Diagnostic Tests of Cross Section Independence for Nonlinear Panel Data Models[*]

Cheng Hsiao
University of Southern California

M. Hashem Pesaran
University of Cambridge, CIMF and
University of Southern California

Andreas Pick
University of Cambridge, CIMF

April 2007

Abstract

In this paper we discuss tests for residual cross section dependence in nonlinear panel data models. The tests are based on average pair-wise residual correlation coefficients. In nonlinear models, the definition of the residual is ambiguous and we consider two approaches: deviations of the observed dependent variable from its expected value and generalized residuals. We show the asymptotic consistency of the cross section dependence (CD) test of Pesaran (2004). In Monte Carlo experiments it emerges that the CD test has the correct size for any combination of N and T whereas the LM test relies on T large relative to N. We then analyze the roll-call votes of the 104th U.S. Congress and find considerable dependence between the votes of the members of Congress.

JEL C12, C33, C35

Keywords Cross-section dependence, nonlinear panel data model

1 Introduction

Many panel data models assume that observations across individuals are independent. However, there could be common shocks that affect all individuals. 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 appealing to first test for cross sectional dependence before one attempts to incorporate cross sectional dependence into a model.

There are essentially two approaches to test for cross sectional dependence. One is to postulate a "connection" or "spatial" matrix, then test if the coefficient of this spatial matrix is zero, e.g. Moran (1948), Kelejian and Prucha (2001). Although under the null of no cross-correlations the coefficient of spatial matrix is zero no matter how this matrix is postulated, the power of this kind of test presumably will depend on the choice of the spatial matrix. Moreover, the computation of the spatial regression model is quite complicated, see Kelejian and Prucha (1999) and Lee (2002).

Another approach to testing cross sectional dependence is to directly test if the cross-correlations of the errors are zero. For example, Breusch and Pagan's (1980) Lagrangian multiplier (LM) test is based on the average of the squared pair-wise correlation coefficients. However, the mean of the squared correlation coefficients is not correctly centered when T is small. When N is large, the incorrect centering of the squared correlation coefficients is likely to be accentuated, resulting in significant size distortions. Pesaran, Ullah and Yamagata (2006) have proposed a bias-adjusted normal approximation version of the LM test for linear regression models with strictly exogenous regressors and normal errors. Small sample evidence based on their Monte Carlo experiments suggests that the bias adjusted LM tests successfully control the size. However, if the model is nonlinear, it does not appear feasible to derive the exact mean and variance based, for example, on the work of Ullah (2004).

As an alternative to the test based on the square of the error correlation coefficients, Pesaran (2004) proposes to use the simple average of all pair-wise correlation coefficients of the least squares residuals from the individual linear regressions in the panel, which is closely related to the C_{AVE} by Frees (1995). The advantage of Pesaran cross section dependence test (CD test) is that it is correctly centered for fixed N and T under the null of cross section independence assuming that the errors are symmetrically

distributed.

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. Such tests are important as parameter estimates may be inconsistent if cross section correlation of errors is not accounted for at the estimation stage. It is, therefore, important to establish whether the cross section error independence can be maintained prior to estimation and inference.

This paper explores the use of the LM and CD tests for nonlinear panel data models. In such models, the calculation of the errors is not as straightforward as in linear models. We consider two approaches to estimate the errors. The first estimate of the errors is the deviation of the observed dependent variable from its expected value, and the second estimate is to predict the errors conditional on the observed dependent variable, the so called generalized residual, see e.g. Gourieroux, Monfort and Trognon (1985). Based on the estimated residuals we introduce the LM and CD test statistics for nonlinear panel data model and, using Monte Carlo experiments, we examine the small sample performance of the tests for the probit and the Tobit model. Using data on voting in the U.S. Congress and campaign contributions by political lobby groups previously analyzed by Wawro (2001) we demonstrate the application of the test.

The next section introduces the nonlinear panel data model. Section 3 discusses the estimation of the residuals and the tests for cross section dependence. The small sample performance is evaluated using Monte Carlo experiments in Section 4, and the tests are applied to the data on voting in the U.S. Congress in Section 5. Finally, Section 6 provides some concluding remarks. Technical details are provided in Appendices A and B, and a bootstrap procedure to approximate the finite sample distribution of the CD test is discussed in Appendix C.

2 The nonlinear panel data model

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

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

where $\mathbf{x}_{it}$ is a $k \times 1$ vector of exogenous variables, $\boldsymbol{\theta}_i$ is a $q \times 1$ vector of parameters, ε_{it} is a scalar disturbance, N is the number of cross section observations, and T is the number of observations in time. The variable y_{it} is observed, which is related to the latent variable via the link function $\mathrm{g}(\cdot)$,

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

This general model encompasses many econometric models. Examples include binary choice models where

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

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. If the distribution of ε_{it} is the logistic, then this constitutes the logit model. If ε_{it} is standard normally distributed, then this is the probit model. The Tobit model is obtained if the latent model is that of equation (3), errors are normal and the link function is

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

The disturbances are assumed to be potentially contemporaneously correlated, $\boldsymbol{\varepsilon}_t \sim (\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{Nt})'$. This paper focuses on testing $\boldsymbol{\Sigma} = \mathbf{D}$ against $\boldsymbol{\Sigma} \neq \mathbf{D}$, where $\mathbf{D}$ is a diagonal matrix . Such an error structure could arise, for example, from the presence of unobserved common factors

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

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

3 Testing for cross section independence

Pesaran (2004) has suggested two approaches to test for cross section dependence using the pair-wise correlation coefficients of the residuals in the regression equations of the ith and j^{th} unit, $\tilde{\rho}_{ij}$. One is the LM test of Breusch and Pagan (1980)

$$\mathrm{LM} = \sqrt{\frac{1}{N(N-1)}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} (T\tilde{\rho}_{ij}^2 - 1). \tag{6}$$

The other is the CD test

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

It is clear that, in contrast to the LM test, the CD test requires the cross section correlation to be different from zero on average to detect deviations from cross-section independence. While we believe that this is not a restrictive assumption for most real life situations, this limitation should be borne in mind when applying the CD test.

For linear models residuals are estimated directly from the underlying linear regressions. For nonlinear models the concept of a residual is ambiguous and can be defined in a number of different ways. One possibility would be to define the disturbances of the nonlinear models analogous to the linear case as the deviation of the observed dependent variable from its expected value

$$u_{it} = y_{it} - \mathrm{E}(y_{it}|\mathbf{x}_{it}, \boldsymbol{\theta}_i) \tag{8}$$

with an estimated counterpart, the residual, given by

$$\tilde{u}_{it} = y_{it} - \mathrm{E}\left(y_{it}|\mathbf{x}_{it}, \tilde{\boldsymbol{\theta}}_i\right), \tag{9}$$

where $\tilde{\boldsymbol{\theta}}_i$ is a consistent estimator of $\boldsymbol{\theta}_i$ under the null of cross section independence. For the probit model, for example, this deviation is given by

$$\tilde{u}_{it} = y_{it} - \Phi(\tilde{\boldsymbol{\beta}}_i'\mathbf{x}_{it}), \tag{10}$$

and for the Tobit model it is

$$\tilde{u}_{it} = y_{it} - \left(\tilde{\boldsymbol{\beta}}_i'\mathbf{x}_{it} + \tilde{\sigma}_i\tilde{\lambda}_{it}\right)\Phi\left(\frac{\tilde{\boldsymbol{\beta}}_i'\mathbf{x}_{it}}{\tilde{\sigma}_i}\right), \tag{11}$$

where

$$\tilde{\lambda}_{it} = \phi\left(\frac{\tilde{\boldsymbol{\beta}}_i'\mathbf{x}_{it}}{\tilde{\sigma}_i}\right)\left[\Phi\left(\frac{\tilde{\boldsymbol{\beta}}_i'\mathbf{x}_{it}}{\tilde{\sigma}_i}\right)\right]^{-1},$$

and $\phi(\cdot)$ is the standard normal probability density function (pdf), and $\Phi(\cdot)$ is the standard normal cumulative distribution function (cdf).

For many models, the residuals will be heteroskedastic. One can transform them into homoskedastic residuals by dividing the estimated residuals by their estimated standard errors. In the case of the probit model the "standardized residual" is defined as

$$\tilde{u}_{it}^* = \frac{y_{it} - \Phi(\tilde{\boldsymbol{\beta}}_i'\mathbf{x}_{it})}{\sqrt{\Phi(\tilde{\boldsymbol{\beta}}_i'\mathbf{x}_{it})(1 - \Phi(\tilde{\boldsymbol{\beta}}_i'\mathbf{x}_{it}))}}, \tag{12}$$

and, in the case of the Tobit model, it is

$$\tilde{u}_{it}^* = \tilde{u}_{it}/\tilde{\omega}_{it} \tag{13}$$

where

$$\tilde{\omega}_{it}^2 = \left[(\tilde{\boldsymbol{\beta}}_i'\mathbf{x}_{it})^2 + \tilde{\boldsymbol{\beta}}_i'\mathbf{x}_{it}\tilde{\sigma}_i\tilde{\lambda}_{it} + \tilde{\sigma}_i^2 - (\tilde{\boldsymbol{\beta}}_i'\mathbf{x}_{it} + \tilde{\sigma}_i\tilde{\lambda}_{it})^2\Phi\left(\frac{\tilde{\boldsymbol{\beta}}_i'\mathbf{x}_{it}}{\tilde{\sigma}_i}\right)\right]\Phi\left(\frac{\tilde{\boldsymbol{\beta}}_i'\mathbf{x}_{it}}{\tilde{\sigma}_i}\right).$$

The derivations of the residuals and their variances is given in Appendix A.

Alternatively, the computation of ρ_{ij} can be based on the generalized residual proposed by Gourieroux, Monfort, and Trognon (1985), Gourieroux,

Monfort, Renault, and Trognon (1987), and Chesher and Irish (1987). The generalized residuals are defined as

$$u_{it}^g = \mathrm{E}_0(\mathrm{f}(y_{it}^*, \mathbf{x}_{it}, \boldsymbol{\theta}_i) \,|y_{it}) = \psi_0(y_{it}, \mathbf{x}_{it}, \boldsymbol{\theta}_i), \qquad (14)$$

where $\mathrm{E}_0(.)$ is the expectation operator under the null hypothesis of no cross section dependence. In contrast to the residual (8), the expectation of the generalized residuals in (14) is conditional on the observed dependent variable, y_{it}. An estimator of $\psi_0(y_{it}, \mathbf{x}_{it}, \boldsymbol{\theta}_i)$ is given by

$$\tilde{u}_{it}^g = \psi_0(y_{it}, \mathbf{x}_{it}, \tilde{\boldsymbol{\theta}}_i). \qquad (15)$$

Using equation (15), the "generalized residual" for the probit models is given by

$$\tilde{u}_{it}^g = \frac{\phi(\tilde{\boldsymbol{\beta}}_i' \mathbf{x}_{it})}{\Phi(\tilde{\boldsymbol{\beta}}_i' \mathbf{x}_{it})[1 - \Phi(\tilde{\boldsymbol{\beta}}_i' \mathbf{x}_{it})]} [y_{it} - \Phi(\tilde{\boldsymbol{\beta}}_i' \mathbf{x}_{it})]. \qquad (16)$$

For the Tobit model we have

$$\tilde{u}_{it}^g = (y_{it} - \tilde{\boldsymbol{\beta}}_i \mathbf{x}_{it})\mathrm{I}(y_{it}) - \tilde{\sigma}_i \frac{\phi(\tilde{\boldsymbol{\beta}}_i \mathbf{x}_{it}/\tilde{\sigma}_i)}{\Phi(-\tilde{\boldsymbol{\beta}}_i \mathbf{x}_{it}/\tilde{\sigma}_i)}[1 - \mathrm{I}(y_{it})], \qquad (17)$$

where $\tilde{\sigma}_i$ is the estimated standard deviation of the error term (Chesher and Irish (1987)).

The generalized residual is also heterskedastic for many models. In Appendix A we give the variances for the probit and the Tobit model, which could be used to obtain standardized versions of the generalized residuals. However, in line with the literature we will use the residuals (16) and (17) below.

In general ρ_{ij} can be estimated using any of the residuals defined above. For example, using the residuals, $\tilde{u}_{it}$, we have

$$\tilde{\rho}_{ji} = \frac{\sum_{t=1}^T (\tilde{u}_{it} - \bar{\bar{u}}_i)(\tilde{u}_{jt} - \bar{\bar{u}}_j)}{\left(\sum_{t=1}^T (\tilde{u}_{it} - \bar{\bar{u}}_i)^2\right)^{1/2} \left(\sum_{t=1}^T (\tilde{u}_{jt} - \bar{\bar{u}}_j)^2\right)^{1/2}}, \qquad (18)$$

where

$$\bar{\bar{u}}_i = T^{-1} \sum_{t=1}^T \tilde{u}_{it}.$$

The estimated correlation coefficient, $\tilde{\rho}_{ji}$, is then used in equations (6) and (7) to obtain the LM and CD test statistics. For large T $\bar{\bar{u}}_i$ will tend to zero and could be ignored but for better small sample performance the mean correction might be desirable.

Under the null hypothesis and for T sufficiently large, $\tilde{\rho}_{ji} \xrightarrow{p} 0$, for each i and j. However, the probability limit of $\tilde{\rho}_{ji}$ will differ from zero in the presence of cross section dependence. Under the null of cross section independence and for sufficiently large N and T, the CD statistic tends to a standard normal variate. See Appendix B for a proof and precise mathematical conditions.

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,

$$y_{it}^{*(r)} = \alpha_i^{(r)} + \beta' x_{it}^{(r)} + \varepsilon_{it}^{(r)}, \tag{19}$$

where $i = 1, 2, \ldots, N$, $t = 1, 2, \ldots, T$, and r, $r = 1, 2, \ldots, R$, denotes the replication index in the Monte Carlo experiments with $R = 1000$, $\beta = 1$. The regressors are generated as

$$x_{it}^{(r)} = \delta f_t^{(r)} + \eta_{it}^{(r)}$$

$\eta_{it}^{(r)} = \lambda \eta_{i,t-1}^{(r)} + \zeta_{it}^{(r)}$, $\zeta_{it}^{(r)} \sim \text{iid N}(0, 1)$, and $f_t^{(r)} \sim \text{iid N}(0, 1)$. We set $\delta = 1$ and $\lambda = 0.5$. Finally,

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

where $\bar{x}_i^{(r)} = \sum_{t=1}^T x_{it}^{(r)}/T$, $\hat{S}_{xi}^{(r)} = \left[(N-1)^{-1} \sum_{i=1}^N (\bar{x}_i^{(r)} - \bar{\bar{x}}^{(r)})^2 \right]^{1/2}$, $\bar{\bar{x}}^{(r)} = \sum_{i=1}^N \bar{x}_i^{(r)}/N$, and $\nu_i^{(r)} \sim \text{iid N}(0, 1)$. 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 estimation of β under a probit specification only makes use of $y_{it}^{(r)} = \text{I}\left(y_{it}^{*(r)}\right)$, and under the Tobit specification $y_{it}^{(r)} = y_{it}^{*(r)} \text{I}\left(y_{it}^{*(r)}\right)$. Hence, without loss of generality the variance of the error term, $u_{it}^{(r)}$, may be set equal to unity. To allow for correlation across the errors of different cross section units we adopt the following standardized one-factor structure

$$\varepsilon_{it}^{(r)} = \frac{\gamma_i^{(r)} f_t^{(r)} + e_{it}^{(r)}}{\sqrt{1 + \gamma_i^{(r)2}}}$$

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

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

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

Using the artificial data, 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, $\tilde{u}^*_{it}$ and $\tilde{u}^g_{it}$, and the CD and LM test statistics are then calculated with mean corrections given in (18).

4.2 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 Tobit model. The results in these tables suggest the following.

(i) There are substantial size distortions for the LM test even when N or T or both are large.

(ii) The empirical size is close to the nominal size for the CD test even for N and T as small as 10. This result holds generally and does not require the fixed effects and the regressors to be uncorrelated.

(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 0.9. On the other hand, when $N = 20$ and $T = 100$, the power of CD test is about $0.6 \sim 0.7$.

(iv) The test results are robust to the way residuals from the nonlinear models are computed.

(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

We illustrate the use the use of the CD test by re-analyzing the data on voting and campaign contributions of Wawro (2001). Using a panel probit model 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 the voting behaviour of members of the US Congress with the unemployment rate in the constituency of each member of Congress as an additional explanatory variable.

The data set, which is available from Prof. Wawro's web page[1], contains data for a selections of the roll-call votes for each session of the 102th, 103rd and 104th Congress, where the selected roll-call votes are those that

---

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

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

$T\backslash N$	Standardized residuals, $\tilde{u}_{it}^*$					Generalized residuals, $\tilde{u}_{it}^g$				
	10	20	30	50	100	10	20	30	50	100
					$\gamma_i^{(r)} = 0,\ \forall i$					
					CD test					
10	0.064	0.059	0.077	0.064	0.068	0.066	0.063	0.078	0.057	0.054
20	0.057	0.056	0.059	0.058	0.075	0.056	0.055	0.051	0.059	0.072
30	0.054	0.052	0.063	0.048	0.049	0.054	0.053	0.062	0.047	0.051
50	0.050	0.045	0.061	0.059	0.061	0.048	0.045	0.060	0.056	0.062
100	0.046	0.051	0.060	0.067	0.062	0.048	0.057	0.060	0.066	0.061
					LM test					
10	0.197	0.459	0.698	0.974	1.000	0.192	0.456	0.712	0.975	1.000
20	0.085	0.203	0.356	0.656	0.985	0.093	0.224	0.419	0.734	0.996
30	0.077	0.132	0.235	0.394	0.858	0.074	0.160	0.290	0.527	0.940
50	0.070	0.087	0.099	0.196	0.491	0.076	0.113	0.152	0.346	0.784
100	0.060	0.083	0.067	0.083	0.195	0.063	0.091	0.095	0.186	0.534
					$\gamma_i^{(r)} \sim \text{U}(0.1, 0.3)$					
					CD test					
10	0.097	0.153	0.235	0.381	0.702	0.098	0.141	0.231	0.365	0.674
20	0.128	0.220	0.361	0.596	0.903	0.136	0.235	0.371	0.596	0.897
30	0.140	0.281	0.473	0.736	0.973	0.149	0.295	0.487	0.738	0.978
50	0.190	0.404	0.639	0.908	0.998	0.188	0.424	0.657	0.909	0.999
100	0.246	0.634	0.891	0.993	1.000	0.266	0.660	0.898	0.994	1.000
					LM test					
10	0.193	0.475	0.728	0.971	1.000	0.186	0.485	0.733	0.973	1.000
20	0.126	0.244	0.383	0.759	0.991	0.130	0.264	0.429	0.821	0.995
30	0.104	0.175	0.261	0.559	0.943	0.105	0.191	0.311	0.652	0.974
50	0.097	0.142	0.240	0.437	0.868	0.092	0.154	0.283	0.548	0.952
100	0.095	0.145	0.238	0.440	0.880	0.102	0.158	0.290	0.548	0.957
					$\gamma_i^{(r)} \sim \text{U}(-0.2, 0.6)$					
					CD test					
10	0.104	0.126	0.206	0.347	0.642	0.101	0.129	0.200	0.329	0.615
20	0.124	0.195	0.298	0.545	0.838	0.132	0.203	0.295	0.543	0.836
30	0.136	0.270	0.394	0.661	0.929	0.141	0.280	0.396	0.667	0.934
50	0.170	0.358	0.550	0.808	0.992	0.181	0.366	0.561	0.818	0.993
100	0.243	0.501	0.766	0.938	1.000	0.251	0.502	0.773	0.941	1.000
					LM test					
10	0.206	0.450	0.730	0.974	1.000	0.208	0.465	0.737	0.974	1.000
20	0.147	0.286	0.466	0.822	0.996	0.145	0.306	0.518	0.871	0.996
30	0.114	0.222	0.402	0.681	0.988	0.114	0.268	0.457	0.774	0.996
50	0.136	0.252	0.403	0.681	0.975	0.140	0.281	0.481	0.786	0.996
100	0.176	0.373	0.613	0.854	0.999	0.198	0.427	0.687	0.919	1.000

The Table reports the percentage of rejections of the null of no cross-section dependence for the CD-statistic (7) and the LM-statistic (6) from the standardized and the generalized residuals at the 5% significance level for 1000 repetitions of the experiment.

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

$T\backslash N$	Standardized residuals, $\tilde{u}_{it}^{*}$					Generalized residuals, $\tilde{u}_{it}^{g}$				
	10	20	30	50	100	10	20	30	50	100
					$\gamma_i^{(r)}=0,\ \forall i$					
					CD test					
10	0.063	0.080	0.060	0.066	0.072	0.062	0.048	0.068	0.054	0.057
20	0.059	0.055	0.047	0.055	0.066	0.048	0.048	0.070	0.053	0.056
30	0.069	0.066	0.063	0.067	0.059	0.049	0.064	0.059	0.055	0.047
50	0.048	0.052	0.052	0.043	0.064	0.050	0.055	0.045	0.057	0.060
100	0.038	0.046	0.062	0.060	0.053	0.045	0.065	0.048	0.038	0.046
					LM test					
10	0.184	0.461	0.737	0.984	1.000	0.201	0.498	0.765	0.982	1.000
20	0.122	0.253	0.406	0.722	0.985	0.137	0.288	0.501	0.822	0.996
30	0.114	0.160	0.260	0.467	0.893	0.102	0.214	0.374	0.668	0.983
50	0.069	0.129	0.153	0.245	0.589	0.080	0.126	0.246	0.537	0.941
100	0.067	0.078	0.092	0.134	0.255	0.084	0.070	0.182	0.362	0.882
					$\gamma_i^{(r)}\sim \mathrm{U}(0.1,0.3)$					
					CD test					
10	0.144	0.267	0.388	0.600	0.881	0.133	0.261	0.391	0.613	0.884
20	0.173	0.379	0.545	0.786	0.971	0.200	0.371	0.588	0.808	0.978
30	0.174	0.474	0.674	0.892	0.999	0.221	0.497	0.710	0.924	0.996
50	0.261	0.620	0.846	0.984	1.000	0.301	0.682	0.883	0.990	1.000
100	0.402	0.841	0.983	1.000	1.000	0.481	0.894	0.989	1.000	1.000
					LM test					
10	0.225	0.542	0.784	0.986	1.000	0.254	0.550	0.783	0.994	1.000
20	0.193	0.368	0.596	0.871	0.999	0.173	0.391	0.591	0.894	1.000
30	0.132	0.316	0.467	0.788	0.991	0.147	0.327	0.521	0.845	0.996
50	0.149	0.298	0.492	0.761	0.985	0.138	0.316	0.525	0.851	0.997
100	0.189	0.367	0.557	0.848	1.000	0.195	0.403	0.653	0.931	1.000
					$\gamma_i^{(r)}\sim \mathrm{U}(-0.2,0.6)$					
					CD test					
10	0.144	0.243	0.356	0.574	0.821	0.143	0.264	0.360	0.572	0.845
20	0.151	0.326	0.448	0.725	0.945	0.180	0.351	0.527	0.745	0.964
30	0.178	0.397	0.569	0.821	0.988	0.205	0.431	0.628	0.865	0.995
50	0.264	0.511	0.707	0.975	0.999	0.285	0.557	0.765	0.935	0.999
100	0.361	0.693	0.874	0.989	1.000	0.429	0.730	0.907	0.992	1.000
					LM test					
10	0.228	0.581	0.829	0.991	1.000	0.230	0.559	0.838	0.989	1.000
20	0.183	0.443	0.675	0.925	1.000	0.209	0.474	0.714	0.954	0.999
30	0.216	0.410	0.648	0.914	1.000	0.224	0.464	0.746	0.956	1.000
50	0.267	0.498	0.708	0.913	1.000	0.274	0.572	0.820	0.979	1.000
100	0.345	0.685	0.878	0.985	1.000	0.438	0.822	0.964	0.999	1.000

The specification of the observed dependent variable is given in (5). Otherwise see footnote of Table 1.

the lobby groups themselves deemed important. We test for cross section independence in the sessions of the 104th Congress for the roll-call votes suggested by the USCC, which has the largest number of votes of the data sets. We exclude members of Congress that do not change their vote more than three times as the estimation of the regression equation for each individual separately would otherwise be computationally unstable. This leads to two data sets with $M = 19$ each, and $N = 139$ for the first session and $N = 145$ for the second session, where M is the number of motions that are put before Congress and are recorded in Prof. Wawro's data set—as the timing of the votes is not obvious we refer to the different motions for which the roll-call votes were recorded as $m = 1, 2, \ldots, M$ instead of t.

Wawro (2001, p.570) includes a motion-specific intercept to account for the particular political context around the roll-call votes. We do not include such a motion-specific intercept. However, if the political context around a motion influences the voting behaviour, then omitting a motion specific intercept is tantamount to introducing cross-correlation among the residuals. The CD test can therefore be interpreted as a test for the necessity to include a motion-specific intercept.

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

$$\mathcal{L}_i = \prod_{m=1}^{M} \Phi(\boldsymbol{\beta}_i \mathbf{x}_{im})^{y_{im}} (1 - \Phi(\boldsymbol{\beta}_i \mathbf{x}_{im}))^{(1-y_{im})}, \text{ for } i = 1, 2, \ldots, N \qquad (20)$$

where y_{im} is a binary indicator for the votes of the member of Congress ("aye" or "nay"), $\mathbf{x}_{im}$ contains an intercept, the contributions of the USCC, the contributions of the AFL-CIO, and the unemployment rate of the constituency of the member of Congress. Note that this model is more general than the random effects model estimated by Warwo (2001) as we also allow the slope parameters to vary between individuals.

Using the parameter vector $\tilde{\boldsymbol{\beta}}_i$, we calculate the conditional generalized residual as given in equation (16) and the unconditional residual of equation (12). From these we obtain the pairwise correlation coefficient. The average correlation coefficient for the conditional residuals are 0.283 for the first session and 0.310 for the second session. For the unconditional residuals, they are 0.286 and 0.310.

Using pairwise correlation coefficients we can calculate the CD test statistic. However, both panels are unbalanced as some observations are missing and, therefore, the statistic has to be adjusted to

$$\text{CD} = \sqrt{\frac{2}{N(N-1)}} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \sqrt{M_{ij}} \tilde{\rho}_{ij} \right), \qquad (21)$$

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

Table 3: CD test for roll-call votes in the 104th Congress

	$\tilde{u}_{it}^g$	Bootstrap 5% crit. val.		$\tilde{u}_{it}^*$	Bootstrap 5% crit. val.	
All motions						
1st session	151.488	$[-1.670$	$2.545]$	151.470	$[-1.677$	$2.539]$
2nd session	169.300	$[-1.655$	$2.392]$	169.310	$[-1.658$	$2.393]$
Subset of motions						
1st session	104.488	$[-1.727$	$2.092]$	104.496	$[-1.726$	$2.092]$
2nd session	118.742	$[-1.733$	$2.078]$	118.839	$[-1.726$	$2.078]$

The results for the subset of motions only use motions with less than 90% unanimity, which excludes two motions of each session of Congress. The bootstrap critical values were computed using 1000 iterations. The first number gives the 2.5% lower critical value and the second number the 2.5% upper critical value.

While the Monte Carlo results suggest that the CD test has the correct size for all combinations of N and T, it might be worthwhile to double check that the results for possible departures from the asymptotic test results. Hence, we also calculate critical values of the CD test using a bootstrap procedure, the details of which are given in Appendix C.

The upper half of Table 3 reports the results for the CD test for the entire sample. The test statistics are very large and clearly reject the null of cross section independence. The values in brackets are the 5% critical bootstrap values. The bootstrap test results are generally in line with asymptotic test results, and confirm the existence of statistically significant evidence of cross section error correlations in Warwo's application.

In order to address the question of whether our test results are driven by a few motions with near unanimity of the votes, we eliminated the motions where more than 90% of the votes are in agreement. This reduces the number of motions being considered, M, from 19 to 17 for each of the two sessions. We apply the CD tests to this subset and report the results in the lower half of Table 3. While the test statistics are reduced in size, they remain statistically highly significant. The CD test therefore still rejects the null hypothesis of cross section error independence in this empirical application.

6   Conclusion

In this paper, we have generalized Pesaran's (2004) CD test for cross section independence to nonlinear models. Our Monte Carlo studies show that there are substantial size distortions of the Lagrangian multiplier type tests. On the other hand, CD tests perform well even in small N and T cases. The

empirical size of the CD test is close to the nominal size. The test also has good power, in particular when N is large, even when T is relatively small.

The CD test is simple to implement. 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 exist 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 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 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 unconditional residual for the probit model

$$
\begin{aligned}
u_{it} &= y_{it} - \mathrm{E}(y_{it}) \\
&= y_{it} - \mathrm{E}(y_{it}|y_{it}=1)\mathrm{Pr}(y_{it}=1) - \mathrm{E}(y_{it}|y_{it}=0)\mathrm{Pr}(y_{it}=0) \\
&= y_{it} - \Phi(\boldsymbol{\beta}_i'\mathbf{x}_{it}),
\end{aligned}
$$

where for notational convenience the fact that the moments are conditional on $\mathbf{x}_{it}$ and the parameters is not stated explicitly. The variance is

$$
\begin{aligned}
\mathrm{Var}(u_{it}) &= \mathrm{E}\left(y_{it}^2\right) - \Phi(\boldsymbol{\beta}_i'\mathbf{x}_{it})^2 \\
&= \mathrm{E}(y_{it}^2|y=1)\mathrm{Pr}(y_{it}=1) + \mathrm{E}(y_{it}^2|y=0)\mathrm{Pr}(y_{it}=0) - \Phi(\boldsymbol{\beta}_i'\mathbf{x}_{it})^2 \\
&= \Phi(\boldsymbol{\beta}_i'\mathbf{x}_{it})\left[1 - \Phi(\boldsymbol{\beta}_i'\mathbf{x}_{it})\right].
\end{aligned}
$$

The generalized residual for the probit model

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

The variance of the generalized residual is

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

The unconditional residual for the Tobit model

$$
\begin{aligned}
u_{it} &= y_{it} - \mathrm{E}(y_{it}) \\
&= y_{it} - \mathrm{E}(y_{it}|y_{it}>1)\mathrm{Pr}(y_{it}>1) - \mathrm{E}(y_{it}|y_{it}=0)\mathrm{Pr}(y_{it}=0) \\
&= y_{it} - \mathrm{E}(\boldsymbol{\beta}_i'\mathbf{x}_{it} + u_{it}|u_{it} > -\boldsymbol{\beta}_i'\mathbf{x}_{it})\,\Phi\left(\boldsymbol{\beta}_i'\mathbf{x}_{it}/\sigma_i\right) \\
&= y_{it} - \left(\boldsymbol{\beta}_i'\mathbf{x}_{it} + \sigma_i\lambda_{it}\right)\Phi\left(\boldsymbol{\beta}_i'\mathbf{x}_{it}/\sigma_i\right),
\end{aligned}
$$

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

$-\boldsymbol{\beta}_i'\mathbf{x}_{it}/\sigma_i$. The variance is

$$
\begin{aligned}
\mathrm{Var}(u_{it}) &= \mathrm{E}\left(y_{it}^2\right) - \left[\left(\boldsymbol{\beta}_i'\mathbf{x}_{it} + \sigma_i\lambda_{it}\right)\Phi\left(\boldsymbol{\beta}_i'\mathbf{x}_{it}/\sigma_i\right)\right]^2 \\
&= \mathrm{E}(y_{it}^2|u_{it} > -\boldsymbol{\beta}_i'\mathbf{x}_{it})\mathrm{Pr}(y_{it} > 0) - \left[\left(\boldsymbol{\beta}_i'\mathbf{x}_{it} + \sigma_i\lambda_{it}\right)\Phi\left(\boldsymbol{\beta}_i'\mathbf{x}_{it}/\sigma_i\right)\right]^2 \\
&= \sigma_i^2\left\{\mathrm{E}((\boldsymbol{\alpha}_i'\mathbf{x}_{it} + v_{it})^2|v_{it} > -\boldsymbol{\alpha}_i'\mathbf{x}_{it})\Phi\left(\boldsymbol{\alpha}_i'\mathbf{x}_{it}\right)\right. \\
&\quad \left. - \left[\left(\boldsymbol{\alpha}_i'\mathbf{x}_{it} + \lambda_{it}\right)\Phi\left(\boldsymbol{\alpha}_i'\mathbf{x}_{it}\right)\right]^2\right\} \\
&= \sigma_i^2\left\{\left[(\boldsymbol{\alpha}_i'\mathbf{x}_{it})^2 + \boldsymbol{\alpha}_i'\mathbf{x}_{it}\lambda_{it} + 1\right]\Phi\left(\boldsymbol{\alpha}_i'\mathbf{x}_{it}\right)\right. \quad\quad (22)\\
&\quad \left. - \left[\left(\boldsymbol{\alpha}_i'\mathbf{x}_{it} + \lambda_{it}\right)\Phi\left(\boldsymbol{\alpha}_i'\mathbf{x}_{it}\right)\right]^2\right\},
\end{aligned}
$$

where $\boldsymbol{\alpha}_{it} = \boldsymbol{\beta}_{it}/\sigma_i$, $v_{it} = u_{it}/\sigma_i \sim \mathrm{N}(0,1)$, and we used the fact that

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

Substituting $\boldsymbol{\alpha}_{it}$ into (22) gives the variance in (13).

The generalized residual for the Tobit model

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

For the variance we have,

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

Appendix B: Limiting distribution of CD test for nonlinear models

In this appendix we show that the limiting distribution of Pesaran (2004) CD test holds for nonlinear panel data models. We consider a nonlinear model of the form

$$
y_{it} = \mathrm{f}(\mathbf{x}_{it}, \boldsymbol{\theta}_i) + u_{it},
$$

where $\boldsymbol{\theta}_i$ is a $p \times 1$ vector of unknown parameters for cross section unit i. We denote the true value of $\boldsymbol{\theta}_i$ by $\boldsymbol{\theta}_{i0}$, and make the following assumptions.

A1: For each i, the disturbances, u_{it}, are serially independent with zero means and variances, σ_i^2, such that $0 < \sigma_i^2 < \infty$.

A2: Under the null hypothesis defined by $H_0 : u_{it} = \sigma_i \epsilon_{it}$, with $\epsilon_{it} \sim iid(0,1)$ for all i and t, the disturbances, ϵ_{it}, are symmetrically distributed around 0.

A3: The $k \times 1$ explanatory variables, $\mathbf{x}_{it}$, are strictly exogenous such that $\mathrm{E}(u_{it}|\mathbf{x}_i) = 0$ for all i and t, where $\mathbf{x}_i = (\mathbf{x}'_{i1}, \ldots, \mathbf{x}'_{iT})'$, such that $\frac{1}{T}\sum_{t=1}^T u_{it}\frac{\partial f_{jt}}{\partial \boldsymbol{\theta}_{j0}} \xrightarrow{p} 0$, $\forall\, i, j$, and t.

A4: Let Θ_i be an open neighborhood of $\boldsymbol{\theta}_i$ and $f_{it} = f(\mathbf{x}_{it}, \boldsymbol{\theta}_i)$, f_{it} is continuous in $\boldsymbol{\theta}_i \in \Theta_i$ uniformly in t.

A5: $\frac{\partial f_{it}}{\partial \boldsymbol{\theta}_i}$ exists and is continuous on Θ_i and $\frac{1}{T}\sum_{t=1}^T \frac{\partial f_{it}}{\partial \boldsymbol{\theta}_i}\frac{\partial f_{it}}{\partial \boldsymbol{\theta}'_i} \xrightarrow{p} \boldsymbol{\Omega}_{ii}$ and $\frac{1}{T}\sum_{t=1}^T \frac{\partial f_{jt}}{\partial \boldsymbol{\theta}_j}\frac{\partial f_{it}}{\partial \boldsymbol{\theta}'_i} \xrightarrow{p} \boldsymbol{\Omega}_{ij}$, where $\boldsymbol{\Omega}_{ii}$ and $\boldsymbol{\Omega}_{ij}$ are finite, non-stochastic matrices, and convergence is uniformly for all $\boldsymbol{\theta}_i \in \Theta_i$.

A6: $\frac{\partial^2 f_{it}}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}'_i}$ is continuous in $\boldsymbol{\theta}_i \in \Theta_i$ uniformly in t, and $\frac{1}{T}\sum_{t=1}^T \frac{\partial^2 f_{it}}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}'_i}$ converges to a finite nonsingular matrix.

A7: $\sqrt{T}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0}) \overset{a}{\sim} \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a positive definite matrix.

Let $\mathcal{L}_i$ denote the log-likelihood function of the ith cross section unit with the joint pdf $\ell_i = \ell(u_{i1}, \ldots, u_{iT}) = \prod_{t=1}^T \ell(u_{it})$, and denote by $\hat{\boldsymbol{\theta}}_i$ the maximum likelihood estimator of $\boldsymbol{\theta}_i$,

$$\left.\frac{\partial \mathcal{L}_i}{\partial \boldsymbol{\theta}_i}\right|_{\hat{\boldsymbol{\theta}}_i} = \mathbf{0}.$$

Then as $T \longrightarrow \infty$, $\hat{\boldsymbol{\theta}}_i$ is consistent and asymptotically normally distributed with

$$\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0} = -\left(\left.\frac{\partial^2 \mathcal{L}_i}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}'_i}\right|_{\boldsymbol{\theta}_{i0}}\right)^{-1}\left(\left.\frac{\partial \mathcal{L}_i}{\partial \boldsymbol{\theta}_i}\right|_{\boldsymbol{\theta}_{i0}}\right) + O_p\left(\frac{1}{T}\right).$$

The estimated residuals, $\hat{u}_{it}$,

$$\begin{aligned}
\hat{u}_{it} &= y_{it} - f(\mathbf{x}_{it}, \hat{\boldsymbol{\theta}}_i) \\
&= y_{it} - f(\mathbf{x}_{it}, \boldsymbol{\theta}_{i0}) - \left.\frac{\partial f_{it}}{\partial \boldsymbol{\theta}'_i}\right|_{\boldsymbol{\theta}_{i0}}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0}) + O_p\left(\frac{1}{T}\right).
\end{aligned}$$

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \hat{u}_{it} \hat{u}_{jt} \;=\; \frac{1}{\sqrt{T}} \sum_{t=1}^{T} u_{it} u_{jt}$$

$$+ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0})' \left. \frac{\partial \mathrm{f}_{it}}{\partial \boldsymbol{\theta}_i} \right|_{\boldsymbol{\theta}_{i0}} \left. \frac{\partial \mathrm{f}_{jt}}{\partial \boldsymbol{\theta}_j'} \right|_{\boldsymbol{\theta}_{j0}} (\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_{j0}) \right]$$

$$- \frac{1}{\sqrt{T}} \sum_{t=1}^{T} u_{it} \left. \frac{\partial \mathrm{f}_{jt}}{\partial \boldsymbol{\theta}_j'} \right|_{\boldsymbol{\theta}_{j0}} (\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_{j0})$$

$$- \frac{1}{\sqrt{T}} \sum_{t=1}^{T} u_{jt} \left. \frac{\partial \mathrm{f}_{it}}{\partial \boldsymbol{\theta}_i'} \right|_{\boldsymbol{\theta}_{i0}} (\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0}) + O_p\left(\frac{1}{T}\right)$$

$$= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} u_{it} u_{jt}$$

$$+ \frac{1}{\sqrt{T}} \left[\sqrt{T}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0})' \left(\frac{1}{T} \sum_{t=1}^{T} \left. \frac{\partial \mathrm{f}_{it}}{\partial \boldsymbol{\theta}_i} \right|_{\boldsymbol{\theta}_{i0}} \left. \frac{\partial \mathrm{f}_{jt}}{\partial \boldsymbol{\theta}_j'} \right|_{\boldsymbol{\theta}_{j0}} \right) \sqrt{T}(\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_{j0}) \right]$$

$$- \left(\frac{1}{T} \sum_{t=1}^{T} u_{it} \left. \frac{\partial \mathrm{f}_{jt}}{\partial \boldsymbol{\theta}_j'} \right|_{\boldsymbol{\theta}_{j0}} \right) \sqrt{T}(\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_{j0})$$

$$- \left(\frac{1}{T} \sum_{t=1}^{T} u_{jt} \left. \frac{\partial \mathrm{f}_{it}}{\partial \boldsymbol{\theta}_i'} \right|_{\boldsymbol{\theta}_{i0}} \right) \sqrt{T}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0}) + O_p\left(\frac{1}{T}\right)$$

Hence, using assumptions A3, A5, and A6 it follows that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \hat{u}_{it} \hat{u}_{jt} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} u_{it} u_{jt} + O_p\left(\frac{1}{T}\right).$$

Therefore, following the same argument of Pesaran (2004), we can show that the limiting distribution of the CD test continues to hold in the case of nonlinear panel data models as well.

Appendix C: 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(·) in equation (2). 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 $\hat{\varepsilon}_{it} \sim \text{iid}\, \mathrm{F}(0, \tilde{\sigma}_i^2)$ 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 for 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 the, 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

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

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, and Alain Trognon (1985) 'A general approach to serial correlation.' *Econometric Theory* 1, 315–340

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 (1999) 'A generalized moments estimator for the autoregressive parameter in a spatial model.' *International Economic Review*, 40, 509–533

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

Lee, Lung-Fei (2002) 'Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models.' *Econometric Theory* 18, 252–277

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

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

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 (2006) 'A bias adjusted LM test of error cross section independence.'*Cambridge Working Paper in Economics* 0641

Ullah, Aman (2004) *Finite Sample Econometrics*, New York: Oxford University Press

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