Diagnostic Tests of Cross Section Independence for Limited Dependent Variable Panel Data Models*

Cheng Hsiao
University of Southern California, and
Wang Yanan Institute for Studies in Economics
Xiamen University

M. Hashem Pesaran
University of Cambridge, CIMF, and USC

Andreas Pick
Erasmus University Rotterdam,
De Nederlandsche Bank, and CIMF

July 28, 2010

Abstract

This paper considers the problem of testing for cross section independence in limited dependent variable panel data models. It derives a Lagrangian multiplier (LM) test and shows that in terms of generalized residuals of Gourieroux, Monfort, Renault and Trognon (1987) it reduces to the LM test of Breusch and Pagan (1980). Due to the tendency of the LM test to over-reject in panels with large N (cross section dimension), we also consider the application of the cross section dependence test (CD) proposed by Pesaran (2004). In Monte Carlo experiments it emerges that for most combinations of N and T the CD test is correctly sized, whereas the validity of the LM test requires T (time series dimension) to be quite large relative to N. We illustrate the cross-sectional independence tests by an application to a probit panel of roll-call votes in the U. S. Congress and find that the votes display a significant degree of cross section dependence.
JEL: C12, C33, C35
Keywords: Nonlinear panels, cross section dependence, probit and Tobit models

*We would like to thank Qi Li, the editor (Anindya Banerjee), and two anonymous referees for helpful comments. Most of the research for this paper was done while the third author was at the University of Cambridge. He acknowledges financial support from Sinopia, quantitative specialist of HSBC Global Asset Management. The opinions expressed in this paper do not necessarily represent those of DNB.

1 Introduction

Many panel data models assume that observations across individuals are independent. However, there could be common shocks that affect all individuals but with differing degrees. Often economic theories also predict that agents take actions that lead to interdependence among themselves. For example, the prediction that risk-averse agents will make insurance contracts allowing them to smooth idiosyncratic shocks implies dependence in consumption across individuals. If observations are dependent across individuals, estimators that are based on the assumption of cross sectional independence may be inconsistent. Since contrary to time series data, there is no natural ordering for cross sectional indices, i, appropriate modeling and estimation of cross sectional dependence can be difficult, in particular if the dimension of cross sectional observations, N, is large and the time series dimension, T, is small. Therefore, it is prudent to first test for cross sectional dependence before embarking on estimation and inference.[1]

A popular approach to test for cross sectional independence is to prespecify the strength of cross-sectional correlation through an $N \times N$ spatial weighting matrix $\boldsymbol{W}$, and then test if the proportional factor, ρ, is equal to zero, see Moran (1948) and Kelejian and Prucha (2001). Under the null of cross-sectional independence, ρ is equal to zero for any choice of $\boldsymbol{W}$. However, the power of the test will be sensitive to the specification of $\boldsymbol{W}$. An alternative approach that does not depend on a particular choice of the spatial weighting matrix is the Lagrangian multiplier (LM) test proposed by Breusch and Pagan (1980) to test the diagonality of the error covariance matrix of a seemingly unrelated equation system. In the context of linear panels, the LM statistic is N times the average of the squared pair-wise correlation coefficients of the residuals. However, there is a fundamental difference between limited dependent variable type models and linear models. There is a one-to-one correspondence between the residuals and the observed variables in linear models, which is absent for limited dependent variable models.

In this paper, we derive an LM test of cross section independence for limited dependent variable panel data models and show that an asymptotic version of the test can be written as the LM test of Breusch and Pagan, but in terms of the generalized residuals defined by Gourieroux, Monfort, Renault and Trognon (1987). However, as noted by Pesaran (2004) and further elaborated by Pesaran, Ullah and Yamagata (2008), when N is large relative

[1] In a recent paper, Ng (2006) employs spacing variance ratio statistics to test the severity of cross section correlation in panels by partitioning the pair-wise cross-correlations into groups from high to low. The proposed statistics are intended as agnostic tools for identifying and characterizing correlations across groups. However, they cannot be used as diagnostic tests of cross section independence that underlie the standard analysis of panel data.

to T the LM test is not correctly centered and can lead to serious over-reject of the null of cross section independence. To deal with this problem Pesaran, Ullah and Yamagata (2008) propose a bias-adjusted version of the LM test for linear models that seem to work well when the errors are normally distributed and the regressors are strictly exogenous. However, if the model is nonlinear, it does not appear feasible to derive the exact mean and variance of the squared pair-wise correlation coefficients of the generalized residuals.

As an alternative to the test based on the square of the error correlation coefficients, Pesaran (2004) proposes a cross-sectional dependence test that uses the simple average of all pair-wise correlation coefficients (CD test), which is closely related to the C_{AVE} by Frees (1995). Pesaran shows that for the linear model the CD test is correctly centered for fixed N and T under the null of cross section independence assuming that the errors are symmetrically distributed. We explore the use of the CD test in the limited dependent variable model and derive asymptotic results for non-linear models with additive error terms. Using Monte Carlo experiments we find that the CD test performs well even in cases where N and T are relatively small. The CD test shows some size distortion only in the cases where $N = 500$ and T relatively small around 20 or less. In contrast, the size distortion of the LM test is significant except for the case when T is much larger than N.

Since for some non-linear models it is rather complicated to derive the generalized residual, in the Monte Carlo experiments we also consider using in-sample forecast errors, standardized to allow for the heteroskedastic nature of such errors in limited dependent variable models. Following McCullagh and Nelder (1989), we refer to the latter as the Pearson residual. We investigate the small sample performance of the LM and CD tests using both types of residuals and find that there is little to choose between the generalized and the Pearson residuals.

Finally, we illustrate the CD test by an application to a probit panel data model of the voting behavior of the members of the U. S. Congress previously analyzed by Wawro (2001). The explanatory variables considered are campaign contributions of business and labor lobby groups and the unemployment rate in the constituency of the voting member. We find clear evidence against the null hypothesis of cross section independence. This result is confirmed using bootstrap critical values for the CD test.

The rest of the paper is set out as follows. Section 2 introduces the limited dependent variable panel data model. Section 3 discusses the tests for cross section independence. The small sample performance of the tests are evaluated using Monte Carlo experiments in Section 4. Section 5 illustrates the use of the tests in the empirical application. Section 6 provides some concluding remarks. Technical details of some of the derivations are provided in Appendices A to C. Appendix D describes the bootstrap procedure used to approximate the finite sample distribution of the CD test in the empirical application.

2 The limited dependent variable panel data model

Suppose that the latent variable, y_{it}^*, is generated by the following panel data model,

$$\mathrm{f}(y_{it}^*, \boldsymbol{x}_{it}, \boldsymbol{\theta}_i) = \varepsilon_{it}, \quad \text{for } i = 1, 2, \ldots, N, \; t = 1, 2, \ldots, T, \tag{1}$$

where $\boldsymbol{x}_{it}$ is a $k \times 1$ vector of exogenous variables, $\boldsymbol{\theta}_i$ is a $q \times 1$ vector of unknown parameters that may or may not be identical across i, ε_{it} is a scalar disturbance, N is the number of cross section observations, and T is the number of time series observations. The variable y_{it} is observed, which is related to the latent variable, y_{it}^*, via the link function $\mathrm{g}(\cdot)$,

$$y_{it} = \mathrm{g}(y_{it}^*). \tag{2}$$

This specification encompasses many econometric models. Examples include the probit model where

$$\mathrm{f}(y_{it}^*, \boldsymbol{x}_{it}, \boldsymbol{\beta}_i) = y_{it}^* - \boldsymbol{\beta}_i' \boldsymbol{x}_{it} = \varepsilon_{it}, \tag{3}$$

ε_{it} follows a standard normal distribution, and

$$\mathrm{g}(y_{it}^*) = \mathrm{I}(y_{it}^*), \tag{4}$$

where $\mathrm{I}(A)$ is the indicator function which is unity if $A > 0$ and zero otherwise.

The Tobit model is obtained if the latent model is that of equation (3), errors are normal and the link function is

$$\mathrm{g}(y_{it}^*) = y_{it}^* \mathrm{I}(y_{it}^*). \tag{5}$$

In what follows we assume that $\boldsymbol{\varepsilon}_t$, $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{Nt})'$, conditional on $\boldsymbol{x}_{it}$ is independently distributed over time with mean $\mathbf{0}$ and the covariance matrix $\boldsymbol{\Sigma}$, and focus on testing $\boldsymbol{\Sigma} = \boldsymbol{D}$ against $\boldsymbol{\Sigma} \neq \boldsymbol{D}$, where $\boldsymbol{D}$ is a diagonal matrix. A non-diagonal $\boldsymbol{\Sigma}$ could arise, for example, from the presence of unobserved common factors

$$\varepsilon_{it} = \boldsymbol{\gamma}_i' \boldsymbol{f}_t + e_{it}, \tag{6}$$

where $\boldsymbol{\gamma}_i$ is the vector of factor loadings, $\boldsymbol{f}_t \sim \text{iid}\,(\mathbf{0}, \boldsymbol{\Sigma}_f)$, and $e_{it} \sim \text{iid}\,(0, \sigma_{ie}^2)$.

3 Testing for cross-sectional independence

Let $\boldsymbol{y}_t = (y_{1t}, y_{2t}, \ldots, y_{Nt})'$, and denote the probability of observing $\boldsymbol{y}_t$, by P_t. We have

$$P_t = \int_{A(\boldsymbol{\varepsilon}_t|\boldsymbol{y}_t)} f(\boldsymbol{\varepsilon}_t) d\boldsymbol{\varepsilon}_t, \tag{7}$$

where

$$f(\boldsymbol{\varepsilon}_t) = (2\pi)^{-\frac{N}{2}} \mid \boldsymbol{\Sigma} \mid^{-\frac{1}{2}} \exp\{-\frac{1}{2}(\boldsymbol{\varepsilon}_t' \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}_t)\},$$

and $A(\boldsymbol{\varepsilon}_t|\boldsymbol{y}_t)$ denotes the region of integration which is determined by the realized $\boldsymbol{y}_t$ and the form of the link function. For instance, in the case of the iid probit model $A(\boldsymbol{\varepsilon}_t|\boldsymbol{y}_t)$ denotes the region $\{a_{it} < \varepsilon_{it} < b_{it}\}$. When $y_{it}^* = \boldsymbol{\beta}_i' \boldsymbol{x}_{it} + \varepsilon_{it}$ and $y_{it} = 1$ then $a_{it} = -\boldsymbol{\beta}_i' \boldsymbol{x}_{it}, b_{it} = \infty$, and when $y_{it} = 0$ then $a_{it} = -\infty$, and $b_{it} = -\boldsymbol{\beta}_i' \boldsymbol{x}_{it}$.

Cross-sectional independence implies that all off-diagonal elements of $\boldsymbol{\Sigma}$ are zero under joint normality assumption. Since the variance of each cross-sectional unit is just a scale factor and does not affect the limiting distribution of the LM statistic or its variant, for ease of notation we set the variance of ε_{it} to 1, $\text{Var}(\varepsilon_{it}) = \sigma_i^2 = 1$, then $\boldsymbol{\Sigma} = \boldsymbol{R}$ where the diagonal elements of $\boldsymbol{R}$ are all equal to 1 and the off-diagonal elements, ρ_{ij}, denote the correlation coefficients between the two errors, ε_{it} and ε_{jt}. Then under the null hypothesis of cross-sectional independence,

$$H_0 : \boldsymbol{R} = \boldsymbol{I}_N,$$

where $\boldsymbol{I}_N$ is an identity matrix of order N, and the alternative is

$$H_1 : \boldsymbol{R} \neq \boldsymbol{I}_N.$$

It will be assumed that $A(\boldsymbol{\varepsilon}_t|\boldsymbol{y}_t)$ does not depend on $\boldsymbol{R}$.

3.1 The Lagrange-multiplier test

In the case where N is fixed and $T \to \infty$, the test of H_0 can be based on the Lagrange multiplier (LM) statistic defined by,

$$\begin{aligned} \text{LM} \quad = \quad & \left(T^{-1/2} \left. \frac{\partial \ell_{NT}(\boldsymbol{\rho})}{\partial \boldsymbol{\rho}} \right|_{\mathbf{R}=\mathbf{I}_N} \right)' \left(-\text{E} \left[T^{-1} \left. \frac{\partial^2 \ell_{NT}(\boldsymbol{\rho})}{\partial \boldsymbol{\rho} \partial \boldsymbol{\rho}'} \right|_{\mathbf{R}=\mathbf{I}_N} \right] \right)^{-1} \\ & \left(T^{-1/2} \left. \frac{\partial \ell_{NT}(\boldsymbol{\rho})}{\partial \boldsymbol{\rho}} \right|_{\mathbf{R}=\mathbf{I}_N} \right), \end{aligned} \tag{8}$$

where ℓ_{NT} is the log likelihood of $\boldsymbol{y}_t$, which is given by

$$\ell_{NT} = \sum_{t=1}^{T} \ln \; P_t. \tag{9}$$

Taking the first partial derivatives of the log likelihood in (9) with respect to ρ_{ij} yields

$$\begin{aligned} \frac{\partial \ell_{NT}}{\partial \rho_{ij}} &= \sum_{t=1}^{T} \frac{1}{P_t} \cdot \frac{\partial P_t}{\partial \rho_{ij}} \\ \frac{\partial P_t}{\partial \rho_{ij}} &= \int_{A(\boldsymbol{\varepsilon}_t|\boldsymbol{y}_t)} \frac{\partial \ln f(\boldsymbol{\varepsilon}_t)}{\partial \rho_{ij}} f(\boldsymbol{\varepsilon}_t) d\boldsymbol{\varepsilon}_t, \end{aligned} \tag{10}$$

where

$$\frac{\partial \ln f(\boldsymbol{\varepsilon}_t)}{\partial \rho_{ij}} = -\frac{1}{2}\mathrm{tr}(\boldsymbol{R}^{-1}\boldsymbol{A}_{ij}) + \frac{1}{2}\boldsymbol{\varepsilon}_t' \boldsymbol{R}^{-1}\boldsymbol{A}_{ij}\boldsymbol{R}^{-1}\boldsymbol{\varepsilon}_t, \tag{11}$$

and $\boldsymbol{A}_{ij}$ is an $N \times N$ matrix with all the elements equal to zero except for the (i, j) and (j, i) elements which are equal to 1.

Under the null of $\boldsymbol{R} = \boldsymbol{I}_N$, $f(\boldsymbol{\varepsilon}_t) = \prod_{i=1}^{N} \phi(\varepsilon_{it})$, where $\phi(\cdot)$ denotes the standard normal density, $\partial \ln f(\boldsymbol{\varepsilon}_t)/\partial \rho_{ij}|_{\mathbf{R}=\mathbf{I}_N} = \varepsilon_{it}\varepsilon_{jt}$, and (note that $\mathrm{tr}(\boldsymbol{A}_{ij}) = 0$)

$$\frac{1}{P_t} \cdot \frac{\partial P_t}{\partial \rho_{ij}}\bigg|_{\mathbf{R}=\mathbf{I}_N} = u_{it}u_{jt}, \tag{12}$$

where u_{it} is the conditional expectation of ε_{it} given y_{it} and $\boldsymbol{x}_{it}$, namely,

$$u_{it} = \frac{\int_{a_{it}}^{b_{it}} \varepsilon_{it}\phi(\varepsilon_{it})d\varepsilon_{it}}{\int_{a_{it}}^{b_{it}} \phi(\varepsilon_{it})d\varepsilon_{it}}. \tag{13}$$

Hence

$$\frac{\partial \ell_{NT}}{\partial \rho_{ij}}\bigg|_{\mathbf{R}=\mathbf{I}_N} = \sum_{t=1}^{T} u_{it}u_{jt}. \tag{14}$$

In general, u_{it} depends on the unknown parameters entering the expressions for a_{it} and b_{it}. When these parameters are replaced by their estimators (say $\tilde{a}_{it}$ and $\tilde{b}_{it}$) the estimate of u_{it} which we denote by $\tilde{u}_{it}$ is the so called generalized residual, $\mathrm{E}(\varepsilon_{it} \mid y_{it}, \tilde{a}_{it}, \tilde{b}_{it})$, originally introduced by Gourieroux, Monfort, Renault and Trognon (1987, GMRT). To avoid confusion we refer to u_{it} as the generalized error.

The exact expression for u_{it} clearly depends on the form of the link function. For instance, in the case of the probit model

$$\begin{aligned} u_{it} &= \frac{\phi(a_{it}) - \phi(b_{it})}{\Phi(b_{it}) - \Phi(a_{it})}, \\ &= \frac{\phi(\boldsymbol{\beta}_i'\boldsymbol{x}_{it})}{\Phi(\boldsymbol{\beta}_i'\boldsymbol{x}_{it})[1 - \Phi(\boldsymbol{\beta}_i'\boldsymbol{x}_{it})]}[y_{it} - \Phi(\boldsymbol{\beta}_i'\boldsymbol{x}_{it})] \end{aligned} \tag{15}$$

and $\Phi(\cdot)$ denotes the integrated standard normal. See Appendix A for details. In the case of the Tobit model we have

$$u_{it} = (y_{it} - \boldsymbol{\beta}_i'\boldsymbol{x}_{it})\mathrm{I}(y_{it}) - \sigma_i \frac{\phi(\boldsymbol{\beta}_i'\boldsymbol{x}_{it}/\sigma_{iu})}{\Phi(-\boldsymbol{\beta}_i'\boldsymbol{x}_{it}/\sigma_{iu})}[1 - \mathrm{I}(y_{it})], \tag{16}$$

where σ_{iu} is the standard deviation of the error term (Chesher and Irish, 1987).

Using the results in Appendix B we have that

$$\left.\frac{\partial^2 \ell_{NT}}{\partial \rho_{ij}^2}\right|_{\boldsymbol{R}=\boldsymbol{I}_N} = -\sum_{t=1}^{T} u_{it}^2 u_{jt}^2 + \sum_{t=1}^{T} \left(1-\eta_{it}^2\right)\left(1-\eta_{jt}^2\right),$$

where

$$\eta_{it}^2 = \frac{\int_{a_{it}}^{b_{it}} \varepsilon_{it}^2 \phi(\varepsilon_{it}) d\varepsilon_{it}}{\int_{a_{it}}^{b_{it}} \phi(\varepsilon_{it}) d\varepsilon_{it}} = 1 - \left[\frac{b_{it}\phi(b_{it}) - a_{it}\phi(a_{it})}{\Phi(b_{it}) - \Phi(a_{it})}\right].$$

For example, in the case of the probit model (where $a_{it} = -\boldsymbol{\beta}_i'\boldsymbol{x}_{it}, b_{it} = \infty$, if $y_{it} = 1$ and $a_{it} = -\infty$, and $b_{it} = -\boldsymbol{\beta}_i'\boldsymbol{x}_{it}$, if $y_{it} = 0$), we have $1 - \eta_{it}^2 = u_{it}\boldsymbol{\beta}_i'\boldsymbol{x}_{it}$. Therefore, when $N = 2$ the correlation matrix $\boldsymbol{R}$ is a 2×2 matrix and testing the null hypothesis reduces to testing the off-diagonal element $\rho_{12} = 0$, where

$$-\frac{1}{T}\frac{\partial^2 \ell_{NT}}{\partial \rho_{12}^2} = \frac{1}{T}\sum_{t=1}^{T} u_{1t}^2 u_{2t}^2 - \frac{1}{T}\sum_{t=1}^{T} \left(\boldsymbol{\beta}_1'\boldsymbol{x}_{1t}\boldsymbol{x}_{2t}'\boldsymbol{\beta}_2\right) u_{1t}u_{2t}. \tag{17}$$

When $N = 3$ the test involves the vector of correlation coefficients, $\boldsymbol{\rho} = (\rho_{12}, \rho_{13}, \rho_{23})$. The Hessian matrix is now a full 3×3 matrix with diagonal elements

$$-\frac{1}{T}\frac{\partial^2 \ell_{NT}}{\partial \rho_{ij}^2} = \frac{1}{T}\sum_{t=1}^{T} u_{it}^2 u_{jt}^2 - \frac{1}{T}\sum_{t=1}^{T} u_{it}u_{jt}\left(\boldsymbol{\beta}_i'\boldsymbol{x}_{it}\boldsymbol{x}_{jt}'\boldsymbol{\beta}_j\right) \text{ for } i \neq j$$

and the off-diagonal elements

$$\begin{aligned} -\frac{1}{T}\frac{\partial^2 \ell_{NT}}{\partial \rho_{12}\partial \rho_{13}} &= -\frac{1}{T}\sum_{t=1}^{T} u_{2t}u_{3t}, \quad -\frac{1}{T}\frac{\partial^2 \ell_{NT}}{\partial \rho_{12}\partial \rho_{23}} = -\frac{1}{T}\sum_{t=1}^{T} u_{1t}u_{3t}, \\ -\frac{1}{T}\frac{\partial^2 \ell_{NT}}{\partial \rho_{13}\partial \rho_{23}} &= -\frac{1}{T}\sum_{t=1}^{T} u_{1t}u_{2t}. \end{aligned}$$

Similarly, using the above results the Hessian matrix can be set up for any fixed N.

A number of different versions of the LM test can be constructed, depending on how the various terms in $T^{-1} \left.\frac{\partial^2 \ell_{NT}(\boldsymbol{\rho})}{\partial \boldsymbol{\rho}\partial \boldsymbol{\rho}'}\right|_{\boldsymbol{R}=\boldsymbol{I}_N}$ are evaluated/estimated. For example, in Section 3.7 of their paper GMRT consider the case of $N = 2$, leave the term $T^{-1}\sum_{t=1}^{T} u_{1t}^2 u_{2t}^2$ as is but (implicitly) replace the second term in (17) , namely $-T^{-1}\sum_{t=1}^{T}\left(\boldsymbol{\beta}_1'\boldsymbol{x}_{1t}\boldsymbol{x}_{2t}'\boldsymbol{\beta}_2\right) u_{1t}u_{2t}$, by its asymptotic value of zero under H_0. Hence suggest the following score statistic

$$\text{LM}_{\text{GMRT}} = \frac{\left(T^{-1/2}\sum_{t=1}^{T} \tilde{u}_{1t}\tilde{u}_{2t}\right)^2}{T^{-1}\sum_{t=1}^{T} \tilde{u}_{1t}^2 \tilde{u}_{2t}^2}, \tag{18}$$

where $\tilde{u}_{it}$ for $i = 1$ and 2 are the generalized residuals computed under the null hypothesis, see also Chesher and Irish (1987).

However, for the general case it is more convenient to replace the Hessian matrix by its expectations (or probability limit) under the null hypothesis. Since for each i, u_{it} is serially uncorrelated and under H_0, $\mathrm{E}(u_{it}u_{jt}) = 0$ for all $i \neq j$, it then follows that

$$\underset{\mathrm{T}\to\infty}{\mathrm{plim}} \left(T^{-1} \sum_{t=1}^{T} u_{it}u_{jt} \right) = 0, \text{ for } i \neq j.$$

Also,

$$\underset{\mathrm{T}\to\infty}{\mathrm{plim}} \left(T^{-1} \frac{\partial^2 \ell_{NT}}{\partial \rho_{ij}^2} \right) = \sigma_{iu}^2 \sigma_{ju}^2, \text{ for } i \neq j,$$

where $\sigma_{iu}^2 = \mathrm{p}\lim_{\mathrm{T}\to\infty} T^{-1} \sum_{t=1}^{T} u_{it}^2 = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \mathrm{E}(u_{it}^2)$. Hence, $\underset{T\to\infty}{\mathrm{plim}} \left[-T^{-1} \left. \frac{\partial^2 \ell_{NT}(\boldsymbol{\rho})}{\partial \boldsymbol{\rho} \partial \boldsymbol{\rho}'} \right|_{\boldsymbol{R}=\boldsymbol{I}_N} \right]$ reduces to an $N(N-1)/2$ by $N(N-1)/2$ diagonal matrix with the elements

$$\sigma_{1u}^2\sigma_{2u}^2, \sigma_{1u}^2\sigma_{3u}^2, \ldots, \sigma_{1u}^2\sigma_{Nu}^2; \sigma_{2u}^2\sigma_{3u}^2, \sigma_{2u}^2\sigma_{4u}^2, \ldots, \sigma_{2u}^2\sigma_{Nu}^2; \ldots; \sigma_{N-1,u}^2\sigma_{Nu}^2.$$

In the construction of the LM test, σ_{iu}^2, can be consistently estimated by $T^{-1}\sum_{t=1}^{T} \tilde{u}_{it}^2$.

Proposition 1 *Consider models of the form* (1) *and* (2)*. When the* $N \times 1$ *vector* $\boldsymbol{\varepsilon}_t$ *is independently distributed with mean* $\mathbf{0}$ *and covariance matrix* $\boldsymbol{I}_N$*,* $\sqrt{T}\tilde{\rho}_{ij} \to \mathrm{N}(0,1)$*, where* $\tilde{\rho}_{ij}$ *is the Pearson correlation coefficient computed using the generalized residuals estimated under the null hypothesis, namely*

$$\tilde{\rho}_{ij} = \frac{T^{-1}\sum_{t=1}^{T} \tilde{u}_{it}\tilde{u}_{jt}}{\sqrt{T^{-1}\sum_{t=1}^{T} \tilde{u}_{it}^2}\sqrt{T^{-1}\sum_{t=1}^{T} \tilde{u}_{jt}^2}}, \tag{19}$$

In the case where N *is fixed and* $T \to \infty$*, the LM statistic*

$$\mathrm{LM} = T \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \tilde{\rho}_{ij}^2 \tag{20}$$

converges to a χ^2*-distribution with* $N(N-1)/2$ *degrees of freedom.*

The LM test, being based on the likelihood function, has a number of attractive features and is consistent against a wide class of alternatives. However, a typical panel data set has N much larger than T. While for large N the scaled LM statistic, $\sqrt{\frac{2}{N(N-1)}}\mathrm{LM}$, continues to be asymptotically normally distributed as long as $\tilde{u}_{it}$ is independently distributed across i,

$\mathrm{E}(T\tilde{\rho}_{ij}^2) \neq 0$ for all T. As a result, the scaled LM statistic will not be properly centered when N is large relative to T. Results from our Monte Carlo experiments reported in Section 4 confirm this. As a result, in panels where $N > T$, the LM test tends to over-reject, often substantially.

To correct for the bias in large N and finite T panels in the case of linear models, Pesaran, Ullah and Yamagata (2008) propose the following bias-adjusted version of the LM test

$$\mathrm{NLM}_{\mathrm{adj}} = \sqrt{\frac{2}{N(N-1)}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{(T-k)\tilde{\rho}_{ij}^2 - \mu_{Tij}}{\upsilon_{Tij}}, \tag{21}$$

where $\mu_{Tij} = \mathrm{E}\left[(T-k)\,\tilde{\rho}_{ij}^2\right]$ and $\upsilon_{Tij}^2 = \mathrm{Var}\left[(T-k)\,\tilde{\rho}_{ij}^2\right]$. It is shown that this statistic is asymptotically distributed as $\mathrm{N}(0,1)$ for all $T > k+8$ and as $N \to \infty$, where k is the number of regressors in the model. These authors are able to derive exact expressions for μ_{Tij} and υ_{Tij}^2 when the regressors are strictly exogenous and the errors are normally distributed. In principal, the same procedure can be applied to nonlinear panels but an exact analytical derivation of μ_{Tij} and υ_{Tij}^2 does not seem possible. Numerical techniques can be used, but could be quite time consuming and will not be pursued in this paper.

3.2 CD test

Given the over-rejection of the LM test when N is large and the complications of the bias-adjusted LM test, we consider the CD test proposed by Pesaran (2004) which is shown to have desirable small sample properties in the context of linear models.

Proposition 2 *Consider models of the form* (1) *and* (2). *When the* $N \times 1$ *vector* $\boldsymbol{\varepsilon}_t$ *is independently distributed with mean* $\mathbf{0}$ *and covariance matrix* $\boldsymbol{I}_N$, $\sqrt{T}\tilde{\rho}_{ij} \to \mathrm{N}(0,1)$, *where* $\tilde{\rho}_{ij}$ *is the Pearson correlation coefficient computed using the generalized residuals estimated under the null hypothesis given in* (19). *When* $N, T \to \infty$, *the CD statistic*

$$CD = \sqrt{\frac{2T}{N(N-1)}} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \tilde{\rho}_{ij} \right), \tag{22}$$

converges to a standard normal distribution, $N(0,1)$.

We provide a proof for a general nonlinear model in Appendix C. The power of the CD test approaches unity if the average of the pair-wise correlation coefficients is different from zero. LM tests, on the other hand, will have power approaching one without requiring the average of the cross-correlations to be non-zero. Thus there is a trade-off between LM and CD tests. The LM test tends to over-reject if N is large but has power against a wider class of alternatives than the CD test. On the other hand, we expect the CD test to have little size distortions even for large N, but it is likely to lack power if

$$\lim_{N\to\infty}\left(\frac{2}{N(N-1)}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\rho_{ij}\right)=0.$$

3.3 Pearson residuals

The above statistics are defined in terms of the generalized residuals, that is, conditional expectations of ε_{it} given y_{it}. For some nonlinear models $\mathrm{E}(\varepsilon_{it}|y_{it})$ may be tedious to derive. An alternative will be simply to define the residual as the deviation between y_{it} and its conditional expectation (evaluated at $\boldsymbol{\theta}_{it}=\tilde{\boldsymbol{\theta}}_{it}$)

$$\tilde{v}_{it}=y_{it}-\mathrm{E}\Big(y_{it}|\boldsymbol{x}_{it},\tilde{\boldsymbol{\theta}}_{it}\Big).$$

Given that for many models the residuals of this form will not be homoskedastic, one should consider standardizing these residuals. For the probit model, for example, the standardized version of $\tilde{v}_{it}$ is

$$\tilde{v}_{it}^{*}=\frac{y_{it}-\Phi(\tilde{\boldsymbol{\beta}}_i'\boldsymbol{x}_{it})}{\sqrt{\Phi(\tilde{\boldsymbol{\beta}}_i'\boldsymbol{x}_{it})(1-\Phi(\tilde{\boldsymbol{\beta}}_i'x_{it}))}},\tag{23}$$

where $\tilde{\boldsymbol{\beta}}_i$ is the ML estimator of β_i under the null hypothesis. Following McCullagh and Nelder (1989, p.37), we shall refer to v_{it}^{*}, generically defined by $\tilde{v}_{it}^{*}=\Big[y_{it}-\mathrm{E}\Big(y_{it}|\boldsymbol{x}_{it},\tilde{\boldsymbol{\theta}}_{it}\Big)\Big]/\sqrt{\mathrm{Var}\left(y_{it}|\boldsymbol{x}_{it},\tilde{\boldsymbol{\theta}}_{it}\right)}$, as the Pearson residual. We will investigate the performance of the LM and CD tests in (20) and (22) when the correlation coefficient is constructed from the Pearson residual and compare the results to the case when the generalized residuals are used.

For the Tobit model we have

$$\tilde{v}_{it}=y_{it}-\left(\tilde{\boldsymbol{\beta}}_i'\mathbf{x}_{it}+\tilde{\sigma}_{iu}\lambda_{it}\right)\Phi\left(\tilde{\boldsymbol{\beta}}_i'\mathbf{x}_{it}/\tilde{\sigma}_{iu}\right).\tag{24}$$

with the associated Pearson residual is given by

$$\tilde{v}_{it}^{*}=\frac{y_{it}-\left(\tilde{\boldsymbol{\beta}}_i'\mathbf{x}_{it}+\tilde{\sigma}_{iu}\lambda_{it}\right)\Phi\left(w_{it}\right)}{\sqrt{\tilde{\sigma}_{iu}^{2}\left\{\left[w_{it}^{2}+w_{it}\lambda_{it}+1\right]\Phi\left(w_{it}\right)-\left[\left(w_{it}+\lambda_{it}\right)\Phi\left(w_{it}\right)\right]^{2}\right\}}}\tag{25}$$

where $w_{it}=\tilde{\boldsymbol{\beta}}_i'\mathbf{x}_{it}/\sigma_{iu}$, $\tilde{\sigma}_{iu}$ is the estimator of the standard error of the disturbance term in the Tobit model and $\lambda_{it}=\phi(w_{it})/\Phi(w_{it})$ is the inverse Mills ratio with argument $-w_{it}$. See Appendix A for details.

4 Small sample properties: Monte Carlo evidence

4.1 Data generating process

The Monte Carlo experiments are based on the following data generating process (DGP) for the latent variable,

$$y_{it}^{*} = \alpha_i + \beta_i x_{it} + \varepsilon_{it}, \tag{26}$$

where $i = 1, 2, \ldots, N$, and $t = 1, 2, \ldots, T$, with $\beta_i = 1$. The regressors are generated as

$$x_{it} = \delta f_{xt} + \eta_{it}$$

$\eta_{it} = \lambda \eta_{i,t-1} + \zeta_{it}$, $\zeta_{it} \sim \text{iid}\,\text{N}(0,1)$, and $f_{xt} \sim \text{iid}\,\text{N}(0,1)$. We set $\delta = 1$ and $\lambda = 0.5$. Finally, the fixed effects are generated as

$$\alpha_i = \bar{x}_i + \hat{S}_{xi}\nu_i,$$

where $\bar{x}_i = \sum_{t=1}^{T} x_{it}/T$, $\hat{S}_{xi} = \left[(N-1)^{-1}\sum_{i=1}^{N}(\bar{x}_i - \bar{\bar{x}})^2\right]^{1/2}$, $\bar{\bar{x}} = \sum_{i=1}^{N}\bar{x}_i/N$, and $\nu_i \sim \text{iid}\,\text{N}(0,1)$. Hence, the setup covers the case, where the individual specific effects are allowed to be correlated with the explanatory variables. This is an important consideration in the analysis of micro panels, as noted, for example, by Chamberlain (1980). The results are based on 2000 replications per experiment.

The estimation of β under a probit specification only makes use of $y_{it} = \text{I}\left(y_{it}^{*}\right)$, and under the Tobit specification $y_{it} = y_{it}^{*}\text{I}\left(y_{it}^{*}\right)$. Hence, without loss of generality the variance of the error term, ε_{it}, may be set equal to unity. To investigate the power of the proposed tests we allow for correlation across the errors of different cross section units by adopting the following standardized one-factor structure

$$\varepsilon_{it} = \frac{\gamma_i f_{\varepsilon t} + e_{it}}{\sqrt{1+\gamma_i^2}}$$

where γ_i is a scalar, $f_{\varepsilon t} \sim \text{iid}\,\text{N}(0,1)$, and $e_{it} \sim \text{iid}\,\text{N}(0,1)$. Under these assumptions we have $\text{E}(\varepsilon_{it}) = 0$ and $\text{Var}(\varepsilon_{it}) = 1$. The pair-wise correlation coefficient of the errors is given by

$$\text{Corr}\left(\varepsilon_{it}, \varepsilon_{jt}\right) = \frac{\gamma_i \gamma_j}{\sqrt{\left(1+\gamma_i^2\right)\left(1+\gamma_j^2\right)}}.$$

In the experiments reported below we use $\gamma_i = 0$, $\forall i$, $\gamma_i \sim \text{U}(0.1, 0.3)$, and $\gamma_i \sim \text{U}(-0.2, 0.6)$, where $\text{U}(a,b)$ denotes the uniform distribution with lower bound a and upper bound b.

Using the artificial data, α_i, β_i, (and σ_i in the case of the Tobit model) are estimated under the assumption of cross section independence by maximum likelihood for each i separately. Then, $\tilde{\rho}_{ij}$ is computed using the two alternative residuals, namely the generalized and the Pearson residuals ($\tilde{u}_{it}$ and $\tilde{v}_{it}^{*}$, respectively) as set out in Appendix A, and the LM and CD test statistics are then calculated using (20) and (22), respectively.

4.2 Monte Carlo results

Table 1 presents the size and power of CD and LM tests for the probit models, and Table 2 presents the size and power of CD and LM tests for the Tobit model. The results in these tables suggest the following.

(i) There are substantial size distortions for the LM test unless T is much larger than N.

(ii) The empirical size is close to the nominal size for the CD test even for N and T as small as 10. For $N = 500$ the size of the LM test is 1 for all T, whereas the CD test is slightly oversized only when T is much smaller than N. This result holds generally and does not require the fixed effects to be uncorrelated with the regressors. When N is much larger than T we recommend using bootstrap method to approximate critical values. See the empirical illustration and Appendix D.

(iii) The power of CD test improves as either N or T increases. However, the power improves much faster when N increases than when T increases. When $T = 20$ and $N = 100$, the power is about 0.7 for the probit model and greater than 0.9 for the Tobit model. On the other hand, when $N = 20$ and $T = 100$, the power of CD test is about 0.5 for the probit model and less than 0.9 for the Tobit model.

(iv) The test results are reasonably robust to the way residuals from the nonlinear models are computed. There is little to choose between the two CD tests based on the generalized or the Pearson residuals.

(v) Even when the LM test has the correct size, as in the case where $T = 100$ and $N = 10$, the CD test continues to exhibit a higher power.

5 Application to an analysis of campaign contributions and roll-call votes

Here we provide an empirical application of the CD test to the probit model of voting behavior investigated by Wawro (2001). Using data on the voting behavior of members of the US Congress, Wawro analyzes the influence of campaign contributions of a business lobby group (the US Chamber of Commerce, USCC) and a labor lobby group (the American Federation of Labor-Congress of Industrial Organizations, AFL-CIO) on voting outcomes

Table 1: Size and power of CD and LM tests: The probit model

	Generalized residuals, $\tilde{u}_{it}$						Pearson residuals, $\tilde{v}^*_{it}$					
$T\backslash N$	10	20	30	50	100	500	10	20	30	50	100	500
						Size: $\gamma_i = 0,\ \forall i$						
						CD test						
10	0.065	0.063	0.071	0.061	0.063	0.129	0.062	0.061	0.070	0.059	0.064	0.155
20	0.057	0.059	0.062	0.061	0.061	0.072	0.059	0.061	0.068	0.058	0.062	0.101
30	0.053	0.054	0.062	0.045	0.053	0.054	0.051	0.055	0.062	0.045	0.053	0.078
50	0.050	0.050	0.056	0.050	0.059	0.063	0.053	0.054	0.059	0.051	0.059	0.069
100	0.052	0.059	0.051	0.063	0.052	0.045	0.049	0.053	0.050	0.060	0.056	0.048
						LM test						
10	0.200	0.447	0.722	0.969	1.000	1.000	0.205	0.456	0.724	0.972	1.000	1.000
20	0.094	0.216	0.390	0.688	0.988	1.000	0.090	0.218	0.381	0.690	0.987	1.000
30	0.078	0.139	0.254	0.459	0.903	1.000	0.082	0.135	0.252	0.459	0.906	1.000
50	0.067	0.097	0.130	0.272	0.637	1.000	0.062	0.098	0.118	0.260	0.632	1.000
100	0.060	0.082	0.085	0.137	0.351	1.000	0.064	0.082	0.086	0.135	0.368	0.996
						Power: $\gamma_i \sim$ U$(0.1, 0.3)$						
						CD test						
10	0.087	0.117	0.171	0.256	0.466	0.983	0.086	0.120	0.172	0.265	0.469	0.986
20	0.113	0.180	0.268	0.423	0.707	1.000	0.108	0.174	0.265	0.422	0.705	1.000
30	0.119	0.227	0.356	0.574	0.843	1.000	0.117	0.222	0.349	0.576	0.837	1.000
50	0.157	0.356	0.536	0.765	0.941	1.000	0.163	0.350	0.530	0.771	0.943	1.000
100	0.233	0.552	0.791	0.947	0.997	1.000	0.231	0.544	0.797	0.951	0.999	1.000
						LM test						
10	0.187	0.476	0.737	0.973	1.000	1.000	0.189	0.478	0.740	0.972	1.000	1.000
20	0.124	0.251	0.399	0.772	0.991	1.000	0.122	0.252	0.402	0.776	0.992	1.000
30	0.095	0.177	0.279	0.583	0.948	1.000	0.096	0.176	0.290	0.596	0.955	1.000
50	0.100	0.130	0.236	0.439	0.850	1.000	0.101	0.138	0.242	0.437	0.882	1.000
100	0.099	0.149	0.234	0.437	0.814	1.000	0.095	0.149	0.227	0.431	0.879	1.000
						Power: $\gamma_i \sim$ U$(-0.2, 0.6)$						
						CD test						
10	0.083	0.109	0.156	0.224	0.426	0.971	0.085	0.106	0.157	0.230	0.434	0.976
20	0.108	0.153	0.218	0.378	0.642	1.000	0.107	0.153	0.220	0.383	0.634	1.000
30	0.109	0.220	0.304	0.493	0.784	1.000	0.105	0.217	0.305	0.494	0.778	1.000
50	0.152	0.305	0.450	0.670	0.928	1.000	0.147	0.302	0.452	0.671	0.929	1.000
100	0.213	0.431	0.643	0.856	0.994	1.000	0.215	0.440	0.645	0.859	0.993	1.000
						LM test						
10	0.210	0.456	0.729	0.970	1.000	1.000	0.209	0.454	0.732	0.971	1.000	1.000
20	0.136	0.275	0.448	0.793	0.995	1.000	0.140	0.277	0.453	0.801	0.996	1.000
30	0.105	0.213	0.358	0.645	0.967	1.000	0.107	0.199	0.360	0.650	0.983	1.000
50	0.130	0.214	0.363	0.615	0.916	1.000	0.124	0.212	0.359	0.632	0.960	1.000
100	0.160	0.327	0.513	0.746	0.942	1.000	0.148	0.319	0.522	0.769	0.984	1.000

The table reports the percentage of rejections of the null of cross section independence for the LM and the CD test statistics defined by (20) and (22), using generalized and Pearson residuals as set out in Section 3. The tests are carried out at the 5 per cent significance level, with 2000 replications per experiment.

Table 2: Size and power of CD and LM tests: The Tobit model

$T\backslash N$	Generalized residuals, $\tilde{u}_{it}$						Pearson residuals, $\tilde{v}^*_{it}$					
	10	20	30	50	100	500	10	20	30	50	100	500
						Size: $\gamma_i = 0,\ \forall i$						
						CD test						
10	0.062	0.055	0.065	0.060	0.063	0.103	0.057	0.071	0.057	0.063	0.067	0.176
20	0.057	0.054	0.062	0.057	0.059	0.060	0.063	0.056	0.050	0.056	0.061	0.123
30	0.050	0.060	0.058	0.059	0.056	0.062	0.064	0.060	0.060	0.064	0.066	0.085
50	0.047	0.057	0.049	0.061	0.065	0.057	0.049	0.060	0.056	0.056	0.063	0.076
100	0.047	0.062	0.055	0.044	0.046	0.060	0.046	0.054	0.063	0.051	0.051	0.061
						LM test						
10	0.193	0.492	0.769	0.983	1.000	1.000	0.187	0.482	0.765	0.985	1.000	1.000
20	0.134	0.267	0.464	0.771	0.993	1.000	0.129	0.274	0.449	0.779	0.991	1.000
30	0.111	0.192	0.316	0.572	0.932	1.000	0.108	0.189	0.312	0.579	0.938	1.000
50	0.075	0.118	0.199	0.401	0.764	1.000	0.072	0.140	0.195	0.390	0.757	1.000
100	0.083	0.079	0.141	0.247	0.572	1.000	0.074	0.097	0.138	0.259	0.562	0.996
						Power: $\gamma_i \sim$ U$(0.1, 0.3)$						
						CD test						
10	0.135	0.265	0.388	0.598	0.877	1.000	0.141	0.267	0.391	0.601	0.880	1.000
20	0.190	0.379	0.570	0.796	0.973	1.000	0.183	0.399	0.574	0.798	0.971	1.000
30	0.208	0.466	0.688	0.908	0.995	1.000	0.199	0.476	0.690	0.903	0.998	1.000
50	0.290	0.647	0.870	0.985	1.000	1.000	0.301	0.658	0.871	0.988	1.000	1.000
100	0.458	0.884	0.987	1.000	1.000	1.000	0.450	0.877	0.991	1.000	1.000	1.000
						LM test						
10	0.228	0.544	0.806	0.991	1.000	1.000	0.208	0.546	0.803	0.990	1.000	1.000
20	0.180	0.397	0.582	0.887	1.000	1.000	0.187	0.383	0.607	0.892	1.000	1.000
30	0.146	0.309	0.513	0.809	0.997	1.000	0.137	0.300	0.513	0.811	0.996	1.000
50	0.149	0.297	0.487	0.810	0.988	1.000	0.153	0.283	0.509	0.818	0.990	1.000
100	0.191	0.371	0.595	0.876	0.999	1.000	0.186	0.379	0.602	0.893	1.000	1.000
						Power: $\gamma_i \sim$ U$(-0.2, 0.6)$						
						CD test						
10	0.142	0.257	0.361	0.559	0.851	1.000	0.142	0.247	0.369	0.569	0.838	1.000
20	0.165	0.343	0.499	0.734	0.958	1.000	0.161	0.344	0.477	0.739	0.953	1.000
30	0.213	0.415	0.609	0.841	0.989	1.000	0.202	0.410	0.602	0.827	0.988	1.000
50	0.279	0.529	0.758	0.934	0.999	1.000	0.282	0.529	0.746	0.961	0.999	1.000
100	0.405	0.703	0.891	0.989	1.000	1.000	0.382	0.712	0.889	0.989	1.000	1.000
						LM test						
10	0.245	0.566	0.831	0.993	1.000	1.000	0.237	0.577	0.836	0.992	1.000	1.000
20	0.201	0.454	0.691	0.933	1.000	1.000	0.191	0.452	0.688	0.940	1.000	1.000
30	0.216	0.449	0.700	0.928	1.000	1.000	0.220	0.445	0.693	0.937	1.000	1.000
50	0.278	0.521	0.758	0.944	1.000	1.000	0.287	0.536	0.761	0.944	1.000	1.000
100	0.411	0.749	0.916	0.992	1.000	1.000	0.401	0.744	0.918	0.992	1.000	1.000

The generalized and Pearson residuals for the Tobit model are derived in Appendix A and set out in Section 3. Also see the notes to Table 1.

with the unemployment rate in the constituency of the voting member as an additional explanatory variable.

The data set, which is available from Prof. Wawro's web page[2], contains data for a selection of the roll-call votes for the 102th, 103rd and 104th Congress, where the selected roll-call votes are those that the lobby groups themselves deemed important. Hence, we have six data sets, two for each Congress where one contains the votes selected by the USCC and the other the votes selected by the AFL-CIO. We test for cross section independence in each of the six data sets.

The data sets have between $T = 13$ and 15 motions, where T is the number of motions that are put before Congress and are recorded in Wawro's data set. Given that the number of roll-call votes, viewed as the cross section dimension, N, is much larger than the number of motions, taken as the time series dimension, T, the LM test will suffer from serious size distortions. The Monte Carlo results above demonstrate that the CD test has the correct size in such cases. We will therefore restrict our attention to the CD test.

Wawro (2001, p.570) includes motion-specific dummies to account for the particular political context around each roll-call vote. Clearly, if we cannot reject cross section independence a motion specific intercept would not be necessary. We test for cross section independence using a probit model under slope homogeneity and when the slope coefficients are allowed to vary across individual members of Congress. In this way the sensitivity of the test outcomes to slope heterogeneity can be assessed.

We proceed as follows. We estimate the parameters of the probit model by maximizing the likelihood

$$\mathcal{L}_i = \prod_{t=1}^{T} \Phi(\boldsymbol{z}_{it})^{y_{it}} [1 - \Phi(\boldsymbol{z}_{it})]^{(1-y_{it})}, \text{ for } i = 1, 2, \ldots, N, \tag{27}$$

where $\boldsymbol{z}_{it} = \boldsymbol{\beta}'\mathbf{x}_{it}$ under slope homogeneity (pooled specification), and $\boldsymbol{z}_{it} = \boldsymbol{\beta}_i'\mathbf{x}_{it}$ in the specification with individual specific parameters, y_{it} is a binary indicator for the votes of the i^{th} member of Congress ("aye" or "nay") in the t^{th} motion, $\mathbf{x}_{it}$ contains an intercept, the contributions of the USCC, the contributions of the AFL-CIO, and the unemployment rate in the constituency of the voting member.

Using the ML estimator of $\boldsymbol{\beta}$ or $\boldsymbol{\beta}_i$ we calculate the generalized residual as given in (15) and the Pearson residual defined by (23). From these we obtain the pair-wise correlation coefficients, $\tilde{\rho}_{ij}$. Due to the unbalanced nature of the panel under consideration, the CD statistic is computed as

$$\text{CD} = \sqrt{\frac{2}{N(N-1)}} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \sqrt{T_{ij}} \tilde{\rho}_{ij} \right), \tag{28}$$

[2] http://www.columbia.edu/~gjw10/panelprobit.html

where T_{ij} is the number of motions where observations on votes are available for both i and j individuals.

The Monte Carlo results in the pervious section suggest that the CD test could have a slightly larger than nominal size for some of the combinations of N and T in this application. We therefore calculated bootstrap critical values for the CD test. Details of the bootstrap procedure are given in Appendix D.

Table 3 reports the results for the pooled probit model with the votes deemed important by the AFL-CIO in the upper half and those deemed important by the USCC in the lower half of the table. The average pair-wise correlation coefficients for the generalized residuals computed for different roll-call votes are between 0.126 and 0.198. Their significance can be evaluated using the CD test statistics, which are all substantially larger than the 99% critical value of the standard normal distribution. The values in brackets are the bootstrap 5% critical values. The bootstrap test results are in line with asymptotic test results, and reject the null of cross section independence. The difference between the results using the generalized and the Pearson residuals is very small and the tests lead to the same conclusion.

The results for the probit model with individual specific parameters are reported in Table 4. The average pair-wise correlation coefficients of the residuals is slightly smaller yet still considerable, ranging from 0.106 to 0.146. The CD test statistics are also slightly smaller. This is likely to be due to the greater efficiency of the pooled estimation. Additionally, for the 103rd congress a number of members of congress had to be excluded to achieve convergence of the individual specific estimations. That said the results point unanimously in the same direction as those from the pooled estimation. This suggests that the influence of potential parameter heterogeneity and efficiency due to pooling do not change the result, namely that the null of cross section independence is clearly rejected.

6 Conclusion

In this paper, we have derived a Lagrangian multiplier test of cross section independence for nonlinear panel data models, and have proposed a generalized version of Pesaran's CD test. Our Monte Carlo studies show that the LM test is subject to serious size distortions when the cross section dimension is large. On the other hand, the CD test performs well even in small N and T cases. The empirical size of the CD test is close to the nominal size for most combinations of N and T, except for cases where N is much larger than T, for example when $N = 500$, and T is 20 or less. In cases where N is much larger than T, we suggest using bootstrap method to approximate critical values, as we did in the empirical illustration. The test also has good power, in particular when N is large, even when T is relatively small so long

Table 3: CD test for roll-call votes in the U. S. Congress, pooled estimation

Congr.	N/T	Generalized residuals				Pearson residuals			
		CD	Bootstrap 5% crit. val.		$\bar{\rho}$	CD	Bootstrap 5% crit. val.		$\bar{\rho}$
AFL-CIO									
102nd	82/13	26.153	[−1.643	2.458]	0.127	26.106	[−1.638	2.450]	0.126
103rd	203/14	74.955	[−1.575	2.106]	0.141	74.787	[−1.582	2.132]	0.140
104th	65/14	31.632	[−1.623	2.330]	0.185	31.753	[−1.632	2.280]	0.186
USCC									
102nd	130/15	69.685	[−1.673	2.385]	0.198	69.380	[−1.669	2.363]	0.198
103rd	276/14	120.300	[−1.593	2.220]	0.166	120.286	[−1.602	2.212]	0.166
104th	193/15	106.675	[−1.719	2.282]	0.181	106.620	[−1.730	2.304]	0.181

The column with heading CD gives the values of the CD test statistic. The bootstrap critical values were computed using 1000 iterations. The first number in square brackets gives the 2.5% lower critical value and the second number the 2.5% upper critical value. Details of the bootstrap procedure are given in Appendix D. The columns with heading $\bar{\rho}$ gives the average pair-wise correlation coefficients. N is the number of voting members of Congress, and T is the number of motions.

Table 4: CD test for roll-call votes in the U. S. Congress, individual specific estimation

Congr.	N/T	Generalized residuals				Pearson residuals			
		CD	Bootstrap 5% crit. val.		$\bar{\rho}$	CD	Bootstrap 5% crit. val.		$\bar{\rho}$
AFL-CIO									
102nd	82/13	21.922	[−1.653	2.232]	0.106	21.930	[−1.659	2.122]	0.106
103rd	183/14	60.379	[−1.473	2.961]	0.138	60.350	[−1.463	3.156]	0.138
104th	65/14	24.499	[−1.564	2.654]	0.144	22.174	[−1.555	2.686]	0.146
USCC									
102nd	130/15	50.455	[−1.589	2.826]	0.143	50.924	[−1.603	2.954]	0.145
103rd	258/14	86.760	[−1.608	2.348]	0.141	88.700	[−1.572	2.509]	0.143
104th	193/15	85.497	[−1.643	2.420]	0.145	86.300	[−1.671	2.428]	0.146

See notes to Table 3. When estimating individual specific parameters observations on a few members of the 103rd Congress had to be removed from the sample for convergence reason.

as the average pair-wise correlation of the residuals does not converge to zero, which corresponds to the case of weak cross section dependence in the sense defined by Chudik, Pesaran and Tosetti (2010).

The CD test is simple to implement and can be readily adapted to unbalanced panels. As is well known in panel data literature when T is small and N is large, the presence of individual-specific effects introduces the classical incidental parameter problems (Neyman and Scott 1948). The estimation of structural parameters are often entangled with the estimation of incidental parameters. To obtain a consistent estimator of structural parameters, one often has to impose stringent conditions on the data and the estimation becomes complicated (see e.g. Hsiao 2003). The problem can only become more unwieldy if there exists cross section dependence. A nice feature of Pesaran CD test is that one can estimate model parameters under cross section independence, and the presence of individual-specific effects (possibly correlated with the regressors) does not affect the performance of the test because each cross sectional unit parameter is estimated using that unit's time series observation alone.

In this paper, we assume that all the explanatory variables in a nonlinear model are exogenous. In the case where one or more explanatory variables are endogenous, as long as estimated parameters of the model are $\sqrt{T}$ consistent under cross section independence, say by N2SLS (Amemiya 1974), the estimated generalized or Pearson residuals can still be obtained and the asymptotics of the CD test remain valid.

In cases, such as the application in this paper, where cross section error independence is rejected, one may wish to investigate the nature of the dependence, possibly along the lines of Ng (2006). Also the estimation of the structural parameters in the model will need to take the cross section dependence into account. While these two topics are beyond the scope of the current paper and are left for future research, this paper proposes a simple yet reasonably powerful test for the detection of cross section error dependence, which is the starting point for any such endeavor.

Appendix A: Derivations of the residuals

The generalized residual for the probit model

$$
\begin{aligned}
u_{it} &= \mathrm{E}(\varepsilon_{it}|y_{it}, \mathbf{x}_{it}) \\
&= \mathrm{E}(\varepsilon_{it}|y_{it}=1, \mathbf{x}_{it})y_{it} + \mathrm{E}(\varepsilon_{it}|y_{it}=0, \mathbf{x}_{it})(1-y_{it}) \\
&= \frac{\phi(\boldsymbol{\beta}_i'\mathbf{x}_{it})}{\Phi(\boldsymbol{\beta}_i'\mathbf{x}_{it})\left[1-\Phi(\boldsymbol{\beta}_i'\mathbf{x}_{it})\right]}\left[y_{it}-\Phi(\boldsymbol{\beta}_i'\mathbf{x}_{it})\right].
\end{aligned}
$$

The variance of the generalized residual is

$$
\begin{aligned}
\mathrm{Var}(u_{it}) &= \frac{\phi(\boldsymbol{\beta}_i'\mathbf{x}_{it})^2}{\Phi(\boldsymbol{\beta}_i'\mathbf{x}_{it})^2\left[1-\Phi(\boldsymbol{\beta}_i'\mathbf{x}_{it})\right]^2}\mathrm{E}\left[(y_{it}-\Phi(\boldsymbol{\beta}_i'\mathbf{x}_{it})\right]^2 \\
&= \frac{\phi(\boldsymbol{\beta}_i'\mathbf{x}_{it})^2}{\Phi(\boldsymbol{\beta}_i'\mathbf{x}_{it})\left[1-\Phi(\boldsymbol{\beta}_i'\mathbf{x}_{it})\right]}.
\end{aligned}
$$

The Pearson residual for the probit model

$$
\begin{aligned}
v_{it} &= y_{it} - \mathrm{E}(y_{it}\,|\mathbf{x}_{it}) \\
&= y_{it} - \Phi(\boldsymbol{\beta}_i'\mathbf{x}_{it}).
\end{aligned}
$$

The variance is

$$
\mathrm{Var}(v_{it}) = \Phi(\boldsymbol{\beta}_i'\mathbf{x}_{it})\left[1-\Phi(\boldsymbol{\beta}_i'\mathbf{x}_{it})\right].
$$

The generalized residual for the Tobit model

$$
\begin{aligned}
u_{it} &= \mathrm{E}(\varepsilon_{it}|y_{it}>0, \mathbf{x}_{it})\mathrm{I}(y_{it}) + \mathrm{E}(\varepsilon_{it}|y_{it}=0, \mathbf{x}_{it})\left[1-\mathrm{I}(y_{it})\right] \\
&= (y_{it}-\boldsymbol{\beta}_i'\mathbf{x}_{it})\mathrm{I}(y_{it}) - \sigma_i\frac{\phi(\boldsymbol{\beta}_i'\mathbf{x}_{it}/\sigma_{iu})}{\Phi(-\boldsymbol{\beta}_i'\mathbf{x}_{it}/\sigma_{iu})}\left[1-\mathrm{I}(y_{it})\right].
\end{aligned}
$$

For the variance we have,

$$
\begin{aligned}
\mathrm{Var}(u_{it}) &= \mathrm{E}(\varepsilon_{it}^2|\varepsilon_{it} > -\boldsymbol{\beta}_i'\mathbf{x}_{it})\Phi(\boldsymbol{\beta}_i'\mathbf{x}_{it}/\sigma_{iu}) \\
&\quad +\sigma_{iu}^2\frac{\phi(\boldsymbol{\beta}_i'\mathbf{x}_{it}/\sigma_{iu})^2}{\Phi(-\boldsymbol{\beta}_i'\mathbf{x}_{it}/\sigma_{iu})^2}\left[1-\Phi(\boldsymbol{\beta}_i'\mathbf{x}_{it}/\sigma_{iu})\right] \\
&= \sigma_{iu}^2\left[(1-\lambda_{it}\boldsymbol{\beta}_i'\mathbf{x}_{it}/\sigma_{iu})\Phi(\boldsymbol{\beta}_i'\mathbf{x}_{it}/\sigma_{iu}) + \frac{\phi(\boldsymbol{\beta}_i'\mathbf{x}_{it}/\sigma_{iu})^2}{\Phi(-\boldsymbol{\beta}_i'\mathbf{x}_{it}/\sigma_{iu})}\right],
\end{aligned}
$$

where $\lambda_{it} = \phi(\boldsymbol{\beta}_i'\mathbf{x}_{it}/\sigma_{iu})/\Phi(\boldsymbol{\beta}_i'\mathbf{x}_{it}/\sigma_{iu})$ is the inverse Mills ratio with argument $-\boldsymbol{\beta}_i'\mathbf{x}_{it}/\sigma_{iu}$.

The Pearson residual for the Tobit model The in-sample forecast error for the Tobit model is defined by

$$\begin{aligned} v_{it} &= y_{it} - \mathrm{E}(y_{it}|\mathbf{x}_{it}) \\ &= y_{it} - \left(\boldsymbol{\beta}_i'\mathbf{x}_{it} + \sigma_{iu}\lambda_{it}\right)\Phi\left(\boldsymbol{\beta}_i'\mathbf{x}_{it}/\sigma_{iu}\right). \end{aligned}$$

The variance is

$$\begin{aligned} \mathrm{Var}(v_{it}) &= \mathrm{E}\left(y_{it}^2|\mathbf{x}_{it}\right) - \left[\left(\boldsymbol{\beta}_i'\mathbf{x}_{it} + \sigma_{iu}\lambda_{it}\right)\Phi\left(\boldsymbol{\beta}_i'\mathbf{x}_{it}/\sigma_{iu}\right)\right]^2 \\ &= \sigma_{iu}^2\left\{\left[(\boldsymbol{\alpha}_i'\mathbf{x}_{it})^2 + \boldsymbol{\alpha}_i'\mathbf{x}_{it}\lambda_{it} + 1\right]\Phi\left(\boldsymbol{\alpha}_i'\mathbf{x}_{it}\right)\right. \\ &\quad \left. - \left[\left(\boldsymbol{\alpha}_i'\mathbf{x}_{it} + \lambda_{it}\right)\Phi\left(\boldsymbol{\alpha}_i'\mathbf{x}_{it}\right)\right]^2\right\}, \end{aligned} \tag{29}$$

where $\boldsymbol{\alpha}_{it} = \boldsymbol{\beta}_{it}/\sigma_{iu}$, $\xi_{it} = \varepsilon_{it}/\sigma_{iu} \sim \mathrm{N}(0,1)$, and we used the fact that

$$\begin{aligned} \mathrm{E}(\xi_{it}^2|\xi_{it} > -\boldsymbol{\alpha}_i'\mathbf{x}_{it}) &= \mathrm{Var}(\xi_{it}|\xi_{it} > -\boldsymbol{\alpha}_i'\mathbf{x}_{it}) + \mathrm{E}(\xi_{it}|\xi_{it} > -\boldsymbol{\alpha}_i'\mathbf{x}_{it})^2 \\ &= [1 - \lambda_{it}(\lambda_{it} + \boldsymbol{\alpha}_i'\mathbf{x}_{it})] + \lambda_{it}^2 \\ &= 1 - \lambda_{it}\boldsymbol{\alpha}_i'\mathbf{x}_{it}. \end{aligned}$$

Substituting $\boldsymbol{\alpha}_{it}$ into (29) gives the variance used in (25) to obtain the Pearson residual of the Tobit model.

Appendix B: Mathematical details for the LM test

Taking the second partial derivative of the log likelihood function using (10) we have

$$\frac{\partial^2 \ell_{NT}}{\partial\rho_{ij}\partial\rho_{rs}} = -\sum_{t=1}^{T}\frac{\partial \ln P_t}{\partial\rho_{ij}}\frac{\partial \ln P_t}{\partial\rho_{rs}} + \sum_{t=1}^{T}\frac{1}{P_t}\frac{\partial^2 P_t}{\partial\rho_{ij}\partial\rho_{rs}}, \tag{30}$$

where

$$\frac{\partial^2 P_t}{\partial\rho_{ij}\partial\rho_{rs}} = \int_{A(\boldsymbol{\varepsilon}_t|\boldsymbol{y}_t)}\frac{\partial \ln f(\boldsymbol{\varepsilon}_t)}{\partial\rho_{ij}}\frac{\partial \ln f(\boldsymbol{\varepsilon}_t)}{\partial\rho_{rs}}f(\boldsymbol{\varepsilon}_t)d\boldsymbol{\varepsilon}_t + \int_{A(\boldsymbol{\varepsilon}_t|\boldsymbol{y}_t)}\frac{\partial^2 \ln f(\boldsymbol{\varepsilon}_t)}{\partial\rho_{ij}\partial\rho_{rs}}f(\boldsymbol{\varepsilon}_t)d\boldsymbol{\varepsilon}_t.$$

Furthermore,

$$\begin{aligned} \frac{\partial^2 \ln f(\boldsymbol{\varepsilon}_t)}{\partial\rho_{ij}\partial\rho_{rs}} &= \frac{1}{2}\mathrm{tr}(\boldsymbol{R}^{-1}\boldsymbol{A}_{rs}\boldsymbol{R}^{-1}\boldsymbol{A}_{ij}) - \frac{1}{2}\boldsymbol{\varepsilon}_t'\boldsymbol{R}^{-1}\boldsymbol{A}_{rs}\boldsymbol{R}^{-1}\boldsymbol{A}_{ij}\boldsymbol{R}^{-1}\boldsymbol{\varepsilon}_t \\ &\quad -\frac{1}{2}\boldsymbol{\varepsilon}_t'\boldsymbol{R}^{-1}\boldsymbol{A}_{ij}\boldsymbol{R}^{-1}\boldsymbol{A}_{rs}\boldsymbol{R}^{-1}\boldsymbol{\varepsilon}_t, \end{aligned}$$

and under the null hypothesis

$$\left.\frac{\partial^2 \ln f(\boldsymbol{\varepsilon}_t)}{\partial\rho_{ij}\partial\rho_{rs}}\right|_{\boldsymbol{R}=\boldsymbol{I}_N} = \frac{1}{2}\mathrm{tr}(\boldsymbol{A}_{rs}\boldsymbol{A}_{ij}) - \frac{1}{2}\boldsymbol{\varepsilon}_t'\left(\boldsymbol{A}_{rs}\boldsymbol{A}_{ij} + \boldsymbol{A}_{ij}\boldsymbol{A}_{rs}\right)\boldsymbol{\varepsilon}_t. \tag{31}$$

Note that $\boldsymbol{A}_{ij}\boldsymbol{A}_{rs} = \mathbf{0}$ if $i \neq r, s$ and $j \neq r, s$, $\boldsymbol{A}_{ij}\boldsymbol{A}_{rs} = \boldsymbol{D}_{lk}$ if either $i = r$, $i = s$, $j = r$ or $j = s$ and l is the subscript i or j that is not equal to either r or s and k is the subscript r or s that is not equal to either i or j, so as an example, $\boldsymbol{A}_{5,8}\boldsymbol{A}_{3,8} = \boldsymbol{D}_{5,3}$, and $\boldsymbol{D}_{lk}$ is a matrix of zeros except for the (l, k)th element, which is equal to 1. Also, $\boldsymbol{A}_{ij}^2 = \boldsymbol{C}_{ij}$, where $\boldsymbol{C}_{ij}$ is an $N \times N$ matrix with all elements equal to zero except for the ith and jth diagonal elements that are equal to 1, which yields $\mathrm{tr}(\mathbf{C}_{ij}) = 2$, and $\boldsymbol{\varepsilon}_t'\mathbf{C}_{ij}\boldsymbol{\varepsilon}_t = \varepsilon_{it}^2 + \varepsilon_{jt}^2$

Hence, for $(i, j) = (r, s)$,

$$\left.\frac{\partial^2 \ln f(\boldsymbol{\varepsilon}_t)}{\partial \rho_{ij}^2}\right|_{\boldsymbol{R}=\boldsymbol{I}_N} = 1 - \varepsilon_{it}^2 - \varepsilon_{jt}^2,$$

and (recalling that $\partial \ln f(\boldsymbol{\varepsilon}_t)/\partial \rho_{ij}|_{\mathbf{R}=\mathbf{I}_N} = \varepsilon_{it}\varepsilon_{jt}$)

$$\begin{aligned}\left.\frac{1}{P_t}\frac{\partial^2 P_t}{\partial \rho_{ij}^2}\right|_{\boldsymbol{R}=\boldsymbol{I}_N} &= \frac{\int_{A(\boldsymbol{\varepsilon}_t|\boldsymbol{y}_t)} \varepsilon_{it}^2\varepsilon_{jt}^2 \prod_{k=1}^N \phi(\varepsilon_{kt})d\boldsymbol{\varepsilon}_t}{\int_{A(\boldsymbol{\varepsilon}_t|\boldsymbol{y}_t)} \prod_{k=1}^N \phi(\varepsilon_{kt})d\boldsymbol{\varepsilon}_t} \\ &\quad + \frac{\int_{A(\boldsymbol{\varepsilon}_t|\boldsymbol{y}_t)} \left(1 - \varepsilon_{it}^2 - \varepsilon_{jt}^2\right) \prod_{k=1}^N \phi(\varepsilon_{kt})d\boldsymbol{\varepsilon}_t}{\int_{A(\boldsymbol{\varepsilon}_t|\boldsymbol{y}_t)} \prod_{k=1}^N \phi(\varepsilon_{kt})d\boldsymbol{\varepsilon}_t}.\end{aligned}$$

Therefore, using the above result and (12) in (30) for $(i, j) = (r, s)$ we have

$$\left.\frac{\partial^2 \ell_{NT}}{\partial \rho_{ij}^2}\right|_{\boldsymbol{R}=\boldsymbol{I}_N} = -\sum_{t=1}^T u_{it}^2 u_{jt}^2 + \sum_{t=1}^T \left(1 - \eta_{it}^2\right)\left(1 - \eta_{jt}^2\right), \qquad (32)$$

where u_{it} is defined by (13) and

$$\eta_{it}^2 = \frac{\int_{a_{it}}^{b_{it}} \varepsilon_{it}^2 \phi(\varepsilon_{it})d\varepsilon_{it}}{\int_{a_{it}}^{b_{it}} \phi(\varepsilon_{it})d\varepsilon_{it}} = 1 - \left[\frac{b_{it}\phi(b_{it}) - a_{it}\phi(a_{it})}{\Phi(b_{it}) - \Phi(a_{it})}\right].$$

For $(i, j) \neq (r, s)$ we can rewrite (30) using the results in (11) and (31) as

$$\begin{aligned}\frac{\partial^2 \ell_{NT}}{\partial \rho_{ij}\partial \rho_{rs}} &= -\sum_{t=1}^T \frac{1}{P_t^2}\frac{\partial P_t}{\partial \rho_{ij}}\frac{\partial P_t}{\partial \rho_{rs}} + \sum_{t=1}^T \frac{1}{P_t}\int_{A(\boldsymbol{\varepsilon}_t|\boldsymbol{y}_t)} f(\boldsymbol{\varepsilon}_t)\left\{\varepsilon_{rt}\varepsilon_{st}\varepsilon_{it}\varepsilon_{jt}\right. \\ &\quad \left. -\frac{1}{2}\mathrm{tr}(\boldsymbol{A}_{rs}\boldsymbol{A}_{ij}) - \frac{1}{2}\boldsymbol{\varepsilon}_t'\boldsymbol{A}_{rs}\boldsymbol{A}_{ij}\boldsymbol{\varepsilon}_t - \frac{1}{2}\boldsymbol{\varepsilon}_t'\boldsymbol{A}_{ij}\boldsymbol{A}_{rs}\boldsymbol{\varepsilon}_t\right\} d\boldsymbol{\varepsilon}_t.\end{aligned}$$

From (12) we have that

$$\sum_{t=1}^T \frac{1}{P_t^2}\frac{\partial P_t}{\partial \rho_{ij}}\frac{\partial P_t}{\partial \rho_{rs}} = \sum_{t=1}^T u_{it}u_{jt}u_{rt}u_{st}.$$

Furthermore,

$$\begin{aligned}
&\sum_{t=1}^{T}\frac{1}{P_t}\int_{A(\boldsymbol{\varepsilon}_t|\boldsymbol{y}_t)} f(\boldsymbol{\varepsilon}_t)\left\{\varepsilon_{rt}\varepsilon_{st}\varepsilon_{it}\varepsilon_{jt}\right\}d\boldsymbol{\varepsilon}_t \\
&= \sum_{t=1}^{T}\frac{\int_{a_{it}}^{b_{it}}\varepsilon_{it}\phi(\varepsilon_{it})d\varepsilon_{it}}{\int_{a_{it}}^{b_{it}}\phi(\varepsilon_{it})d\varepsilon_{it}}\frac{\int_{a_{jt}}^{b_{jt}}\varepsilon_{jt}\phi(\varepsilon_{jt})d\varepsilon_{jt}}{\int_{a_{jt}}^{b_{jt}}\phi(\varepsilon_{jt})d\varepsilon_{jt}}\frac{\int_{a_{rt}}^{b_{rt}}\varepsilon_{rt}\phi(\varepsilon_{rt})d\varepsilon_{rt}}{\int_{a_{rt}}^{b_{rt}}\phi(\varepsilon_{rt})d\varepsilon_{rt}}\frac{\int_{a_{st}}^{b_{st}}\varepsilon_{st}\phi(\varepsilon_{st})d\varepsilon_{st}}{\int_{a_{st}}^{b_{st}}\phi(\varepsilon_{st})d\varepsilon_{st}} \\
&= \sum_{t=1}^{T}u_{it}u_{jt}u_{rt}u_{st}.
\end{aligned}$$

Note that $\mathrm{tr}(\boldsymbol{A}_{rs}\boldsymbol{A}_{ij})=0$ if $(i,j)\neq(r,s)$, and that if either $i=r$, $i=s$, $j=r$ or $j=s$

$$\frac{1}{2}\boldsymbol{\varepsilon}_t'\boldsymbol{A}_{rs}\boldsymbol{A}_{ij}\boldsymbol{\varepsilon}_t+\frac{1}{2}\boldsymbol{\varepsilon}_t'\boldsymbol{A}_{ij}\boldsymbol{A}_{rs}\boldsymbol{\varepsilon}_t=\varepsilon_{jt}\varepsilon_{st}$$

and therefore

$$\begin{aligned}
\frac{\partial^2\ell_{NT}}{\partial\rho_{ij}\partial\rho_{rs}} &= \sum_{t=1}^{T}\frac{\int_{a_{it}}^{b_{it}}\varepsilon_{it}\phi(\varepsilon_{it})d\varepsilon_{it}\cdot\int_{a_{st}}^{b_{st}}\varepsilon_{st}\phi(\varepsilon_{st})d\varepsilon_{st}}{\int_{a_{it}}^{b_{it}}\phi(\varepsilon_{it})d\varepsilon_{it}\cdot\int_{a_{st}}^{b_{st}}\phi(\varepsilon_{st})d\varepsilon_{st}} \\
&= \sum_{t=1}^{T}u_{it}u_{st}
\end{aligned}$$

Finally, if $i\neq r,s$ and $j\neq r,s$, $\boldsymbol{A}_{rs}\boldsymbol{A}_{ij}=\boldsymbol{A}_{ij}\boldsymbol{A}_{rs}=\boldsymbol{0}$, and therefore

$$\frac{\partial^2\ell_{NT}}{\partial\rho_{ij}\partial\rho_{rs}}=0.$$

Appendix C: Limiting distribution of CD test for nonlinear models

In most cases of interest the generalized and Pearson residuals can be written as a scaled version of an error from a nonlinear regression equation. For example, the generalized residuals for probit models can be written as

$$\tilde{u}_{it}=H(\boldsymbol{x}_{it},\tilde{\boldsymbol{\theta}}_i)\left[y_{it}-G(\boldsymbol{x}_{it},\tilde{\boldsymbol{\theta}}_i)\right], \tag{33}$$

where $H(\boldsymbol{x}_{it},\tilde{\boldsymbol{\theta}}_i)=\phi(\tilde{\boldsymbol{\theta}}_i'\mathbf{x}_{it})/\left\{\boldsymbol{\Phi}(\tilde{\boldsymbol{\theta}}_i'\mathbf{x}_{it})[1-\Phi(\tilde{\boldsymbol{\theta}}_i'\mathbf{x}_{it})]\right\}$, and $G(\boldsymbol{x}_{it},\tilde{\boldsymbol{\theta}}_i)=\boldsymbol{\Phi}(\tilde{\boldsymbol{\theta}}_i'\mathbf{x}_{it})$. In testing for cross section independence, assuming that $H(\boldsymbol{x}_{it},\tilde{\boldsymbol{\theta}}_i)$ and its first derivatives, $\mathbf{h}_{it}(\theta_i)=\partial H(\boldsymbol{x}_{it},\boldsymbol{\theta}_i)/\partial\boldsymbol{\theta}_i$, are uniformly bounded in $\boldsymbol{x}_{it}$ and $\boldsymbol{\theta}_i$, the scalar function, $H(\boldsymbol{x}_{it},\tilde{\boldsymbol{\theta}}_i)$, does not play a significant role and will be omitted for expositional simplicity. Here we show that the

CD test is asymptotically normally distributed in the case of the following nonlinear model

$$y_{it} = G(\boldsymbol{x}_{it}, \boldsymbol{\theta}_{i0}) + v_{it}, \tag{34}$$

where $G(\boldsymbol{x}_{it}, \boldsymbol{\theta}_{i0}) = \mathrm{E}(y_{it}|\boldsymbol{x}_{it})$, $\boldsymbol{\theta}_{i0}$ is a $p \times 1$ vector of true parameters, which may or may not be identical across i. Let $G_{it} = G(\boldsymbol{x}_{it}, \boldsymbol{\theta}_{i0})$ and $\tilde{G}_{it} = G(\boldsymbol{x}_{it}; \tilde{\boldsymbol{\theta}}_i)$, where

$$\tilde{\boldsymbol{\theta}}_i = \underset{\boldsymbol{\theta}_i}{\operatorname{argmin}} \sum_{t=1}^{T} [y_{it} - G(\boldsymbol{x}_{it}; \boldsymbol{\theta}_i)]^2 \tag{35}$$

be the extremum estimator. As it becomes clear from the proof any other $\sqrt{T}$-consistent estimator of $\boldsymbol{\theta}_{i0}$, such as the ML estimator that allows for the scaler function $H(\boldsymbol{x}_{it}, \boldsymbol{\theta}_{i0})$ in (33), might also be used.

Denote convergence in probability and in distribution by $\xrightarrow{p}$ and $\xrightarrow{d}$, respectively, and assume that

A1 $v_{it} = \sigma_{iv}\varepsilon_{it}$, $\varepsilon_{it} \sim \mathrm{iid}(0, 1)$, for all i and t, and $0 < \sigma_{iv}^2 < \infty$.

A2 Under the null hypothesis defined by H_0 : ε_{it} is distributed independently of ε_{jt} for all $i \neq j$.

A3 The $k \times 1$ explanatory variables, $\mathbf{x}_{it}$ are strictly exogenous such that $\mathrm{E}(\varepsilon_{it} \mid \boldsymbol{x}_i) = 0$ for all i and t, where $\boldsymbol{x}_i = (\boldsymbol{x}'_{i1}, \ldots, \boldsymbol{x}'_{iT})'$.

A4 G_{it} is continuous in $\boldsymbol{\theta}_i \in \Theta_i$ uniformly in t, where Θ_i is an open neighborhood of $\boldsymbol{\theta}_{i0}$.

A5 $\mathbf{g}_{it}(\theta_i) = \frac{\partial G_{it}}{\partial \boldsymbol{\theta}_i}$ exists and is continuous and bounded on Θ_i, and

$$\frac{1}{T}\sum_{t=1}^{T} \mathbf{g}_{it}(\theta_i)\mathbf{g}'_{jt}(\theta_j) \xrightarrow{p} \boldsymbol{\Omega}_{ij},$$

where $\boldsymbol{\Omega}_{ij}$ are finite, non-stochastic matrices, $\boldsymbol{\Omega}_{ii}$ is non-singular, and convergence is uniform for all $\boldsymbol{\theta}_i \in \Theta_i$.

A6 $\frac{\partial^2 G_{it}}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}'_i}$ is continuous in $\boldsymbol{\theta}_i \in \Theta_i$ uniformly in t, with suitably bounded elements such that $\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\partial^2 G_{it}}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}'_i}\right) v_{it} \xrightarrow{p} 0$.

A7 $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \mathbf{g}_{it}(\boldsymbol{\theta}_{i0}) v_{it} \xrightarrow{d} \mathrm{N}(\mathbf{0}, \sigma_i^2 \boldsymbol{\Omega}_{ii})$.

Proposition 3 *Under A1-A7, the statistic*

$$\mathrm{CD} = \sqrt{\frac{2T}{N(N-1)}} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \tilde{\rho}_{ij} \right) \xrightarrow{d} \mathrm{N}(0, 1), \ \text{as } T \to \infty, \tag{36}$$

where $\tilde{\rho}_{ij} = T^{-1}\sum_{t=1}^{T} (\tilde{v}_{it}/\tilde{\sigma}_{iv})(\tilde{v}_{jt}/\tilde{\sigma}_{jv})$, $\tilde{v}_{it} = y_{it} - \tilde{G}_{it}$, *and* $\tilde{\sigma}_{iv}^2 = T^{-1}\sum_{t=1}^{T} \tilde{v}_{it}^2$.

Proof. Let $S_i = \sum_{t=1}^{T}[y_{it} - G(\boldsymbol{x}_{it}, \boldsymbol{\theta}_i)]^2$. Then $\tilde{\boldsymbol{\theta}}_i$ is the solution of

$$\left.\frac{\partial S_i}{\partial \boldsymbol{\theta}_i}\right|_{\tilde{\boldsymbol{\theta}}_i} = -2\sum_{t=1}^{t} v_{it} \left.\frac{\partial G_{it}}{\partial \boldsymbol{\theta}_i}\right|_{\tilde{\boldsymbol{\theta}}_i} = 0. \tag{37}$$

The estimated residual, $\tilde{v}_{it}$, is[3]

$$\begin{aligned}\tilde{v}_{it} &= y_{it} - \tilde{G}_{it} \\ &= v_{it} - \mathbf{g}_{it}'(\boldsymbol{\theta}_{i0})(\tilde{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0}) + O_p\left(\frac{1}{T}\right). \end{aligned} \tag{38}$$

The second equality follows from taking a Taylor series expansion of $\hat{G}_{it}$ around $\boldsymbol{\theta}_{i0}$. Taking a Taylor series expansion of (37) we have

$$\tilde{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0} = \left(T^{-1}\sum_{t=1}^{T}\mathbf{g}_{it}(\boldsymbol{\theta}_{i0})\mathbf{g}_{it}'(\boldsymbol{\theta}_{i0}) - T^{-1}\sum_{t=1}^{T}\frac{\partial^2 G_{it}}{\partial \boldsymbol{\theta}_{i0}\partial \boldsymbol{\theta}_{i0}'}v_{it}\right)^{-1} \tag{39}$$

$$\left[\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}(\boldsymbol{\theta}_{i0})v_{it}\right] + O_p\left(\frac{1}{T}\right). \tag{40}$$

Under Assumptions A5-A7

$$\tilde{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0} = \boldsymbol{\Omega}_{ii}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{it}(\boldsymbol{\theta}_{i0})v_{it}\right) + O_p\left(\frac{1}{T}\right), \tag{41}$$

and hence

$$\sqrt{T}\left(\tilde{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0}\right) \xrightarrow{d} N(\mathbf{0}, \sigma_{iv}^2\boldsymbol{\Omega}_{ii}).$$

Using (41) in (38) and after some algebra

$$\tilde{\sigma}_{iv}^2 = \sigma_{iv}^2 + O_p\left(\frac{1}{T}\right).$$

[3] In the case of the more general nonlinear specification, (33), the estimated generalized residual becomes

$$\begin{aligned}\tilde{v}_{it} &= v_{it} - H(\boldsymbol{x}_{it}, \boldsymbol{\theta}_{i0})\mathbf{g}_{it}'(\boldsymbol{\theta}_{i0})(\tilde{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0}) \\ &\quad -(\tilde{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0})'\mathbf{h}_{it}(\boldsymbol{\theta}_{i0})\mathbf{g}_{it}'(\boldsymbol{\theta}_{i0})(\tilde{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0}) + O_p\left(\frac{1}{T}\right),\end{aligned}$$

where $\mathbf{h}_{it}(\theta_i) = \frac{\partial H_{it}}{\partial \boldsymbol{\theta}_i}$, and if $\tilde{\boldsymbol{\theta}}_i$ is a $\sqrt{T}$-consistent estimator of $\boldsymbol{\theta}_{i0}$ it then readily follows that

$$\sqrt{T}(\tilde{v}_{it} - v_{it}) = -H(\boldsymbol{x}_{it}, \boldsymbol{\theta}_{i0})\mathbf{g}_{it}'(\boldsymbol{\theta}_{i0})\sqrt{T}(\tilde{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0}) + O_p\left(\frac{1}{\sqrt{T}}\right).$$

Hence, the rest of the results set out below equally applies to this more general case so long as $H(\boldsymbol{x}_{it}, \boldsymbol{\theta}_{i0})$ is uniformly bounded in $\mathbf{x}_{it}$ and $\boldsymbol{\theta}_{i0}$.

Similarly, using the above results we have

$$\begin{aligned}\sqrt{T}\tilde{\rho}_{ij} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\tilde{v}_{it}\tilde{v}_{jt}}{\tilde{\sigma}_{iv}\tilde{\sigma}_{jv}} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\tilde{v}_{it}\tilde{v}_{jt}}{\sigma_{iv}\sigma_{jv}} + O_p\left(\frac{1}{T}\right) = \\ &-\frac{1}{\sqrt{T}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{g}_{it}'(\boldsymbol{\theta}_{i0})\varepsilon_{jt}\right)\boldsymbol{\Omega}_{ii}^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{g}_{it}(\boldsymbol{\theta}_{i0})\varepsilon_{it}\right) \\ &-\frac{1}{\sqrt{T}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{g}_{jt}'(\boldsymbol{\theta}_{j0})\varepsilon_{it}\right)\boldsymbol{\Omega}_{jj}^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{g}_{jt}(\boldsymbol{\theta}_{j0})\varepsilon_{jt}\right) \\ &+\frac{1}{\sqrt{T}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{g}_{jt}'(\boldsymbol{\theta}_{j0})\varepsilon_{jt}\right)\boldsymbol{\Omega}_{jj}^{-1}\boldsymbol{\Omega}_{ji}\boldsymbol{\Omega}_{ii}^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{g}_{it}(\boldsymbol{\theta}_{i0})\varepsilon_{it}\right) \\ &+O_p\left(\frac{1}{T}\right)\end{aligned}$$

Consider the second term and note that under Assumptions A7 , $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{g}_{it}'(\boldsymbol{\theta}_{i0})\varepsilon_{jt} = O_p(1)$, and by Assumption A5, all elements of $\boldsymbol{\Omega}_{ii}^{-1}$ and $\boldsymbol{\Omega}_{ji}$ for all i and j, are finite and non-stochastic. Hence

$$\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{g}_{it}'(\boldsymbol{\theta}_{i0})\varepsilon_{jt}\right)\boldsymbol{\Omega}_{ii}^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{g}_{it}(\boldsymbol{\theta}_{i0})\varepsilon_{it}\right) = O_p(1).$$

Furthermore,

$$\begin{aligned}&\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{g}_{it}'(\boldsymbol{\theta}_{i0})\varepsilon_{jt}\right)\boldsymbol{\Omega}_{ii}^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{g}_{it}(\boldsymbol{\theta}_{i0})\varepsilon_{it}\right) \\ &= \frac{1}{T}\sum_{t=1}^{T}\sum_{t'=1}^{T}\left[\mathbf{g}_{it}'(\boldsymbol{\theta}_{i0})\boldsymbol{\Omega}_{ii}^{-1}\mathbf{g}_{it'}(\boldsymbol{\theta}_{i0})\right]\varepsilon_{jt}\varepsilon_{it'}.\end{aligned}$$

where $\mathbf{g}_{it}'(\boldsymbol{\theta}_{i0})\boldsymbol{\Omega}_{ii}^{-1}\mathbf{g}_{it'}(\boldsymbol{\theta}_{i0})$ are bounded in $\boldsymbol{\theta}_{i0}$ and for all i, t and t', and v_{jt} and $v_{it'}$ are independently distributed for all i, j, t and t'. Hence,

$$\mathrm{E}\left\{\frac{1}{T}\sum_{t=1}^{T}\sum_{t'=1}^{T}\left[\mathbf{g}_{it}'(\boldsymbol{\theta}_{i0})\boldsymbol{\Omega}_{ii}^{-1}\mathbf{g}_{it'}(\boldsymbol{\theta}_{i0})\right]\varepsilon_{jt}\varepsilon_{it'}\right\} = 0.$$

Similar arguments also applies to the other two terms in the expression for $T^{-1/2}\sum_{t=1}^{T}\tilde{v}_{it}\tilde{v}_{jt}/(\tilde{\sigma}_{iv}\tilde{\sigma}_{jv})$.

Therefore, $\mathrm{E}\left(\sqrt{T}\tilde{\rho}_{ij}\right) = O(T^{-1})$ and

$$\sqrt{T}\tilde{\rho}_{ij} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it}\varepsilon_{jt} + O_p\left(\frac{1}{\sqrt{T}}\right)$$

and

$$\sqrt{T}\tilde{\rho}_{ij} \xrightarrow{d} \mathrm{N}(0,1). \tag{42}$$

Hence, for a fixed N and as $T \to \infty$

$$\mathrm{CD} = \sqrt{\frac{2}{N(N-1)}} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \sqrt{T}\tilde{\rho}_{ij} \right) \xrightarrow{d} \mathrm{N}(0,1),$$

Using similar lines of reasoning as in Pesaran (2004) it is also easily established that $\mathrm{CD} \xrightarrow{d} \mathrm{N}(0,1)$, for N and T large. ∎

Proposition 4 *Under A1 - A7,*

(i) If N is fixed and $T \to \infty$,

$$\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} T\tilde{\rho}_{ij}^2 \xrightarrow{d} \chi^2_{N(N-1)/2}. \tag{43}$$

(ii) If $T \to \infty$, followed by $N \to \infty$

$$\sqrt{\frac{1}{N(N-1)}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} (T\tilde{\rho}_{ij}^2 - 1) \xrightarrow{d} \mathrm{N}(0,1). \tag{44}$$

Proof. Under A1-A7 and as $T \to \infty$ we have $\sqrt{T}\tilde{\rho}_{ij} \xrightarrow{d} \mathrm{N}(0,1)$, and the arguments in Pesaran (2004) apply. ∎

Appendix D: Bootstrap procedure

A bootstrap approximation might be used to improve the finite sample approximation of the distribution of the CD test. The bootstrap procedure we suggest has previously been employed in different contexts in the literature. Härdle, Mammen and Proença (2001) use the bootstrap approximation to improve the size of the Horowitz-Härdle test for the specification of the link function, $g(\cdot)$ in equation (4). Dikta, Kvesic and Schmidt (2006) call the procedure a "model based resampling scheme" and use it to test for the functional form of the underlying regression model.

For the test at hand the bootstrap procedure works as follows.

1. Using the observed data y_{it} and $\mathbf{x}_{it}$ estimate the parameters for the model and obtain $\tilde{\boldsymbol{\theta}}_i$ for each $i = 1, 2, \ldots, N$.

2. Sample $\hat{\varepsilon}_{it} \sim \mathrm{iid}\,\mathrm{F}(0, \tilde{\sigma}_i^2)$ for $i = 1, 2, \ldots N$ and $t = 1, 2, \ldots T$, where $\mathrm{F}(\cdot)$ is the distribution of the error term implied by the maintained model.

3. Construct $\hat{y}_{it}$ using the model $\mathrm{f}\left(\hat{y}^*_{it}, \mathbf{x}_{it}, \tilde{\boldsymbol{\theta}}_i\right) = \hat{\varepsilon}_{it}$ and $\hat{y}_{it} = \mathrm{g}(\hat{y}^*_{it})$.

4. Using $\hat{y}_{it}$ and $\mathbf{x}_{it}$ estimate the parameters of the model and obtain $\hat{\tilde{\boldsymbol{\theta}}}_i$ for each $i = 1, 2, \ldots, N$. Construct the CD test statistic using $\mathbf{x}_{it}$ and $\hat{\tilde{\boldsymbol{\theta}}}_i$.

5. Repeat step 2–4 B times.

6. The B samples of the test statistic are then used to calculate the critical values against which the test statistic obtained from the data is evaluated. The critical values are, say, the 2.5% lowest and the 2.5% highest values in the sample of the B bootstrap test statistics.

Given that nonlinear models are typically estimated via maximum likelihood, this bootstrap procedure entails considerable computational costs. Härdle, Mammen and Proença (2001) suggest to set the starting values in the estimation of $\hat{\tilde{\theta}}_i$ to $\tilde{\theta}_i$ and use only one iteration to obtain the estimates. In the application in Section 5, however, we let the maximization algorithm run to convergence.

References

Amemiya, Takeshi (1974) 'The nonlinear two-stage least-squares estimator.' *Journal of Econometrics* 2, 105–110.

Breusch, T. S. and Adrian R. Pagan (1980) 'The Lagrange multiplier test and its application to model specifications in Econometrics.' *Review of Economic Studies* 47, 239–253.

Chamberlain, Gary (1980) 'Analysis of covariance with qualitative data.' *Review of Economic Studies* 47, 225–238.

Chesher, Andrew, and Margaret Irish (1987) 'Numerical and graphical residual analysis in the grouped and censored normal linear model.' *Journal of Econometrics* 34, 33–61.

Chudik, Alex, M. Hashem Pesaran, and Elisa Tosetti (2010) 'Weak and strong cross section dependence and estimation of large panels.' *Econometrics Journal* (forthcoming).

Dikta, Gerhard, Marsel Kvesic, and Christian Schmidt (2006) 'Bootstrap approximation in model checks with binary data.' *Journal of the American Statistical Association* 101, 521–530.

Frees, Edward W. (1995) 'Assessing cross sectional correlation in panel data.' *Journal of Econometrics* 69, 393–414.

Gourieroux, Christian, Alain Monfort, Eric Renault, and Alain Trognon (1987) 'Generalised residuals.' *Journal of Econometrics* 34, 5–32.

Härdle, Wolfgang, Enno Mammen, and Isabel Proença (2001) 'A bootstrap test for single index models.' *Statistics* 35, 427–452.

Hsiao, Cheng (2003) *Analysis of Panel Data*, 2nd ed., Cambridge: Cambridge University Press.

Kelejian, Harry H. and Ingmar R. Prucha (2001) 'On the asymptotic distribution of the Moran I test statistic with applications.' *Journal of Econometrics* 104, 219–257.

McCullagh, P. and J. A. Nelder (1989) *Generalized Linear Models*, 2nd ed., Boca Raton: Chapman & Hall/CRC.

Moran, P. A. P. (1948) 'The interpretation of statistical maps.' *Biometrika* 35, 255–260.

Neyman, J. and Elizabeth R. Scott (1948) 'Consistent estimates based on partially consistent observations.' *Econometrica* 16, 1-32.

Ng, Serena (2006) 'Testing cross-section correlation in panel data using spacing.' *Journal of Business and Economic Statistics* 24, 12–23.

Pesaran, M. Hashem (2004) 'General diagnostic tests for cross section dependence in panels.' *Cambridge Working Paper in Economics* 0435.

Pesaran, M. Hashem, Aman Ullah and Takashi Yamagata (2008) 'A bias adjusted LM test of error cross section independence.'*Econometrics Journal* 11, 105–127.

Wawro, Gregory (2001) 'A panel probit analysis of campaign contributions and roll-call votes.' *American Journal of Political Science* 45, 563–579