Journal of Urban Economics 80 (2014) 76–86

Contents lists available at ScienceDirect

Journal of Urban Economics

www.elsevier.com/locate/jue

Spatial lag models with nested random effects: An instrumental variable procedure with an application to English house prices

Badi H. Baltagi [a,*], Bernard Fingleton [b], Alain Pirotte [c,d]

[a] Syracuse University, Department of Economics and Center for Policy Research, United States
[b] University of Cambridge, Department of Land Economy, United Kingdom
[c] CRED-TEPP (CNRS), University of Panthéon-Assas Paris II, France
[d] IFSTTAR-DEST, French Institute of Science and Technology for Transport, Development and Networks, France

ARTICLE INFO

Article history:
Received 21 February 2013
Revised 24 October 2013
Available online 6 November 2013

JEL classification:
C33

Keywords:
House prices
Panel data
Spatial lag
Nested random effects
Instrumental variables
Spatial dependence

ABSTRACT

This paper sets up a nested random effects spatial autoregressive panel data model to explain annual house price variation for 2000–2007 across 353 local authority districts in England. The estimation problem posed is how to allow for the endogeneity of the spatial lag variable producing the simultaneous spatial spillover of prices across districts together with the nested random effects in a panel data setting. To achieve this, the paper proposes new estimators based on the instrumental variable approaches of Kelejian and Prucha (1998) and Lee (2003) for the cross-sectional spatial autoregressive model. Monte Carlo results show that our estimators perform well relative to alternative approaches and produces estimates based on real data that are consistent with the theoretical house price model underpinning the reduced form.

1. Introduction

Recently a huge literature has emerged which studies the causes of spatial variation in house prices, often involving the inclusion of the real estate sector within contemporary spatial economics theory (Helpman, 1998; Hanson, 2001; Hanson, 2005; Brakman et al., 2004; Glaeser, 2008; Fingleton, 2009). As Behrens and Robert-Nicoud (2009) observe, writing in the context of NEG, or the New Economic Geography (Fujita et al., 1999), housing represents the single most important expenditure item and asset for households throughout the world.

In this paper, we focus on the estimation of a theoretical house price model in which spatio-temporal variations in house prices are driven by supply and demand conditions, with spatial effects coming from two distinct sources. One is the direct dependence of house prices in a given locality on house prices in nearby localities. The other is *via* hierarchical error components typical of multilevel models. Direct dependence is the net effect of what we refer to as displaced demand and displaced supply. Displaced demand occurs where, *ceteris paribus*, high prices nearby cause demand to increase, because the negative relationship between prices and demand leads to purchasers switching away from high price nearby locations. Displaced supply occurs where high prices nearby cause supply to fall as a result of suppliers of housing switching to where higher prices give better investment returns. The supply and demand equations lead to a reduced form in which prices depend directly on prices nearby. We refer to this as a spillover effect.

The second source of spatial heterogeneity comes from the presence of hierarchical error components which represent the impact of local (district) effects embedded within wider (county) effects. Intuitively, local effects can be thought of as postcode effects, where particular postcodes are associated with more or less prestige. Thus a given postcode which is known to be expensive (inexpensive) is assumed to cause an increase (or decrease) in house prices which is uniform across the postcode district. Likewise we envisage an independent county effect (a number of districts together are nested within a county). County A, which is a prestigious address, will tend to be associated with higher prices than less prestigious county B. The difference between these two sources of spatial heterogeneity in house prices is that the hierarchical district and county effects are constant within counties and districts, and terminate abruptly at county or district boundaries. We can think of this in terms of within distances being equal,

* Corresponding author. Fax: +1 315 443 1081.
E-mail address: bbaltagi@maxwell.syr.edu (B.H. Baltagi).

0094-1190/$ - see front matter
http://dx.doi.org/10.1016/j.jue.2013.10.006

between distances being zero. In contrast, the spillover effects have an autoregressive specification, so that they extend across space with diminishing effect as distance increases. In this case within distances are zero, and between distances unequal. Thus spatial heterogeneity and autocorrelation is partly accounted for by the discrete non-overlapping effects of the components of the hierarchical level, and partly by spatial dependence operating, *via* the spatial autoregressive process, simultaneously across all areas. To illustrate this, consider two identical houses on either side of a street bisected by a district boundary. In terms of spillover, there will be a tendency for the closely proximate and indentical houses to converge on a single equilibrium price, but a significant postcode effect could induce differential pricing.

Our solution to the problem of estimating the panel data model with spatial spillover effects and random hierachical effects, which we refer to as a nested random effects spatial autoregressive panel data model, is to propose a novel methodological approach. While the focus of the paper is squarely on the theory and application in relation to house prices, this has been introduced elsewhere in the literature. In contrast, our proposed estimator is new. Moreover despite the housing context of the present paper, the estimator could be applied to similar situations in which spatial heterogeneity and autocorrelation occurs as a consequence of the dual effects of spatial spillovers and hierarchical error components. For example, wage variations across regions nested within countries are commonly highly spatially autocorrelated. Part of this could be attributed to spillover effects across neighboring regions, as commuting and workforce migration cause wages to move to some equilibrium level, with the force behind this tendency diminishing as distance increases. The hierarchical component in wage determination typically comes in the form of regional effects (for example wages in capital city regions tend to be higher), and national level effects, perhaps relating to national (as opposed to region-specific) taxation and benefits policies. More generally, the analysis of micro level data does typically reveal cross-sectional correlation, and spatial panel data models are becoming increasingly attractive in empirical economic research. Although the dependence structure usually relates to location and geographical distance, it has a wider application in a more general economic or social network space. Recent developments in econometrics allow us to take into account cross-sectional correlation in a panel data context, as demonstrated for example by Elhorst (2003, 2010a), Anselin et al. (2008), Anselin (2010) and Baltagi (2011). The contribution here combines this with hierarchical panel modeling, building on the sparse earlier literature which suggests combining nested models and spatial autoregressive processes in a cross-sectional context (see Corrado and Fingleton, 2011, 2012).

Early work on hierarchical panels was carried out by Fuller and Battese (1973), Montmarquette and Mahseredjian (1989), Antweiler (2001) and Baltagi et al. (2001, 2002), to mention just a few. For example, Montmarquette and Mahseredjian (1989) study whether schooling matters in educational achievements in Montreal's Francophone public elementary schools. Here data on students are naturally grouped by school. Antweiler (2001) derives the maximum likelihood estimator for an unbalanced nested three-way error component model. This is applied to the problem of explaining the determinants of pollution concentration (measured by the log of atmospheric sulfuric dioxide) at 293 observation stations located in 44 countries over the time period 1971–1996. This data is highly unbalanced in that out of a total of 2621 observations, about a third of these are from stations in one country, the United States. Comparing the results of maximum likelihood for a nested versus a simple (non-nested) unbalanced error component model, Antweiler (2001) finds that the scale elasticity coefficient estimate for the nested model is less than half that for the non-nested model. Baltagi et al. (2001) propose natural extensions of the ANalysis Of Variance (ANOVA), Maximum Likelihood Estimator (MLE) and Minimum Norm Quadratic Unbiased Estimator (MINQUE) and compare their performance using Monte Carlo experiments. The ANOVA methods seem to perform well and are recommended. These estimation methods are also used to investigate the productivity of public capital in private production. In this case, American states are naturally grouped into regions of the United States. In a companion paper, Baltagi et al. (2002) derive Lagrange Multiplier tests that test for random effects in this unbalanced nested error component panel model.

With spatial interdependence, which quite naturally is a feature of house price data, estimation is complicated by the presence of a spatially lagged dependent variable, which is typically correlated with the disturbance terms. From the purely spatial perspective, a common way to proceed is to apply Maximum Likelihood methods, see Anselin (1988) in the spatial context and Antweiler (2001) in the unbalanced nested three-way error component context. However, Maximum Likelihood procedures are often challenging when the sample size is large. Moreover, they call for explicit distributional assumptions. In fact, this is why Kelejian and Prucha (1998) and Lee (2003) proposed an Instrumental Variable (IV) procedure for the cross-sectional spatial autoregressive model, which is computationally simple and less restrictive regarding the distribution of the disturbances. In this paper, we extend this cross-sectional IV approach to the nested random effects spatial autoregressive panel data model.[1]

The plan of the paper is as follows: Section 2 presents the theoretical house price framework. Section 3 describes the spatial autoregressive model with nested random effects, and Section 4 outlines an IV procedure to estimate this model (details of which are given in Appendix A). Section 5 describes the house price data, while Section 6 discusses the empirical results. Section 7 gives our concluding remarks. Appendix B sets out our Monte Carlo design used to study the small sample performance of the proposed estimator, Appendix C gives the results of these Monte Carlo experiments, and Appendix D gives the log-likelihood functions of the nested random effects spatial model and the random effects spatial model.

2. Theoretical framework

In spatial econometrics, one sometimes sees all-encompassing specifications involving various autoregressive spatial lags. For example, Elhorst (2001) presented a dynamic first order autoregressive distributed lag model, thus

$$y_{it} = \gamma y_{it-1} + \rho_1 \sum_{j=1}^{N} w_{ij} y_{jt} + \rho_2 \sum_{j=1}^{N} w_{ij} y_{jt-1} + x_{it}\beta_1 + \sum_{j=1}^{N} w_{ij} x_{jt}\beta_2 + x_{it-1}\beta_3 + \sum_{j=1}^{N} w_{ij} x_{jt-1}\beta_4 + \varepsilon_{it}, \quad (1)$$

where y_{it} denotes the dependent variable at location i in the tth period, x_{it}, the exogenous explanatory variable at location i in the tth period, w_{ij}, the i,j element of a spatial matrix $\mathbf{W}$ of dimension $(N \times N)$, $\gamma,\ \rho_1,\ \rho_2,\ \beta_1,\ \beta_2,\ \beta_3,\ \beta_4$, the scalar parameters to be estimated and $\varepsilon \sim$i.i.d.$(0, \sigma_\varepsilon^2)$.

However, while quite general, specification (1) presents problems of identification (Anselin et al., 2008; Manski, 1993; Elhorst, 2010b). Based on a theoretical framework that assumes the existence of equilibrium housing supply and demand

[1] In a similar spirit, Baltagi and Liu (2011) derived an IV estimator in the context of spatial autoregressive random effects panel data model. However, this estimator does not deal with the unbalancedness or the nested structure of the panel data.