varepsilon_i$ has an F distribution with 1 and $T - k - 1$ degrees of freedom. In the case where ε_{it} are non-Guassian, the moment conditions in (18) follow from standard application of results in Lieberman (1994), assuming the regressors are suitably bounded.

Theorem 1 *Consider the regression model, (1), and suppose that Assumptions 1-3 hold, and the idiosyncratic errors, ε_{it}, are symmetrically distributed around 0, then under $H_0 : \boldsymbol{\gamma}_i = \mathbf{0}$, and for all $N > 1$ and $T > k + 1$ we have*

$$E\left(\hat{\rho}_{ij}\right) = 0, \ \text{for all } i \neq j, \tag{19}$$

$$E\left(\hat{\rho}_{ij}\hat{\rho}_{is}\right) = 0, \ \text{for all } i \neq j \neq s, \tag{20}$$

$$E\left(CD\right) = 0, \tag{21}$$

$$Var(CD) = 1 + \frac{T \ \bar{a}_N}{(T - k - 1)^2} - \frac{(k + 1)^2}{(T - k - 1)^2}, \tag{22}$$

$$\bar{a}_N = \frac{2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} Tr\left(\mathbf{A}_i\mathbf{A}_j\right)}{N(N - 1)} < k + 1, \tag{23}$$

where $\mathbf{A}_i = \mathbf{X}_i(\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'$, $\hat{\rho}_{ij}$ and CD are defined by (2) and (10), respectively.[4]

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

$$\hat{\rho}_{ij} = T^{-1} \sum_{t=1}^{T} \xi_{it}\xi_{jt}, \tag{24}$$

where ξ_{it} are the scaled residuals defined by

$$\xi_{it} = \frac{e_{it}}{(T^{-1}\mathbf{e}_i'\mathbf{e}_i)^{1/2}}, \tag{25}$$

e_{it} is the OLS residuals from the individual-specific regressions, defined by (3), and $\mathbf{e}_i = (e_{i1}, e_{i2}, ..., e_{iT})'$. Also under H_0, $\mathbf{e}_i = \omega_i\mathbf{M}_i\boldsymbol{\varepsilon}_i$, where $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, ..., \varepsilon_{iT})'$. 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

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

Hence, unconditionally we also have

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

Using this result in (24) now yields (recall that under H_0 the errors, ε_{it}, are cross-sectionally independent),

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

which in turn establishes that (using (10))

$$E(CD) = 0,$$

---

[4]Similar 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.

for any N, and all $T > k + 1$. Under H_0 and Assumptions 1-3, $\hat{\rho}_{ij}$ and $\hat{\rho}_{is}$ are cross-sectionally uncorrelated for i, j and s, such that $i \neq j \neq s$. More specifically

$$
\begin{aligned}
E\left(\hat{\rho}_{ij}\hat{\rho}_{is}\right) &= T^{-2} \sum_{t=1}^{T} \sum_{t'=1}^{T} E\left(\xi_{it}\xi_{jt}\xi_{it'}\xi_{st'}\right) \\
&= T^{-2} \sum_{t=1}^{T} \sum_{t'=1}^{T} E\left(\xi_{it}\xi_{it'}\right) E\left(\xi_{jt}\right) E\left(\xi_{st'}\right) = 0, \text{ for } i \neq j \neq s.
\end{aligned}
$$

Also since the regressors are assumed to be strictly exogenous, we further have[5]

$$
Var\left(\sqrt{T}\hat{\rho}_{ij}\right) = E\left(T\hat{\rho}_{ij}^2\right) = T \, Tr(\mathbf{M}_i\mathbf{M}_j)/(T - k - 1)^2.
$$

Using this result in (10) we have

$$
\begin{aligned}
Var(CD) &= \frac{2T}{N(N-1)(T-k-1)^2} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} [T - 2(k+1) + Tr\left(\mathbf{A}_i\mathbf{A}_j\right)] \right) \\
&= \frac{T\left[T - 2(k+1)\right]}{(T-k-1)^2} + \frac{2T}{N(N-1)(T-k-1)^2} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} Tr\left(\mathbf{A}_i\mathbf{A}_j\right) \right).
\end{aligned}
$$

Hence

$$
Var(CD) = 1 + \frac{T \, \bar{a}_N}{(T-k-1)^2} - \frac{(k+1)^2}{(T-k-1)^2} = 1 + O\left(\frac{1}{T}\right),
$$

where

$$
\bar{a}_N = \frac{2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} Tr\left(\mathbf{A}_i\mathbf{A}_j\right)}{N(N-1)}.
$$

But $Tr\left(\mathbf{A}_i\mathbf{A}_j\right) < \left[Tr(\mathbf{A}_i^2)Tr(\mathbf{A}_j^2)\right]^{1/2} = k+1$, and we must also have $\bar{a}_N < k+1$. This completes the proof of the theorem. $\blacksquare$

The above results also suggest the following modified version of CD,

$$
\widetilde{CD} = \frac{CD}{\left[1 + \frac{T \, \bar{a}_N}{(T-k-1)^2} - \frac{(k+1)^2}{(T-k-1)^2}\right]^{1/2}}, \tag{26}
$$

which is distributed *exactly* with a zero mean and a unit variance. In cases where $T - k - 1$ is relatively large, and the regressors, $\mathbf{x}_{it}$, are cross-sectionally weakly correlated, the term involving $\bar{a}_N$ in the expression for the variance of CD will be small and both statistics are likely to perform very similarly, and the CD test is recommended on grounds of its simplicity. To keep the analysis simple, and without of loss generality, in what follows we shall focus on the CD test.

---

[5] For a proof see Appendix A.2 in Pesaran, Ullah, and Yamagata, (2008, pp. 123-124).

4.1 The distribution of the CD test under H_0

Consider now the distribution of the CD test. As shown in (20), the elements in the double summation that forms the CD statistic are uncorrelated but they need not be independently distributed when T is finite. Therefore, when T is finite the standard central limit theorems can not be exploited in order to derive the distribution of the CD statistic.[6] To resolve the problem we first re-write the CD statistic (defined by (10)) as

$$CD_{NT} = \sqrt{\frac{2}{N(N-1)}} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \sqrt{T}\hat{\rho}_{ij} \right), \tag{27}$$

and recall that $\hat{\rho}_{ij} = T^{-1} \sum_{t=1}^{T} \xi_{it}\xi_{jt}$, where ξ_{it} is defined by (25). Now under $H_0 : \boldsymbol{\gamma}_i = 0$, using standard results from regression analysis, we have

$$\xi_{it} = \zeta_{it} + T^{-1/2}h_{it,T}, \tag{28}$$

$$h_{it,T} = -\mathbf{x}'_{it} \left(\frac{\mathbf{X}'_i\mathbf{X}_i}{T} \right)^{-1} \left(\frac{\mathbf{X}'_i\boldsymbol{\zeta}_{i,T}}{\sqrt{T}} \right), \tag{29}$$

where $\zeta_{it,T} = \varepsilon_{it}/(\boldsymbol{\varepsilon}'_i\mathbf{M}_i\boldsymbol{\varepsilon}_i/T)^{1/2}$, and $\boldsymbol{\zeta}_{iT} = (\zeta_{i1}, \zeta_{i2}, ..., \zeta_{iT})'$. To simplify the notations we abstract from the dependence of $\zeta_{it,T}$ on T. It will also prove helpful to recall that under Assumptions 1, 2a and 2b, $E(\zeta_{it,T}) = 0$, $Cov(\zeta_{it,T}, \zeta_{jt,T}) = 0$, for all $i \neq j$, and for each i, $Var(\zeta_{it,T}) = \sigma^2_{\zeta_i} = 1 + O(T^{-1})$.

Before deriving the asymptotic distribution of CD_{NT} we present some preliminary results in the following lemma and provide the proofs in the Appendix.

Lemma 1: *Consider*

$$w_{t,NT} = N^{-1/2} \sum_{i=1}^{N} \zeta_{it,T}, \text{ and } h_{t,NT} = N^{-1/2} \sum_{i=1}^{N} h_{it,T}, \tag{30}$$

where $\zeta_{it,T}$ and $h_{it,T}$ are defined by (14) and (16), and suppose that Assumptions 1 and 2 hold. Then

$$E(w_{t,NT}) = 0, \ E(w^2_{t,NT}) = 1 + O\left(\frac{1}{T}\right), \tag{31}$$

$$E(h_{t,NT}) = 0, \ E\left(h^2_{t,NT}\right) = N^{-1} \sum_{i=1}^{N} \mathbf{x}'_{it} \left(\frac{\mathbf{X}'_i\mathbf{X}_i}{T} \right)^{-1} \mathbf{x}_{it} + O\left(\frac{1}{T}\right) = O_p(1), \tag{32}$$

$$Cov(h_{t,NT}, w_{t',NT}) = O(T^{-1/2}), \ Cov(w_{t,NT}, w_{t',NT}) = O(T^{-1}), \text{ for } t \neq t', \tag{33}$$

$$w_{t,NT} = O_p(1), \ h_{t,NT} = O_p(1), \text{ and } \frac{1}{T} \sum_{t=1}^{T} w_{t,NT}h_{t,NT} \rightarrow_p 0. \tag{34}$$

---

[6]This corrects the statement made in error in Pesaran (2004).

9

With these results in mind, we write CD_{NT} as

$$CD_{NT} = \sqrt{\frac{2}{N(N-1)}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \xi_{it}\xi_{jt}, \tag{35}$$

and note that

$$\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \xi_{it}\xi_{jt} = \frac{1}{2}\left[\left(\sum_{i=1}^{N} \xi_{it}\right)^2 - \sum_{i=1}^{N} \xi_{it}^2\right], \tag{36}$$

and hence

$$CD_{NT} = \sqrt{\frac{N}{2(N-1)}} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[\left(\frac{\sum_{i=1}^{N} \xi_{it}}{\sqrt{N}}\right)^2 - \frac{\sum_{i=1}^{N} \xi_{it}^2}{N}\right]. \tag{37}$$

However, using (28),

$$\left(N^{-1/2} \sum_{i=1}^{N} \xi_{it}\right)^2 = \left(N^{-1/2} \sum_{i=1}^{N} \zeta_{it}\right)^2 + \left((NT)^{-1/2} \sum_{i=1}^{N} h_{it,T}\right)^2 \tag{38}$$

$$+ 2\left(N^{-1/2} \sum_{i=1}^{N} \zeta_{it}\right)\left((NT)^{-1/2} \sum_{i=1}^{N} h_{it,T}\right),$$

and

$$N^{-1} \sum_{i=1}^{N} \xi_{it}^2 = N^{-1} \sum_{i=1}^{N} \zeta_{it}^2 + (NT)^{-1} \sum_{i=1}^{N} h_{it,T}^2 + \tag{39}$$

$$2T^{-1/2}\left(N^{-1/2} \sum_{i=1}^{N} h_{it,T}\right)\left(N^{-1/2} \sum_{i=1}^{N} \zeta_{it}\right).$$

Consider now the terms involving $h_{it,T}$, and note that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T}\left((NT)^{-1/2} \sum_{i=1}^{N} h_{it,T}\right)^2 = \frac{1}{\sqrt{T}}\frac{1}{T} \sum_{t=1}^{T} h_{t,NT}^2 = O_p\left(T^{-1/2}\right) \tag{40}$$

$$\frac{1}{\sqrt{T}}\frac{1}{NT} \sum_{t=1}^{T}\sum_{i=1}^{N} h_{it,T}^2 = O_p\left(T^{-1/2}\right). \tag{41}$$

Further,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T}\left(N^{-1/2} \sum_{i=1}^{N} \zeta_{it}\right)\left((NT)^{-1/2} \sum_{i=1}^{N} h_{it,T}\right) = \frac{1}{T} \sum_{t=1}^{T} w_{tN}h_{t,NT}, \tag{42}$$

where w_{tN}, and $h_{t,NT}$ are defined by (30). Therefore, using (40), (41) and (42) in (38) and (39) and then substituting the results back in (37) we have

$$CD_{NT} = Z_{NT} + o_p(1), \tag{43}$$

10

where

$$Z_{NT} = \sqrt{\frac{N}{2(N-1)}} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[\left(\frac{\sum_{i=1}^{N} \zeta_{it}}{\sqrt{N}} \right)^2 - \frac{\sum_{i=1}^{N} \zeta_{it}^2}{N} \right] \tag{44}$$

and $o_p(1)$ indicates terms that tend to zero in probability as N and $T \to \infty$, in any order.

To derive the distribution of Z_{NT}, recall that $w_{t,NT} = N^{-1/2} \sum_{i=1}^{N} \zeta_{it}$, and write Z_{NT} as

$$Z_{NT} = \sqrt{\frac{N}{(N-1)}} \left(U_{NT} - V_{NT} \right), \tag{45}$$

where

$$U_{NT} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(\frac{w_{t,NT}^2 - E(w_{t,NT}^2)}{\sqrt{2}} \right),$$

and

$$V_{NT} = \frac{\sum_{t=1}^{T} \sum_{i=1}^{N} \left[\zeta_{it}^2 - E(\zeta_{it}^2) \right]}{N\sqrt{2T}} = \frac{\sum_{t=1}^{T} \sum_{i=1}^{N} \chi_{it}}{N\sqrt{2T}},$$

where $\chi_{it} = \zeta_{it}^2 - E(\zeta_{it}^2)$, and $E(w_{t,NT}^2) = 1 + O\left(T^{-1}\right)$. Under our assumptions, χ_{it} are cross-sectionally independently distributed with mean 0 and a finite variance $Var\left(\chi_{it}\right) = E(\zeta_{it}^4) - \sigma_{\zeta_i}^4$, such that $\sup_i Var\left(\chi_{it}\right) < K < \infty$. Furthermore, χ_{it} are serially uncorrelated as $T \to \infty$ (see (33)). Hence, it readily follows that

$$E(V_{NT}) = 0, \ Var\left(V_{NT}\right) = \frac{1}{2TN^2} \sum_{t=1}^{T} \sum_{i=1}^{N} Var\left(\chi_{it}\right) < \frac{1}{2N} \sup_i Var\left(\chi_{it}\right) = O\left(N^{-1}\right). \tag{46}$$

Consider now U_{NT}, and recall that $w_{t,NT}^2$ is asymptotically temporally uncorrelated with $E(w_{t,NT}^2) = 0$, and $E(w_{t,NT}^2) = 1 + O\left(T^{-1}\right)$. Hence,

$$E(U_{NT}) = 0,$$
$$Var(U_{NT}) = \frac{1}{2T} \sum_{t=1}^{T} Var\left(w_{t,NT}^2\right) = \frac{1}{2T} \sum_{t=1}^{T} \left[E\left(w_{t,NT}^4\right) - \left[E\left(w_{t,NT}^2\right) \right]^2 \right]. \tag{47}$$

But, noting that $\zeta_{it,T}$ are cross-sectionally independent, we have

$$E\left(w_{t,NT}^4\right) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{r=1}^{N} \sum_{s=1}^{N} E\left(\zeta_{it}\zeta_{jt}\zeta_{rt}\zeta_{st}\right)$$

$$= \frac{3}{N^2} \left[\sum_{i=1}^{N} E\left(\zeta_{it}^2\right) \right]^2 + \frac{1}{N^2} \sum_{i=1}^{N} E\left(\zeta_{it}^4\right)$$

$$= 3 \left[E(w_{t,NT}^2) \right]^2 + \frac{1}{N^2} \sum_{i=1}^{N} E\left(\zeta_{it}^4\right), \tag{48}$$

11

and substituting (48) into (47) we obtain

$$Var(U_{NT}) = \frac{1}{T} \sum_{t=1}^{T} \left[E\left(w_{t,NT}^2\right)\right]^2 + \frac{1}{2N^2T} \sum_{t=1}^{T} \sum_{i=1}^{N} E\left(\zeta_{it}^4\right).$$

Now using (46) in (44), and then in (43), we have

$$CD_{NT} = U_{NT} + O_p\left(N^{-1/2}\right) + o(1).$$

Also, since $E(w_{t,NT}^2) = 1 + O\left(T^{-1}\right)$, we have

$$U_{NT} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(\frac{w_{t,NT}^2 - E(w_{t,NT}^2)}{\sqrt{2}} \right) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(\frac{w_{t,NT}^2 - 1}{\sqrt{2}} \right) + O_p\left(\frac{1}{\sqrt{T}}\right).$$

But for any t and as $N \to \infty$, $w_{t,NT} = N^{-1/2} \sum_{i=1}^{N} \zeta_{it} \to_d N(0,1)$, and therefore $w_{t,NT}^2 \to_d \chi_t^2(1)$, where $\chi_t^2(1)$, for $t = 1,2,...,T$ are independent chi-square variates with 1 degree of freedom. This in turn implies that as $N \to \infty$, $\frac{1}{\sqrt{2}}\left(w_{t,NT}^2 - 1\right)$, for $t = 1,2,...,T$, are independent random variates with mean zero and a unit variance. Hence, $U_{NT} \to_d N(0,1)$, as N and $T \to \infty$, noting also that the term $O_p\left(\sqrt{T}/N\right)$ vanishes with $N \to \infty$, considering that $\sqrt{T}/N = O(N^{-1+\epsilon/2})$ and $\epsilon < 1$. To summarize:

Theorem 2 *Consider the panel data model (1), and suppose that Assumptions 1 to 4 hold. Then under the null hypothesis of cross-sectional error independence,(11), the CD statistic defined by (10) has the limiting $N(0,1)$ distribution as N and $T \to \infty$, 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 7.

We now show that the CD test is in fact applicable even if the errors are weakly correlated, namely the implicit null of the CD test is weak cross-sectional error dependence, rather than independence. We argue that when N is large the null hypothesis that *all* pairs of errors are independently distributed is rather restrictive. What is needed is a less restrictive null that postulate a sufficiently large number of pairs of errors are independently distributed. The analysis below formalizes what is meant by "a sufficiently large number".

5   Asymptotic Distribution of the CD Test Under Weak Cross-Sectional Error Dependence

In this section we consider the asymptotic distribution of the CD statistic under the null of weak cross-sectional dependence, H_0^w defined by (12). To this end we assume that for each i

$$\begin{cases} T^{-1} \sum_{t=1}^{T} \varepsilon_{it}\mathbf{f}_t = O_p\left(T^{-1/2}\right), \ T^{-1} \sum_{t=1}^{T} \mathbf{x}_{it}\mathbf{f}_t' = O_p\left(T^{-1/2}\right), \\ \qquad T^{-1} \sum_{t=1}^{T} \mathbf{f}_t\mathbf{f}_t' = \mathbf{I}_m + O_p\left(T^{-1/2}\right). \end{cases} \tag{49}$$

We also make the following standard assumptions about the regressors[7]

$$\left(\frac{\mathbf{X}_i'\mathbf{X}_j}{T}\right) = \mathbf{\Sigma}_{ij} + O_p(T^{-1/2}), \; \left(\frac{\mathbf{X}_i'\boldsymbol{\varepsilon}_i}{T}\right) = O_p(T^{-1/2}), \text{ for all } i \text{ and } j, \tag{50}$$

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

Consider now the CD test statistic defined by (10) and note that under H_0^w, the vector of the OLS residuals is given by

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

where $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_T)'$, and as before $\mathbf{M}_i = \mathbf{I}_T - \mathbf{X}_i \left(\mathbf{X}_i'\mathbf{X}_i\right)^{-1} \mathbf{X}_i'$. In this case the distribution of $\hat{\rho}_{ij}$ 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, ω_i^2. Under H_0^w, ξ_{it} defined by (25), can be written as

$$\xi_{it} = \frac{\boldsymbol{\gamma}_i'\mathbf{f}_t + \varepsilon_{it}}{(T^{-1}\boldsymbol{\varepsilon}_i'\mathbf{M}_i\boldsymbol{\varepsilon}_i + 2T^{-1}\boldsymbol{\varepsilon}_i'\mathbf{M}_i\mathbf{F}\boldsymbol{\gamma}_i + T^{-1}\boldsymbol{\gamma}_i'\mathbf{F}'\mathbf{M}_i\mathbf{F}\boldsymbol{\gamma}_i)^{1/2}}$$
$$-\frac{\mathbf{x}_{it}' \left(\mathbf{X}_i'\mathbf{X}_i\right)^{-1} \mathbf{X}_i' \left(\mathbf{F}\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i\right)}{(T^{-1}\boldsymbol{\varepsilon}_i'\mathbf{M}_i\boldsymbol{\varepsilon}_i + 2T^{-1}\boldsymbol{\varepsilon}_i'\mathbf{M}_i\mathbf{F}\boldsymbol{\gamma}_i + T^{-1}\boldsymbol{\gamma}_i'\mathbf{F}'\mathbf{M}_i\mathbf{F}\boldsymbol{\gamma}_i)^{1/2}},$$

or more compactly

$$\xi_{it} = \tilde{\zeta}_{it} + T^{-1/2}\tilde{h}_{it,T} + \tilde{g}_{it,T} \tag{51}$$

where

$$\tilde{\zeta}_{it} = \frac{\varepsilon_{it}}{\psi_{iT}}, \; \tilde{h}_{it,T} = -\mathbf{x}_{it}' \left(\frac{\mathbf{X}_i'\mathbf{X}_i}{T}\right)^{-1} \left(\frac{\mathbf{X}_i'\tilde{\zeta}_i}{\sqrt{T}}\right),$$

$$\psi_{iT} = \left(T^{-1}\boldsymbol{\varepsilon}_i'\mathbf{M}_i\boldsymbol{\varepsilon}_i + 2T^{-1}\boldsymbol{\varepsilon}_i'\mathbf{M}_i\mathbf{F}\boldsymbol{\gamma}_i + T^{-1}\boldsymbol{\gamma}_i'\mathbf{F}'\mathbf{M}_i\mathbf{F}\boldsymbol{\gamma}_i\right)^{1/2},$$

and

$$\tilde{g}_{it,T} = \frac{\boldsymbol{\gamma}_i'\mathbf{f}_t - \mathbf{x}_{it}' \left(\mathbf{X}_i'\mathbf{X}_i\right)^{-1} \mathbf{X}_i'\mathbf{F}\boldsymbol{\gamma}_i}{\psi_{iT}}. \tag{52}$$

Using (51) in (37), we now have

$$CD_{NT} = \sqrt{\frac{N}{2(N-1)}} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[\left(\frac{\sum_{i=1}^{N} \tilde{\zeta}_{it} + T^{-1/2}\tilde{h}_{it,T} + \tilde{g}_{it,T}}{\sqrt{N}}\right)^2 - \frac{\sum_{i=1}^{N} \left(\tilde{\zeta}_{it} + T^{-1/2}\tilde{h}_{it,T} + \tilde{g}_{it,T}\right)^2}{N} \right]. \tag{53}$$

Following the derivations in the previous section, it is possible to show that under Assumptions 1-4, (49), and (50), the null of weak cross-sectional dependence given by (12), then the CD_{NT} statistics tends to $N(0,1)$ if

$$N^{-1} \sum_{i=1}^{N} E(\tilde{\zeta}_{it}^2) \to 1, \tag{54}$$

---

[7]These 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.

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(\frac{\sum_{i=1}^{N} \tilde{g}_{it,T}}{\sqrt{N}} \right)^2 \to_p 0, \text{ and } \frac{1}{N\sqrt{T}} \sum_{t=1}^{T} \sum_{i=1}^{N} \tilde{g}_{it,T}^2 \to_p 0. \tag{55}$$

To establish these results, we first note that under Assumptions (49), and (50)

$$\psi_{iT}^2 = 1 + \boldsymbol{\gamma}_i' \boldsymbol{\gamma}_i + O_p(T^{-1/2}).$$

Using this result we have

$$N^{-1} \sum_{i=1}^{N} E(\check{\zeta}_{it}^2) = N^{-1} \sum_{i=1}^{N} E\left(\frac{\varepsilon_{it}^2}{\psi_{iT}^2} \right) \to N^{-1} \sum_{i=1}^{N} \frac{1}{1 + \boldsymbol{\gamma}_i' \boldsymbol{\gamma}_i} = 1 - N^{-1} \sum_{i=1}^{N} \frac{\boldsymbol{\gamma}_i' \boldsymbol{\gamma}_i}{1 + \boldsymbol{\gamma}_i' \boldsymbol{\gamma}_i}.$$

But under (17), $\sum_{i=1}^{N} \frac{\boldsymbol{\gamma}_i' \boldsymbol{\gamma}_i}{1 + \boldsymbol{\gamma}_i' \boldsymbol{\gamma}_i} = O(N^\alpha)$, and $N^{-1} \sum_{i=1}^{N} E(\check{\zeta}_{it}^2) \to 1$, if $\alpha < 1$.

Consider now the two expressions in (55), and note that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(\frac{\sum_{i=1}^{N} \tilde{g}_{it,T}}{\sqrt{N}} \right)^2 = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(\frac{\sum_{i=1}^{N} \tilde{\boldsymbol{\gamma}}_i' \mathbf{f}_t - \sum_{i=1}^{N} \mathbf{x}_{it}' \left(\mathbf{X}_i' \mathbf{X}_i \right)^{-1} \mathbf{X}_i' \mathbf{F} \tilde{\boldsymbol{\gamma}}_i}{\sqrt{N}} \right)^2 \tag{56}$$

$$= \frac{\sqrt{T}}{N} \left(\sum_{i=1}^{N} \tilde{\boldsymbol{\gamma}}_i' \right) \left(\frac{\mathbf{F}'\mathbf{F}}{T} \right) \left(\sum_{i=1}^{N} \tilde{\boldsymbol{\gamma}}_i \right)$$

$$+ \frac{1}{N\sqrt{T}} \sum_{t=1}^{T} \left(\sum_{i=1}^{N} \mathbf{x}_{it}' \left(\mathbf{X}_i' \mathbf{X}_i \right)^{-1} \mathbf{X}_i' \mathbf{F} \tilde{\boldsymbol{\gamma}}_i \right)^2 \tag{57}$$

$$- 2 \frac{1}{N\sqrt{T}} \left(\sum_{i=1}^{N} \tilde{\boldsymbol{\gamma}}_i' \right) \sum_{i=1}^{N} \left(\mathbf{F}' \mathbf{X}_i \right) \left(\mathbf{X}_i' \mathbf{X}_i \right)^{-1} \left(\mathbf{X}_i' \mathbf{F} \right) \tilde{\boldsymbol{\gamma}}_i$$

where $\tilde{\boldsymbol{\gamma}}_i = \tilde{\boldsymbol{\gamma}}_i / \psi_{iT}$. But under (49) and (17) and setting $T = O(N^\epsilon)$, we have

$$\frac{\sqrt{T}}{N} \left(\sum_{i=1}^{N} \tilde{\boldsymbol{\gamma}}_i' \right) \left(\frac{\mathbf{F}'\mathbf{F}}{T} \right) \left(\sum_{i=1}^{N} \tilde{\boldsymbol{\gamma}}_i \right) = O_p \left(N^{2\alpha - 1 + \epsilon/2} \right),$$

and $\frac{\sqrt{T}}{N} \left(\sum_{i=1}^{N} \tilde{\boldsymbol{\gamma}}_i' \right) \left(\frac{\mathbf{F}'\mathbf{F}}{T} \right) \left(\sum_{i=1}^{N} \tilde{\boldsymbol{\gamma}}_i \right) \to_p 0$, as $N \to \infty$, if $2\alpha - 1 + \epsilon/2 < 0$, or if $\alpha < (2 - \epsilon)/4$.
Similarly,

$$\frac{1}{N\sqrt{T}} \sum_{t=1}^{T} \left(\sum_{i=1}^{N} \mathbf{x}_{it}' \left(\mathbf{X}_i' \mathbf{X}_i \right)^{-1} \mathbf{X}_i' \mathbf{F} \tilde{\boldsymbol{\gamma}}_i \right)^2$$

$$= \frac{1}{N\sqrt{T}} \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} \tilde{\boldsymbol{\gamma}}_i' \mathbf{F}' \mathbf{X}_i \left(\mathbf{X}_i' \mathbf{X}_i \right)^{-1} \mathbf{x}_{it} \mathbf{x}_{jt}' \left(\mathbf{X}_j' \mathbf{X}_j \right)^{-1} \mathbf{X}_j' \mathbf{F} \tilde{\boldsymbol{\gamma}}_j$$

$$= \frac{1}{N\sqrt{T}} \sum_{i=1}^{N} \sum_{j=1}^{N} \tilde{\boldsymbol{\gamma}}_i' \left(\mathbf{F}' \mathbf{X}_i \right) \left(\mathbf{X}_i' \mathbf{X}_i \right)^{-1} \left(\mathbf{X}_i' \mathbf{X}_j \right) \left(\mathbf{X}_j' \mathbf{X}_j \right)^{-1} \left(\mathbf{X}_j' \mathbf{F} \right) \tilde{\boldsymbol{\gamma}}_j$$

$$= \frac{1}{N\sqrt{T}} \left(\sum_{i=1}^{N} \tilde{\boldsymbol{\gamma}}_i' \mathbf{F}' \mathbf{A}_i \right) \left(\sum_{i=1}^{N} \tilde{\boldsymbol{\gamma}}_i' \mathbf{F}' \mathbf{A}_i \right)',$$

14

where $\mathbf{A}_i = \mathbf{X}_i \left(\mathbf{X}_i' \mathbf{X}_i \right)^{-1} \mathbf{X}_i$. But (using the norm $\|\mathbf{A}\|^2 = Tr(\mathbf{A}'\mathbf{A})$)

$$\left\| \sum_{i=1}^{N} \tilde{\boldsymbol{\gamma}}_i' \mathbf{F}' \mathbf{A}_i \right\| \leq \sum_{i=1}^{N} \|\tilde{\boldsymbol{\gamma}}_i'\| \|\mathbf{F}' \mathbf{A}_i\| \leq \sum_{i=1}^{N} \left(\tilde{\boldsymbol{\gamma}}_i' \tilde{\boldsymbol{\gamma}}_i \right)^{1/2} \left[Tr(\mathbf{F}' \mathbf{A}_i \mathbf{F}) \right]^{1/2},$$

and $\mathbf{F}' \mathbf{A}_i \mathbf{F} = \left(\frac{\mathbf{F}' \mathbf{X}_i}{\sqrt{T}} \right) \left(\frac{\mathbf{X}_i' \mathbf{X}_i}{T} \right)^{-1} \left(\frac{\mathbf{X}_i \mathbf{F}}{\sqrt{T}} \right) = O_p(1)$, by Assumption (49). Hence $\sum_{i=1}^{N} \tilde{\boldsymbol{\gamma}}_i' \mathbf{F}' \mathbf{A}_i = O_p \left(\sum_{i=1}^{N} \left(\tilde{\boldsymbol{\gamma}}_i' \tilde{\boldsymbol{\gamma}}_i \right)^{1/2} \right) = O_p \left(N^\alpha \right)$, and for $T = O(N^\epsilon)$,[8]

$$\frac{1}{N\sqrt{T}} \sum_{t=1}^{T} \left(\sum_{i=1}^{N} \mathbf{x}_{it}' \left(\mathbf{X}_i' \mathbf{X}_i \right)^{-1} \mathbf{X}_i' \mathbf{F} \tilde{\boldsymbol{\gamma}}_i \right)^2 = \frac{1}{N\sqrt{T}} \left(\sum_{i=1}^{N} \tilde{\boldsymbol{\gamma}}_i' \mathbf{F}' \mathbf{A}_i \right) \left(\sum_{i=1}^{N} \tilde{\boldsymbol{\gamma}}_i' \mathbf{F}' \mathbf{A}_i \right)' = O \left(N^{2\alpha - 1 - \epsilon/2} \right).$$

Thus, the second term of (57) vanishes if $\alpha < (2 + \epsilon)/4$, which is satisfied if $\alpha < (2 - \epsilon)/4$. It is also easily established that the third term in (56) will also vanish if $\alpha < (2 + \epsilon)/4$.

Finally, consider the second expression in (55) and note that

$$\frac{1}{N\sqrt{T}} \sum_{t=1}^{T} \sum_{i=1}^{N} \tilde{g}_{it,T}^2$$

$$= \frac{1}{N\sqrt{T}} \sum_{t=1}^{T} \sum_{i=1}^{N} \left[\tilde{\boldsymbol{\gamma}}_i' \mathbf{f}_t - \mathbf{x}_{it}' \left(\mathbf{X}_i' \mathbf{X}_i \right)^{-1} \mathbf{X}_i' \mathbf{F} \tilde{\boldsymbol{\gamma}}_i \right]^2,$$

$$= \frac{1}{N\sqrt{T}} \sum_{t=1}^{T} \sum_{i=1}^{N} \left[\begin{array}{c} \tilde{\boldsymbol{\gamma}}_i' \mathbf{f}_t \mathbf{f}_t' \tilde{\boldsymbol{\gamma}}_i + \tilde{\boldsymbol{\gamma}}_i' \mathbf{F}' \mathbf{X}_i \left(\mathbf{X}_i' \mathbf{X}_i \right)^{-1} \mathbf{x}_{it} \mathbf{x}_{it}' \left(\mathbf{X}_i' \mathbf{X}_i \right)^{-1} \mathbf{X}_i' \mathbf{F} \tilde{\boldsymbol{\gamma}}_i - \\ 2 \tilde{\boldsymbol{\gamma}}_i' \mathbf{f}_t \mathbf{x}_{it}' \left(\mathbf{X}_i' \mathbf{X}_i \right)^{-1} \mathbf{X}_i' \mathbf{F} \tilde{\boldsymbol{\gamma}}_i \end{array} \right],$$

$$= \frac{1}{N\sqrt{T}} \sum_{i=1}^{N} \left(\tilde{\boldsymbol{\gamma}}_i' \mathbf{F}' \mathbf{F} \tilde{\boldsymbol{\gamma}}_i - \tilde{\boldsymbol{\gamma}}_i' \mathbf{F}' \mathbf{A}_i \mathbf{F} \tilde{\boldsymbol{\gamma}}_i \right).$$

Using similar lines of reasoning as above, it is easily established that $\frac{1}{N\sqrt{T}} \sum_{t=1}^{T} \sum_{i=1}^{N} \tilde{g}_{it,T}^2 \to_p 0$, if $\alpha < (2 - \epsilon)/2$, which is satisfied if $\alpha < (2 - \epsilon)/4$, considering that $\epsilon \leq 1$.

The above results are summarized in the following theorem:

Theorem 3 *Consider the panel data model (1), and suppose that Assumptions 1 to 4, (49), and (50) hold. Suppose further that N and $T \to \infty$, such that $T/N^\epsilon \to \kappa$, where ϵ lies in the range $(0, 1]$ and κ is a finite positive non-zero constant. Then the CD statistic defined by (10) has the limiting $N(0, 1)$ distribution as N and $T \to \infty$, so long as α, the exponent of cross-sectional dependence of the errors, u_{it}, is less than $(2 - \epsilon)/4$. In the case where N and T tend to infinity at the same rate the CD statistic tends to $N(0, 1)$ if $\alpha < 1/4$. The CD test is consistent for all values of $\alpha > 1/2$, with the power of the test rising in α and $N\sqrt{T}$.*

This theorem provides a full characterization of the distribution of the CD test under different degrees of cross-sectional dependence, ranging from independence, weak dependence, to strong dependence. The theorem also establishes the asymptotic power distribution of the CD test and shows that the test has sufficient power when $\alpha > 1/2$, with the power rising in $N\sqrt{T}$.

---

[8]Note that since $\psi_{iT} > 0$ then the order of $\sum_{i=1}^{N} \boldsymbol{\gamma}_i$ and $\sum_{i=1}^{N} \tilde{\boldsymbol{\gamma}}_i$ will be the same.

6 Cross Section Dependence in Heterogeneous Dynamic Panels

The analyses of the previous sections readily extend to models with lagged dependent variables. As an example consider the following first-order dynamic panel data model

$$y_{it} = a_i + \lambda_i y_{i,t-1} + \sigma_i u_{it}, \ i = 1, 2, ..., N; \ t = 1, 2, ..., T, \tag{58}$$

where $a_i = \mu_i(1 - \lambda_i)$, $y_{i0} = \mu_i + c_i u_{i0}$, and for each i the errors, u_{it}, $t = 0, 1, ..., T$ 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 $E(y_{it}) = \mu_i$, and if the process has started a long time in the past we would have $c_i = \sigma_i \left(1 - \lambda_i^2\right)^{-1/2}$. In the unit root case where $\lambda_i = 1$, c_i could still differ across i depending on the number of periods that the ith 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 root processes that could prevail in a given panel, testing for cross-sectional dependence is likely to be complicated and in general might require N and T to be large. As it is well known the OLS estimates of a_i and λ_i for the individual series, as well as the fixed and random effects panel estimates used under slope homogeneity ($\lambda_i = \lambda$) are biased when T is small. The bias could be substantial for values of λ_i near unity. Nevertheless, as it turns out in the case of pure autoregressive panels (without exogenous regressors) the CD test is still valid for all values of λ_i including those close to 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 u_{it}, $t = 0, 1, ..., T$ are symmetrically distributed. To see this we first write the individual AR(1) processes, (58), in matrix notation as

$$\boldsymbol{\Lambda}_i(\mathbf{y}_i^* - \mu_i \boldsymbol{\tau}_{T+1}) = \mathbf{D}_i \mathbf{u}_i^*, \tag{59}$$

where $\mathbf{y}_i^* = (y_{i0}, y_{i1}, ..., y_{iT})'$, $\mathbf{u}_i^* = (u_{i0}, u_{i1}, ..., u_{iT})'$, $\boldsymbol{\tau}_{T+1}$ 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 c_i and the remaining elements equal to σ_i, and

$$\boldsymbol{\Lambda}_i = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -\lambda_i & 1 & 0 & & 0 & 0 \\ 0 & -\lambda_i & 1 & & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & -\lambda_i & 1 \end{pmatrix}.$$

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

$$\hat{\lambda}_i = \frac{\mathbf{u}_i^{*\prime} \mathbf{H}_i' \mathbf{G}_1' \mathbf{M}_\tau \mathbf{G}_0 \mathbf{H}_i \mathbf{u}_i^*}{\mathbf{u}_i^{*\prime} \mathbf{H}_i' \mathbf{G}_1' \mathbf{M}_\tau \mathbf{G}_1 \mathbf{H}_i \mathbf{u}_i^*},$$

$$\hat{a}_i = \mu_i (1 - \lambda_i) + \left(\frac{\boldsymbol{\tau}_T' \mathbf{G}_0 \mathbf{H}_i \mathbf{u}_i^*}{T}\right) - \left(\frac{\boldsymbol{\tau}_T' \mathbf{G}_1 \mathbf{H}_i \mathbf{u}_i^*}{T}\right) \hat{\lambda}_i,$$

where $\mathbf{M}_\tau = \mathbf{I}_T - \tau(\tau'\tau)^{-1}\tau$, $\mathbf{H}_i = \mathbf{\Lambda}_i^{-1}\mathbf{D}_i$, $\mathbf{G}_0 = (\mathbf{0}_{T\times 1}, \mathbf{I}_T)$, $\mathbf{G}_1 = (\mathbf{I}_T, \mathbf{0}_{T\times 1})$, 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, $e_{it} = y_{it} - \hat{a}_i - \hat{\lambda}_i y_{i,t-1}$, for $t = 1, 2, ..., T$,

$$e_{it} = -\left(\hat{\lambda}_i - \lambda_i\right)(y_{i,t-1} - \mu_i) + \sigma_i u_{it} - \left(\frac{\tau_T'\mathbf{G}_0\mathbf{H}_i\mathbf{u}_i^*}{T}\right) + \left(\frac{\tau_T'\mathbf{G}_1\mathbf{H}_i\mathbf{u}_i^*}{T}\right)\hat{\lambda}_i.$$

Using (59) we also note that $y_{i,t-1} - \mu_i = \mathbf{s}_{t-1}'\mathbf{H}_i\mathbf{u}_i^*$, where $\mathbf{s}_{t-1}$ is a $(T+1)\times 1$ selection vector with zero elements except for its tth element which is unity. Therefore, e_{it}, and hence $\xi_{it} = \left(T^{-1}\mathbf{e}_i'\mathbf{e}_i\right)^{-1/2}e_{it}$ will be an odd function of $\mathbf{u}_i^*$, and we have $E(\xi_{it}) = 0$, $t = 1, 2, .., T$, under the assumption that $\mathbf{u}_i^*$ has a symmetric distribution. Thus, under the null hypothesis that u_{it} and u_{jt} are cross-sectionally independent we have $E(\hat{\rho}_{ij}) = 0$, and the CD test continues to hold for pure dynamic heterogeneous panel data models. Under weak cross-sectionally dependent errors it is easily seen that the conditions (54) and (55) are satisfied under (12) as N and $T \to \infty$. Finally, the CD 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. For proofs and further discussions see Pesaran (2004).[9]

7   Small Sample Evidence

In investigating the small sample properties of the CD test we consider two basic panel data regression models, a static model with a single exogenous regressor, and a dynamic second-order autoregressive specification. Both models allow for heterogeneity of slopes and error variances and include two unobserved factors for modelling different degrees of cross-sectional dependence in the errors, as measured by the maximal cross-sectional exponents of the unobserved factors.

The observations for the static panel are generated as

$$y_{it} = \delta_i + \beta_i x_{it} + u_{it}, \text{ for } i = 1, 2, ..., N; \; t = 1, 2, ..., T,$$

where $\delta_i \sim IIDN(1, 1)$,

$$x_{it} = \rho_{ix} x_{it-1} + \nu_{it}, \; i = 1, 2, ...N \text{ for } t = 1, 2, ..., T,$$

$\nu_{it} \sim IIDN(0, 1)$, and $x_{i0} = (1 - \rho_x^2)^{-1/2}\nu_{i0}$, for $i = 1, 2, ...N$. We do not expect the small sample properties of the CD test to depend on the nature of the regressors, and throughout the experiments we set $\rho_{ix} = 0.9$. We allow for heterogeneous slopes by generating them as $\beta_i \sim IIDN(1, 1)$, for $i = 1, 2, ..., N$.

The errors, u_{it}, are generated as a serially uncorrelated multi-factor process:

$$u_{it} = \gamma_{i1}f_{1t} + \gamma_{i2}f_{2t} + \varepsilon_{it},$$

with $\varepsilon_{it} \sim IIDN(0, \sigma_{i\varepsilon}^2)$, $\sigma_{i\varepsilon}^2 \sim IID \; \chi^2(2)/2$, for $i = 1, 2, ..., N$. The factors are generated as $f_{jt} \sim IIDN(0, 1)$, for $j = 1$ and 2. The factor loadings are generated as:

$$\gamma_{ji} = v_{ji}, \text{ for } i = 1, 2, ..., M_j \text{ and } j = 1, 2,$$
$$\gamma_{ji} = \rho_\gamma^{i-M_j}, \text{ for } i = M_j + 1, M_j + 2, ..., N \text{ and } j = 1, 2$$

---

[9] In the more general case where the panel data model contains lagged dependent variables as well as exogenous regressors, the symmetry of error distribution does not seem to be sufficient for the symmetry of the residuals, and the problem requires further investigation.

where $M_j = [N^{\alpha_j}]$ for $j = 1, 2$, $v_{ji} \sim IIDU(\mu_{v_j} - 0.5, \mu_{v_j} + 0.5)$. We set $\mu_{v_j} = 1$ for $j = 1, 2$. We set $\rho_\gamma = 0$, since our preliminary analysis suggested that the results are not much affected by the choice of ρ_γ, although one would expect that the performance of the CD test to deteriorate if values of ρ_γ close to unity are considered. In such cases larger sample sizes (N) are needed. Here by setting $\rho_\gamma = 0$, we are also able to consider the baseline case where the errors are cross-sectionally independent, which corresponds to setting the exponent of cross-sectional dependence to zero, $\alpha = 0$ if $\rho_\gamma = 0$. But if $\rho_\gamma \neq 0$ one does not obtain error cross-sectional independence by setting $\alpha = 0$.

We considered a one-factor as well as a two-factor specification. In the one-factor case we set

$$\alpha = (0, 0.1, 0.2, 0.25, 0.35, 0.5, 0.65, 0.75, 0.85, 0.9, 1).$$

In the two-factor case $\alpha_1 \geq \alpha_2$, and $\alpha = \max(\alpha_1, \alpha_2)$. More specifically, we set

$$(\alpha_1, \alpha_2) = \left[\begin{array}{c} (0,0), (0.1,0), (0.2,0.1), (0.25,0.15), (0.35,0.25), (0.50,0.25), \\ (0.65,0.25), (0.75,0.25), (0.85,0.25), (0.90,0.25), (1.0,0.25) \end{array} \right],$$

so that in the case of the two-factor model we also have

$$\alpha = \max(\alpha_1, \alpha_2) = (0, 0.1, 0.2, 0.25, 0.35, 0.5, 0.65, 0.75, 0.85, 0.9, 1).$$

The dynamic panel data model was generated as a second-order autoregressive process with heterogeneous slopes:

$$y_{it} = (1 - \lambda_{i1} - \lambda_{i2}) \mu_i + \lambda_{i1} y_{i,t-1} + \lambda_{i2} y_{i,t-2} + u_{it}.$$

μ_i and u_{it} were generated exactly as in the case of the static specification. The autoregressive coefficients, λ_{i1} and λ_{i2}, were generated as $\lambda_{i1} \sim IIDU(0, 0.4)$, and $\lambda_{i2} = 0.2$, for all i, and fixed across replications.

All experiments were carried out for $N = 20, 50, 100, 250, 500$ and $T = 20, 50, 100$, to evaluate the applicability of the CD test to panels where N is much larger than T. The number of replications was set to $2,000$.

The results are summarized in Tables 1 and 2 for the static and dynamic specifications, respectively. The tables give the rejection frequencies of the CD test for different values of α, sample sizes, N and T. The left panels of the tables refer to the one-factor error models and the right panels to the two-factor case. For all values of N and T the rejection frequencies are around 5% (the nominal size of the CD test) when $\alpha < 1/4$ and start to rise significantly as α approaches and exceed the 0.5 threshold, and attains its maximum of unity for $\alpha \geq 0.75$. These findings hold equally for static and dynamic models. However, at $\alpha = 1/4$, there is some evidence of over rejection (7% as compared to 5%) when N is small relative to T, namely for $N = 20$ and $T = 100$.

The Monte Carlo evidence matches the asymptotic theory remarkably well, and suggests that the test can be used fruitfully as a prelude to the estimation and inference concerning the values of α in the range $[0.70, 1]$ which are typically identified with strong factor dependence. See also Bailey, Kapetanios and Pesaran (2012).

8 Concluding Remarks

This paper provides a rigorous proof of the validity of the CD test proposed in Pesaran (2004), and further establishes that the CD test is best viewed as a test of weak cross-sectional dependence.

The null hypothesis of the CD test is shown to be $\alpha < (2 - \epsilon)/4$, where α is the exponent of cross-sectional dependence introduced in Bailey, Kapetanios and Pesaran (2012), and ϵ measures the degree to which T expands relative to N, as defined by $T = O(N^\epsilon)$, for values of $0 < \epsilon \leq 1$. It is shown that the CD test is particularly powerful against $\alpha > 1/2$, and its power rises with α and in $\sqrt{T}N$. As a test of weak cross-sectional dependence, the CD test continues to be valid under fairly general conditions even when T is small and N large. The test 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 Monte Carlo evidence reported in the paper shows that the CD statistic provides a simple and powerful test of weak cross-sectional dependence in the case of static as well as dynamic panels.

As a possible area of further research it would be interesting to investigate if the test of cross-sectional independence proposed in Hsiao, Pesaran and Pick (2012) for non-linear panel data models can also be viewed as a test of weak-cross-sectional dependence, and in particular determine the range of values of α for which the test has power.

Appendix: Proof of the Lemma 1

Using (16) and (14) we first note that (conditional on $\mathbf{X}_i$)

$$E(h_{it,T}) = 0, \; Var(h_{it,T}) = \sigma_{\zeta_i}^2 \mathbf{x}_{it}' \left(\frac{\mathbf{X}_i' \mathbf{X}_i}{T} \right)^{-1} \mathbf{x}_{it} < K < \infty, \text{ for all } i, \tag{60}$$

$$Cov(h_{it,T}, h_{jt,T}) = O(T^{-1}), \text{ for all } i \text{ and } j,$$

$$Cov(h_{it,T}, \zeta_{jt,T}) = 0, \text{ for all } i \neq j,$$

$$Cov(h_{it,T}, \zeta_{it,T}) = E\left(h_{it,T}\zeta_{it,T}\right) = -T^{-1/2}\sigma_{\zeta_i}^2 \mathbf{x}_{it}' \left(\frac{\mathbf{X}_i' \mathbf{X}_i}{T} \right)^{-1} \mathbf{x}_{it} = O\left(\frac{1}{\sqrt{T}} \right), \tag{61}$$

and by (18), we have

$$\sigma_{\zeta_i}^2 = 1 + O\left(\frac{1}{T} \right).$$

Furthermore, since conditional on $\mathbf{X}_i$, ζ_{it} and $h_{it,T}$ are cross-sectionally independent we also have

$$Var(h_{t,NT}) = N^{-1} \sum_{i=1}^{N} Var\left(h_{it,T}\right) = N^{-1} \sum_{i=1}^{N} \mathbf{x}_{it}' \left(\frac{\mathbf{X}_i' \mathbf{X}_i}{T} \right)^{-1} \mathbf{x}_{it} + O(T^{-1}) = O(1), \tag{62}$$

$$Var\left(w_{t,NT}\right) = N^{-1} \sum_{i=1}^{N} Var\left(\zeta_{it,T}\right) = 1 + O(T^{-1}) = O(1), \tag{63}$$

and similarly

$$E\left(w_{t,NT} h_{t,NT}\right) = N^{-1} \sum_{i=1}^{N} E\left(h_{it,T} \zeta_{it,T}\right) = O\left(T^{-1/2}\right), \tag{64}$$

$$Cov\left(h_{t,NT}, w_{t',NT}\right) = E\left(h_{t,NT} w_{t',NT}\right) = N^{-1} \sum_{i=1}^{N} E\left(h_{it,T} \zeta_{it'}\right) = O\left(T^{-1/2}\right), \text{ for } t \neq t',$$

$$Cov\left(w_{t,NT}, w_{t',NT}\right) = E\left(w_{t,NT} w_{t',NT}\right) = N^{-1} \sum_{i=1}^{N} E\left(\zeta_{it} \zeta_{it'}\right) = N^{-1} \sum_{i=1}^{N} E\left(\frac{\varepsilon_{it}\varepsilon_{it'}}{\varepsilon_i' \mathbf{M}_i \varepsilon_i / T} \right) = O(T^{-1}), \text{ for } t \neq t'.$$

Now let $q_{t,NT} = w_{t,NT} h_{t,NT}$, and consider the limiting properties of

$$\bar{q}_{NT} = \frac{1}{T} \sum_{t=1}^{T} q_{t,NT}.$$

First, using (64) we note that $E\left(q_{t,NT}\right) = O\left(T^{-1/2}\right)$, and $E\left(\bar{q}_{NT}\right) = O\left(T^{-1/2}\right)$. Also

$$
\begin{aligned}
Cov\left(q_{t,NT}, q_{t',NT}\right) &= E\left(w_{t,NT} h_{t,NT} w_{t',NT} h_{t',NT}\right) - E\left(w_{t,NT} h_{t,NT}\right) E\left(w_{t',NT} h_{t',NT}\right) \\
&= E\left(w_{t,NT} h_{t,NT}\right) E\left(w_{t',NT} h_{t',NT}\right) + E\left(w_{t,NT} w_{t',NT}\right) E\left(h_{t,NT} h_{t',NT}\right) \\
&\quad + E\left(w_{t,NT} h_{t',NT}\right) E\left(h_{t,NT} w_{t',NT}\right) - E\left(w_{t,NT} h_{t,NT}\right) E\left(w_{t',NT} h_{t',NT}\right) \\
&= E\left(w_{t,NT} w_{t',NT}\right) E\left(h_{t,NT} h_{t',NT}\right) + E\left(w_{t,NT} h_{t',NT}\right) E\left(h_{t,NT} w_{t',NT}\right), \text{ for } t \neq t'. \\
&= O(T^{-1}),
\end{aligned}
$$

and

$$
\begin{aligned}
E\left(q_{t,NT}^2\right) &= E\left(w_{t,NT}^2 h_{t,NT}^2\right) = N^{-2} \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{r=1}^{N}\sum_{s=1}^{N} E\left(\zeta_{it,T}\zeta_{jt,T}h_{rt,T}h_{st,T}\right) \\
&= N^{-2}\sum_{i=1}^{N} E\left(\zeta_{it,T}^2 h_{it,T}^2\right) + \left[N^{-1}\sum_{i=1}^{N} E\left(\zeta_{it,T}^2\right)\right]\left[N^{-1}\sum_{i=1}^{N} E\left(h_{it,T}^2\right)\right].
\end{aligned}
$$

But by assumption (see (18)), $E\left(\zeta_{it,T}^2 h_{it,T}^2\right) \leq \sqrt{E\left(\zeta_{it,T}^4\right)E(h_{it,T}^4)} < K < \infty$, and using (62) and (63) we have $E\left(q_{t,NT}^2\right) = O(1)$. Recalling also that $E\left(q_{NT}\right) = O\left(T^{-1/2}\right)$ it follows that $Var(q_{t,NT}) = O(1)$. Using this result in conjunction with $Cov\left(q_{t,NT}, q_{t',NT}\right) = O(T^{-1})$, established we have

$$
Var\left(\bar{q}_{NT}\right) = \frac{1}{T^2}\sum_{t=1}^{T} Var(q_{t,NT}) + \frac{2}{T^2}\sum_{t=1}^{T}\sum_{t'=1}^{T} Cov\left(q_{t,NT}, q_{t',NT}\right) = O(T^{-1}),
$$

Therefore, recalling that $E\left(\bar{q}_{NT}\right) = O\left(T^{-1/2}\right)$, we finally have

$$
\bar{q}_{NT} \to_p 0, \tag{65}
$$

as N and T tend to infinity (in any order). We also note that ,

$$
T^{-1/2}\left(N^{-1/2}\sum_{i=1}^{N}\zeta_{it}\right)\left(N^{-1/2}\sum_{i=1}^{N} h_{it,T}\right) = T^{-1/2}q_{t,TN} = O_p\left(T^{-1/2}\right). \tag{66}
$$

21

References

[1] Anselin, L. (1988), Spatial Econometrics: Methods and Models, Dorddrecht: *Kluwer Academic Publishers*.

[2] Anselin, L. (2001), "Spatial Econometrics", in B. Baltagi (ed.), *A Companion to Theoretical Econometrics*, Blackwell, Oxford.

[3] Bailey, N., G. Kapetanios, and Pesaran, M.H. (2012), "Exponent of Cross-sectional Dependence: Estimation and Inference", *University of Cambridge Working Papers in Economics* 1206, Faculty of Economics, University of Cambridge.

[4] Baltagi, B. Q. Feng, and Kao, C. (2011), "Testing for Sphericity in a Fixed Effects Panel Data Model. *The Econometrics Journal* 14, 25-47.

[5] 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.

[6] Chamberlain, G. (1983), "Funds, factors and diversification in Arbitrage pricing theory", *Econometrica* 51, 1305-1323.

[7] Chudik, A., Pesaran, M. H. and Tosetti, E. (2011), "Weak and strong cross-section dependence and estimation of large panels", *The Econometrics Journal*, 14, C45–C90.

[8] Cliff, A. and Ord, J.K. (1973), *Spatial Aurocorrection*, 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] Frees, E. W. (1995). Assessing cross-sectional correlation in panel data. *Journal of Econometrics* 69, 393–414.

[12] Haining, R.P. (2003), Spatial data Analysis: Theory and Practice, *Cambridge University Press*, Cambridge.

[13] Hsiao, C., M.H. Pesaran, Pick, A. (2012), "Diagnostic Tests of Cross-section Independence for Limited Dependent Variable Panel Data Models", *Oxford Bulletin of Economics and Statsitics*, forthcoming.

[14] Lieberman, O. (1994), "A Laplace approximation to the moments of a ratio of quadratic forms", *Biometrika* 81, 681-690.

[15] Moscone, F. and Tosetti, E. (2009), "A Review and Comparisons of Tests of Cross-Section Independence in Panels", *Journal of Economic Surveys*, 23, 528–561.

[16] Moran, P.A.P. (1948), "The Interpretation of Statistical Maps", *Biometrika*, 35, 255-60.

[17] Pesaran, M.H. (2004), "General Diagnostic Tests for Cross Section Dependence in Panels", *CESifo Working Paper Series No. 1229*; *IZA Discussion Paper No. 1240*. Available at SSRN: http://ssrn.com/abstract=572504.

[18] 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.

[19] Pesaran, M.H., A. Ullah, and T. Yamagata, (2008), "A Bias-Adjusted LM Test Of Error Cross Section Independence", *The Econometrics Journal*, 11, 105–127.

[20] Sarafidis, V., T. Yamagata, D. Robertson (2009), "A Test of Cross Section Dependence for a Linear Dynamic Panel Model with Regressors", *Journal of Econometrics*, 148, 149-161.

[21] Sarafidis, V. and T. Wansbeek, (2012), "Cross-sectional Dependence in Panel Data Analysis", *Econometric Reviews*, forthcoming.

[22] Zellner, A. (1962), "An Efficient Method for Estimating Seemingly Unrelated Regressions and Tests of Aggregation Bias", *Journal of American Statistical Association*, 58, 977-992.

Table 1: Rejection frequencies of the CD test at 5% significance level for static heterogeneous panels with one exogenous regressor

		One factor					Two factors		
α	N\T=	20	50	100	α	N\T=	20	50	100
0.00	20	0.060	0.049	0.061	0.00	20	0.056	0.057	0.054
0.10		0.060	0.049	0.061	0.10		0.056	0.057	0.054
0.20		0.060	0.049	0.061	0.20		0.056	0.057	0.054
0.25		0.082	0.063	0.091	0.25		0.071	0.072	0.080
0.35		0.082	0.063	0.091	0.35		0.080	0.086	0.097
0.50		0.263	0.428	0.664	0.50		0.236	0.364	0.598
0.65		0.831	0.996	1.000	0.65		0.781	0.973	1.000
0.75		0.988	1.000	1.000	0.75		0.981	0.999	1.000
0.85		1.000	1.000	1.000	0.85		1.000	1.000	1.000
0.90		1.000	1.000	1.000	0.90		1.000	1.000	1.000
1.00		1.000	1.000	1.000	1.00		1.000	1.000	1.000
0.00	50	0.059	0.049	0.055	0.00	50	0.063	0.051	0.055
0.10		0.059	0.049	0.055	0.10		0.063	0.051	0.055
0.20		0.069	0.054	0.067	0.20		0.074	0.060	0.056
0.25		0.069	0.054	0.067	0.25		0.074	0.060	0.056
0.35		0.083	0.074	0.096	0.35		0.085	0.078	0.086
0.50		0.383	0.546	0.756	0.50		0.297	0.576	0.716
0.65		0.907	0.997	1.000	0.65		0.883	0.995	1.000
0.75		1.000	1.000	1.000	0.75		1.000	1.000	1.000
0.85		1.000	1.000	1.000	0.85		1.000	1.000	1.000
0.90		1.000	1.000	1.000	0.90		1.000	1.000	1.000
1.00		1.000	1.000	1.000	1.00		1.000	1.000	1.000
0.00	100	0.062	0.056	0.061	0.00	100	0.048	0.048	0.049
0.10		0.062	0.056	0.061	0.10		0.048	0.048	0.049
0.20		0.064	0.062	0.066	0.20		0.054	0.054	0.052
0.25		0.069	0.071	0.079	0.25		0.054	0.060	0.060
0.35		0.094	0.125	0.178	0.35		0.078	0.107	0.135
0.50		0.306	0.646	0.886	0.50		0.302	0.647	0.774
0.65		0.955	1.000	1.000	0.65		0.960	1.000	1.000
0.75		1.000	1.000	1.000	0.75		1.000	1.000	1.000
0.85		1.000	1.000	1.000	0.85		1.000	1.000	1.000
0.90		1.000	1.000	1.000	0.90		1.000	1.000	1.000
1.00		1.000	1.000	1.000	1.00		1.000	1.000	1.000
0.00	250	0.059	0.049	0.050	0.00	250	0.055	0.052	0.052
0.10		0.059	0.049	0.050	0.10		0.055	0.052	0.052
0.20		0.061	0.052	0.053	0.20		0.061	0.058	0.056
0.25		0.061	0.052	0.053	0.25		0.059	0.059	0.055
0.35		0.085	0.084	0.090	0.35		0.075	0.079	0.095
0.50		0.317	0.541	0.816	0.50		0.302	0.464	0.841
0.65		0.994	1.000	1.000	0.65		0.996	1.000	1.000
0.75		1.000	1.000	1.000	0.75		1.000	1.000	1.000
0.85		1.000	1.000	1.000	0.85		1.000	1.000	1.000
0.90		1.000	1.000	1.000	0.90		1.000	1.000	1.000
1.00		1.000	1.000	1.000	1.00		1.000	1.000	1.000
0.00	500	0.054	0.052	0.050	0.00	500	0.061	0.060	0.058
0.10		0.054	0.052	0.050	0.10		0.061	0.060	0.058
0.20		0.056	0.053	0.047	0.20		0.062	0.055	0.058
0.25		0.056	0.053	0.050	0.25		0.064	0.064	0.063
0.35		0.074	0.081	0.087	0.35		0.081	0.091	0.091
0.50		0.402	0.690	0.849	0.50		0.350	0.622	0.827
0.65		1.000	1.000	1.000	0.65		1.000	1.000	1.000
0.75		1.000	1.000	1.000	0.75		1.000	1.000	1.000
0.85		1.000	1.000	1.000	0.85		1.000	1.000	1.000
0.90		1.000	1.000	1.000	0.90		1.000	1.000	1.000
1.00		1.000	1.000	1.000	1.00		1.000	1.000	1.000

α is maximal cross-sectional exponent of the errors u_{it} in the panel data model
$y_{it} = \mu_i + \beta_i x_{it} + u_{it}$, $u_{it} = \gamma_{i1} f_{1t} + \gamma_{i2} f_{2t} + \sigma_{i\varepsilon} \varepsilon_{it}$, $i = 1, ..., N$, $t = 1, ..., T$.
$\alpha = \max(\alpha_j)$, where α_j corresponds to the rate at which $\sum_{i=1}^{N} \gamma_{ij}^2$ rises with N $(O(N^{\alpha_j}))$,
for $j = 1, 2$ factors.

Table 2: Rejection frequencies of the CD test at 5% significance level for AR(2) heterogeneous panels

		One factor					Two factors		
α	N\T=	20	50	100	α	N\T=	20	50	100
0.00	20	0.058	0.052	0.050	0.00	20	0.051	0.047	0.056
0.10		0.058	0.052	0.050	0.10		0.051	0.047	0.056
0.20		0.058	0.052	0.050	0.20		0.051	0.047	0.056
0.25		0.070	0.072	0.076	0.25		0.059	0.071	0.088
0.35		0.070	0.072	0.076	0.35		0.064	0.078	0.115
0.50		0.213	0.370	0.577	0.50		0.188	0.345	0.583
0.65		0.809	0.993	0.999	0.65		0.758	0.979	0.999
0.75		0.977	1.000	1.000	0.75		0.971	1.000	1.000
0.85		0.999	1.000	1.000	0.85		0.999	1.000	1.000
0.90		1.000	1.000	1.000	0.90		1.000	1.000	1.000
1.00		1.000	1.000	1.000	1.00		1.000	1.000	1.000
0.00	50	0.055	0.045	0.049	0.00	50	0.047	0.051	0.055
0.10		0.055	0.045	0.049	0.10		0.047	0.051	0.055
0.20		0.059	0.052	0.057	0.20		0.049	0.054	0.061
0.25		0.059	0.052	0.057	0.25		0.049	0.054	0.061
0.35		0.080	0.070	0.087	0.35		0.062	0.076	0.092
0.50		0.261	0.599	0.771	0.50		0.217	0.492	0.750
0.65		0.815	0.997	1.000	0.65		0.758	0.996	1.000
0.75		0.997	1.000	1.000	0.75		0.995	1.000	1.000
0.85		1.000	1.000	1.000	0.85		1.000	1.000	1.000
0.90		1.000	1.000	1.000	0.90		1.000	1.000	1.000
1.00		1.000	1.000	1.000	1.00		1.000	1.000	1.000
0.00	100	0.059	0.046	0.051	0.00	100	0.054	0.050	0.054
0.10		0.059	0.046	0.051	0.10		0.054	0.050	0.054
0.20		0.065	0.050	0.051	0.20		0.059	0.054	0.053
0.25		0.067	0.051	0.061	0.25		0.068	0.067	0.066
0.35		0.098	0.089	0.139	0.35		0.094	0.113	0.155
0.50		0.319	0.523	0.860	0.50		0.264	0.610	0.867
0.65		0.957	1.000	1.000	0.65		0.866	1.000	1.000
0.75		1.000	1.000	1.000	0.75		1.000	1.000	1.000
0.85		1.000	1.000	1.000	0.85		1.000	1.000	1.000
0.90		1.000	1.000	1.000	0.90		1.000	1.000	1.000
1.00		1.000	1.000	1.000	1.00		1.000	1.000	1.000
0.00	250	0.060	0.047	0.044	0.00	250	0.056	0.048	0.052
0.10		0.060	0.047	0.044	0.10		0.056	0.048	0.052
0.20		0.057	0.053	0.047	0.20		0.059	0.055	0.055
0.25		0.057	0.053	0.047	0.25		0.059	0.052	0.055
0.35		0.076	0.081	0.088	0.35		0.071	0.077	0.089
0.50		0.311	0.577	0.808	0.50		0.254	0.525	0.851
0.65		0.996	1.000	1.000	0.65		0.992	1.000	1.000
0.75		1.000	1.000	1.000	0.75		1.000	1.000	1.000
0.85		1.000	1.000	1.000	0.85		1.000	1.000	1.000
0.90		1.000	1.000	1.000	0.90		1.000	1.000	1.000
1.00		1.000	1.000	1.000	1.00		1.000	1.000	1.000
0.00	500	0.050	0.047	0.053	0.00	500	0.049	0.050	0.050
0.10		0.050	0.047	0.053	0.10		0.049	0.050	0.050
0.20		0.052	0.046	0.056	0.20		0.049	0.050	0.053
0.25		0.058	0.047	0.060	0.25		0.050	0.052	0.055
0.35		0.066	0.077	0.099	0.35		0.064	0.069	0.083
0.50		0.308	0.571	0.803	0.50		0.313	0.533	0.789
0.65		0.998	1.000	1.000	0.65		0.998	1.000	1.000
0.75		1.000	1.000	1.000	0.75		1.000	1.000	1.000
0.85		1.000	1.000	1.000	0.85		1.000	1.000	1.000
0.90		1.000	1.000	1.000	0.90		1.000	1.000	1.000
1.00		1.000	1.000	1.000	1.00		1.000	1.000	1.000

α is maximal cross-sectional exponent of the errors u_{it} in the panel data model
$y_{it} = (1 - \lambda_{i1} - \lambda_{i2}) \mu_i + \lambda_{i1} y_{i,t-1} + \lambda_{i2} y_{i,t-2} + u_{it}$, $u_{it} = \gamma_{i1} f_{1t} + \gamma_{i2} f_{2t} + \sigma_{i\varepsilon} \varepsilon_{it}$,
$i = 1, ..., N$, $t = 1, ..., T$. $\alpha = \max(\alpha_j)$, where α_j corresponds to the rate at which
$\sum_{i=1}^{N} \gamma_{ij}^2$ rises with N ($O(N^{\alpha_j})$), for $j = 1, 2$ factors.