Journal of Econometrics 148 (2009) 149–161

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

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

Vasilis Sarafidis[a], Takashi Yamagata[b,*], Donald Robertson[c]

[a] Discipline of Econometrics and Business Statistics, University of Sydney, Australia

[b] Department of Economics and Related Studies, University of York, United Kingdom

[c] Faculty of Economics, University of Cambridge, United Kingdom

ARTICLE INFO

Article history:
Received 8 May 2006
Received in revised form
23 September 2008
Accepted 10 October 2008
Available online 28 October 2008

JEL classifications:
C12
C13
C15
C33

Keywords:
Cross section dependence
Generalised method of moments
Dynamic panel data
Overidentifying restrictions test

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, R., Bond, S., 1998. Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics 87, 115–143] 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. 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 et al. (2007), Lee (2004, 2007), and others). There are several contributions in the literature that allow for time-varying individual effects (Holtz-Eakin et al., 1988; Ahn et al., 2001; Han et al., 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]

☆ We would like to thank Tom Flavin, Hashem Pesaran and Neville Weber for helpful discussions. We have also benefited from the constructive and most helpful comments of Cheng Hsiao and two anonymous referees. Sarafidis gratefully acknowledges full financial support from the ESRC during his Ph.D. studies at Cambridge University (PTA-030-2002-00328). Yamagata gratefully acknowledges the financial support from the ESRC (Grant No. RES-000-23-0135).

* Corresponding author. Tel.: +44 0 1904 433758.
E-mail address: ty509@york.ac.uk (T. Yamagata).

[1] Other examples are provided by Ahn et al. (2001).

0304-4076/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2008.10.006

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 (forthcoming) and in the current paper.

Several tests for cross section dependence have been proposed in the econometric 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 et al. (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.

2. Existing tests for cross section dependence

Consider a panel data model

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

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

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

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 and 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

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

For sufficiently large N and T, the CD test statistic tends to a standard normal variate under the null of cross section