A test of cross section dependence for a linear dynamic panel model with regressors

Sarafidis, Vasilis; Yamagata, Takashi; Robertson, Donald

Postprint / Postprint
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
www.peerproject.eu

Accepted Manuscript

A test of cross section dependence for a linear dynamic panel model with regressors

Vasilis Sarafidis, Takashi Yamagata, Donald Robertson

Please cite this article as: Sarafidis, V., Yamagata, T., Robertson, D., A test of cross section dependence for a linear dynamic panel model with regressors. *Journal of Econometrics* (2008), doi:10.1016/j.jeconom.2008.10.006

A Test of Cross Section Dependence for a Linear Dynamic Panel Model with Regressors[*]

Vasilis Sarafidis[◇]   Takashi Yamagata[△†]

Donald Robertson[♣]

[◇]Discipline of Econometrics and Business Statistics, University of Sydney
[△]Department of Economics and Related Studies, University of York
[♣]Faculty of Economics, University of Cambridge

23 September 2008

Abstract

This paper proposes a new testing procedure for detecting error cross section dependence after estimating a linear dynamic panel data model with regressors using the generalised method of moments (GMM). The test is valid when the cross-sectional dimension of the panel is large relative to the time series dimension. Importantly, our approach allows one to examine whether any error cross section dependence remains after including time dummies (or after transforming the data in terms of deviations from time-specific averages), which will be the case under heterogeneous error cross section dependence. Finite sample simulation-based results suggest that our tests perform well, particularly the version based on the Blundell and Bond (1998) system GMM estimator. In addition, it is shown that the system GMM estimator, based only on partial instruments consisting of the regressors, can be a reliable alternative to the standard GMM estimators under heterogeneous error cross section dependence. The proposed tests are applied to employment equations using UK firm data and the results show little evidence of heterogeneous error cross section dependence.

1

1 Introduction

During the past decade a substantial literature has been developed analysing the effects of cross section dependence as well as advancing ways of dealing with it in panel data models. Cross section dependence may arise for several reasons − often, due to spatial correlations, economic distance and common unobserved shocks. In the case of spatial attributes, where a natural immutable distance measure is available, the dependence may be captured through spatial lags using techniques that are familiar from the time series literature (Anselin, 1988, 2001). In economic applications, spatial techniques are often adapted using alternative measures of economic distance (see e.g. Conley, 1999, Kapoor, Kelejian and Prucha, 2004, Lee, 2004, Lee, 2007, and others). There are several contributions in the literature that allow for time-varying individual effects (Holtz-Eakin, Newey and Rosen, 1988, Ahn, Lee and Schmidt, 2001 and Han, Orea and Schmidt, 2005). Recently, a number of researchers have modelled cross section dependence by restricting the covariance matrix of the errors using a common factor specification with a fixed number of unobserved factors and individual-specific factor loadings that give rise to heterogenous cross section dependence (see Forni and Reichlin, 1998, Robertson and Symons, 2000, Phillips and Sul, 2003, Stock and Watson, 2002, Bai and Ng, 2004, Moon and Perron, 2004, Pesaran, 2006, among others). The factor structure approach is widely used because it can approximate a wide variety of error cross section dependence. For example, in a panel data set of firms we may think of the factors as capturing fluctuations in economic activity or changes in regulatory policy for the industry as a whole, and so on. The impact of these factors will vary across firms, due to differences in size, liquidity constraints, market share etc. In a macro panel data model, the factors may represent a general demand shock or an oil price shock with the factor loadings reflecting the relative openness of the economies, differences in technological constraints, and so on.[1]

In the literature of estimating linear dynamic panel data models with a large number of cross-sectional units (N) and a moderately small number of time series observations (T), generalised method of moments (GMM) estimators are widely used, such as those proposed by Arellano and Bond (1991), Ahn and Schmidt (1995), Arellano and Bover (1995) and Blundell and Bond (1998). These methods typically assume that the disturbances are cross-sectionally independent. On the other hand, in empirical applications it is common practice to include time dummies, or, equivalently, to transform the observations in terms of deviations from time-specific averages (i.e. to cross-sectionally demean the data) in order to eliminate any common time-varying shocks; see, for example, Arellano and Bond (1991) and Blundell and Bond (1998). This transformation will marginal out these common effects, unless their impact differs across cross-sectional units (heterogeneous cross section dependence). In this case, the standard GMM estimators used in the literature will not be consistent, as shown by Sarafidis and Robertson (2007) and in the current paper.

Several tests for cross section dependence have been proposed in the econometric

---

[1] Other examples are provided by Ahn, Lee and Schmidt (2001).

literature. The most widely used test is perhaps the Lagrange Multiplier (LM) test proposed by Breusch and Pagan (1980), which is based on the squared pair-wise Pearson's correlation coefficients of the residuals. This test requires T being much larger than N. Frees (1995) proposed a cross section dependence test that is based on the squared Spearman rank correlation coefficients and allows N to be larger than T. Recently, Ng (2006) has developed tools for guiding practitioners as to how much residual cross section correlation is in the data and which cross-sectional units are responsible for this in particular, through tests that are based on probability integral transformations of the ordered residual correlations. However, the proposed procedures are valid only in panels for which $\sqrt{T}$-consistent estimates are available. Pesaran, Ullah and Yamagata (2006) developed bias-adjusted normal approximation versions of the LM test of Breusch and Pagan (1980), which are valid for large-N panel data models but with strictly exogenous regressors only. Pesaran (2004) proposed another test for cross section dependence, called the CD test, which is closely related to Friedman's (1937) test statistic. Pesaran showed that the CD test can also be applied to a wide variety of models, including heterogeneous dynamic models with multiple breaks and non-stationary dynamic models with small/large N and T. However, as Frees (1995) implied and Pesaran (2004) pointed out, the problem of the CD test is that in a stationary dynamic panel data model it will fail to reject the null of error cross section independence when the factor loadings have zero mean in the cross-sectional dimension. It follows that the CD test will have poor power properties when it is applied to a regression with time dummies or on cross-sectionally demeaned data.

This paper proposes a new testing procedure for error cross section dependence after estimating a linear dynamic panel data model with covariates by the generalised method of moments. This is valid when N is large relative to T. Importantly, unlike the CD test, our approach allows one to examine whether any error cross section dependence remains after including time dummies, or after transforming the data in terms of deviations from time-specific averages, which will be the case under heterogeneous error cross section dependence.

The small sample performance of our proposed tests is investigated by means of Monte Carlo experiments and we show that they have correct size and satisfactory power for a wide variety of simulation designs. Furthermore, the paper suggests a consistent GMM estimator under heterogeneous error cross section dependence. Results on the finite sample properties of the estimator are reported and discussed.

Our proposed tests and estimators are applied to employment equations using UK firm data, and it is found that there is little evidence of heterogeneous cross section dependence in this data set.

The remainder of the paper proceeds as follows. Section 2 reviews some relevant existing tests for error cross section dependence. Section 3 proposes a new test for cross section dependence and a consistent GMM estimator under these circumstances. Section 4 reports the results of our Monte Carlo experiments. Section 5 illustrates an empirical application of our approach. Finally, Section 6 contains concluding remarks.

3

2 Existing Tests for Cross Section Dependence

Consider a panel data model

$$y_{it} = \alpha_i + \boldsymbol{\beta'}\mathbf{x}_{it} + u_{it},\, i = 1, 2, \ldots, N,\, t = 1, 2, \ldots, T, \tag{1}$$

where the u_{it} may exhibit cross section dependence. The hypothesis of interest is

$$H_0 : E(u_{it}u_{jt}) = 0 \quad \forall\, t \text{ for all } i \neq j, \tag{2}$$

vs

$$H_1 : E(u_{it}u_{jt}) \neq 0 \text{ for some } t \text{ and some } i \neq j, \tag{3}$$

where the number of possible pairings (u_{it}, u_{jt}) rises with N. In the literature several tests for error cross section dependence have been proposed, and some relevant ones are discussed in this section.

2.1 Breusch-Pagan (1980) Lagrange Multiplier Test

Breusch and Pagan (1980) proposed a Lagrange multiplier (LM) statistic for testing the null of zero cross-equation error correlations, which is defined as

$$LM = T \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\rho}_{ij}^2, \tag{4}$$

where $\hat{\rho}_{ij}$ is the sample estimate of the pair-wise Pearson correlation coefficient of the residuals

$$\hat{\rho}_{ij} = \hat{\rho}_{ji} = \frac{\sum_{t=1}^{T} e_{it}e_{jt}}{\left(\sum_{t=1}^{T} e_{it}^2\right)^{1/2} \left(\sum_{t=1}^{T} e_{jt}^2\right)^{1/2}}, \tag{5}$$

where e_{it} is the Ordinary Least Squares (OLS) estimate of u_{it} in (1). LM is asymptotically distributed as chi-squared with $N(N-1)/2$ degrees of freedom under the null hypothesis, as $T \to \infty$ with N fixed.

2.2 Pesaran's (2004) CD Test

Recently Pesaran (2004) proposed another test for cross section dependence, called CD test, which allows for a flexible model structure, including fairly general heterogeneous dynamic models and nonstationary models. The test statistic is defined as

$$CD = \sqrt{\frac{2T}{N(N-1)}} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\rho}_{ij} \right). \tag{6}$$

4

For sufficiently large N and T, the CD test statistic tends to a standard normal variate under the null of cross section independence.[2] The finite sample evidence in Pesaran (2004) shows that the estimated size of the test is very close to the nominal level for any combinations of N and T considered. As Pesaran (2004) notes, there are two important cases in which the CD test can be unreliable. Firstly, when the distribution of the errors is not symmetric, the CD test becomes invalid and it may not have correct size.[3] Secondly, the CD test may lack power towards some directions of alternatives. To see this, consider the following single-factor structure for the error process

$$u_{it} = \phi_i f_t + \varepsilon_{it},$$

where ϕ_i is a factor loading that is fixed and bounded, f_t is an unobserved common factor such that $f_t \sim i.i.d.(0,1)$, $\varepsilon_{it} \sim i.i.d.(0,\sigma^2)$ and $E(f_t\varepsilon_{it}) = 0$ for all i and t. The common factor f_t generates error cross section dependence because of the fact that $cov(u_{it}, u_{jt}) = \phi_i\phi_j$, and the power of the CD test hinges on this non-zero covariance. Now suppose that $\phi_i \sim i.i.d.(0,\sigma_\phi^2)$ and ϕ_i is uncorrelated with f_t and ε_{it}. In this case, $cov(u_{it}, u_{jt}) = E(\phi_i)E(\phi_j) = 0$, even if there does exist (potentially large) error cross section dependence.

In the next section, we elaborate on the stochastic properties of the factors and factor loadings, and develop a new cross section dependence test.

3 Sargan's Difference Tests for Heterogeneous Error Cross Section Dependence in a Linear Dynamic Model with Regressors

3.1 Model Specification

Consider the following model

$$y_{it} = \alpha_i + \lambda y_{i,t-1} + \boldsymbol{\beta}'\mathbf{x}_{it} + u_{it}, \ i = 1, 2, ..., N, \ t = 1, 2, ..., T, \tag{7}$$

where $|\lambda| < 1$, $\boldsymbol{\beta}$ is a $(K \times 1)$ parameter vector that is bounded and non-zero, $\mathbf{x}_{it}$ is a $(K \times 1)$ vector of regressors with $\mathbf{x}_{it} = (x_{1it}, x_{2it}, ..., x_{Kit})'$, α_i is a random effect with finite mean and finite variance, and u_{it} has a multi-factor structure such that

$$u_{it} = \boldsymbol{\phi}_i'\mathbf{f}_t + \varepsilon_{it}, \tag{8}$$

where $\boldsymbol{\phi}_i = (\phi_{1i}, \phi_{2i}, ..., \phi_{Mi})'$ is a $(M \times 1)$ vector of factor loadings that is assumed to be $i.i.d.(\boldsymbol{\phi}, \boldsymbol{\Sigma}_\phi)$ with $\boldsymbol{\Sigma}_\phi$ being a positive semi-definite matrix, $\mathbf{f}_t = (f_{1t}, f_{2t}, ..., f_{Mt})'$ is a

---

[2] As Frees (1995) pointed out, test statistics similar to the CD test of Pesaran were proposed by Friedman (1937), based on the Spearman rank correlation coefficient (which is expected to be robust against non-normality). Although we do not consider the Friedman test in this paper, results that are similar to the CD test would apply for this test.

[3] However, the experimental results of Pesaran (2004) illustrate that the CD test is robust to skewed errors.

$(M \times 1)$ vector of time-varying common factors that are assumed to be non-stochastic and bounded, and ε_{it} is an independently distributed random variable over i with zero-mean and finite variance σ_i^2.

The error (multi-) factor structure has been employed extensively in the economic literature.[4] For example, in a macro panel, a common factor could be an unobserved technological shock, and the factor loadings can be thought of as capturing a cross-sectionally heterogeneous response to such shock. Note that since our asymptotic is $N \to \infty$ with T fixed, $\mathbf{f}_t$ is treated as non-stochastic here. In this paper we explicitly employ a random coefficient assumption for the factor loadings.

The process for $\mathbf{x}_{it}$ is defined as

$$\mathbf{x}_{it} = \boldsymbol{\mu}_i + \boldsymbol{\rho} \odot \mathbf{x}_{i,t-1} + \boldsymbol{\Gamma}_{xi} \mathbf{f}_t + \boldsymbol{\pi} \varepsilon_{i,t-1} + \mathbf{v}_{it}, \ i = 1, 2, ..., N, \ t = 1, 2, ..., T, \qquad (9)$$

where $\boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2}, ..., \mu_{iK})'$ is a $(K \times 1)$ vector of random effects with finite mean and finite variance, $\odot$ signifies the Hadamard product, $\boldsymbol{\rho} = (\rho_1, \rho_2, ..., \rho_K)'$ such that $|\rho_k| < 1$ for $k = 1, 2, ..., K$, $\boldsymbol{\Gamma}_{xi} = (\boldsymbol{\gamma}_{1i}, \boldsymbol{\gamma}_{2i}, ..., \boldsymbol{\gamma}_{Ki})'$ with $\boldsymbol{\gamma}_{ki} = (\gamma_{ki1}, \gamma_{ki2}, ..., \gamma_{kiM})'$, such that $\boldsymbol{\Gamma}_{xi} \sim i.i.d.(\boldsymbol{\Gamma}_x, \boldsymbol{\Sigma}_{\Gamma x})$, $\boldsymbol{\pi} = (\pi_1, \pi_2, ..., \pi_K)'$, and $\mathbf{v}_{it}$ is a vector of independently distributed random variables over i with mean vector zero and a finite variance matrix $\Sigma_{vi} = diag(\sigma_{vki}^2)$, $k = 1, 2, ..., K$. The model defined by (7), (8), and (9) is general enough to allow for a large variety of plausible specifications that are widely used in the economic literature. Furthermore, this model accommodates more simple processes for $\mathbf{x}_{it}$, such as those where $\mathbf{x}_{it}$ is strictly exogenous, or exogenous with respect to $\mathbf{f}_t$.

The null hypothesis of interest is then

$$H_0 : var(\boldsymbol{\phi}_i) = \boldsymbol{\Sigma}_\phi = \mathbf{0} \qquad (10)$$

against the alternative

$$H_1 : \boldsymbol{\Sigma}_\phi \neq \mathbf{0}, \qquad (11)$$

as opposed to (2) and (3).

Remark 1 *Observe that error cross section dependence may occur under the null hypothesis when $\boldsymbol{\phi}_i = \boldsymbol{\phi}$ for all i, since $E(u_{it} u_{jt}) = \boldsymbol{\phi}' \mathbf{f}_t \mathbf{f}_t' \boldsymbol{\phi}$, which is not zero unless $\boldsymbol{\phi} = \mathbf{0}$. However, such error cross section dependence can be eliminated simply by including time dummies, or equivalently by cross-sectionally demeaning the data. This implies that the null hypothesis in (10) can be interpreted as saying that the cross section dependence is homogeneous across pairs of cross-sectional units, against the alternative hypothesis (11) of heterogeneous error cross section dependence.*

We make the following assumptions:

---

[4]See e.g. Robertson and Symons (2000), Phillips and Sul (2003), Bai and Ng (2004), Moon and Perron (2004), and Pesaran (2006) among others. Notice that these methods are only justified when T is large. For related work that allows for time-varying individual effects in the fixed T, large N case, see Holtz-Eakin, Newey and Rosen (1988), Ahn, Lee, and Schmidt (2001) and Han, Orea and Schmidt (2005).

Assumption 1: $E(\alpha_i \varepsilon_{it}) = 0$, $E(\boldsymbol{\mu}_i \varepsilon_{it}) = \mathbf{0}$, and $E(\alpha_i \mathbf{v}_{it}) = \mathbf{0}$, for all i, t. Also $E(\mathbf{v}_{it} \varepsilon_{is}) = \mathbf{0}$ for all t and s.

Assumption 2: $E(\varepsilon_{it} \varepsilon_{is}) = 0$ and $E(\mathbf{v}_{it} \mathbf{v}_{is}') = \mathbf{0}$ for all i and $t \neq s$.

Assumption 3: $E(y_{i0} \varepsilon_{it}) = 0$ and $E(x_{i1} \varepsilon_{it}) = 0$ for all i and $t = 1, 2, ..., T$.

Assumption 4: Time-varying common factors, $f_{m,t}$, $m = 1, 2, ..., M; t = 1, 2, ..., T$ are non-stochastic and bounded.

Assumption 5:

(a) $E(\varepsilon_{it} \boldsymbol{\phi}_i) = \mathbf{0}$, $E(\mathbf{v}_{it} \boldsymbol{\phi}_i') = \mathbf{0}$, $E(\varepsilon_{it} \boldsymbol{\Gamma}_{xi}) = \mathbf{0}$, $cov(\alpha_i, \boldsymbol{\phi}_i) = \mathbf{0}$, $cov(\alpha_i, \boldsymbol{\Gamma}_{xi}) = \mathbf{0}$, $cov(\boldsymbol{\mu}_i, \boldsymbol{\phi}_i') = \mathbf{0}$, $cov(\boldsymbol{\Gamma}_{xi}', \boldsymbol{\mu}_i) = \mathbf{0}$, for all i and t.

(b) $cov(\boldsymbol{\Gamma}_{xi}, \boldsymbol{\phi}_i) = \mathbf{0}$ and $cov(\boldsymbol{\phi}_i, \mathbf{x}_{i1}') = \mathbf{0}$.

Assumption 6: $\boldsymbol{\beta} \neq \mathbf{0}$.

Assumptions 1-3 are standard in the GMM literature; see, for example, Ahn and Schmidt (1995). Assumption 4 ensures that the initial observations of y and $\mathbf{x}$ are bounded.

Assumption 5(a), a random coefficient type assumption on the factor loadings, allows cross-sectionally heterogeneous inter-dependence among both level and first-differenced variables $(y_{it}, x_{1it}, x_{2it}, ..., x_{Kit})$, as well as cross section dependence of each variable through the common factor $\mathbf{f}_t$. This contrasts to a simple time effects assumption, namely $\boldsymbol{\phi}_i = \boldsymbol{\phi}$ and $\boldsymbol{\Gamma}_{xi} = \boldsymbol{\Gamma}_x$ for all i, which is stronger than ours. On the other hand, Assumption 5(a) is stronger than assuming $\boldsymbol{\phi}_i$ and $\boldsymbol{\Gamma}_{xi}$ to be merely bounded, in that the cross-sectionally demeaned variables of $(y_{it}, x_{1it}, x_{2it}, ..., x_{Kit})$ become cross-sectionally uncorrelated to each other, asymptotically.

Assumption 5(b) implies that the cross-sectionally demeaned $\mathbf{x}_{i,t-1}$ and $(u_{it} - u_{i,t-1})$ or the cross-sectionally demeaned $(\mathbf{x}_{i,t-1} - \mathbf{x}_{i,t-2})$ and u_{it} are uncorrelated under the alternative hypothesis (11). This will obviously hold true if the common factor components have no impact on $\mathbf{x}_{it}$ − namely, when $\boldsymbol{\Gamma}_{xi} = \mathbf{0}$ for all i in equation (9). This restriction can be relaxed provided that the direct effect of the factors on y_{it} is uncorrelated with the effect of the factors on $\mathbf{x}_{it}$, which satisfies Assumption 5(b).[5]

Assumption 6, $\boldsymbol{\beta} \neq \mathbf{0}$, is also necessary in order to obtain a consistent estimator under the alternative of heterogeneous error cross section dependence, although this is probably not a restrictive assumption in many applications.

For further discussion, stacking (7) for each i yields

$$\mathbf{y}_i = \alpha_i \boldsymbol{\tau}_T + \lambda \mathbf{y}_{i,-1} + \mathbf{X}_i \boldsymbol{\beta} + \mathbf{u}_i, \ i = 1, 2, ..., N, \tag{12}$$

$$\mathbf{u}_i = \mathbf{F} \boldsymbol{\phi}_i + \boldsymbol{\varepsilon}_i, \tag{13}$$

---

[5]For sufficiently large N, the overidentifying restrictions test (defined by equation (28) below) is expected to have enough power to detect the violation of Assumption 5(b).

7

where $\mathbf{y}_i = (y_{i1}, y_{i2}, ..., y_{iT})'$, $\boldsymbol{\tau}_g$ is a $(g \times 1)$ vector of unity, $\mathbf{y}_{i,-1} = (y_{i0}, y_{i1}, ..., y_{iT-1})'$, $\mathbf{X}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{iT})'$, $\mathbf{u}_i = (u_{i1}, u_{i2}, ..., u_{iT})'$, $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_T)'$, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, ..., \varepsilon_{iT})'$, and the cross-sectionally demeaned and differenced equation is defined by

$$\Delta\underline{\mathbf{y}}_i = \lambda\Delta\underline{\mathbf{y}}_{i,-1} + \Delta\underline{\mathbf{X}}_i\boldsymbol{\beta} + \Delta\underline{\mathbf{u}}_i, \ i = 1, 2, ..., N, \tag{14}$$

$$\Delta\underline{\mathbf{u}}_i = \Delta\mathbf{F}\underline{\boldsymbol{\phi}}_i + \Delta\underline{\boldsymbol{\varepsilon}}_i, \tag{15}$$

or

$$\Delta\underline{\mathbf{y}}_i = \Delta\underline{\mathbf{W}}_i\boldsymbol{\theta} + \Delta\underline{\mathbf{u}}_i, \tag{16}$$

where the underline signifies that the variables are cross-sectionally demeaned, and "Δ" denotes the first-differenced operator. For example, $\Delta\underline{\mathbf{y}}_i = (\Delta\underline{y}_{i2}, \Delta\underline{y}_{i3}, ..., \Delta\underline{y}_{iT})'$, $\Delta\underline{y}_{it} = (\underline{y}_{it} - \underline{y}_{it-1})$, $\underline{y}_{it} = (y_{it} - \bar{y}_t)$ with $\bar{y}_t = N^{-1}\sum_{i=1}^{N} y_{it}$, $\Delta\underline{\mathbf{W}}_i = (\Delta\underline{\mathbf{y}}_{i,-1}, \Delta\underline{\mathbf{X}}_i)$, $\boldsymbol{\theta} = (\lambda, \boldsymbol{\beta}')'$, with obvious notation.

Remark 2 *Pesaran's CD test statistic as defined in (6) can fail to reject the null hypothesis of homogeneous error cross section dependence given in (11), when $E(\boldsymbol{\phi}_i) = \mathbf{0}$. This arises when time dummies are introduced or the data are cross-sectionally demeaned, since $E(\underline{\boldsymbol{\phi}}_i) = \mathbf{0}$. On the other hand, the Breusch and Pagan (1980) LM test defined as in (4) will have power, although its use is justified only when T is much larger relative to N.*

3.2 Sargan's Difference Tests based on the First-differenced GMM Estimator

Define the matrices of instruments

$$\underline{\mathbf{Z}}_{Yi} = \begin{bmatrix} \underline{y}_{i0} & 0 & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \underline{y}_{i0} & \underline{y}_{i1} & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & 0 & \underline{y}_{i0} & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ldots & \ldots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \underline{y}_{i0} & \cdots & \underline{y}_{iT-2} \end{bmatrix}, \ (T-1 \times h_y) \tag{17}$$

where $h_y = T(T-1)/2$, and

$$\underline{\mathbf{Z}}_{Xi} = \begin{bmatrix} \underline{\mathbf{x}}'_{i1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \cdots & \cdots & \mathbf{0} \\ \mathbf{0} & \underline{\mathbf{x}}'_{i1} & \underline{\mathbf{x}}'_{i2} & \mathbf{0} & \cdots & \cdots & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \underline{\mathbf{x}}'_{i1} & \cdots & \cdots & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ldots & \ldots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \underline{\mathbf{x}}'_{i1} & \cdots & \underline{\mathbf{x}}'_{iT-1} \end{bmatrix}, \ (T-1 \times h_x) \tag{18}$$

where $h_x = KT(T-1)/2$. If $\underline{\mathbf{x}}_{it}$ is strictly exogenous − that is, $\boldsymbol{\pi} = \mathbf{0}$ in (9), $\underline{\mathbf{Z}}_{Xi} = \boldsymbol{\tau}_{T-1} \otimes vec(\mathbf{X}_i)'$.[6]

---

[6]One can construct $\mathbf{Z}_{Xi}$ such that both strictly exogenous and predetermined regressors are present. See, for example, Arellano (2003).

Proposition 3 *Consider two sets of moment conditions*

$$\text{Moment Conditions I: } E[\underline{\mathbf{Z}}_{Yi}^{'}\Delta\underline{\mathbf{u}}_i] = \mathbf{0} \tag{19}$$

and

$$\text{Moment Conditions II: } E[\underline{\mathbf{Z}}_{Xi}^{'}\Delta\underline{\mathbf{u}}_i] = \mathbf{0}, \tag{20}$$

where $\underline{\mathbf{Z}}_{Yi}$, $\underline{\mathbf{Z}}_{Xi}$, and $\Delta\underline{\mathbf{u}}_i$ are defined as in (17), (18), and (15), respectively. Under Assumptions 1-4, 5(a)-(b), 6 and model (7), both sets of moment conditions given in (19) and (20) hold under the null hypothesis (10). However, under the alternative hypothesis given in (11), (20) holds but (19) does not.

Proof. See Appendix A. ∎

Now define the full set of moment conditions

$$E[\underline{\mathbf{Z}}_i^{'}\Delta\underline{\mathbf{u}}_i] = \mathbf{0}, \tag{21}$$

where

$$\underline{\mathbf{Z}}_i = \begin{bmatrix} \underline{\mathbf{Z}}_{Yi} & \underline{\mathbf{Z}}_{Xi} \end{bmatrix} \ (T-1 \times h), \tag{22}$$

and $h = h_y + h_x$.

Based on Proposition 3, the hypothesis testing for (10) reduces to a test for the validity of the subset of the moment conditions in (21), as given in (19). There are at least three possible approaches. The first one is to take a Hausman-type (1978) approach. Newey (1985) showed that the Hausman test may be inconsistent against some local alternatives. In our application, this may happen when $h_y = T(T-1)/2 > K+1$, which is the most likely empirical situation.[7] In addition, Arellano and Bond (1991) examined a Hausman-type test for serial correlation and found that it has poor finite sample performance, in that it rejects the null hypothesis far too often. The second approach is to adopt Newey's (1985) optimal test. The third approach is to use Sargan's (1988) difference test, which we are adopting, since it is relatively easier to compute than the Newey's test statistic, and furthermore, they are asymptotically equivalent. We consider two versions of Sargan's difference test. One is based on the Arellano and Bond (1991) two-step first-differenced GMM estimator − hereafter DIF − and the other is based on the Blundell and Bond (1998) two-step system GMM estimator − hereafter, SYS.

Initially, the test based on DIF is considered. This is defined as

$$\widetilde{\widetilde{\boldsymbol{\theta}}}_{DIF2} = \left(\sum_{i=1}^{N} \Delta\underline{\mathbf{W}}_i^{'}\underline{\mathbf{Z}}_i \dot{\widetilde{\Omega}}^{-1} \sum_{i=1}^{N} \underline{\mathbf{Z}}_i^{-1}\Delta\underline{\mathbf{W}}_i \right)^{-1} \sum_{i=1}^{N} \Delta\underline{\mathbf{W}}_i^{'}\underline{\mathbf{Z}}_i^{-1}\dot{\widetilde{\Omega}}^{-1} \sum_{i=1}^{N} \underline{\mathbf{Z}}_i^{'}\Delta\underline{\mathbf{y}}_i, \tag{23}$$

where $\dot{\widetilde{\Omega}} = N^{-1}\sum_{i=1}^{N} \underline{\mathbf{Z}}_i^{'}\Delta\dot{\widetilde{\underline{\mathbf{u}}}}_i\Delta\dot{\widetilde{\underline{\mathbf{u}}}}_i^{'}\underline{\mathbf{Z}}_i$ with $\Delta\dot{\widetilde{\underline{\mathbf{u}}}}_i = \Delta\underline{\mathbf{y}}_i - \Delta\underline{\mathbf{W}}_i\dot{\widetilde{\boldsymbol{\theta}}}_{DIF1}$, which is the residual vector based on the one-step DIF estimator (denoted by $DIF1$),

$$\widetilde{\dot{\boldsymbol{\theta}}}_{DIF1} = \left(\sum_{i=1}^{N} \Delta\underline{\mathbf{W}}_i^{'}\underline{\mathbf{Z}}_i^{-1}\widetilde{\Omega}^{-1} \sum_{i=1}^{N} \underline{\mathbf{Z}}_i^{'}\Delta\underline{\mathbf{W}}_i \right)^{-1} \sum_{i=1}^{N} \Delta\underline{\mathbf{W}}_i^{'}\underline{\mathbf{Z}}_i\widetilde{\Omega}^{-1} \sum_{i=1}^{N} \underline{\mathbf{Z}}_i^{'}\Delta\underline{\mathbf{y}}_i, \tag{24}$$

---

[7]See Proposition 5 and its proof in Newey (1985), under the assumption of i.i.d. errors.

9

where $\widetilde{\Omega} = N^{-1} \sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{H} \mathbf{Z}_i$, and $\mathbf{H}$ is the square matrix of order $T-1$, with 2's on the main diagonal, -1's on the first off-diagonals, and zeros elsewhere.

Now, the two-step DIF estimator based only on the restricted set of moment conditions in (20), which is consistent both under the null and the alternative, is defined as

$$\widehat{\widetilde{\boldsymbol{\theta}}}_{DIFX2} = \left(\sum_{i=1}^{N} \Delta \underline{\mathbf{W}}_i' \mathbf{Z}_{Xi} \widehat{\widetilde{\Omega}}_X^{-1} \sum_{i=1}^{N} \mathbf{Z}_{Xi}' \Delta \underline{\mathbf{W}}_i \right)^{-1} \sum_{i=1}^{N} \Delta \underline{\mathbf{W}}_i' \mathbf{Z}_{Xi} \widehat{\widetilde{\Omega}}_X^{-1} \sum_{i=1}^{N} \mathbf{Z}_{Xi}' \Delta \underline{\mathbf{y}}_i, \quad (25)$$

where $\widehat{\widetilde{\Omega}}_X = N^{-1} \sum_{i=1}^{N} \underline{\mathbf{Z}}_{Xi}' \Delta \widehat{\underline{\mathbf{u}}}_i \Delta \widehat{\underline{\mathbf{u}}}_i' \underline{\mathbf{Z}}_{Xi}$ with $\Delta \widehat{\underline{\mathbf{u}}}_i = \Delta \underline{\mathbf{y}}_i - \Delta \underline{\mathbf{W}}_i \widehat{\boldsymbol{\theta}}_{DIFX1}$, where $\Delta \widehat{\underline{\mathbf{u}}}_i$ is the residual vector based on the one-step DIF estimator that exploits (20), denoted by $DIFX1$,

$$\widehat{\boldsymbol{\theta}}_{DIFX1} = \left(\sum_{i=1}^{N} \Delta \underline{\mathbf{W}}_i' \mathbf{Z}_{Xi} \widehat{\Omega}_X^{-1} \sum_{i=1}^{N} \mathbf{Z}_{Xi}' \Delta \underline{\mathbf{W}}_i \right)^{-1} \sum_{i=1}^{N} \Delta \underline{\mathbf{W}}_i' \mathbf{Z}_{Xi} \widehat{\Omega}_X^{-1} \sum_{i=1}^{N} \underline{\mathbf{Z}}_{Xi}' \Delta \underline{\mathbf{y}}_i, \quad (26)$$

where $\widehat{\Omega}_X = N^{-1} \sum_{i=1}^{N} \underline{\mathbf{Z}}_{Xi}' \mathbf{H} \underline{\mathbf{Z}}_{Xi}$.

Sargan's (1958), or Hansen's (1982) test statistic of overidentifying restrictions for the full set of moment conditions in (21) is based on $DIF2$ and is given by

$$S_{DIF2} = N^{-1} \left(\sum_{i=1}^{N} \Delta \widetilde{\underline{\mathbf{u}}}_i' \underline{\mathbf{Z}}_i \right) \widetilde{\Omega}^{-1} \left(\sum_{i=1}^{N} \underline{\mathbf{Z}}_i' \Delta \widetilde{\underline{\mathbf{u}}}_i \right), \quad (27)$$

where $\Delta \widetilde{\underline{\mathbf{u}}}_i = \Delta \underline{\mathbf{y}}_i - \Delta \underline{\mathbf{W}}_i \widetilde{\boldsymbol{\theta}}_{DIF2}$. Sargan's statistic for (20) that is based on $DIFX2$ is given by

$$S_{DIFX2} = N^{-1} \left(\sum_{i=1}^{N} \Delta \widehat{\widetilde{\underline{\mathbf{u}}}}_i' \underline{\mathbf{Z}}_{Xi} \right) \widehat{\widetilde{\Omega}}_X^{-1} \left(\sum_{i=1}^{N} \underline{\mathbf{Z}}_{Xi}' \Delta \widehat{\widetilde{\underline{\mathbf{u}}}}_i \right), \quad (28)$$

where $\Delta \widehat{\widetilde{\underline{\mathbf{u}}}}_i = \Delta \underline{\mathbf{y}}_i - \Delta \underline{\mathbf{W}}_i \widehat{\widetilde{\boldsymbol{\theta}}}_{DIFX2}$.

We are now ready to state a proposition on Sargan's difference test of heterogeneous cross section dependence:

Proposition 4 *Under Assumptions 1-4, 5(a)-(b), 6 and model (7), $S_{DIF2} \xrightarrow{d} \chi^2_{h-(K+1)}$ under the null hypothesis, $S_{DIFX2} \xrightarrow{d} \chi^2_{h_x-(K+1)}$ under the null hypothesis and the alternative hypothesis, and*

$$D_{DIF2} = (S_{DIF2} - S_{DIFX2}) \xrightarrow{d} \chi^2_{h_y} \quad (29)$$

under the null hypothesis, where S_{DIF2}, S_{DIFX2} are as defined in (27) and (28) respectively, while $h = h_y + h_x$ with h_y and h_x defined as in (17) and (18) respectively.

Proof. See Appendix B. ∎

10

Remark 5 *The cross section correlation of the sample moment conditions $\underline{\mathbf{Z}}_i' \Delta \underline{\tilde{\tilde{\mathbf{u}}}}_i$, which is induced as a result of the cross-sectional demeaning of the series, is asymptotically negligible, and standard central limit theorems can be applied to the asymptotic expansion of $N^{-1/2} \sum_{i=1}^{N} \underline{\mathbf{Z}}_i' \Delta \underline{\tilde{\tilde{\mathbf{u}}}}_i$; See proof in Appendix B.*

Remark 6 *Sargan's difference test has non-trivial asymptotic local power; See proof in Appendix B.*

Remark 7 *Proposition 4 holds with unbalanced panel data, so long as $\min N_t \to \infty$, where N_t is the number of cross-sectional units for a given time t. In this case, the cross-sectional demeaning for $\underline{y}_t$, say, is now defined as $\underline{y}_t = y_t - \bar{y}_t$, where $\bar{y}_t = N_t^{-1} \sum_{i=1}^{N_t} y_{it}$.*[8]

3.3 Sargan's Difference Tests based on the System GMM Estimator

It has been well documented (see, for example, Blundell and Bond, 1998) that the DIF estimator can suffer from a weak instruments problem when λ is close to unity and/or the variance of the individual effects is large relative to that of the idiosyncratic errors. Thus, we also consider another version of Sargan's difference test based on the Blundell and Bond (1998) SYS estimator, which is known to be more robust to the problem of weak instruments under certain conditions.

Arellano and Bover (1995) proposed the use of lagged differences as possible instruments for the equations in levels, $E(u_{it} \Delta \mathbf{w}_{it}) = \mathbf{0}$ with $\mathbf{w}_{it} = (y_{t-1}, \mathbf{x}_t')'$ for $t = 3, 4, ..., T$, which is valid under the null hypothesis and Assumptions 1-3. In addition, Blundell and Bond (1998, 2000) proposed using an additional condition $E(u_{i2} \Delta \mathbf{w}_{i2}) = \mathbf{0}$ under mild conditions upon the initial observations, which would follow from joint stationarity of the y and x processes. Accordingly, we add Assumption 5(c):

Assumption 5(c): $cov(\alpha_i, \Delta y_{i1}) = \mathbf{0}$ and $cov(\alpha_i, \Delta \mathbf{x}_{i2}) = \mathbf{0}$.

Define

$$\underline{\mathbf{Z}}_{Yi}^+ = \begin{bmatrix} \mathbf{Z}_{Yi} & \mathbf{0} \\ \mathbf{0} & \underline{\mathbf{Z}}_{Yi}^L \end{bmatrix} \ (2(T-1) \times h_{ys}), \tag{30}$$

where $\underline{\mathbf{Z}}_{Yi}^L = diag(\Delta \underline{y}_{i,t-1})$ for $t = 2, 3, ..., T$, $h_{ys} = h_y + (T-1)$, and

$$\underline{\mathbf{Z}}_{Xi}^+ = \begin{bmatrix} \underline{\mathbf{Z}}_{Xi} & \mathbf{0} \\ \mathbf{0} & \underline{\mathbf{Z}}_{Xi}^L \end{bmatrix} \ (2(T-1) \times h_{xs}), \tag{31}$$

where $\underline{\mathbf{Z}}_{Xi}^L$ is a $(T-1 \times K(T-1))$ matrix whose s^{th} diagonal raw vector is $\Delta \underline{\mathbf{x}}_{is}'$, $s = 2, 3, ..., T$, otherwise zeros, and $h_{xs} = h_x + K(T-1)$.

Proposition 8 *Consider two sets of moment conditions*

$$\text{Moment Conditions } I^*: \ E[\underline{\mathbf{Z}}_{Yi}^{+'} \boldsymbol{v}_i] = \mathbf{0} \tag{32}$$

---

[8] For more details on the computations of the Arellano and Bond (1991) and Blundell and Bond (1998) estimators with unbalanced panels, see Arellano and Bond (1999).

and

$$Moment\ Conditions\ II^*\colon\ E[\underline{\mathbf{Z}}_{Xi}^{+\prime}\underline{\boldsymbol{v}}_i] = \mathbf{0}, \tag{33}$$

where $\underline{\mathbf{Z}}_{Yi}^{+}$ and $\underline{\mathbf{Z}}_{Xi}^{+}$ are defined as in (30) and (31) respectively,

$$\underline{\boldsymbol{v}}_i = (\Delta\underline{\mathbf{u}}_i', \underline{\mathbf{u}}_i^{+\prime})', \ \ and \ \underline{\mathbf{u}}_i^+ = (\underline{u}_{i2}, \underline{u}_{i3}, ..., \underline{u}_{iT},)'. \tag{34}$$

Under Assumptions 1-4, 5(a)-(c), 6 and model (7), both (32) and (33) hold under the null hypothesis in (10). However, under the alternative hypothesis given in (11), (33) holds but (32) does not.

Proof. See Appendix C. ■

In the same manner as D_{DIF2}, we define the Sargan's difference test statistic based on SYS estimator

$$D_{SYS2} = (S_{SYS2} - S_{SYSX2}) \tag{35}$$

where S_{SYS2} is Sargan's test statistic of overidentifying restrictions based on the two-step SYS estimator that makes use of both sets of moment conditions I* and II* in Proposition 8 (denoted by $SYS2$), namely

$$E[\underline{\mathbf{Z}}_i^{+\prime}\underline{\boldsymbol{v}}_i] = \mathbf{0}, \tag{36}$$

where

$$\underline{\mathbf{Z}}_i^+ = \begin{bmatrix} \underline{\mathbf{Z}}_{Yi}^+ & \underline{\mathbf{Z}}_{Xi}^+ \end{bmatrix} \ (2(T-1) \times h_s), \tag{37}$$

$h_s = h_{sy} + h_{sx}$, and S_{SYSX2} is Sargan's test statistic based on the two-step SYS estimator that exploits only Moment Conditions II*, as defined in (33), denoted by $SYSX2$.[9] One-step and two-step SYS estimators based on the moment conditions (36), $\widetilde{\boldsymbol{\theta}}_{SYS1}$ and $\widehat{\widetilde{\boldsymbol{\theta}}}_{SYS2}$, and those based on (33), $\widetilde{\widehat{\boldsymbol{\theta}}}_{SYSX1}$ and $\widehat{\widetilde{\boldsymbol{\theta}}}_{SYSX2}$, are defined accordingly.[10] It is straightforward to see that $D_{SYS2} \xrightarrow{d} \chi^2_{h_{ys}}$ as $N \to \infty$, under the null hypothesis.

3.4 Discussion

The overidentifying restrictions test can be regarded as a misspecification test, in a sense that it is designed to detect violations of moment conditions, which are the heart of GMM methods. Thus, it will have power under the alternative hypothesis of heterogenous error cross section dependence. Nonetheless, the proposed Sargan's difference test is expected to have higher power than the overidentifying restrictions test, so long as Assumption 6, $\boldsymbol{\beta} \neq \mathbf{0}$, holds, since the former exploits extra information about the validity of the moment conditions under the alternative hypothesis, which the latter does not use. This

---

[9]With small samples D_{DIF2} may not be positive, but it can be patched easily. See, for example, Hayashi (2000, p.220). However, we did not adopt this modification here since one of our aims is to show the properties of a consistent estimator based only on orthogonality conditions $E[\underline{\mathbf{Z}}_{Xi}'\underline{\mathbf{u}}_i] = \mathbf{0}$ or $E[\underline{\mathbf{Z}}_{Xi}^{+\prime}\underline{\boldsymbol{v}}_i] = \mathbf{0}$.

[10]The initial weighting matrix for one-step GMM-SYS estimator is defined as a block diagnoal matrix of order $2(T-1)$, whose diagnoal blocks are $\mathbf{H}$ and $\mathbf{I}_{T-1}$.

also implies that, when $\boldsymbol{\beta} = \mathbf{0}$, the overidentifying restrictions test should replace our approach.[11]

Now consider a violation of Assumption 2, $E(\varepsilon_{it}\varepsilon_{is}) = 0$ for $t \neq s$, no error serial correlation. Under the alternative of error cross section dependence, the composite error $\underline{u}_{it} = \underline{\boldsymbol{\phi}}'_i \mathbf{f}_t + \underline{\varepsilon}_{it}$ will be serially correlated, since $E(\underline{u}_{it}\underline{u}_{is}) = \mathbf{f}'_t E(\underline{\boldsymbol{\phi}}_i\underline{\boldsymbol{\phi}}'_i)\mathbf{f}_s \neq 0$ for all i. This means that the second-order serial correlation test based on DIF2 or SYS2, the m_2 test, proposed by Arellano and Bond (1991), is likely to reject the hypothesis of no error serial correlation, under the alternative of heterogeneous error cross section dependence.[12] Then, a question that may arise is how to distinguish between error cross section dependence and serial correlation in the idiosyncratic errors. To answer this question, consider two scenarios. First, suppose that there is first-order autoregressive serial correlation but no heterogeneous error cross section dependence, such that $\varepsilon_{it} = \rho_\varepsilon \varepsilon_{it-1} + \nu_{it}$ with $|\rho_\varepsilon| < 1$ and $\nu_{it} \sim i.i.d.(0, \sigma^2_\nu)$. In this case, the problem can be solved in a straightforward manner by adding a further lag of the dependent variable on the right hand side of (7) and using (up to) y_{it-3} as instruments for Δy_{it-1} and Δy_{it-2}. Second, suppose there is both first-order autoregressive error serial correlation as above and heterogeneous error cross section dependence. Clearly, the m_2 test based on DIFX2 or SYSX2 is likely to reject the null even when $E(\varepsilon_{it}\varepsilon_{is}) = 0$ for $t \neq s$. Meanwhile, the probability of rejecting the null by the overidentifying restrictions test for the restrictions based only on the subset of $\mathbf{X}_i$ (defined by (20) or (33)), tends to its significance level when $E(\varepsilon_{it}\varepsilon_{is}) = 0$ for $t \neq s$, but such a probability goes to one when $E(\varepsilon_{it}\varepsilon_{is}) \neq 0$ for $t \neq s$. Therefore, the solution given in the first case applies, but the test statistic to employ is the overidentifying restrictions test based only on the subset of $\mathbf{X}_i$, not the m_2 test based on DIFX2 or SYSX2.

Finally, we have shown that the moment conditions (20) and (33) hold under the alternative of error cross section dependence, therefore, the DIFX2 and SYSX2 estimators are consistent. However, in finite samples there could be a trade-off between efficiency and bias. If the degree of heterogeneity of the error cross section dependence is relatively small, then the bias of the standard GMM estimators exploiting moment conditions including (19) or (20) which are invalid, may be small enough so that these estimators are preferred (in root mean square errors terms) to a consistent estimator based only on the valid moment conditions (20) or (33). We will investigate the finite sample performance of these estimators in the next section.[13]

---

[11]In Appendix B, it is formally shown that when $h_x > (K+1)$, the D_{DIF2} test is asymptotically more powerful than the S_{DIF2} test under the local alternatives.

[12]Heterogeneity of λ would also render the error term serially correlated, as discussed in Pesaran and Smith (1995).

[13]Other solutions have been proposed in the literature, such as a panel feasible generalized median unbiased estimator, proposed by Phillips and Sul (2003), or the common correlated effects (CCE) estimator proposed by Pesaran (2006). However, both estimators require a larger value for T than that considered in this paper.

13

4 Small Sample Properties of Cross Section Dependence Tests

This section investigates by means of Monte Carlo experiments the finite sample performance of our tests, the Breusch and Pagan (1980) LM test and Pesaran's (2004) CD test, all based on cross-sectionally demeaned variables. Our main focus is on the effects of (i) the degree and heterogeneity of error cross section dependence, (ii) the relative importance of the variance of the factor loadings and the idiosyncratic errors, and (iii) different values of λ and $\boldsymbol{\beta}$. In order to make the results comparable across experiments, we control the population signal-to-noise ratio and the impact of the ratio between the variance of the individual-specific time-invariant effects and the variance of the idiosyncratic errors and the common factor on y_{it}. To this end, we extend the Monte Carlo design of Kiviet (1995) and Bun and Kiviet (2006) to accommodate a factor structure in the error process.

Recently Bowsher (2002) reports finite sample evidence that Sargan's overidentifying restrictions test exploiting all moment conditions available can reject the null hypothesis too infrequently in linear dynamic panel models. Thus, we only make use of $\underline{y}_{it-2}$ and $\underline{y}_{it-3}$ as instruments for $\Delta \underline{y}_{it-1}$ and we use $\underline{x}_{it-2}$ and $\underline{x}_{it-3}$ as instruments for $\Delta \underline{x}_{it-1}$.

4.1 Design

The data generating process (DGP) we consider is given by

$$
\begin{aligned}
y_{it} &= \alpha_i + \lambda y_{i,t-1} + \beta x_{it} + u_{it}, \\
u_{it} &= \phi_i f_t + \varepsilon_{it},\ i = 1, 2, ..., N; t = -48, -47, ..., T,
\end{aligned}
\tag{38}
$$

where $\alpha_i \sim i.i.d.N(1, \sigma_\alpha^2)$ and $f_t \sim i.i.d.N(0, \sigma_f^2)$.[14] ε_{it} is drawn from (i) $i.i.d.N(0, \sigma_\varepsilon^2)$ and (ii) $i.i.d.(\chi_1^2 - 1)/\sqrt{2}$, in order to investigate the effect of non-normal errors. $y_{i,-49} = 0$ and the first 49 observations are discarded.[15] To control the degree and heterogeneity of cross section dependence three specifications for the distribution of ϕ_i are considered:

$$
\left\{
\begin{array}{lll}
\text{Low cross section dependence:} & \phi_i \sim i.i.d.U[-0.3, 0.7] \\
\text{Medium cross section dependence:} & \phi_i \sim i.i.d.U[-1, 2] \\
\text{High cross section dependence:} & \phi_i \sim i.i.d.U[-1, 4].
\end{array}
\right.
$$

Also, as we change the value of $\lambda = 0.2, 0.5, 0.8$, β is equated to $1 - \lambda$ in order to keep the long run effect of x on y constant.

The DGP of x_{it} considered here is given by

$$
x_{it} = \rho x_{i,t-1} + \pi \varepsilon_{i,t-1} + \gamma_i f_t + v_{it},\ i = 1, 2, ..., N; t = -48, -47, ..., T,
\tag{39}
$$

---

[14]Note that the unobserved common factor, f_t, is randomly drawn to control the signal-to-noise ratio without loss of generality.

[15]We do not report the results based on non-normal errors in this paper, since the results were very similar. They are available from the authors upon request.

14

where $\rho = 0.5$, $\gamma_i \sim i.i.d.U[-1, 2]$, $v_{it} \sim i.i.d.N(0, \sigma_v^2)$. π is set to 0.5.[16] $x_{i,-49} = 0$ and the first 50 observations are discarded.

Since our focus is on the performance of the tests and estimators, we pay careful attention to the main factors that affect it − namely, (i) the signal-to-noise ratio, (ii) the relative importance of the variance of the factor loadings and the idiosyncratic errors, and (iii) the impact of the ratio between the variance of the individual-specific effects and the variance of the idiosyncratic error and factor loadings on y_{it}. To illustrate, we define the signal as $\sigma_s^2 = var\left(y_{it}^* - u_{it}\right)$, where $y_{it}^* = y_{it} - \alpha_i/(1 - \lambda)$. Then, denoting the variance of the composite error by $\sigma_u^2 = var(u_{it}) = \mu_\phi^{(2)}\sigma_f^2 + \sigma_\varepsilon^2$ with $\mu_\phi^{(2)} = E(\phi_i^2)$, we define the signal-to-noise ratio as $\zeta = \sigma_s^2/\sigma_u^2$. We set $\zeta = 3$. The relative importance in terms of the magnitude of the variance of $\phi_i f_t$ and ε_{it}, as measured by $\mu_\phi^{(2)}\sigma_f^2/\sigma_\varepsilon^2$, is thought in the literature to be an important factor to control for and we achieve this by changing $\mu_\phi^{(2)}$ and applying the normalisation $\sigma_f^2 = \sigma_\varepsilon^2 = 1$. As it has been discussed by Kiviet (1995), Blundell and Bond (1998), and Bun and Kiviet (2006), in order to compare the performance of estimators across different experimental designs it is important to control the relative importance of α_i and (ε_{it}, f_t). We choose σ_α^2 such that the ratio of the impact on $var(y_{it})$ of the two variance components α_i and (ε_{it}, f_t) is constant across designs.[17]

We consider all combinations of $N = 50, 100, 200, 400$, and $T = 5, 9$. All experiments are based on 2,000 replications.

4.2 Results

Tables 1 reports the size and power of the tests for $T = 5$.[18] LM denotes Breusch and Pagan's LM test, as defined in (4), and CD denotes Pesaran's CD test, defined in (6), both of which are based on the fixed effects estimator. D_{DIF2} is Sargan's difference test based on the two-step DIF estimator defined in (29), and D_{SYS2} is Sargan's difference test based on the two-step SYS estimator defined in (35). The size of the LM test is always indistinguishable from 100% and therefore it is not recommended. The CD test does not reject the null in all experiments, and has no power across experiments. On the other hand, although the size of D_{DIF2} and D_{SYS2} is below the nominal level for $N = 50$ (especially for the latter), as N becomes larger the size quickly approaches its nominal size. In addition, our proposed tests have satisfactory power. D_{SYS2} has more power than D_{DIF2} in general, unless D_{SYS2} rejects the null too infrequently. Different values of λ seem to have very little effect on the performance of D_{DIF2} and D_{SYS2}.

We now turn our attention to the performance of the estimators. Table 2 reports the bias of the estimators for λ.[19] $DIF1$ and $DIF2$ are the one-step and two-step DIF

---

[16]x_{it} has zero mean without loss of generality, since we cross section-demean all data before computing the statistics.

[17]See Appendix D for the details of the way of controlling these parameters.

[18]We do not report the results for $T = 9$ in this paper, since these were similar to those for $T = 5$. They are available from the authors upon request.

[19]We do not report the performance of the estimators for β, since it has a similar pattern to that for λ, although it is not as much affected by error cross section dependence.

estimators respectively, defined by (24) and (23), and they are based on the full set of moment conditions I and II in Proposition 3. $DIFX1$ and $DIFX2$ denote the one-step and two-step DIF estimators defined by (26) and (25), and they are based only on the subset of the moment conditions II. $SYS1$ and $SYS2$ are the one-step and two-step SYS estimators respectively, and they are based on the full set of moment conditions I* and II* in Proposition 8, and $SYSX1$ and $SYSX2$ denote the one-step and two-step SYS estimators based only on the subset of moment conditions II*. The bias of all GMM estimators under low cross section dependence is not noticeably different from that under zero cross section dependence. As the degree of error cross section dependence rises, the bias of the GMM estimators based on the full set of moment conditions increases, which is expected as only those estimators based on Moment Conditions II or II* are consistent. As a result, the relative bias between those estimators that use the full set of moment conditions and those that use only Moment Conditions II or II* increases. Table 3 reports root mean square errors of the estimators for λ. Under no error cross section dependence and low cross section dependence, $DIF2$ and $SYS2$ outperform $DIFX2$ and $SYSX2$ respectively in terms of root mean square error. However, under moderate and high cross section dependence, $DIFX2$ and $SYSX2$ have a smaller root mean square error compared to $DIF2$ and $SYS2$ respectively, in most cases.

5 An Empirical Example: Employment Equations of U.K. Firms

In this section we examine the homogeneity of error cross section dependence of the employment equations using (unbalanced) panel data for a sample of UK companies, which is an updated version of that used by Arellano and Bond (1991), and it is contained in the DPD-Ox package.[20] Briefly, these authors select a sample of 140 companies that operate mainly in the UK with at least 7 continuous observations during the period 1976-1984.

We apply our test to the model specifications of Blundell and Bond (1998). The model we estimated is given by

$$y_{it} = \alpha_i + \beta_1 y_{it-1} + \delta_0 w_{it} + \delta_1 w_{it-1} + \varphi_0 \kappa_{it} + \varphi_1 \kappa_{it-1} + u_{it},$$

where y_{it} is log of the number of employees of company i, w_{it} is log of real product wage, κ_{it} is the log of gross capital stock.

Table 4 presents estimation and test results.[21] Observe that year dummies are included to remove possible time effects, therefore no cross-sectional demeaning of the series is implemented. Our estimation results based on the full sets of instruments, $\mathbf{Z}_i$ and $\mathbf{Z}_i^+$, as defined in (22) and (37) but without cross-sectional demeaning, resemble

---

[20]The data set used is available at http://www.doornik.com/download/dpdox121.zip

[21]The GMM estimates of the parameters have been obtained using the xtabond2 command in Stata; see Roodman, D., (2005). xtabond2: Stata module to extend xtabond dynamic panel data estimator. Center for Global Development, Washington. http://econpapers.repec.org/software/bocbocode/s435901.htm

16

those reported in the last two columns of Table 4 in Blundell and Bond (1998), although the values do not match exactly due to differences in computations and the data set used. First, all m_2 tests suggest that there is no evidence of error serial correlation and this implies possibly no heterogeneous error cross section dependence. This is confirmed by the fact that both Sargan's difference tests based on DIF and SYS for heterogeneous error cross section dependence safely fail to reject the null hypothesis of homogeneous error cross section dependence. The estimation results based only on partial instruments consisting of the covariates, $\mathbf{Z}_{Xi}$ and $\mathbf{Z}_{Xi}^+$, as defined by (18) and (31) but without cross-sectional demeaning, are largely downward biased for the DIF estimator and less so for the SYS estimator. This indicates that the efficiency loss of SYS that does not contain $\mathbf{Z}_{Yi}^+$ in the instrument set is much smaller compared to the efficiency loss of DIF. This feature seems to have some effect on the testing results. For example, the p-value of the first-order serial correlation test, m_1, for DIF with the full set of instruments, $\mathbf{Z}_i$, is zero up to three decimal points, but it goes up to 0.028 with the subset of instruments $\mathbf{Z}_{Xi}$. On the other hand, the p-value of m_1 for SYS is zero up to three decimal points in both cases.

6   Concluding Remarks

This paper has proposed a new testing procedure for error cross section dependence after estimating a linear dynamic panel data model with regressors by the generalised method of moments (GMM). The procedure is valid when the cross-sectional dimension is large and the time series dimension of the panel is small. Importantly, our approach allows one to examine whether any error cross section dependence remains after including time dummies, or after transforming the data in terms of deviations from time-specific averages, which will be the case under heterogeneous error cross section dependence. The finite sample simulation-based results suggest that our tests perform well, particularly the version based on the Blundell and Bond (1988) system GMM estimator. On the other hand, the LM test of Breusch and Pagan (1980) overrejects the null hypothesis substantially and Pesaran's (2004) CD test lacks power. Also it is shown that the system GMM estimator, based only on partial instruments consisting of the regressors, can be a reliable alternative to the standard GMM estimators under heterogeneous error cross section dependence. The proposed tests are applied to employment equations using UK firm data, and the results show little evidence of heterogeneous error cross section dependence.

Appendices

A   Proof of Proposition 3

To simplify the analysis without loss of generality we consider the case where $K = 1$. For $t \geq 2 + j$, given j such that $0 \leq j \leq t - 1$, it can be shown that

$$\underline{x}_{it-j} = \underline{\mu}_i \sum_{\ell=0}^{t-j-2} \rho^\ell + \rho^{t-j-1} \underline{x}_{i1} + \underline{\gamma}_i' \sum_{\ell=0}^{t-j-2} \rho^\ell \mathbf{f}_{t-j-\ell} + \pi \sum_{\ell=0}^{t-j-2} \rho^\ell \underline{\varepsilon}_{i,t-j-1-\ell} + \sum_{\ell=0}^{t-j-2} \rho^\ell \underline{v}_{i,t-j-\ell}, \quad (40)$$

and for $t \geq 1$

$$\underline{y}_{it} = \underline{\alpha}_i \sum_{j=0}^{t-1} \lambda^j + \lambda^t \underline{y}_{i0} + \beta \sum_{j=0}^{t-1} \lambda^j \underline{x}_{i,t-j} + \underline{\phi}_i' \sum_{j=0}^{t-1} \lambda^j \mathbf{f}_{t-j} + \sum_{j=0}^{t-1} \lambda^j \underline{\varepsilon}_{i,t-j}. \quad (41)$$

Firstly we consider $E(\underline{y}_{i,t-s} \Delta \underline{u}_{it})$ for $s \leq t \leq T$, given $2 \leq s \leq T$, under the alternative hypothesis of $\phi_i \neq \phi$. But this is equivalent to considering $E(\underline{y}_{it} \Delta \underline{u}_{i,t+s})$ for $0 \leq t \leq T - s$, given $2 \leq s \leq T$. Initially we focus on the case of $s = 2$. When $t = 0$, $E(\underline{y}_{i0} \Delta \underline{u}_{i2}) = E(\underline{y}_{i0} \underline{\phi}_i') \Delta \mathbf{f}_2$ by Assumption 3, which is not necessarily zero under the alternative. When $t \geq 1$, using (40) and (41), together with $\underline{u}_{it} = \underline{\phi}_i' \mathbf{f}_t + \underline{\varepsilon}_{it}$,

$$E(\underline{y}_{it} \Delta \underline{u}_{i,t+2}) = \lambda^t E\left(\underline{y}_{i0} \underline{\phi}_i'\right) \Delta \mathbf{f}_{t+2} + \Delta \mathbf{f}_{t+2}' E\left(\underline{\phi}_i \underline{\phi}_i'\right) \sum_{j=0}^{t-1} \lambda^j \mathbf{f}_{t-j} \neq 0, \quad (42)$$

under Assumptions 1-4, 5(a)-(b), 6 and model (7). A similar approach for the case of $s > 2$ leads to the conclusion that $E(\underline{y}_{it} \Delta \underline{u}_{i,t+s}) \neq 0$ for $t \geq 0$, $s \geq 2$ under the alternative, as required. Under the null of $\phi_i = \phi$, it follows immediately that $E(\underline{y}_{it} \Delta \underline{u}_{i,t+s}) = 0$ for $t \geq 0$, $s \geq 2$.

Now we consider $E(\underline{x}_{i,t-s} \Delta \underline{u}_{it})$ for $s < t \leq T$, given $1 \leq s \leq T - 1$, under the alternative, which is equivalent to considering $E(\underline{x}_{it} \Delta \underline{u}_{it+s})$ for $1 \leq t < T - s$, given $1 \leq s \leq T - 1$. Initially we focus on the case of $s = 1$. When $t = 1$, $E(\underline{x}_{i1} \Delta \underline{u}_{i2}) = 0$ due to Assumption 3 and 5(b). For $t \geq 2$ and using (40), together with $\underline{u}_{it} = \underline{\phi}_i' \mathbf{f}_t + \varepsilon_{it}$, we have $E(\underline{x}_{it} \Delta \underline{u}_{it+1}) = 0$, under Assumptions 1-4, 5(a)-(b), 6 and model (7), under the alternative. A similar approach for the case where $s > 1$ leads to the conclusion that $E(\underline{x}_{it} \Delta \underline{u}_{i,t+s}) \neq 0$ for $t \geq 0$, $s \geq 1$ under the alternative, as required. Under the null of $\phi_i = \phi$, it follows immediately that $E(\underline{y}_{it} \Delta \underline{u}_{i,t+s}) = 0$ for $t \geq 0$, $s \geq 1$, which completes the proof. In addition, it is straightforward to show that $E(\underline{x}_{i,t-1} \Delta \underline{y}_{i,t-1}) \neq 0$.

B   Proof of Proposition 4

Firstly we establish that $S_{DIFX2} \xrightarrow{d} \chi^2_{h_x - (K+1)}$ under the alternative of $H_1 : \mathbf{\Sigma}_\phi \neq \mathbf{0}$, where S_{DIFX2} is defined in (28). Rewriting $\frac{1}{\sqrt{N}} \sum_{i=1}^N \underline{\mathbf{Z}}_{Xi}' \left(\Delta \underline{\mathbf{u}}_i - \Delta \widehat{\underline{\widetilde{\mathbf{u}}}}_i\right)$ gives

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \underline{\mathbf{Z}}_{Xi}' \Delta \widehat{\underline{\widetilde{\mathbf{u}}}}_i = \frac{1}{\sqrt{N}} \sum_{i=1}^N \underline{\mathbf{Z}}_{Xi}' \Delta \underline{\mathbf{u}}_i - N^{-1} \sum_{i=1}^N \underline{\mathbf{Z}}_{Xi}' \Delta \underline{\mathbf{W}}_i \sqrt{N} \left(\widehat{\widetilde{\boldsymbol{\theta}}}_{DIF2} - \boldsymbol{\theta}\right). \quad (43)$$

Next, orthogonally decomposing $\widehat{\mathbf{\Omega}}_X^{-1/2} \sum_{i=1}^N \underline{\mathbf{Z}}_{Xi}' \Delta \widehat{\underline{\widetilde{\mathbf{u}}}}_i$ gives

$$\begin{aligned}
\frac{1}{\sqrt{N}} \widehat{\mathbf{\Omega}}_X^{-1/2} \sum_{i=1}^N \underline{\mathbf{Z}}_{Xi}' \Delta \widehat{\underline{\widetilde{\mathbf{u}}}}_i &= \frac{1}{\sqrt{N}} \widehat{\mathbf{P}}_B \widehat{\mathbf{\Omega}}_X^{-1/2} \sum_{i=1}^N \underline{\mathbf{Z}}_{Xi}' \Delta \widehat{\underline{\widetilde{\mathbf{u}}}}_i + \frac{1}{\sqrt{N}} \widehat{\mathbf{M}}_B \widehat{\mathbf{\Omega}}_X^{-1/2} \sum_{i=1}^N \underline{\mathbf{Z}}_{Xi}' \Delta \widehat{\underline{\widetilde{\mathbf{u}}}}_i \\
&= \frac{1}{\sqrt{N}} \widehat{\mathbf{M}}_B \widehat{\mathbf{\Omega}}_X^{-1/2} \sum_{i=1}^N \underline{\mathbf{Z}}_{Xi}' \Delta \widehat{\underline{\widetilde{\mathbf{u}}}}_i, \quad (44)
\end{aligned}$$

18

where $\hat{\mathbf{P}}_B = \hat{\mathbf{B}}\left(\hat{\mathbf{B}}'\hat{\mathbf{B}}\right)^{-1}\hat{\mathbf{B}}'$, $\hat{\mathbf{M}}_B = \mathbf{I} - \hat{\mathbf{P}}_B$, $\hat{\mathbf{B}} = -\widehat{\boldsymbol{\Omega}}_X^{-1/2} N^{-1} \sum_{i=1}^{N} \underline{\mathbf{Z}}'_{Xi} \Delta\underline{\mathbf{W}}_i$, and the last line follows from $\hat{\mathbf{B}}'\widehat{\boldsymbol{\Omega}}_X^{-1/2} \sum_{i=1}^{N} \underline{\mathbf{Z}}'_{Xi} \Delta\underline{\hat{\hat{\mathbf{u}}}}_i = \mathbf{0}$ by the definition of the GMM estimator. Substituting (43) into (44) yields

$$\frac{1}{\sqrt{N}}\widehat{\boldsymbol{\Omega}}_X^{-1/2} \sum_{i=1}^{N} \underline{\mathbf{Z}}'_{Xi} \Delta\underline{\hat{\hat{\mathbf{u}}}}_i = \frac{1}{\sqrt{N}}\hat{\mathbf{M}}_B \widehat{\boldsymbol{\Omega}}_X^{-1/2} \sum_{i=1}^{N} \underline{\mathbf{Z}}'_{Xi} \Delta\underline{\mathbf{u}}_i, \tag{45}$$

since $\hat{\mathbf{M}}_B \hat{\mathbf{B}} = \mathbf{0}$.

We can express the instruments as deviations from their cross-sectional averages:

$$\underline{\mathbf{Z}}_{Xi} = \mathbf{Z}_{Xi} - \bar{\mathbf{Z}}_X, \tag{46}$$

where $\mathbf{Z}_{Xi}$ is defined similarly to $\underline{\mathbf{Z}}_{Xi}$ but all $\underline{\mathbf{x}}'_{it}s$ are replaced with $\mathbf{x}_{it}$, and $\bar{\mathbf{Z}}_X$ is defined similarly but all non-zero elements are replaced with their cross-sectional averages, $\bar{\mathbf{x}}_t$. Also define the instruments in terms of deviations from their mean as

$$\underline{\mathbf{Z}}_{Xi}^o = \mathbf{Z}_{Xi} - \mathbf{m}_{Z_X}, \tag{47}$$

and

$$\Delta\underline{\mathbf{u}}_i^o = \Delta\mathbf{F}\left(\boldsymbol{\phi}_i - \boldsymbol{\phi}\right) + \Delta\boldsymbol{\varepsilon}_i \tag{48}$$

where $\mathbf{m}_{Z_X} = E(\mathbf{Z}_{Xi})$, $\boldsymbol{\phi} = E(\boldsymbol{\phi}_i)$. Using (46)-(48)

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \underline{\mathbf{Z}}'_{Xi} \Delta\underline{\mathbf{u}}_i = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \underline{\mathbf{Z}}_{Xi}^{o\prime} \left(\Delta\mathbf{u}_i - \Delta\bar{\mathbf{u}}\right) \tag{49}$$

$$- \left(\bar{\mathbf{Z}}_X - \mathbf{m}_{Z_X}\right)' \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \left(\Delta\mathbf{u}_i - \Delta\bar{\mathbf{u}}\right).$$

It is easily seen that the second term of (49) is asymptotically negligible. Consider the first term of (49). Reminding ourselves that $\Delta\mathbf{u}_i = \Delta\mathbf{F}\boldsymbol{\phi}_i + \Delta\boldsymbol{\varepsilon}_i$, we have

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \underline{\mathbf{Z}}_{Xi}^{o\prime} \left(\Delta\mathbf{u}_i - \Delta\bar{\mathbf{u}}\right)$$

$$= \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \underline{\mathbf{Z}}_{Xi}^{o\prime} \Delta\mathbf{F}\left[\left(\boldsymbol{\phi}_i - \boldsymbol{\phi}\right) - \left(\bar{\boldsymbol{\phi}} - \boldsymbol{\phi}\right)\right] - \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \underline{\mathbf{Z}}_{Xi}^{o\prime} \left(\Delta\boldsymbol{\varepsilon}_i - \Delta\bar{\boldsymbol{\varepsilon}}\right).$$

By the assumptions $\boldsymbol{\phi}_i \sim i.i.d.(\boldsymbol{\phi}, \boldsymbol{\Sigma}_\phi)$, $\varepsilon_{it} \sim i.i.d.(0, \sigma_i^2)$ above, $\left(\bar{\boldsymbol{\phi}} - \boldsymbol{\phi}\right) = O_p(N^{-1/2})$ and $\Delta\bar{\boldsymbol{\varepsilon}} = O_p(N^{-1/2})$ as well as $N^{-1/2}\sum_{i=1}^{N} \underline{\mathbf{Z}}_{Xi}^o = O_p(1)$. Then it follows that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \underline{\mathbf{Z}}'_{Xi} \Delta\underline{\mathbf{u}}_i = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \underline{\mathbf{Z}}_{Xi}^{o\prime} \Delta\underline{\mathbf{u}}_i^o + o_p(1). \tag{50}$$

Since $\widehat{\boldsymbol{\Omega}}_X - \boldsymbol{\Omega}_X = o_p(1)$, $\hat{\mathbf{B}} - \mathbf{B} = o_p(1)$ where $\mathbf{B} = -\boldsymbol{\Omega}_X^{-1/2} N^{-1} \sum_{i=1}^{N} \underline{\mathbf{Z}}_{Xi}^{o\prime} \Delta\underline{\mathbf{W}}_i^o$ and $\boldsymbol{\Omega}_X = p\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N} E(\underline{\mathbf{Z}}_{Xi}^{o\prime} \Delta\underline{\mathbf{u}}_i^o \Delta\underline{\mathbf{u}}_i^{o\prime} \underline{\mathbf{Z}}_{Xi}^o)$ with obvious notations, together with (50), (45) can be written as

$$\frac{1}{\sqrt{N}}\widehat{\boldsymbol{\Omega}}_X^{-1/2} \sum_{i=1}^{N} \underline{\mathbf{Z}}'_{Xi} \Delta\underline{\hat{\hat{\mathbf{u}}}}_i = \frac{1}{\sqrt{N}}\mathbf{M}_B \boldsymbol{\Omega}_X^{-1/2} \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \underline{\mathbf{Z}}_{Xi}^{o\prime} \Delta\underline{\mathbf{u}}_i^o + o_p(1). \tag{51}$$

As $\underline{\mathbf{Z}}_{Xi}^{o\prime} \Delta\underline{\mathbf{u}}_i^o$ are independent across i, a suitable Central Limit Theorem ensures that

$$\boldsymbol{\Omega}_X^{-1/2} \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \underline{\mathbf{Z}}_{Xi}^{o\prime} \Delta\underline{\mathbf{u}}_i^o \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_{h_x}).$$

19

Noting that $rank(\mathbf{M}_B) = h_x - (K+1)$ we have

$$S_{DIFX2} = \frac{1}{N}\left(\sum_{i=1}^{N}\Delta\underline{\mathbf{u}}_i^{o\prime}\mathbf{Z}_{Xi}^o\right)\mathbf{\Omega}_X^{-1/2}\mathbf{M}_B\mathbf{\Omega}_X^{-1/2}\left(\sum_{i=1}^{N}\mathbf{Z}_{Xi}^{o\prime}\Delta\underline{\mathbf{u}}_i^o\right) + o_p(1) \qquad (52)$$
$$\xrightarrow{d}\chi^2_{h_x-(K+1)},$$

under the alternative hypothesis of $H_1 : \mathbf{\Sigma}_\phi \neq \mathbf{0}$, as required. Under the null hypothesis of $H_0 : \mathbf{\Sigma}_\phi = \mathbf{0}$, (52) follows immediately. Also it is straightforward to establish that $S_{DIF2}\xrightarrow{d}\chi^2_{h-(K+1)}$, where S_{DIF2} is defined as in (27), in line with the proof provided for (52).

Now we provide the asymptotic distribution of $S_{DIF2} - S_{DIFX2}$. Consider the local alternative

$$H_N : \boldsymbol{\phi}_i = \boldsymbol{\phi} + \frac{\boldsymbol{\eta}_i}{N^{1/4}},$$

where $0 < ||\boldsymbol{\eta}_i|| < \infty$ for all i, which are assumed to be non-stochastic for expositional convenience. Here the analysis is based on the instruments in terms of deviations from their true mean, rather than from the cross-sectional average, since we have already shown that the effect of such replacement is asymptotically negligible. Without loss of generality, consider $\underline{\mathbf{Z}}_{Yi}^o = diag(\underline{y}_{i,t-2}^o)$, $t = 2, 3, ..., T$, $h_y = T - 1$. Also define

$$\underline{\mathbf{Z}}_i^o = \begin{bmatrix} \underline{\mathbf{Z}}_{Yi}^o & \underline{\mathbf{Z}}_{Xi}^o \end{bmatrix}, \; \mathbf{\Omega} = p\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}E\left[\underline{\mathbf{Z}}_i^{o\prime}\Delta\underline{\mathbf{u}}_i^o\Delta\underline{\mathbf{u}}^{o\prime}\underline{\mathbf{Z}}_i^o\right], \qquad (53)$$

where

$$\mathbf{\Omega} = \begin{bmatrix} \mathbf{\Omega}_Y & \mathbf{\Omega}_{YX} \\ \mathbf{\Omega}_{XY} & \mathbf{\Omega}_X \end{bmatrix}, \qquad (54)$$

with block elements that are conformable with $\underline{\mathbf{Z}}_{Yi}^o$ and $\underline{\mathbf{Z}}_{Xi}^o$. By using (42) we have

$$E\left(N^{-1/2}\sum_{i=1}^{N}\underline{\mathbf{Z}}_{Yi}^{o\prime}\Delta\underline{\mathbf{u}}_i^o\right) = \boldsymbol{\delta}_N,$$

where $\boldsymbol{\delta}_N = O(1)$ is a $(T-1\times 1)$ vector whose first element is $N^{-1}\sum_{i=1}^{N}E(\underline{y}_{i0}\boldsymbol{\eta}_i{}')\Delta\mathbf{f}_2$ and the $(t-1)^{th}$ elements are $N^{-1}\sum_{i=1}^{N}\lambda^t E\left(\underline{y}_{i0}\boldsymbol{\eta}_i{}'\right)\Delta\mathbf{f}_{t+2} + \Delta\mathbf{f}_{t+2}'\boldsymbol{\eta}_i\boldsymbol{\eta}_i{}'\sum_{j=0}^{t-1}\lambda^j\mathbf{f}_{t-j}$, for $t = 2, 3, ..., T-1$. Define

$$\underline{\mathbf{Z}}_i^{o*} = \underline{\mathbf{Z}}_i^o\mathbf{L}', \text{ with } \mathbf{L} = \begin{bmatrix} \mathbf{I}_{h_y} & -\mathbf{\Omega}_{YX}\mathbf{\Omega}_X^{-1} \\ \mathbf{0} & \mathbf{I}_{h_x} \end{bmatrix}$$

where $\mathbf{L}$ is non-singular, so that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\underline{\mathbf{Z}}_i^{o*\prime}\Delta\underline{\mathbf{u}}_i^o - \begin{bmatrix} \boldsymbol{\delta}_N \\ \mathbf{0} \end{bmatrix}\xrightarrow{d}N\left(\mathbf{0}, \mathbf{\Omega}^*\right),$$

where

$$\mathbf{\Omega}^* = \mathbf{L}\mathbf{\Omega}\mathbf{L}' = \begin{bmatrix} \mathbf{\Omega}_Y^* & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega}_X \end{bmatrix},$$

with $\mathbf{\Omega}_Y^* = \mathbf{\Omega}_Y - \mathbf{\Omega}_{YX}\mathbf{\Omega}_X^{-1}\mathbf{\Omega}_{XY}$. It follows that

$$S_{DIF2} = \frac{1}{N}\left(\sum_{i=1}^{N}\Delta\underline{\mathbf{u}}_i^{o\prime}\underline{\mathbf{Z}}_i^{o*}\right)\mathbf{\Omega}^{*-1/2}\mathbf{M}_B^*\mathbf{\Omega}^{*-1/2}\left(\sum_{i=1}^{N}\underline{\mathbf{Z}}_i^{o*\prime}\Delta\underline{\mathbf{u}}_i^o\right) + o_p(1)$$

with $\mathbf{M}_B^* = \mathbf{I} - \mathbf{B}^*\left(\mathbf{B}^{*\prime}\mathbf{B}^*\right)^{-1}\mathbf{B}^{*\prime}$,

$$\mathbf{B}^* = \begin{pmatrix} \mathbf{B}_Y^* \\ \mathbf{B} \end{pmatrix} = -\mathbf{\Omega}^{*-1/2}N^{-1}\sum_{i=1}^{N}\underline{\mathbf{Z}}_i^{o*\prime}\Delta\underline{\mathbf{W}}_i^o,$$

20

so that

$$S_{DIF2} - S_{DIFX2} = \frac{1}{N} \left(\sum_{i=1}^{N} \Delta \underline{\mathbf{u}}_i^{o\prime} \underline{\mathbf{Z}}_i^{o*} \right) \boldsymbol{\Omega}^{*-1/2} \mathbf{M}_B^* \boldsymbol{\Omega}^{*-1/2} \left(\sum_{i=1}^{N} \underline{\mathbf{Z}}_i^{o*\prime} \Delta \underline{\mathbf{u}}_i^{o} \right) \tag{55}$$

$$- \frac{1}{N} \left(\sum_{i=1}^{N} \Delta \underline{\mathbf{u}}_i^{o\prime} \underline{\mathbf{Z}}_{Xi}^{o} \right) \boldsymbol{\Omega}_X^{-1/2} \mathbf{M}_B \boldsymbol{\Omega}_X^{-1/2} \left(\sum_{i=1}^{N} \underline{\mathbf{Z}}_{Xi}^{o\prime} \Delta \underline{\mathbf{u}}_i^{o} \right) + o_p(1)$$

$$= \frac{1}{N} \left(\sum_{i=1}^{N} \Delta \underline{\mathbf{u}}_i^{o\prime} \underline{\mathbf{Z}}_i^{o*} \right) \boldsymbol{\Omega}^{*-1/2} \mathbf{M} \boldsymbol{\Omega}^{*-1/2} \left(\sum_{i=1}^{N} \underline{\mathbf{Z}}_i^{o*\prime} \Delta \underline{\mathbf{u}}_i^{o} \right) + o_p(1)$$

where

$$\mathbf{M} = \mathbf{M}_B^* - \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_B \end{bmatrix} = \begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{bmatrix}, \tag{56}$$

which is a symmetric and idempotent matrix of rank h_y. Finally, we have

$$S_{DIF2} - S_{DIFX2} = \frac{1}{N} \left(\sum_{i=1}^{N} \Delta \underline{\mathbf{u}}_i^{o\prime} \underline{\mathbf{Z}}_i^{o*\prime} \right) \boldsymbol{\Omega}^{*-1/2} \mathbf{M} \boldsymbol{\Omega}^{*-1/2} \left(\sum_{i=1}^{N} \underline{\mathbf{Z}}_i^{o*\prime} \Delta \underline{\mathbf{u}}_i^{o} \right) + o_p(1)$$

$$\xrightarrow{d} \chi^2 (h_y, \xi)$$

which is a non-central chi-squared distribution with h_y degrees of freedom and non-centrality parameter $\xi = p \lim_{N \to \infty} \boldsymbol{\delta}_N' \boldsymbol{\Omega}_Y^{*-1/2} \mathbf{M}_{11} \boldsymbol{\Omega}_Y^{*-1/2} \boldsymbol{\delta}_N > 0$, so long as $\boldsymbol{\delta}_N \neq \mathbf{0}$. Therefore, the result $S_{DIF2} - S_{DIFX2} \xrightarrow{d} \chi_{h_y}^2$ readily follows under the null hypothesis, as required.

Furthermore, using (55) and (56), it is easily seen that $S_{DIF2} \xrightarrow{d} \chi^2 (h - (K + 1), \xi)$ under the local alternatives, where ξ is the same non-centrality parameter of the asymptotic distribution of $(S_{DIF2} - S_{DIFX2})$. Therefore, when $h - (K + 1) > h_y$, or − subtracting h_y from both sides − $h_x > (K + 1)$, the Sargan's difference test is locally more powerful than the overidentifying restrictions test.

C   Proof of Proposition 8

In line with the proof of Proposition 3 in Appendix A, consider $E(\Delta \underline{x}_{it} (\underline{\alpha}_i + \underline{u}_{it}))$. For $t \geq 3$ and using (40) together with $\underline{u}_{it} = \boldsymbol{\phi}_i' \mathbf{f}_t + \underline{\varepsilon}_{it}$, we have under the alternative hypothesis of $\phi_i \neq \phi$, $E(\Delta \underline{x}_{it} (\underline{\alpha}_i + \underline{u}_{it})) = 0$, under Assumptions 1-4, 5(a)-(c), 6 and model (7). A similar line of argument proves that $E(\Delta \underline{x}_{i2} (\underline{\alpha}_i + \underline{u}_{i2})) = 0$. However, for $t \geq 2$ and using (40) and (41) we have under the alternative

$$E(\Delta \underline{y}_{it} (\underline{\alpha}_i + \underline{u}_{it+1})) = \lambda^t E \left(\Delta \underline{y}_{i1} \boldsymbol{\phi}_i' \right) \mathbf{f}_{t+1} + \mathbf{f}_{t+1}' E \left(\boldsymbol{\phi}_i \boldsymbol{\phi}_i' \right) \sum_{j=0}^{t-1} \lambda^j \Delta \mathbf{f}_{t-j} \neq 0,$$

under Assumptions 1-4, 5(a)-(c), 6 and model (7). A similar line of argument will prove that $E(\Delta \underline{y}_{i1} (\underline{\alpha}_i + \underline{u}_{i2})) \neq 0$ under the alternative. Finally, under the null hypothesis, it is also easily seen that $E(\Delta \underline{y}_{it} (\underline{\alpha}_i + \underline{u}_{it+1})) = 0$ for $t \geq 1$ and $E(\Delta \underline{x}_{it} (\underline{\alpha}_i + \underline{u}_{it})) = 0$ for $t \geq 2$, which completes the proof. Furthermore, it is straightforward to show that $E(\Delta \underline{x}_{i,t-1} \underline{x}_{it}) \neq 0$, $E(\Delta \underline{x}_{i,t-1} \underline{y}_{i,t-1}) \neq 0$, $E(\Delta \underline{y}_{i,t-1} \underline{x}_{it}) \neq 0$, $E(\Delta \underline{y}_{i,t-1} \underline{y}_{i,t-1}) \neq 0$.

D   Derivations of Parameters in Monte Carlo Experiments

Using the lag operator, L, we can write y_{it} and x_{it} as

$$y_{it} = \frac{\alpha_i}{1 - \lambda} + \frac{\beta}{1 - \lambda L} x_{it} + \frac{\phi_i}{1 - \lambda L} f_t + \frac{1}{1 - \lambda L} \varepsilon_{it} \tag{57}$$

$$x_{it} = \frac{\pi L}{1 - \rho L} \varepsilon_{it} + \frac{\gamma_i}{1 - \rho L} f_t + \frac{1}{1 - \rho L} v_{it} \tag{58}$$

21

and thereby substituting (58) into (57) yields

$$y_{it} = \frac{\alpha_i}{1-\lambda} + \frac{\beta v_{it}}{(1-\lambda L)(1-\rho L)} + \frac{\beta \gamma_i + \phi_i (1-\rho L)}{(1-\lambda L)(1-\rho L)} f_t + \frac{1+(\beta \pi - \rho)L}{(1-\lambda L)(1-\rho L)} \varepsilon_{it}. \tag{59}$$

Define $y_{it}^* = y_{it} - \alpha_i/(1-\lambda)$, such that (57) can be rewritten as

$$y_{it}^* = \lambda y_{it-1}^* + \beta x_{it} + u_{it} \tag{60}$$

and let the signal-to-noise ratio be denoted by σ_s^2/σ_u^2, where σ_s^2 is the variance of the signal,

$$\sigma_s^2 = var(y_{it}^* - u_{it}) = var(y_{it}^*) + var(u_{it}) - 2cov(y_{it}^*, u_{it}). \tag{61}$$

σ_s^2 varies across designs, with the aim being to keep the signal-to-noise ratio constant over changes in λ and the distribution of ϕ_i, so that the explanatory power of the model does not change. In particular, we set $\sigma_s^2/\sigma_u^2 = \zeta = 3$, where $\sigma_u^2 = var(u_{it})$. We normalise $\sigma_f^2 = \sigma_\varepsilon^2 = 1$ and we keep the total signal-to-noise ratio fixed by modifying σ_s^2 accordingly through changes in σ_v^2. It can be shown that

$$\sigma_s^2 = \left\{ \beta^2 \sigma_v^2 + \sigma_f^2 b_1 + \sigma_\varepsilon^2 b_2 \right\} a_1 - 2\beta \mu_\gamma \mu_\phi \sigma_f^2 - \sigma_u^2,$$

where $a_1 = \frac{(1+\lambda\rho)}{(1-\rho^2)(1-\lambda^2)(1-\lambda\rho)}$, $b_1 = \left(\beta \mu_\gamma + \mu_\phi \right)^2 + \left(\mu_\phi \rho \right)^2 - \frac{2(\mu_\phi \rho)(\lambda+\rho)}{1+\lambda\rho}$ and $b_2 = 1 + (\beta\pi - \rho)^2 + \frac{2(\beta\pi-\rho)(\lambda+\rho)}{1+\lambda\rho}$.

Applying the normalisation $\sigma_f^2 = \sigma_\varepsilon^2 = 1$, substituting $\sigma_s^2 = \sigma_u^2 \zeta$, and solving for σ_v^2 yields

$$\sigma_v^2 = \beta^{-2} \left\{ \left[\sigma_u^2 (1+\zeta) + 2\beta \mu_\gamma \mu_\phi \right] / a_1 - (b_1 + b_2) \right\}.$$

In line with the simulation design of Bun and Kiviet (2006), we choose σ_α^2 such that the ratio of the impacts on $var(y_{it})$ of the two variance components α_i and (ε_{it}, f_t) is ψ^2. By (59)

$$
\begin{aligned}
var(y_{it}) &= var\left(\frac{1}{1-\lambda} \alpha_i \right) + var\left(\frac{\beta}{(1-\lambda L)(1-\rho L)} v_{it} \right) \\
&\quad + var\left[\frac{\beta \gamma_i + \phi_i (1-\rho L)}{(1-\lambda L)(1-\rho L)} f_t \right] + var\left[\frac{1+(\beta\pi-\rho)L}{(1-\lambda L)(1-\rho L)} \varepsilon_{it} \right] \\
&= \frac{\sigma_\alpha^2}{(1-\lambda)^2} + (\beta^2 \sigma_v^2 + \sigma_f^2 b_1 + \sigma_\varepsilon^2 b_2) a_1.
\end{aligned}
$$

Now define ψ^2 such that $\frac{\sigma_\alpha^2}{(1-\lambda)^2} = \psi^2 (\sigma_f^2 b_1 + \sigma_\varepsilon^2 b_2) a_1$. By applying the normalisation $\sigma_\varepsilon^2 = \sigma_f^2 = 1$, we set $\sigma_\alpha^2 = \psi^2 (1-\lambda)^2 (b_1 + b_2) a_1$. We choose $\psi = 1$.

References

[1] Ahn, S.G., Lee, Y.H. and P. Schmidt, 2001, GMM estimation of linear panel data models with time-varying individual effects. Journal of Econometrics 102, 219-255.

[2] Ahn, S.C and P. Schmidt, 1995, Efficient Estimations of Models for Dynamic Panel Data. Journal of Econometrics 68, 5-28.

[3] Anselin, L., 1988, Spatial Econometrics: Method and Models, Kluwer Academic Publishers, Oxford.

[4] Anselin, L., 2001, Spatial Econometrics, in: B. Baltagi, (Eds.), A Companion to Theoretical Econometrics, Blackwell, Oxford, pp. 310–330.

[5] Arellano, M., 2003, Panel Data Econometrics, Oxford University Press, Oxford.

[6] Arellano, M. and S. Bond, 1991, Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Review of Economic Studies 58, 277-297.

[7] Arellano, M. and S. Bond, 1999, Dynamic panel data estimation using DPD98 for Gauss: A guide for users, unpublished manuscript.

[8] Arellano, M. and O. Bover, 1995, Another look at the instrumental-variable estimation of error components models. Journal of Econometrics 68, 29-52.

[9] Bai, J. and S. Ng, 2004, A PANIC attack on unit roots and cointegration. Econometrica 72, 1127-1177.

[10] Blundell, R. and S. Bond, 1998, Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics 87, 115-143.

[11] Blundell, R. and S. Bond, 2000, GMM estimation with persistent panel data: an application to production functions. Econometric Reviews 19, 321-340.

[12] Bond, S., Bowsher, C. and F. Windmeijer, (2001). Criterion-based inference for GMM in autoregressive panel data models. Economics Letters 73, 379-388.

[13] Bowsher, C.G., 2002, On testing overidentifying restrictions in dynamic panel data models. Economics Letters 77, 211-220.

[14] Bun, M.J.G. and J.F. Kiviet, 2006, The effects of dynamic feedbacks on LS and MM estimator accuracy in panel data models. Journal of Econometrics 132, 409-444.

[15] Breusch, T.S. and A.R. Pagan, 1980, The Lagrange Multiplier test and its application to model specifications in econometrics. Review of Economic Studies 47, 239-53.

23

[16] Conley, T.G., 1999, GMM estimation with cross-sectional dependence. Journal of Econometrics 92, 1-45.

[17] Forni, M. and L. Reichlin, 1998, Let's Get Real: A Factor Analytical Approach to Disaggregated Business Cycle Dynamics. Review of Economic Studies 65, 453–473.

[18] Frees, E.W., 1995, Assessing cross-sectional correlation in panel data. Journal of Econometrics 69, 393-414.

[19] Friedman, M., 1937, The use of ranks to avoid the assumption of normality implicit in the analysis of variance. Journal of the American Statistical Association 32, 675-701.

[20] Hahn, J., Hausman, J. and G. Kuersteiner, 2006, Long difference instrumental variables estimation for dynamic panel models with fixed effects. Journal of Econometrics 140, 574-617.

[21] Han, C., Orea, L. and P. Schmidt, 1995, Estimation of a panel data model with parametric temporal variation in individual effects. Journal of Econometrics 126, 241-267.

[22] Hansen, L.P., 1982, Large sample properties of generalized method of moments estimators. Econometrica 50, 1029-1054.

[23] Hausman, J.A., 1978, Specification tests in econometrics. Econometrica 46, 1251-1272.

[24] Holtz-Eakin, D., Newey, W.K. and H. Rosen, 1988, Estimating Vector Autoregressions with Panel Data. Econometrica 56, 1371–1395.

[25] Hayashi, F., 2000, Econometrics, Princeton University Press, Princeton.

[26] Kapoor, M., Kelejian, H.H. and I.R. Prucha, 2007, Panel data models with spatially correlated error components. Journal of Econometrics 140, 97-130.

[27] Kiviet, J.F., 1995, On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. Journal of Econometrics 68, 53-78.

[28] Lee, L., 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica, 72, 1899-1926.

[29] Lee, L., 2007. GMM and 2SLS estimation of mixed regressive spatial autoregressive models. Journal of Econometrics, 137, 489-514.

[30] Moon, H.R. and B. Perron, 2004, Testing for A Unit Root in Panels with Dynamic Factors. Journal of Econometrics 122, 81-126.

[31] Newey, W., 1985, Generalized method of moments specification testing. Journal of Econometrics 29, 229-256.

24

[32] Ng, S., 2006, Testing cross-section correlation in panel data using spacings. Journal of Business & Economic Statistics 24, 12-23.

[33] Pesaran, M.H., 2004, General diagnostic tests for cross section dependence in panels. Unpublished manuscript, Faculty of Economics and Politics, University of Cambridge.

[34] Pesaran, M.H., 2006, Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure. Econometrica 74, 967-1012.

[35] Pesaran, M.H. and R.P. Smith, 1995, Estimating long-run relationships from dynamic heterogeneous panels. Journal of Econometrics 68, 79-113.

[36] Pesaran, M.H., Ullah, A. and T. Yamagata, 2006, A Bias-Adjusted LM Test of Error Cross Section Independence. Econometrics Journal 11, 105-127.

[37] Phillips, P. and D. Sul, 2003, Dynamic panel estimation and homogeneity testing under cross section dependence. The Econometrics Journal 6, 217-259.

[38] Robertson, D. and J. Symons, 2000, Factor residuals in SUR regressions: estimating panels allowing for cross-sectional correlation, Unpublished manuscript, Faculty of Economics and Politics, University of Cambridge.

[39] Sarafidis, V. and D. Robertson, 2008, On The Impact of Cross-sectional Dependence in Short Dynamic Panel Estimation, mimeo.

[40] Sargan, J.D., 1958, The estimation of economic relationships using instrumental variables. Econometrica 26, 393-415.

[41] Sargan, J.D., 1988, Testing for misspecification after estimating using instrumental variables, in: E. Maasoumi, (Eds.), Contributions to Econometrics: John Denis Sargan, Vol. 1. Cambridge University Press.

[42] Stock, J. and M.W. Watson, 2002, Macroeconomic Forecasting Using Diffusion Indexes. Journal of Business & Economic Statistics 20, 147–162.

Table 1: Size and Power of the Cross Section Dependence Tests in the Case with Predetermined Regressors with $T = 5$

	$\lambda = 0.2$				$\lambda = 0.5$				$\lambda = 0.8$			
Test,N	50	100	200	400	50	100	200	400	50	100	200	400
	Size: $\phi_i = \phi = 0$											
LM	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
CD	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
D_{DIF2}	3.55	4.20	5.30	5.35	4.40	4.45	4.60	5.40	6.45	7.05	6.30	5.40
D_{SYS2}	3.65	5.20	5.30	5.15	4.15	5.35	5.55	5.80	5.10	6.25	6.35	6.30
	Power: $\phi_i \sim i.i.d.U[-0.3, 0.7]$											
LM	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
CD	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
D_{DIF2}	4.80	9.10	18.15	29.25	6.00	9.90	18.70	32.10	7.65	13.95	21.75	35.30
D_{SYS2}	4.00	10.90	21.65	36.40	5.35	10.75	22.95	41.60	5.50	13.55	28.05	46.55
	Power: $\phi_i \sim i.i.d.U[-1, 2]$											
LM	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
CD	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
D_{DIF2}	38.65	70.10	86.40	94.00	43.25	74.45	88.45	95.25	46.25	77.85	91.30	95.85
D_{SYS2}	29.80	77.25	92.35	96.95	34.50	79.75	93.30	97.80	36.80	83.30	95.90	98.80
	Power: $\phi_i \sim i.i.d.U[-1, 4]$											
LM	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
CD	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
D_{DIF2}	63.30	89.70	95.15	98.20	67.70	90.30	96.40	98.55	69.00	92.65	98.00	99.25
D_{SYS2}	55.25	93.80	98.60	99.30	56.35	94.10	98.30	99.65	62.20	96.35	98.95	99.80

Notes: LM and CD denote the Breusch-Pagan LM test and Pesaran's (2004) CD test, respectively. Both are based on the residuals of the Fixed Effects estimator. D_{DIF2} and D_{SYS2} denote Sargan's difference tests based on the two-step Arellano Bond (1991) DIF estimator, and on the two-step Blundell and Bond (1998) SYS estimator respectively. The data generating process (DGP) is $y_{it} = \alpha_i + \lambda y_{i,t-1} + \beta x_{it} + \phi_i f_t + \varepsilon_{it}$, $i = 1, 2, ..., N$, $t = -48, -47, ..., T$ with $y_{i,-49} = 0$. The initial 49 observations are discarded. $\alpha_i \sim i.i.d.N(1, \sigma_\alpha^2)$, λ and ϕ_i are as specified in the Table, $\beta = 1 - \lambda$, $f_t \sim i.i.d.N(0, 1)$, $\varepsilon_{it} \sim i.i.d.N(0, 1)$; $x_{it} = \rho x_{i,t-1} + \pi \varepsilon_{i,t-1} + \gamma_i f_t + v_{it}$, $i = 1, 2, ..., N$, $t = -48, -47, ..., T$ with $x_{i,-49} = 0$ and the initial 50 observations being discarded. $\rho = \lambda = 0.5$, $\pi = 0.5$, $\gamma_i \sim i.i.d.U[-1, 2]$, $v_{it} \sim i.i.d.N(0, \sigma_v^2)$, σ_v^2 is chosen such that the signal-to-noise ratio equals 3. σ_α^2 is chosen such that the impact of the two variance components α_i and (f_t, ε_{it}) on $var(y_{it})$ is constant. All variables are cross-sectionally demeaned before computing statistics. All experiments are based on 2,000 replications.

26

Table 2: Bias ($\times 1000$) of Fixed Effects and GMM estimators for λ, in the Case with Predetermined Regressors with $T = 5$

Test,N	$\lambda = 0.2$				$\lambda = 0.5$				$\lambda = 0.8$			
	50	100	200	400	50	100	200	400	50	100	200	400
	No Cross Section Dependence: $\phi_i = \phi = 0$											
FE	-115.62	-114.70	-116.66	-115.46	-187.03	-183.89	-185.06	-184.63	-392.65	-389.05	-388.91	-388.59
DIF1	-23.68	-10.76	-7.76	-2.46	-46.94	-24.22	-10.82	-5.29	-205.25	-105.79	-59.39	-26.46
DIF2	-22.83	-10.11	-7.65	-1.98	-48.67	-24.47	-10.48	-5.25	-223.06	-115.40	-61.99	-27.85
DIFX1	-20.58	-10.07	-7.70	-2.30	-65.99	-32.32	-15.45	-7.80	-438.04	-326.91	-209.79	-116.42
DIFX2	-19.93	-10.18	-7.62	-2.11	-66.16	-33.28	-14.50	-7.79	-457.19	-339.03	-214.87	-118.03
SYS1	13.43	7.31	1.92	2.04	6.89	2.96	1.59	1.42	-12.35	-3.91	-5.17	-1.46
SYS2	11.58	6.06	0.34	1.51	7.24	4.18	2.20	1.42	-4.91	2.04	-0.20	0.54
SYSX1	9.57	4.71	0.54	1.47	5.67	5.55	3.01	2.37	-50.06	-42.01	-30.40	-13.39
SYSX2	7.94	3.37	-0.59	1.45	10.07	8.00	6.87	4.32	-37.91	-28.31	-15.51	-2.08
	Low Cross Section Dependence: $\phi_i \sim i.i.d.U[-0.3, 0.7]$											
FE	-107.82	-106.72	-108.15	-107.89	-171.34	-168.69	-169.18	-168.50	-325.46	-321.39	-321.04	-319.85
DIF1	-22.29	-10.66	-7.80	-3.68	-42.94	-23.58	-11.97	-6.50	-140.41	-75.11	-45.60	-24.33
DIF2	-20.85	-9.55	-7.44	-2.92	-43.64	-23.37	-11.12	-6.08	-147.90	-77.45	-46.13	-24.42
DIFX1	-18.53	-8.89	-7.33	-2.29	-60.43	-29.10	-13.46	-7.00	-353.16	-251.67	-159.44	-87.67
DIFX2	-17.71	-8.66	-7.26	-2.05	-60.16	-29.79	-12.58	-6.89	-365.60	-254.65	-158.36	-85.55
SYS1	12.24	6.24	1.06	0.53	2.80	-0.67	-1.42	-0.59	-18.63	-11.68	-10.26	-6.72
SYS2	10.12	4.93	-0.41	0.12	3.05	1.22	-0.17	-0.37	-12.05	-5.93	-5.40	-3.88
SYSX1	9.36	4.72	0.38	1.27	2.25	3.82	2.44	1.75	-53.49	-43.00	-30.35	-15.23
SYSX2	7.18	3.42	-0.89	1.29	6.42	6.01	6.09	3.68	-43.81	-31.09	-18.55	-6.48
	Moderate Cross Section Dependence: $\phi_i \sim i.i.d.U[-1, 2]$											
FE	-87.17	-85.41	-85.48	-87.94	-143.05	-141.46	-140.35	-138.30	-281.73	-278.38	-277.75	-274.77
DIF1	-30.36	-21.66	-18.81	-19.59	-60.73	-50.04	-42.90	-35.16	-202.24	-165.42	-151.49	-136.35
DIF2	-27.14	-16.75	-15.14	-14.83	-57.18	-44.98	-36.37	-29.22	-204.32	-163.72	-150.64	-135.37
DIFX1	-14.25	-6.81	-5.99	-1.91	-51.31	-24.03	-11.00	-5.72	-289.33	-199.89	-124.44	-68.86
DIFX2	-12.92	-5.71	-5.65	-1.29	-46.85	-21.95	-8.43	-4.37	-281.22	-184.10	-106.67	-55.71
SYS1	-0.15	-2.22	-6.09	-9.19	-15.90	-19.55	-20.10	-15.63	-50.84	-48.19	-46.00	-48.00
SYS2	-1.04	-1.85	-6.40	-8.35	-13.94	-16.42	-17.16	-13.89	-45.48	-43.31	-41.70	-42.90
SYSX1	6.00	3.00	-0.28	0.67	-3.87	1.04	1.15	0.76	-56.04	-40.74	-28.56	-16.86
SYSX2	3.96	2.14	-1.32	1.02	0.77	3.29	4.38	2.87	-43.59	-25.70	-16.02	-5.77
	High Cross Section Dependence: $\phi_i \sim i.i.d.U[-1, 4]$											
FE	-63.49	-61.52	-60.98	-64.34	-119.29	-118.17	-116.32	-113.69	-374.01	-371.82	-372.45	-366.96
DIF1	-34.91	-29.04	-26.18	-29.00	-83.57	-77.34	-72.34	-64.51	-459.56	-434.21	-438.01	-414.08
DIF2	-30.25	-22.18	-20.73	-22.36	-76.68	-68.73	-62.60	-54.55	-466.73	-441.78	-447.79	-427.26
DIFX1	-9.91	-4.75	-4.25	-1.22	-41.29	-19.10	-8.88	-4.72	-453.74	-344.77	-244.99	-148.10
DIFX2	-8.17	-3.38	-3.57	-0.52	-33.56	-14.40	-4.84	-2.42	-436.91	-313.41	-198.66	-106.04
SYS1	-0.25	-3.56	-8.64	-13.43	-17.11	-26.10	-29.33	-25.59	-79.86	-95.60	-100.75	-114.51
SYS2	-1.63	-3.71	-8.97	-12.56	-15.61	-24.64	-27.43	-24.05	-77.14	-94.71	-100.97	-113.50
SYSX1	9.23	4.76	1.29	1.22	11.12	9.72	5.95	3.11	-5.01	2.03	0.03	1.38
SYSX2	4.93	2.10	-0.78	0.87	12.32	8.56	6.71	3.56	4.83	18.18	17.69	22.97

Notes: See notes to Table 1. FE is the fixed effects estimator, DIF1 and DIF2 are the Arellano and Bond (1991) one-step and two-step first differenced GMM (DIF) estimators, respectively. DIFX1 and DIFX2 are the one-step and two-step DIF estimators, respectively, which are based on the instruments consisting of subsets of $\mathbf{X}_i$ only. SYS1 and SYS2 are the Blundell and Bond (1998) one-step and two-step system GMM (SYS) estimators, respectively. SYSX1 and SYSX2 are the one-step and two-step SYS estimators, respectively, which are based on the instruments consisting of subsets of $\mathbf{X}_i$ only.

27

Table 3: Root Mean Square Errors ($\times 1000$) of Fixed Effects and GMM estimators for λ, in the Case with Predetermined Regressors with $T = 5$

Test,N	$\lambda = 0.2$				$\lambda = 0.5$				$\lambda = 0.8$			
	50	100	200	400	50	100	200	400	50	100	200	400
	No Cross Section Dependence: $\phi_i = \phi = 0$											
FE	15.45	14.24	14.27	13.75	37.79	35.41	35.12	34.69	159.49	153.85	152.75	151.86
DIF1	4.42	1.98	1.09	0.47	9.06	4.18	1.81	0.92	78.33	29.52	12.20	5.26
DIF2	5.03	2.34	1.18	0.50	11.00	4.82	1.97	0.97	98.83	37.00	14.03	5.66
DIFX1	7.01	3.48	1.91	0.89	22.98	11.60	5.92	3.16	315.62	204.34	106.17	52.48
DIFX2	8.05	3.91	1.99	0.91	27.12	12.61	6.30	3.23	362.43	230.90	115.28	54.88
SYS1	3.91	1.84	0.94	0.48	4.82	2.48	1.23	0.65	8.64	5.38	3.16	1.66
SYS2	4.11	1.84	0.88	0.39	5.14	2.50	1.18	0.58	9.23	5.41	2.68	1.26
SYSX1	7.09	3.44	1.83	0.95	15.09	8.64	4.80	2.58	41.18	32.94	22.05	13.32
SYSX2	7.79	3.66	1.80	0.84	16.82	9.33	4.83	2.49	47.67	37.96	23.18	12.41
	Low Cross Section Dependence: $\phi_i \sim i.i.d.U[-0.3, 0.7]$											
FE	13.70	12.54	12.43	12.12	32.16	30.03	29.56	29.09	110.86	105.91	104.80	103.60
DIF1	4.26	1.99	1.13	0.57	8.33	4.04	1.94	1.16	44.25	19.17	9.75	5.98
DIF2	4.72	2.29	1.17	0.56	9.97	4.47	2.02	1.11	53.34	21.98	10.06	5.60
DIFX1	6.38	3.17	1.70	0.81	20.57	9.96	5.19	2.73	219.80	133.09	71.42	33.30
DIFX2	7.27	3.53	1.74	0.83	24.14	10.70	5.50	2.74	248.84	143.90	74.07	33.74
SYS1	3.78	1.95	1.04	0.63	4.58	2.55	1.44	0.95	6.91	5.07	3.37	2.46
SYS2	3.95	1.88	0.95	0.50	4.88	2.41	1.25	0.71	7.32	4.65	2.64	1.71
SYSX1	6.57	3.19	1.65	0.86	13.60	7.63	4.27	2.29	33.21	29.45	20.42	12.20
SYSX2	7.05	3.31	1.59	0.77	14.81	8.02	4.26	2.19	36.72	29.89	20.00	10.27
	Moderate Cross Section Dependence: $\phi_i \sim i.i.d.U[-1, 2]$											
FE	11.92	10.98	10.77	10.63	27.41	25.10	24.52	23.50	93.31	89.68	88.77	87.15
DIF1	7.20	5.57	4.98	4.40	18.59	14.22	12.47	10.62	98.86	83.78	81.96	72.86
DIF2	6.34	4.24	3.44	2.91	16.57	11.09	8.86	7.00	102.60	81.07	77.90	66.73
DIFX1	5.14	2.54	1.32	0.65	17.80	8.07	4.06	2.12	176.18	98.94	52.60	25.59
DIFX2	5.11	2.40	1.15	0.56	17.92	7.35	3.74	1.78	186.10	97.06	47.46	21.20
SYS1	7.18	6.56	5.99	5.53	12.54	11.26	10.69	10.06	22.59	22.96	24.48	24.15
SYS2	6.26	4.70	3.90	3.38	10.61	7.93	6.74	5.93	20.06	19.19	19.07	18.60
SYSX1	5.23	2.60	1.31	0.67	11.61	6.48	3.54	1.91	32.33	27.07	18.39	12.57
SYSX2	4.87	2.25	1.06	0.53	10.82	5.52	2.95	1.53	32.23	21.75	14.97	8.52
	High Cross Section Dependence: $\phi_i \sim i.i.d.U[-1, 4]$											
FE	9.38	8.73	8.62	8.33	24.88	22.38	21.96	20.60	185.46	183.37	184.96	180.72
DIF1	7.95	6.81	6.48	6.03	28.58	24.28	23.09	19.77	364.79	357.52	380.38	358.02
DIF2	6.24	4.69	4.26	3.92	24.40	19.06	17.33	14.08	378.32	369.79	395.92	376.19
DIFX1	3.80	1.81	0.95	0.47	15.21	6.73	3.29	1.74	362.81	233.76	141.53	69.36
DIFX2	2.86	1.30	0.62	0.30	11.89	4.63	2.31	1.05	370.67	223.07	119.02	48.29
SYS1	9.28	9.28	9.10	8.47	19.20	18.75	18.68	17.70	59.51	67.97	79.84	81.53
SYS2	7.63	6.62	6.10	5.55	16.35	14.20	13.37	12.27	57.14	64.93	74.88	76.46
SYSX1	4.18	2.02	1.02	0.52	10.96	6.02	3.22	1.68	43.40	34.61	28.51	22.76
SYSX2	2.93	1.27	0.60	0.29	8.95	4.09	2.08	1.00	46.88	34.68	26.44	17.26

See Notes to Table 2.

28

Table 4: Homogeneity Error Cross Section Dependence Tests and Estimates of Employment Equation, 140 Firms with 9-Year Observations

A: Two-Step DIF Estimator, 1976-84				
Estimation Results				
	Based on $\mathbf{Z}_i$		**Based on $\mathbf{Z}_{Xi}$**	
	Coef.	Std.Err.	Coef.	Std.Err.
$y_{i,t-1}$	0.679	(0.084)	0.401	(0.124)
w_{it}	-0.720	(0.117)	-0.551	(0.130)
w_{it-1}	0.463	(0.111)	0.347	(0.112)
κ_{it}	0.454	(0.101)	0.447	(0.110)
κ_{it-1}	-0.191	(0.086)	-0.079	(0.105)
cons	0.005	(0.017)	0.003	(0.014)
Test Results				
	statistics	p-values	statistics	p-values
Sargan	88.8 (79)	[0.211]	62.0 (51)	[0.140]
m_1	-4.46	[0.000]	-2.19	[0.028]
m_2	-0.17	[0.866]	-0.47	[0.641]
D_{DIF2}	26.84 (28)	[0.527]	-	-
B: Two-Step SYS Estimator, 1976-84				
Estimation Results				
	Based on $\mathbf{Z}_i^+$		**Based on $\mathbf{Z}_{Xi}^+$**	
	Coef.	Std.Err.	Coef.	Std.Err.
$y_{i,t-1}$	0.873	(0.044)	0.825	(0.071)
w_{it}	-0.780	(0.116)	-0.717	(0.105)
w_{it-1}	0.527	(0.168)	0.560	(0.149)
κ_{it}	0.470	(0.071)	0.395	(0.088)
κ_{it-1}	-0.358	(0.072)	-0.253	(0.092)
cons	0.948	(0.390)	0.720	(0.402)
Test Results				
	statistics	p-values	statistics	p-values
Sargan	111.6 (100)	[0.201]	77.3 (65)	[0.142]
m_1	-5.81	[0.000]	-5.19	[0.000]
m_2	-0.15	[0.883]	-0.12	[0.906]
D_{SYS2}	34.30 (35)	[0.502]	-	-

Notes: The estimated model is $y_{it} = \alpha_i + \delta t + \beta_1 y_{it-1} + \delta_0 w_{it} + \delta_1 w_{it-1} + \varphi_0 \kappa_{it} + \varphi_1 \kappa_{it-1} + u_{it}$, where y_{it} is the log of the number of employees of company i, w_{it} is the log of real product wage and κ_{it} is the log of gross capital stock. Year dummies are included in all specifications. The standard errors reported are those of the robust one-step GMM estimator. The first row of the test results reports Sargan's statistic for overidentifying restrictions. m_1 and m_2 are the first-order and second-order serial correlation tests in the first-differenced residuals. D_{DIF2} denotes Sargan's Difference test for heterogeneous error cross section dependence based on the two-step Arellano and Bond (1991) DIF GMM estimator. D_{SYS2} denotes Sargan's Difference test based on the two-step Blundell and Bond (1998) SYS GMM estimator. Sargan test and Sargan's difference test are distributed as χ^2 under the null with degrees of freedom reported in parentheses. Instruments used in each equation are for DIF: $y_{i,t-2}, y_{i,t-3}, ..., y_{i,0}; w_{i,t-2}, w_{i,t-3}, ..., w_{i,0}; \kappa_{i,t-2}, \kappa_{i,t-3}, ..., \kappa_{i,0}$, and for SYS: $\Delta y_{i,t-1}, \Delta w_{i,t-1}, \Delta \kappa_{i,t-1}$. $\mathbf{Z}_i, \mathbf{Z}_{Xi}, \mathbf{Z}_i^+, \mathbf{Z}_{Xi}^+$ are defined by (22), (18), (37), (31) but without cross-sectional demeaning, respectively.

29