STRATHCLYDE

DISCUSSION PAPERS IN ECONOMICS

MULTILEVEL MODELLING WITH SPATIAL EFFECTS

BY

LUISA CORRADO AND BERNARD FINGLETON

NO. 11-05

DEPARTMENT OF ECONOMICS
UNIVERSITY OF STRATHCLYDE
GLASGOW

Multilevel Modelling with Spatial Effects

L. Corrado* & B. Fingleton**

This Version: February 2011

Abstract

In multilevel modelling, interest in modeling the nested structure of hierarchical data has been accompanied by increasing attention to different forms of spatial interactions across different levels of the hierarchy. Neglecting such interactions is likely to create problems of inference, which typically assumes independence. In this paper we review approaches to multilevel modelling with spatial effects, and attempt to connect the two literatures, discussing the advantages and limitations of various approaches.

Key-Words: Multilevel Modelling, Spatial Effects, Fixed Effects, Random Effects, IGLS, FGS2SLS.
JEL codes : C21,C31,R0

* Marie Curie Fellow, Faculty of Economics, University of Cambridge, UK and University of Rome II, Italy.
E-mail:lc242@econ.cam.ac.uk. Luisa Corrado acknowledges the support of the Marie-Curie IE Fellowship 039326.

**Professor of Economics, Strathclyde University.
E-mail: bf100@cam.ac.uk.

1 Introduction

Multilevel models are becoming increasingly popular across the range of the social sciences, as researchers come to appreciate that observed outcomes depend on variables organised in a nested hierarchy. [1] We see many applications of multilevel modelling in educational research where there exist a number of well defined groups organized within a hierarchical structure, such as the teacher-pupil relationship, leading to the analysis of effects on individual pupil behaviour coming from different hierarchical levels. For example, irrespective of personal attributes and other factors, being in the same classroom will tend to cause pupil performance to be more similar than it otherwise would be. This suggests that once the grouping has been established, even if its establishment is random, the group itself will tend to become differentiated. This implies that the group and its members can both influence and be influenced by the composition of the group (Goldstein (1998)). In geographical studies, we can often envisage a hierarchy of effects from cities, regions containing cities, and countries containing regions. Failure to recognise these effects emanating from different hierarchical levels can lead to incorrect inference. However while the standard approaches to multilevel analysis are well established, there is none the less much scope for the refinement and development of this extremely useful methodology. In this paper we focus on interdependencies beyond those intra-class correlations that exist because individuals are taught in the same class room, or coexist within the same region. The innovation in this paper is that we take this wider perspective on interdependencies, drawing particularly on the burgeoning literature of spatial econometrics (Anselin (1988a)). This accommodates spatial dependence within cross-sectional regression models, and also within panel data analysis. In our review, we explore the connections between conventional multilevel models and the kinds of models proposed by spatial econometricians. Naturally, as economists, most of our examples and motivation are drawn from economics, and from economic geography, although we believe that the approaches we consider have much wider potential application.

In economics, considerable recent attention has been given to spatial economics and international trade, particularly with the advent of the 'new economic geography' (Fujita and Krugman (1999)). The increasingly spatial perspective means that very often we are faced with cross-sectional data indexed by location rather than time, or panel data in which each time layer comprises a data set of location-specific observations. In economic geography, with a hierarchy of local, regional and national effects typically influencing outcomes, the obvious starting point is multilevel modelling, in which individual level cross-sectional (spatial) data within the same local administrative area, for example, are subject to an effect because of their common location. Perhaps local property taxes are different across local administrative units, and properties, which are the units of observation, have prices partly reflecting these local tax differences. Additional spatial effects may arise at different levels of a nested hierarchy; for instance we may wish also to control for the effects of being located within the same region, perhaps because policy instruments having an effect on property prices are applied at the regional level and are different from the effects of local tax differentiation.

The contribution of the spatial econometrics perspective is that it introduces the notion that the analysis of spatial effects simply via the effects of multilevel group membership (local, regional, national) may be inadequate as a means of totally capturing the true spatial dependencies in the data, and this can produce misleading outcomes. For example, real estate prices may be affected

[1] Over the past decade there has been a development of methods which have enabled researchers to model hierarchical data. Examples of these methods include multilevel models (see, for example, Goldstein (1998)), random coefficient models (Longford (1993)) and hierarchical multilevel models proposed by Goldstein (1986) based on iterative generalized least squares (IGLS).

by individual level real estate attributes (size of plot etc.) and by hierarchical location effects (local administrative area effects, regional effects etc.), but there could also be an interaction between neighbouring properties that depends on, say, distance, so that the intra-area correlation between individual properties is unequal, due to the different distances between properties within areas, and moreover, the correlation does not terminate abruptly at area boundaries, but spills over. This spilling over of externalities across space will lead to unmodelled effects (the error), or indeed prices directly, being spatially autocorrelated. Modelling such externalities comes within the realm of spatial econometrics. By not properly modelling these spatial interdependencies, with positive dependence, the outcome could be that, since the actual information content is smaller than that implied by the actual sample size, using the data as though they were independent leads to standard errors that are biased downwards. Therefore alongside the nested structure of the hierarchical data increasing attention has been paid to different forms of interactions and externalities in the hierarchical system (Durlauf (2003), Manski (2000), Brock (2001), Akerlof (1997)). As another example, Bénabou (1993), Durlauf (1996), Fernandez and Rogerson (1997) consider the effects of residential neighbourhood on education. Typically we will find that a child's education is determined, at least in part, by factors such as school quality, but the characteristics of other pupils with whom the pupil interacts, captured possibly by a social network structure, will also be of relevance. The importance of such externalities has led researchers to define different concepts of membership and neighborhood effects relying on notions of distance in social space (Akerlof (1997), Anselin (2002)). Widespread use has been extended to health research where for example multilevel modelling with spatial dependence has been applied to examine the geographical distribution of diseases, since diseases often spread due to contagion via contact networks (see, Langford, Leyland, Rasbash, and Goldstein (1999)).

Recognition of the different form of interactions between variables which affect each individual unit of the system and the groups they belong to has important empirical implications. In fact, regardless of spatial autocorrelation, the assumption of independence is usually incorrect when data are drawn from a population with a grouped structure since this adds a common element to otherwise independent errors, thereby inducing correlated within group errors. Moulton (1986) finds that it is usually necessary to account for the grouping either in the error term or in the specification of the regressors. Apart from within- group errors, it is also possible that errors between groups will be correlated. For example, if the groups are geographical regions then regions that are neighbours might display greater similarity than regions that are distant. Again, Moulton (1990) shows that even with a small level of correlation, the use of Ordinary Least Squares (OLS), will lead to standard errors with substantial downward bias and to spurious findings of statistical significance.

One way of incorporating the group effect in a multilevel framework is to evaluate the impact of higher level variables on the individual which measure one or more aspects of the composition of the group to which the individual belongs. Bryk and Raudenbush (1992) consider different ways of doing this, such as using a simple mean covariate over the higher level units as an explanatory variable. The mean covariate characterizes group effects which are measurable and in this respect differs from the use of dummy variables which capture the net effect of several variables. Note that it is possible that having controlled for these measurable compositional effects there are still unobservable spatial effects.

Such correlated unobservables can be modeled either as fixed or random effects. If we have data grouped by geographic area with all the areas represented in the sample then a fixed effects specification is appropriate. When only some of the areas are represented in the sample or there is pattern of dependence involving unknown spatial effects, we might opt for random effects in a hierarchical model operating through the error term. This is achieved by way of an unrestricted non-diagonal covariance matrix. As with unconditional ANOVA this will provide the decomposition of the variance for the random effect into an individual component and a group component. Under a spatial dependence process acting at the level of random group effects, the random components are typically affected by those of neighbouring groups. This assumption is usually a relaxation of the main hypothesis in hierarchical modelling, i.e. independence between groups. As we have seen, especially when the groups are geographical areas, this might often be unrealistic.

The paper is organized as follows. The following section presents a diagrammatic representation of the multilevel modelling approach. Section two describes how to model spatial effects through the error term; it also describes different ways of defining spatial and multidimensional weighting matrices. Sections three illustrates how to estimate linear random intercept models. Sections four and five illustrate how to accommodate correlation between predictors and (group) spatial effects using fixed or random effects in multilevel models. Section six links multilevel to spatial models, illustrates how to identify the relevant parameters for the group random effects and the endogenous spatial lag and connects multilevel modelling to recent developments in panel data analysis, highlighting where progress might be made in estimation methodology. Section seven concludes.

1.1 A Representation of a Multilevel Structure

Figure 1 represents a diagrammatic representation of a multilevel hierarchy. Here with R denoting the top level, N the second level, and I the individual level. There is a variable number of individuals per second level group, and varying numbers of second level groups in each category at the top level of the hierarchy. In the context of spatial data we might consider a geographical grouping of individuals with the highest level being R regions (r) each of which nests a total of G smaller geographical sub-regions (g). These sub-regions may be either specific areas of residence or some other relevant geographical units. Located within each sub-region there are individuals (I) with a variable number of individuals per sub-region.

We associate with each individual a response Y_i which is dependent upon a set of covariates X_i. However, in assessing whether we might assign any causal relationships between one or more elements of X_i and the individual response Y_i, it is necessary to consider the hierarchical structure of the data, and in particular within- and between-group effects.

There are a number of advantages in taking a multilevel approach. First, in standard unilevel OLS estimation the presence of nested groups of observations may be dealt with the use of dummy variables. However, the large number of levels result in a dramatic reduction in degrees of freedom. Second, this approach helps to analyze the effect of heterogenous groups in the small sample situation. In fact with unbalanced data, while OLS estimates of the coefficients give equal weights to each cluster, the variance-components model acknowledges the fact that estimates for the fixed coefficients can change according to the cluster size. It is therefore possible to adjust both the estimates and the inference according to the precision associated with each group, which is determined by the number of surveyed individuals (this is technically referred to as shrinkage). In most applications, shrinkage is desirable as it only affects clusters that provide little information and effectively downplays their influence, borrowing strength from other larger clusters. The hierarchical multilevel method, which is sample size dependent, seems to have a distinct advantage over other methods in eliminating bias.

Compared to other approaches such as Clustered-Standard-Error OLS, Multilevel Modelling (MLM) has some advantages: first, while CSE techniques treat the random variation as a simple nuisance the objective of MLM is to estimate and decompose the total random variation in an individual component and a group component. Second, while CSE adjusts only standard errors for non-independence while having no impact on point estimates, MLM provides us with estimates of the variance components at each level and these affect point estimates directly. In turn, variances and covariances constitute valuable information on the contribution of non-observable factors at each level to the variation of the dependent variable (Aslam and Corrado (2007)).

2 Modelling spatial effects through the error term

Below we introduce the basic multilevel model focussing on the two level case. We write this model as:

$$Y_{ij} = \beta_0 + \mathbf{X}_{ij}\boldsymbol{\beta}_1 + \mathbf{Z}_j\boldsymbol{\gamma} + \varepsilon_{ij}, \tag{1}$$

where Y_{ij} denotes the response, $\mathbf{X}_{ij}$ is a $1 \times k$ vector of covariates, $\mathbf{Z}_j$ is a $(1 \times q\ )$ vector of group level 2 explanatory variables (invariant within groups) and ε_{ij} is the random disturbance; i, $i = 1, \ldots, N$ and j, $j = 1, \ldots, G$, denote, respectively, level 1 and level 2 units. We write the error term as:

$$\varepsilon_{ij} = e_{ij} + u_j, \tag{2}$$

where ε_{ij} denotes an additive error term composed of a random error term e_{ij} for the i^{th} unit belonging to level j and a random effect u_j for each level 2 unit. We make the following assumptions:

$$\begin{array}{ll} e_{ij}|X_{ij} \sim N(0, \sigma_e) & \mathrm{Cov}(e_{ij}, e_{i'j}) = 0,\ \forall i \neq i' \\ u_j|X_{ij} \sim N(0, \sigma_u) & \mathrm{Cov}(u_j, e_{ij}) = 0 \end{array} \tag{3}$$

We let σ_e^2 (σ_u^2) denote the variance of e_{ij} (u_j) such that for $cov(e_{ij}, u_j) = 0$, then $\sigma_\varepsilon^2 = \sigma_e^2 + \sigma_u^2$ represents the sum, respectively of the within- and between-group variances. Based upon the above, we may write the (equicorrelated) intra-class correlation as:

$$\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}. \tag{4}$$

This correlation measures the proportion of the variance explained at the group level. In single-level models $\sigma_u^2 = 0$ and $\sigma_\varepsilon^2 = \sigma_e^2$ becomes the standard single level residual variance.

Following Anselin, Le Gallo, and Jayet (2007), who write from a spatial panel data perspective, there are four ways we might wish to model spatial effects operating through the error term, namely i) direct representation, which originates from the geostatistical literature (Cressie (2003)); as noted by Anselin (2003), this requires exact specification of a smooth decay with distance and a parameter space commensurate with a positive definite error variance-covariance matrix. Alternatively, as in Conley (1999), a looser definition of the distance decay may be implemented, leading to non-parametric estimation; ii) spatial error processes typified by much work in spatial econometrics (Anselin (1988a)), based on a so-called **W** matrix defining indirectly the spatial structure of the non-zero elements of the error variance-covariance matrix. The **W** matrix comprises non-negative values representing the a priori assumption about interaction strength between location pairs defined by specific rows and columns of **W**, normally with zeros on the main diagonal. Typically but not necessarily **W** is normalized to sum to 1 across rows; iii) common factor models originating from the time series literature (Hsiao and Pesaran (2004), Pesaran (2007), and Kapetanios and Pesaran (2005)) and iv) spatial error components models (Kelejian and Robinson (1995), Anselin and Moreno (2003)) combining local (e_{ij}) and spillover (u_j) error components. To accommodate these effects we rewrite (2) as

$$\varepsilon_{ij} = e_{ij} + \sum_{j \neq h} u_j W_{jh}, \tag{5}$$

where $\mathbf{W} = \{W_{jh}\}$ allows us to specify the way neighbouring areas affect u_j. The matrix $\mathbf{W}$ is a matrix of distances between the G entities as discussed below. The intra-class correlation is now given by:

$$\rho = \frac{\sigma_u^2}{\sum_{j \neq h} \sigma_u^2 W_{jh}^2 + \sigma_e^2}. \tag{6}$$

One simple way to allow for these effects is to set $W_{jh} = 1(j \wedge h)$ where $1(.)$ is the indicator function and $j \wedge h$ denotes the contiguity of area j with area h. Trivially if $W = I$,where I is the identity matrix, then we have the standard random effects model ignoring any between group effect.

If u_j are treated as fixed parameters then we need to assume that $cov(e_{ij}, u_j) = \sigma_{eu} = 0$, that is transient individual-level random effects are uncorrelated with, say, a level 2 variable such as the area of residence. If u_j and e_{ij} are not independent the Generalised Least Squares (GLS) estimator would be biased and inconsistent. If u_j are permanent random effects we also assume independence between these and the covariates such that $cov(X_{ij}, u_j) = \sigma_{ux} = 0$ (Blundell (1997)). Relaxing the constraints $\sigma_{ux} \neq 0$ and $\sigma_{eu} \neq 0$ is discussed in sections five and six.

2.1 Spatial versus Multidimensional Weighting Matrices

In the following paragraphs we illustrate some of the traditional distance-based unidimensional measures adopted in spatial econometrics (Anselin (1988b)) and introduce other multidimensional measures based on various notions of social or economic distance. Typically, isotropy is assumed, so that only distance between j and h is relevant, not the direction j to h. These may provide the basis for direct or indirect estimation of the error variance-covariance matrix, including the spillover in error components models.

Spatial externalities can sometimes reflect not only pure spatial interaction but other important substantive multiple phenomena at the economic, political, cultural and institutional level operating at the group (or area) level (Tienda (1991)). For example, in community psychology (O'Campo (2003)) often the definition of neighbourhood is based on respondents' perception of their own neighborhood as well as on economic and census data (Aronson and Brodsky (1999); Ross, Reynolds, and Geis (2000); Shumow, Vandell, and Posner (1998)). Analysis of social exclusion (Muntaner, Lynch, and Oates (1999); Ross, Reynolds, and Geis (2000)) and social segregation Goldstein and Noden (2003) often considers geographical position as one among several factors that weaken the links between individuals and the rest of society.

Crude measures of between group spatial 'distance' include simple notions of proximity and contiguity, concepts which have motivated the work of Cliff and Ord (1973) and Cliff and Ord (1981) and specifically the measure of spatial autocorrelation. Cliff and Ord (1973) combines distance and length of the common border thus :

$$W_{jh} = (d_{jh})^a \left(\chi_{jh}\right)^b \tag{7}$$

where d_{jh} denotes the distance between locations j and h and χ_{jh} is the proportion of the boundary of j shared with h whereas a and b are parameters.

There are numerous alternatives, for example we might assume that

$$W_{jh} = \exp(-\eta d_{jh}) \tag{8}$$

with η controlling the strength of a distance-decay effect.

Dacey (1968) suggests that

$$W_{jh} = b_{jh}\alpha_j \chi_{jh} \tag{9}$$

in which b_{jh} is a binary contiguity factor, α_j is the share of unit j in the total area of all the spatial units in the system, and χ_{jh} is the same boundary measure as used in (7).

The above measures apply mostly to physical features of geographical units. However, they are less useful when spatial units consist of points and when the spatial interaction is determined by purely economic variables which may have little to do with spatial configuration of boundaries. Economic distance has been a feature of work by Conley (1999), Pinkse, Slade, and Brett (2002), Conley and Topa (2002), Conley and Ligon (2002) and Slade (2005). For example Conley and Ligon (2002) estimate the costs of moving factors of production. Physical capital transport costs are related to inter-country package delivery rates, and the cost of transporting embodied human capital is based on airfares between capital cities (the correlations with great circle distances are not perfect). In their analysis, for practical reasons, they confine their analysis to single distance metrics, but they prefer multiple distance measures. Taking the wider perspective of the industrial organization literature, distances may be in terms of trade openness space, regulatory space, commercial space, industrial structure space or product characteristics space. These developments in the conceptualization of economic distance have been surveyed in Greenhut, Norman, and Hung (1987). More general distance measures include multidimensional indicator functions. For example, Bodson (1975) use a general accessibility weight (calibrated between 0 and 1) which combines in a logistic function several channels of communication between regions such as railways, motorways etc.:

$$W_{jh} = \sum_{j=1}^{J} p_j \left(a/1 + b\exp(-c_j d_{jh})\right) \tag{10}$$

where p_j indicates the relative importance of the means of communication j. The sum is over the J means of communication with d_{jh} equal to the distance from j to h; a, b and c_j are parameters which need to be estimated.

Alternatively, if we have categorical data based on geographical and other socio-demographic indicators and we want to define general multidimensional binary measures of spatial and social distance across the different units we may use a more general weighting scheme:

$$W_{jh}^{h} = \sum_{j=1}^{G}\sum_{h\neq 1}^{G} W_{jh}^{h_1} \times W_{jh}^{h_2} \times ....W_{jh}^{h_S} \tag{11}$$

where $W^h = \{W_{jh}^{h_s}\}$ denotes the artificially constructed binary matrix based on indicators h_s where $s = 1, 2, ..., S$. Elements in $W_{jh}^{h_s}$ equal 1 if units j and h belong to the same category. Hence W^h is a multidimensional binary measure based on S different spatial and social categorical classifications of the individual units.

3 Linear Multilevel Model

In the random coefficient model the level of the individual response varies according to location. For example, individuals' income levels, controlling for individual level covariates such as educational attainment, vary if they reside in different areas. Part of the reason will be the effect of, say, fixed level two contextual factors (Z), and partly because level two specific random effects $\{u_j\}$. However these fixed and random effects, while accounting for heterogeneity across residential areas, are not spatially correlated, a topic we address subsequently. With this in mind, our specification is as in our general multilevel model, which we rewrite here in general matrix notation. Consequently equation (1) becomes:

$$\mathbf{Y} = \boldsymbol{\beta}_0 + \mathbf{X}\boldsymbol{\beta}_1 + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \tag{12}$$

where $\mathbf{Z} = \{Z_{ij}\}$ is a set of contextual factors at level-two, $\mathbf{Y} = \{Y_{ij}\}$, $\mathbf{X} = \{X_{ij}\}$ and $\boldsymbol{\varepsilon} = \{e_{ij}\} + \{u_j\}$. The dimension of $\mathbf{Y}$ and $\mathbf{X}$ and $\mathbf{Z}$ are $(N \times 1)$, $(N \times k)$ and $(N \times q)$ respectively. The vectors $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_1$ and $\boldsymbol{\gamma}$ denote the vectors of fixed effect coefficients.

We first rewrite (12) in compact form as:

$$\mathbf{Y} = \mathbf{J}\boldsymbol{\lambda} + \boldsymbol{\theta}_{\boldsymbol{\varepsilon}}\boldsymbol{\varepsilon} \tag{13}$$

where $\mathbf{J}$ is $(N \times (k + q + 1))$ and $\boldsymbol{\lambda}$ is $((k + q + 1) \times 1)$ and $\boldsymbol{\theta}_{\boldsymbol{\varepsilon}}$ is the design matrix of the random parameters that will be used in the estimation to derive the estimates for $\widehat{\sigma}_e^2$ and $\widehat{\sigma}_u^2$. The hierarchical two-stage method for estimating the fixed and random parameters (the variance and covariances of the random coefficients) originally proposed by Goldstein (1986),[2] is based upon an Iterative Least Squares (IGLS) method that results in consistent and asymptotically efficient estimates of $\boldsymbol{\lambda}$.

First we obtain starting values for $\boldsymbol{\lambda}$, $\tilde{\boldsymbol{\lambda}}$ by performing OLS in a standard single level system assuming the variance at higher level of the model to be zero. Conditioned upon $\tilde{\boldsymbol{\lambda}}$, we form the vector of residuals which we use to construct an initial estimate, $\mathbf{V}$, the covariance matrix for the response variable $\mathbf{Y}$. Then we iterate the following procedure first estimating the fixed parameters in a GLS regression as:

$$\widehat{\boldsymbol{\lambda}} = (\mathbf{J}^T\mathbf{V}^{-1}\mathbf{J})^{-1}\left(\mathbf{J}^T\mathbf{V}^{-1}\mathbf{Y}\right) \tag{14}$$

and again calculating residuals $\widehat{\mathbf{r}} = \mathbf{Y} - \mathbf{J}\widehat{\boldsymbol{\lambda}}$. We can rearrange this cross-product matrix as a vector by stacking the columns one on top of another into a vector, i.e. $\mathbf{r}^* = vec(\widehat{\mathbf{r}}\,\widehat{\mathbf{r}}^T)$ which is then used in the next level of estimation to obtain consistent estimates of $\widehat{\sigma}_e^2$ and $\widehat{\sigma}_u^2$. Hence, we can estimate the random parameters as:

$$\widehat{\boldsymbol{\theta}}_{\boldsymbol{\varepsilon}} = (\boldsymbol{\varepsilon}^{*T}\mathbf{V}^{*-1}\boldsymbol{\varepsilon}^*)^{-1}\left(\boldsymbol{\varepsilon}^{*T}\mathbf{V}^{*-1}\mathbf{r}^*\right) \tag{15}$$

where $\mathbf{V}^*$ is the Kronecker product of $\mathbf{V}$, namely $\mathbf{V}^* = \mathbf{V} \otimes \mathbf{V}$ and the covariance matrix is given by $\mathbf{V} = E(\widehat{\mathbf{r}}\,\widehat{\mathbf{r}}^T)$. The matrix $\boldsymbol{\varepsilon}^*$ is the design matrix of the random parameters. From $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\varepsilon}}$ we derive estimates of $\widehat{\sigma}_e^2$ and $\widehat{\sigma}_u^2$ which are used to construct the covariance matrix of the response variable $\mathbf{Y}$ at each iteration using a GLS estimation of the fixed parameters. Once the fixed coefficients are obtained, updated residuals are formed

$$\begin{bmatrix} \widehat{\mathbf{r}}_e \\ \widehat{\mathbf{r}}_u \end{bmatrix} = \begin{bmatrix} \widehat{\sigma}_e^2\mathbf{e}^T \\ \widehat{\sigma}_u^2\mathbf{u}^T \end{bmatrix} \mathbf{V}^{-1}(\mathbf{Y} - \mathbf{J}\boldsymbol{\lambda}) \tag{16}$$

[2] The method is currently implemented in the software MLwiN.

and the random parameters estimated once again. This procedure is repeated until some convergence criteria are met.[3]

As Goldstein (1989) has stressed, the IGLS used in the context of random multilevel modelling is equivalent to a maximum likelihood method under multivariate normality which in turn may lead to biased estimates. To produce unbiased estimates a Restricted Iterative Generalized Least Squares (RIGLS) method may be used which, after the convergence is achieved, turns out to be equivalent to a Restricted Maximum Likelihood Estimate (REML). One advantage of the latter method is that, in contrast to IGLS, estimates of the variance components take into account the loss of the degrees of freedom resulting from the estimation of the regression parameters. Hence, while the IGLS estimates for the variance components have a downward bias, the RIGLS estimates don't.

4 Multilevel Models with Spatial Fixed Effects

Often in a multilevel model a regressor can be regarded as endogenous as it is not independent from the random effects in the model. In such circumstances a basic assumption of modelling is not met and obtaining consistent estimators of the parameters is not straightforward.

For example in a UK study of children performance in National Curriculum Key Stage 1 tests, Fielding (1999) stresses that performance may be related to unmeasured school effects u_j. The common influences may be such things as the locality in which the school is situated and from which the pupils generally come.

Goldstein, Jones, and Rice (2002) show that when $cov(u_j, X_{ij}) = \sigma_{ux} \neq 0$ and for small group size, the IGLS estimator is biased and inconsistent. In the panel data literature, the standard test for this is the Hausman test (Hausman (1978)). Considering multilevel models as extensions of random effects panel data models to the case of hierarchical data, we solve the problem of inconsistent estimators due to correlation between the regressors and the random components in a similar way.

The first solution is the Least Squares Dummy Variable Estimator (LSDV). Consistent estimation of $\boldsymbol{\beta}$ can be achieved by specifying (12) as a fixed effect model, removing group variables, $\mathbf{Z}$, specifying dummy variables for group membership, $\mathbf{D}$, and estimating by OLS:

$$\mathbf{Y} = \boldsymbol{\beta}_0 + \mathbf{X}\boldsymbol{\beta}_1 + \mathbf{D}\boldsymbol{\eta} + \boldsymbol{\varepsilon} \tag{17}$$

where $\mathbf{D}$ is a $(N \times (G-1)\ )$ vector of dummy variables and $\boldsymbol{\eta}$ is a $((G-1) \times 1\ )$ vector of coefficients. However the group level variables, $\mathbf{Z}$, and the parameter vector, $\boldsymbol{\gamma}$, as defined in (12), are not now identifiable so the LSDV of $\boldsymbol{\beta}_1$ is not fully efficient compared to a random effects model.

Alternatively we may adopt a within groups (CV) estimator. Given our matrix of dummy variables $\mathbf{D}$, which is supposed to be of full column rank, we define a projection matrix onto the columns space of $\mathbf{D}$, denoted $\mathbf{P}_G = \mathbf{D}(\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T$, whereas $\mathbf{Q}_D = \mathbf{I} - \mathbf{P}_D$ is the projection onto the space orthogonal to $\mathbf{D}$. If we premultiply (17) by the idempotent matrix $\mathbf{Q}_D$ we obtain

$$\mathbf{Q}_D\mathbf{Y} = \mathbf{Q}_D\,\mathbf{X}\,\boldsymbol{\beta}_1 + \mathbf{Q}_D\boldsymbol{\varepsilon} \tag{18}$$

The within estimator of (18) can be treated as an Instrumental Variable (IV) estimator of $\boldsymbol{\beta}_1$, determined by projecting (17) onto the null space of $\mathbf{D}$ by the matrix $\mathbf{Q}_D$ through the instrument set $\mathbf{Q}_D\mathbf{X}$. By applying OLS to (18) we obtain a consistent estimator of $\boldsymbol{\beta}_1$, namely $\hat{\boldsymbol{\beta}}_W$:

[3] Assuming multivariate normality the estimated covariance matrix for the fixed parameter is $\mathrm{cov}(\hat{\boldsymbol{\lambda}}) = (\mathbf{J}^T\mathbf{V}^{-1}\mathbf{J})^{-1}$ and for the random parameters (Goldtsein and Rasbash (1992)) is $\mathrm{cov}(\hat{\boldsymbol{\theta}}) = 2(\boldsymbol{\varepsilon}^T\mathbf{V}^{*-1}\boldsymbol{\varepsilon})^{-1}$.

$$\hat{\boldsymbol{\beta}}_W = (\mathbf{X}^T \mathbf{Q}_D \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{Q}_D \mathbf{Y}) \tag{19}$$

Both the LSDV and the CV estimators do not allow identification of group level coefficients, $\boldsymbol{\gamma}$, as specified in (12). However, these may be retrieved by a two step process (Hausman and Taylor (1981)) where, first, we compute residuals, $\widehat{\mathbf{r}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_W$ using as a first step the estimator in (19), and then regressing the residuals on the group level variables $\mathbf{Z}$:

$$\widehat{\mathbf{r}} = \mathbf{Z}\boldsymbol{\gamma} + \mathbf{u} \tag{20}$$

If all elements of $\mathbf{Z}$ are uncorrelated with $\mathbf{u}$ OLS will be consistent for $\boldsymbol{\gamma}$. If the columns of $\mathbf{Z}$ are correlated with $\mathbf{u}$, instruments may be found for $\mathbf{Z}$, and estimation of (20) can proceed by Two-Stage Least Squares (2SLS). If we have consistent estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ then we may proceed to have consistent estimates of the variance components σ_u^2 and σ_e^2.

5 Multilevel Models with Spatial Random Effects

Let's now assume a specific form of spatial dependence where the dependent variable, $\mathbf{Y}$, depends on its spatial lag as in traditional spatial autoregressive (SAR) models:

$$\mathbf{Y} = \boldsymbol{\beta}_0 + \rho\mathbf{W}\mathbf{Y} + \mathbf{X}\boldsymbol{\beta}_1 + \mathbf{u} + \mathbf{e} \tag{21}$$

where $\mathbf{W}$ is an $N \times N$ matrix with G groups/areas each containing w_j units so that $\sum_{j=1}^{G} w_j = N$. Let's start by assuming that $\mathbf{W}$ is a block diagonal matrix:

$$\begin{aligned} \mathbf{W} &= \text{Diag}(\mathbf{W}_1, ..., \mathbf{W}_G) \\ \mathbf{W}_j &= \frac{1}{w_j - 1}(\mathbf{l}_{w_j}\mathbf{l}_{w_j}^{'} - \mathbf{I}_{w_j}) \quad j = 1, .., G \end{aligned} \tag{22}$$

where $\mathbf{l}_{w_j}$ is the w_j-dimensional column vector of ones and $\mathbf{I}_{w_j}$ is the w_j-dimensional identity matrix. Elements on the diagonal of the matrix $\mathbf{W}$ indicate that individuals within a group are affected by the behaviour of other units residing in the same location, in other words by other members of the same group.

Consider within-group effects and assume $\bar{Y}_j = \frac{1}{w_j - 1}\sum_{i=1}^{w_j} Y_{ij}$ so that each unit within group j has the same weight. This means that, somewhat differently from conventional spatial econometrics, the inter-individual interactions do not spill across group boundaries, and within groups no account is taken of differential location leading to different weights according to distance between individuals. With the assumptions, we can rewrite:

$$Y_{ij} = \beta_0 + \rho_1\bar{Y}_j + \beta_1 X_{ij} + e_{ij} + u_j, \tag{23}$$

Note that because of endogeneity induced by the spatial lag then $Cov\,(u_j \mid X_{ij}) \neq 0$, and we find that the unobserved heterogeneity is correlated with the explanatory variable, X_{ij}. The reason for this is that X_{ij} determines Y_{ij} which determines $\bar{Y}_j$, so X_{ij} correlates with $\bar{Y}_j$ and therefore it correlates with u_j because u_j is in Y_{ij} which is in $\bar{Y}_j$.

If we adopt a fixed effect specification, illustrated in the previous section, we proceed to demean the system to eliminate the fixed effect. To do so we consider the Within-Group and Between-Group specifications:

$$WG \quad : \quad Y_{ij} = \beta_0 + \rho_1 \bar{Y}_j + \beta_1 X_{ij} + e_{ij} + u_j$$
$$BG \quad : \quad \bar{Y}_j = \beta_0 + \rho_1 \bar{Y}_j + \beta_1 \bar{X}_j + \bar{e}_j + u_j$$

and subtracting WG from BG gives:

$$Y_{ij} - \bar{Y}_j = \beta_1 \left(X_{ij} - \bar{X}_j \right) + e_{ij} - \bar{e}_j \tag{24}$$

and then proceed to use OLS (assuming that $E\left[\left(X_{ij} - \bar{X}_j \right)' \left(e_{ij} - \bar{e}_j \right) \right] = 0$) to obtain the within-group coefficient for β_1. However, as for the coefficient, $\boldsymbol{\gamma}$, on the contextual effect, $\mathbf{Z}$, described in the previous section, this would leave the fixed effect ρ_1 unknown.

We therefore transform the original specification (23) and use as an instrument for $\bar{Y}_j$ the following obtained from the BG equation:

$$\bar{Y}_j = \frac{1}{1-\rho_1} \left(\beta_0 + \beta_1 \bar{X}_j \right) + u_j^{'} \tag{25}$$

where $u_j^{'} = \frac{\bar{e}_j + u_j}{1-\rho_1}$. We know that the parameter ρ_1 cannot be identified from (25) as it cannot be isolated from $u_j^{'}$ so the possible identification rests on the within equation relationship. Hence, this is done by substituting (25) into (23) giving:

$$Y_{ij} = \frac{\beta_0}{1-\rho_1} + \beta_1 X_{ij} + \frac{\rho_1}{1-\rho_1} \beta_1 \bar{X}_j + e_{ij} + u_j^{''} \tag{26}$$

where $u_j^{''} = u_j + \rho_1 u_j^{'}$. Therefore, we can efficiently estimate ρ_1, but this relies on the assumption that the dependence between u_j and X_{ij} is given by (25). However, the overall specification still suffers from an additional problem as the covariates X_{ij} and $\bar{X}_j$ may be correlated, thus generating a second type of dependence in which $Cov\left(e_{ij}, X_{ij} \right) \neq 0$.

We solve this additional problem by subtracting the group mean from the individual level (level-one) covariate to produce a centred variable, which is referred to as the within-group deviation. We know that the group mean is constant within each group and that the random group effects, like the group mean, vary only between groups. So they would be uncorrelated with the centered variable, which varies within groups, which can therefore be used as an additional instrumental variable.

In this case, as suggested in the multilevel literature (Snijders and Berkhof (2008); Rabe-Hesketh and Skrondal (2008)), a parametrization (26), with centered effects $\left(X_{ij} - \bar{X}_j \right)$, may be preferable:

$$Y_{ij} = \frac{\beta_0}{1-\rho_1} + \frac{\beta_1}{1-\rho_1} \bar{X}_j + \beta_1 (X_{ij} - \bar{X}_j) + e_{ij} + u_j^{''} \tag{27}$$

Hence by applying a MLM model with within and between group effects we can identify the autoregressive parameter ρ_1 while accounting for spatial dependence in the error term $u_j^{''}$. We have therefore shown the equivalence between a centered two-level multilevel model and a (specific form of) SAR model, where the individual $i's$ response in group j is affected by, and simultaneously affects, the response of other individuals so long as they share the same group. Estimation of fixed and random effects can then proceed by using RIGLS conditional on $\bar{X}_j$.[4]

[4] A further method is proposed by Lee (2007) who considers the within equation:

6 Multilevel Models with Interaction Effects

With the emergence of interaction-based models (Manski (2000), Brock (2001), Akerlof (1997)) research has gradually moved from a pure spatial definition of neighborhood towards a multidimensional measure based on different forms of social distance and spillovers (Anselin (2002), Anselin (1999)). In this setting multilevel models with group-effects are generally defined as economic environments where the payoff function of a given agent takes as direct arguments the choice of other agents (Brock (2001)). Typical example is the emergence of social network where it is often observed that persons belonging to the same group tend to behave similarly (Manski (2000)) and that the propensity of a person to behave in a certain way varies positively with the dominant behaviour in the group as in the case of social norms (Bernheim (1994), Kandori (1992)).[5]

For example Graham (2004) suggests modelling the determinants of student achievement in different schools as an outcome of social interactions and individual heterogeneity as follows:

$$Y_{ij} = \beta_0 + \rho_1 \bar{Y}_j + \beta_1 X_{ij} + \gamma Z_j + e_{ij} + u_j \tag{29}$$

where $\bar{Y}_j$ is the mean achievement level in the $j-th$ classroom and Z_j represents contextual effects. Equation (29) states that student achievement Y_{ij} varies with mean peer group j achievement, $\bar{Y}_j$, and individual-level characteristics X_{ij}. In fact the only difference between (29) and (23) is the presence of Z_j. In particular, endogenous group effects $\bar{Y}_j$ capture the impact of peers on learning; exogenous or contextual effects Z_j arise when peer group background characteristics directly affect students' achievement; correlated or group effects arise because group members share a common environment. Group interaction effects are present when either ρ_1 or γ differ from zero. Following Manski (1993), ρ_1 captures the strength of endogenous group effects, γ is the exogenous or contextual effect, u_j are random group effects and e_{ij} is an individual specific random component capturing other unmodelled sources of variation in Y_{ij}.

By allowing for the possibility that the conditional mean of group effect and the individual effect vary with group size, Graham (2004) also allows for the possibility that peer-group effects may differ according to class size, being stronger in bigger classes. As another example, consider workers in firms, with wages Y_{ij} dependent on individual worker attributes, X_{ij}, firm level contextual effects Z_j (such as sector, company wages policy, level of research and development activity, investment etc). In addition we may have other unmeasured causes of wage variation represented by random group (company) effects u_j and individual random effects e_{ij}, and with $\rho_1 \neq 0$ worker wages may also be

---

$$Y_{ij} - \bar{Y}_j = \frac{(w_j - 1)\beta_1}{(w_j - 1 + \rho_1)} \left(X_{ij} - \bar{X}_j\right) + \frac{(w_j - 1)}{(w_j - 1 + \rho_1)} \left(e_{ij} - \bar{e}_j\right) \tag{28}$$

where w_j is the number of member in the j-th group. In this case the identification of ρ_1 relies on various degrees of deviations across groups. This is possible when different groups have different number of units. However this specification of the within-group equation, differently from (27) cannot help to identify ρ_1 when all groups have the same number of units, i.e. when w_j is a constant for all groups. Also when w_j are all large, ρ_1 cannot be, again, identified as the factor $\frac{(w_j-1)}{(w_j-1+\rho_1)}$ is close to one and the specification (28) can be approximated by the conventional (24) and estimated by OLS. For all other intermediate cases and varying group size Lee uses a conditional maximum likelihood (CML) estimation. The main limitation of the proposed CML is that the fixed group effects which are assumed to be uncorrelated with the regressors. The multilevel specification (27) goes one step forward by considering group random effects and in taking into account the correlation information between group random effects and the regressors in the estimation of the coefficients.

[5] Other influences are the so called peer influence effects which have been extensively examined both in education (Bénabou (1993)), in the psychology literature (Brown (1990) and Brown, Clasen, and Eicher (1986)) and in the occurrence of social pathologies (Bauman (1986); Krosnick and Judd (1982); Jones (1994)).

endogenously determined so that a higher level of productivity/wage by one worker spills over (via $\bar{Y}_j$) to other workers in the firm.

As with (**??**), taking group means of both sides of (29) and solving for $\bar{Y}_j$ (assuming $\rho_1 \neq 1$) results in the between group variation:

$$\bar{Y}_j = \frac{\beta_0}{1-\rho_1} + \frac{\beta_1}{1-\rho_1}\bar{X}_j + \frac{\gamma}{1-\rho_1}Z_j + u_j^{'} \tag{30}$$

where $u_j^{'} = \frac{\bar{e}_j + u_j}{1-\rho_1}$. This is simply equation (25) with the additional variable Z_j. If $\bar{X}_j = Z_j$ (what Manski calls reflection) from (30) and (29) and centering:

$$Y_{ij} = \frac{\beta_0}{1-\rho_1} + \frac{\beta_1 + \gamma}{1-\rho_1}\bar{X}_j + \beta_1(X_{ij} - \bar{X}_j) + e_{ij} + u_j^{''} \tag{31}$$

with $u_j^{''} = u_j + \rho_1 u_j^{'}$. In this case the number of coefficients in the reduced form (31) is not sufficient to identify the coefficients in the structural equation (29). Note that without group effects ($\rho_1 = \gamma = 0$) the reduced form simplifies to the basic one-way error component model $Y_{ij} = \beta_0 + \beta_1 X_{ij} + u_j + e_{ij}$. In other words group effects generate excess between group variance by, for instance, introducing mean peer characteristics, $\bar{X}_j$, as an effect on outcomes.[6]

In the following section we will show how spatial models and multilevel models with group interactions ($\rho_1 \neq 0, \gamma \neq 0$) can be connected and the relevant parameters identified using simple instruments for the endogenous effects, $\bar{Y}_j$.

6.1 Linking Spatial Models to Multilevel Models: Identifying Interaction Effects

Extending the work by Cohen-Cole (2006) to the area of modelling interaction effects in a multilevel setting we can rewrite (29) in a way that takes into account possible interdependencies across groups, where a group may be those living in a specific district or region. We assume intra-group effects $\bar{Y}_j = \frac{1}{w_j - 1}\sum_{i=1}^{w_j} Y_{ij}$ where $\bar{Y}_j$ is an average of all w_j unit responses within group j and out-groups effects $\bar{Y}_l = \frac{1}{w_l - 1}\sum_{l \neq j}^{w_l} Y_l$ where $\bar{Y}_l$ is an average of the responses across all 'neighbouring' groups, except group j. Typically we might consider $\bar{Y}_l$ to be based only on those groups that are spatially or socially proximate. Note that we implicitly assume that all surrounding groups enter with the same weight, $\frac{1}{w_l - 1}$, but we could also adopt other weighting schemes based on relative distance or on other criteria as described in section 2.1. Similarly for the out-group contextual effects we have $\bar{Z}_l = \frac{1}{w_l - 1}\sum_{l \neq j}^{w_l} Z_l$, leading to the model :

$$Y_{ij} = \beta_0 + \rho_1 \bar{Y}_j + \rho_2(\bar{Y}_j - \bar{Y}_l) + \beta_1 X_{ij} + \gamma(Z_j - \bar{Z}_l) + e_{ij} + u_j, \tag{32}$$

in which the outcome of individual i in region j, Y_{ij}, depends on the average outcome of group j, where the implicit assumption is that unit i is affected equally by all other units living in the same location j. We also assume that the individual outcome depends on the average outcome, $\bar{Y}_l$, and average contextual effects, $\bar{Z}_l$, of other regions 'surrounding' group j. We represent the endogenous out-groups spillover variable in deviation form, as the mean within the group minus the mean in regions 'nearby', hence the variable is $\bar{Y}_j - \bar{Y}_l$. Likewise the contextual spillover variable is specified

[6] A simple way to detect the presence of group effects is to measure the excess between-group variance (Graham (2004)). If we denote the group size with w_j then excess variance is defined as the ratio of unconditional (scaled) between-group and within group variances $EV = \frac{E\left[w_j(\overline{Y}_j - \mu_y)^2\right]}{E\left[(w_j - 1)^{-1}\sum_{i=1}^{w_j}(Y_{ij} - \overline{Y}_j)^2\right]}$.

as $Z_j - \bar{Z}_l$. Spatial autocorrelation in the error term exists because Y_{ij} depends on u_j which also affects $\bar{Y}_j$ and $(\bar{Y}_j - \bar{Y}_l)$ through the coefficients ρ_1 and ρ_2.

In order to identify all the relevant parameters in the model we consider the instrumental relationship derived from (32):

$$\bar{Y}_j = \frac{1}{1-\rho_1-\rho_2}\left(\beta_0 + \beta_1\bar{X}_j + \gamma(Z_j - \bar{Z}_l) - \rho_2\bar{Y}_l\right) + u_j^{'} \tag{33}$$

where $u_j^{'} = \frac{\bar{e}_j + u_j}{1-\rho_1-\rho_2}$.

Replacing (33) in (32) gives:

$$\begin{aligned} Y_{ij} &= \frac{\beta_0}{1-\rho_1-\rho_2} + \beta_1 X_{ij} - \frac{\rho_2}{1-\rho_1-\rho_2}\bar{Y}_l + \frac{(\rho_1+\rho_2)}{1-\rho_1-\rho_2}\beta_1\bar{X}_j \\ &\quad + \frac{\gamma}{1-\rho_1-\rho_2}Z_j - \frac{\gamma}{1-\rho_1-\rho_2}\bar{Z}_l + e_{ij} + u_j^{''} \end{aligned} \tag{34}$$

and assuming reflection $\bar{X}_j = Z_j$:

$$\begin{aligned} Y_{ij} &= \frac{\beta_0}{1-\rho_1-\rho_2} + \beta_1 X_{ij} + \frac{\beta_1(\rho_1+\rho_2)+\gamma}{1-\rho_1-\rho_2}\bar{X}_j \\ &\quad - \frac{\gamma}{1-\rho_1-\rho_2}\bar{Z}_l - \frac{\rho_2}{1-\rho_1-\rho_2}\bar{Y}_l + e_{ij} + u_j^{''} \end{aligned} \tag{35}$$

where $u_j^{''} = u_j + (\rho_1+\rho_2)u_j^{'}$. It is clear from (35) that in the absence of collinearity we can identify all the parameters in the structural equation (32) if the number of level-one units (typically number of individual people) exceeds the number of groups and $\frac{\partial Y_{ij}}{\partial \bar{Y}_l} \neq 0$ i.e. if for some $j \neq l$ agents in one group are affected by the value of Y or by contextual effects Z in 'neighbouring' groups.

Reparametrising, we obtain an equivalent multilevel model:

$$\begin{aligned} Y_{ij} &= \frac{\beta_0}{1-\rho_1-\rho_2} + \underbrace{(X_{ij} - \bar{X}_j)\beta_1}_{\text{within-group}} + \underbrace{\frac{\gamma+\beta_1}{1-\rho_1-\rho_2}\beta_1\bar{X}_j}_{\text{between-group}} \\ &\quad - \underbrace{\frac{\rho_2}{1-\rho_1-\rho_2}\bar{Y}_l}_{\text{out-group endogenous}} - \underbrace{\frac{\gamma}{1-\rho_1-\rho_2}\bar{Z}_l}_{\text{out-group contextual}} + e_{ij} + u_j^{''} \end{aligned} \tag{36}$$

So a MLM model with out-group effects $\bar{Y}_l$ and $\bar{Z}_l$ facilitates identification (c.f. Manski, 1993) of the spatial model parameters in equation (32). This is a consequence of using the out-group effects $\bar{Y}_l$, $\bar{Z}_l$ and the between-group effects $\bar{X}_j$ as internal instruments for the endogenous variable $\bar{Y}_j$, which maintains the assumption that $Cov\,(u_j, X_{ij}) = 0$. Also centering the variables resolves the potential correlation between the regressors and the individual error term $Cov\,(e_{ij}, X_{ij}) = 0$, again allowing consistent estimation. Estimation is achieved via RIGLS/REML conditional on $\bar{Y}_l$, $\bar{Z}_l$ and $\bar{X}_j$.

6.2 SARAR Models, GMM and FGS2SLS

We now consider alternative estimation routines for multilevel models with both endogenous spatial lags and spatial effects in the error process extending methods currently available with panel data. To do this we start by considering a simple random effects panel specification for time $t = 1, ..., T$ and for individual $i = 1, ..., N$ given by:

$$\begin{aligned} Y_{it} &= \beta_{0t} + \beta_1 X_{it} + \nu_{it} \\ \beta_{0t} &= \beta_0 + \mu_i \end{aligned}$$

with $\nu_{it} \sim iid(0, \sigma_\nu^2)$ and $\mu_i \sim iid(0, \sigma_\mu^2)$ which can be rewritten as:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \mu_i + \nu_{it}$$

where Y_{it} is individual i's response at time t, X_{it} is the exogenous variable, u_i is an error specific to each individual and ν_{it} is a transient error component specific to each time and each individual. We can introduce spatial effects both as an endogenous spatial lag:

$$Y_{it} = \beta_0 + \rho W Y_{it} + \beta_1 X_{it} + \mu_i + \nu_{it}$$

and as an autoregressive error process:

$$\begin{aligned} Y_{it} &= \beta_0 + \rho W Y_{it} + \beta_1 X_{it} + e_{it} \\ e_{it} &= \lambda M e_{it} + \xi_t \end{aligned}$$

and generalizing to k regressors in the panel context this becomes:

$$\mathbf{Y} = \rho(\mathbf{I}_T \otimes \mathbf{W})\mathbf{Y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

in which $\mathbf{Y}$ is a $TN \times 1$ vector of observations obtained by stacking Y_{it} for $i = 1...N$ and $t = 1 \ldots T$, $\mathbf{X}$ is a $TN \times k$ matrix of regressors and $\boldsymbol{\beta}$ is a $k \times 1$ vector of coefficients. Note that in this specification, the autoregressive spatial dependence involving the dependent variable and the errors extends ad infinitum, rather than being confined by group boundaries.

Following Kapoor, Kelejian, and Prucha (2007) and Kelejian and Prucha (1998), all of the diagonal elements of $\mathbf{W}$ are zero, and $\mathbf{I} - \rho\mathbf{W}$ is non-singular. Also $\mathbf{W}$ is uniformly bounded in absolute value, meaning that a constant c exists such that $max_{1 \leq i \leq N} \sum_{j=1}^{N} \mid W_{ij} \mid \leq c \leq \infty$ and $max_{1 \leq j \leq N} \sum_{i=1}^{N} \mid W_{ij} \mid \leq c \leq \infty$. Similarly $\mathbf{M}$ is an $N \times N$ matrix with similar properties to $\mathbf{W}$ and the elements of $\mathbf{X}$ are also uniformly bounded in absolute value.

In addition, given a $TN \times TN$ identity matrix with $1s$, the $NT \times 1$ vector $\mathbf{e}$ is

$$\mathbf{e} = (\mathbf{I}_{NT} - \lambda \mathbf{I}_T \otimes \mathbf{M})^{-1} \boldsymbol{\xi}$$

in which $\boldsymbol{\xi}$ is an $NT \times 1$ vector of innovations, λ is a scalar parameter and $\mathbf{M}$ is an $N \times N$ matrix with similar properties to $\mathbf{W}$. Regarding the error components in space-time, time dependency is introduced into the innovations via the permanent individual error component μ, thus:

$$\boldsymbol{\mu} \sim iid(0, \sigma_\mu^2) \tag{37}$$

$$\boldsymbol{\nu} \sim iid(0, \sigma_\nu^2) \tag{38}$$

$$\boldsymbol{\xi} = (\boldsymbol{\iota}_T \otimes \mathbf{I}_N)\boldsymbol{\mu} + \boldsymbol{\nu} \tag{39}$$

so that $\boldsymbol{\mu}$ is an $N \times 1$ vector of random effects specific to each individual, $\boldsymbol{\nu}$ is the transient error component comprising an $NT \times 1$ vector of errors specific to each individual and time, $\boldsymbol{\iota}_T$ is a $T \times 1$ matrix with 1s , and $\boldsymbol{\iota}_T \otimes \mathbf{I}_N$ is a $TN \times N$ matrix equal to T stacked matrices. The result is that the $TN \times TN$ innovations variance-covariance matrix $\boldsymbol{\Omega}_\xi$ is nonspherical. Also $\sigma_1^2 = \sigma_v^2 + T\sigma_\mu^2$. Note that this differs from the specifications given by Anselin (1988a) and by Baltagi and Li (2006), where the autoregressive error process is confined to $\boldsymbol{\nu}$. In contrast the Kapoor, Kelejian, and Prucha (2007) set-up assumes that the individual effects $\boldsymbol{\mu}$ have the same autoregressive process.

In the multistage context, we cannot apply the Kronecker products because the data no longer comprise the 'same' individual i at T different time points. Hence matrix cells W_{ij} and M_{ij} no longer define the spatial connection between individuals's i and j at time t, but simply the spatial connection between individuals's i and j. Given that time is now constant, with N individuals the model becomes:

$$\mathbf{Y} = \boldsymbol{\beta}_0 + \rho\mathbf{W}\mathbf{Y} + \mathbf{X}\boldsymbol{\beta}_1 + \mathbf{e} \tag{40}$$

in which $\mathbf{Y}$ is an $N \times 1$ vector of observations, $\mathbf{X}$ is a $N\times$ k matrix of regressors, ρ is a scalar parameter and $\mathbf{WY}$ is an $N \times 1$ vector obtained as a result of the matrix product of $N \times N$ matrix $\mathbf{W}$ and $\mathbf{Y}$. In addition, given an $N \times N$ identity matrix with $1s$, the $N \times 1$ vector $\mathbf{e}$ is:

$$\mathbf{e} = (\mathbf{I}_N - \lambda\mathbf{M})^{-1}\boldsymbol{\xi}$$

in which $\boldsymbol{\xi}$ is an $N \times 1$ vector of innovations.

We now consider the error components in which the N individuals are, for example, distributed amongst G level two groups (say neighbourhoods) and these G neighbourhoods are nested within R level three groups (regions), and the regions are nested within countries. Confining attention to the G neighbourhoods and R regions, we envisage the random components thus.

$$\boldsymbol{\xi} = \mathbf{G}\boldsymbol{\mu} + \mathbf{R}\boldsymbol{\eta} + \boldsymbol{\nu} \tag{41}$$

The random component representing the neighbourhood effect is represented by $\mathbf{G}\boldsymbol{\mu}$ in which $\boldsymbol{\mu} \sim iid(0, \sigma_\mu^2)$ is a $G \times 1$ vector specific to the level 2 variable, and $\mathbf{G}$ is a $N \times G$ matrix of $1s$ and $0s$, so that $G_{ij} = 1$ indicates that individual i is affected by the neighbourhood effect μ_j. For the level 3 (sub-region) component $\mathbf{R}\boldsymbol{\eta}$,there are r random draws from $\eta \sim iid(0, \sigma_\eta^2)$, and $\mathbf{R}$ is a $N \times R$ matrix of $1s$ and $0s$.

Returning to the panel model, Kapoor, Kelejian, and Prucha (2007) advocate feasible GLS plus generalised moments (GM) as an approach to consistent estimation. A small step to extend this (Fingleton (2008)) is feasible generalized spatial two stage least squares (FGS2SLS) to allow also the endogenous spatial lag, in which case the first stage of estimation uses 2SLS to produce consistent estimates of the residuals the presence of the endogenous variable $\rho\mathbf{WY}$. In the second stage, the GM estimator of $\sigma_\nu^2, \sigma_\mu^2$ and λ is the solution to moments conditions based on the consistent residuals. In the third stage, since these estimates lead to the estimate of the nonspherical error covariance matrix $\boldsymbol{\Omega}_\xi$, and this is then used to produce robust IV estimates of $\rho, \boldsymbol{\beta}_1$ and their standard errors following Bowden and Turkington (1984) and Greene (2003). However in the multistage context, a comparable GM-based estimator of $\boldsymbol{\Omega}_\xi$ is as far as we know not currently available. Evidently the FGS2SLS plus GMM estimator outlined for panel data with spatial effects is a special case of a multilevel model with a level 2 variable replacing time, with each sub-sample at each value of the level 2 random effect being the same size (a balanced panel) and with spatial effects operating via the endogenous spatial lag $\rho\mathbf{WY}$ and via the error process $\mathbf{e} = \lambda\mathbf{Me} + \boldsymbol{\xi}$ being confined only to those individuals within a specific level 2 group and not spilling across to all other individuals. Additionally, it presupposes that there are no higher levels.

7 Conclusions

In multilevel modelling alongside the nested structure of the hierarchical data increasing attention has been paid to different forms of interactions and spatial externalities in the hierarchical system. Neglecting such interactions is likely to create problems of inference since the existence of spatial dependence adds a common element to otherwise independent errors. The paper has critically reviewed several estimation methods to control for spatial dependence in a multilevel framework. A fixed effect specification often modelled in the mean equation through additive and multiplicative dummy variables may be appropriate when, for example, we have data grouped by area with all the areas from the geographical populations represented in the sample. However, having controlled for these measurable compositional effects may not suffice to have consistent estimates of the fixed and random parameters since there may still be unobservable effects due to a commonality of residence or to neighborood effects. The paper has illustrated several alternative estimation strategies such as the within groups (CV) estimator which accommodate the correlation between predictors and spatial random effects.

The application of the IGLS method to multilevel models where the spatial dependence is modeled as a fixed effect may generate problems: for example even after controlling for correlation between predictors and spatial random effects, as in the CV method of Goldstein, Jones, and Rice (2002), the group level coefficients are now not identifiable and the estimates not fully efficient. We have therefore shown how spatial models and multilevel models with group interactions can be connected and the relevant parameters for the random and spatial lag identified using simple instruments for the endogenous effects. Since such models will be characterised by spatial autocorrelation in the group random effects an alternative route is to model spatial dependence through the error term by allowing for an unrestricted non-diagonal covariance matrix as in the feasible generalized spatial two stage least squares (FGS2SLS) method where the spatial dependence affects both the endogenous variable and the distribution of the random components. As the paper shows the latter strategy may be superior since often spatial externalities at any level of the hierarchical model may render the random components non-Gaussian.

References

Akerlof, G. A. (1997): "Social Distance and Social Decisions," *Econometrica*, 65(5), 1005–1028.

Anselin, L. (1988a): *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers, The Netherlands.

——— (1988b): *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers.

——— (1999): "The Future of Spatial Analysis in the Social Sciences," *Geographic Information Sciences*, 5(2), 67–76.

——— (2003): "Spatial Externalities," *International Regional Science Review*, 26(2), 147–152.

Anselin, L., J. Le Gallo, and J. Jayet (2007): "Panel Data Models: Some Recent Developments," in *The Econometrics of Panel Data, Fundamentals and Recent Developments in Theory and Practice*, ed. by L. Matyas, and P. Sevestre, Dordrecht. Kluwer, 3rd Edition.

Anselin, L., and R. Moreno (2003): "Properties of Tests for Spatial Error Components," *Regional Science and Urban Economics*, 33, 595–618.

Anselin, L. and Cho, W. K. T. (2002): "Spatial Effects and Ecological Inference," *Political Analysis*, 10(3), 276–303.

Aronson, R., and P. Brodsky, A. and O'Campo (1999): "PSOC in Community Context: Multilevel Correlates of a Measure of Psychological Sense of Community in Low Income, Urban Neighborhoods," *Journal of Community Psychology*, 27, 659–680.

Aslam, A., and L. Corrado (2007): "No Man is An Island: The Inter-personal Determinants of Regional Well-Being in Europe," Cambridge Working Papers in Economics 0717, Faculty of Economics, University of Cambridge.

Baltagi, B. H., and D. Li (2006): "Prediction in the Panel Data Model with Spatial Correlation: The Case of Liquor," *Spatial Economic Analysis*, 1, 175–185.

Bauman, K. and Fisher, L. (1986): "On the Measurement of Friend Behavior in Research on Friend Influence and Selection: Findings from Longitudinal Studies of Adolescent Smoking and Drinking," *Journal of Youth and Adolescence*, 15, 345–353.

Bénabou, R. (1993): "Workings of a City: Location, Education, and Production," *Quarterly Journal of Economics*, CVIII, 619–652.

Bernheim, D. (1994): "A Theory of Conformity," *Journal of Political Economy*, 102, 841–877.

Blundell, R. and Windmeijer, F. (1997): "Correlated Cluster Effects and Simultaneity in Multilevel Models," *Health Economics*, 1, 6–13.

Bodson, P. and Peters, D. (1975): "Estimation of the Coefficients of a Linear Regression in the Presence of Spatial Autocorrelation: An Application to a Belgium Labor Demand Function. Environment and Planning," *Environment and Planning*, 7, 455–472.

Bowden, R. J., and D. A. Turkington (1984): *Instrumental Variables*. Cambridge University Press.

Brock, W. A. and Durlauf, S. N. (2001): "Interactions-Based Models," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. Leamer, vol. 5, pp. 3297–3380, Amsterdam. Elsevier Science, Volume 5.

Brown, B. (1990): "Peer Groups and Peer Cultures," in *At the Threshold*, ed. by S. Feldman, and G. Elliott, Cambridge, MA. Harvard University Press.

Brown, B., D. Clasen, and S. Eicher (1986): "Perceptions of Peer Pressure, Peer Conformity Dispositions and Self-Reported Behavior Among Adolescents," *Developmental Psychology*, 22, 521–530.

Bryk, A. S., and S. W. Raudenbush (1992): *Hierarchical Linear Models for Social and Behavioral Research: Applications and Data Analysis Methods*. Sage Publications, Newbury Park, CA.

Cliff, A. D., and J. K. Ord (1973): *Spatial Autocorrelation*. Pion, London.

——— (1981): *Spatial Processes: Models and Applications*. Pion Ltd., London.

Cohen-Cole, E. (2006): "Multiple Groups Identification in the Linear-in-Means Model," *Economic Letters*, 92(2), 157–162.

Conley, T. G. (1999): "GMM Estimation with Cross-Sectional Dependence," *Journal of Econometrics*, 92, 1–45.

Conley, T. G., and E. Ligon (2002): "Economic Distance, Spillovers and Cross Country Comparisons," *Journal of Economic Growth*, 7, 157–187.

Conley, T. G., and G. Topa (2002): "Socio-economic Distance and Spatial Patterns in Unemployment," *Journal of Applied Econometrics*, 17, 303–327.

Cressie, N. (2003): *Statistics for Spatial Data*. Kluwer Academic Publishers.

Dacey, M. (1968): "A Review of Measures of Contiguity for Two and K-Color Maps," in *Spatial Analysis: A Reader in Statistical Geography*, ed. by B. Berry, and D. Marble, pp. 479–485, Englewood Cliffs, N.J. Prentice-Hall.

Durlauf, S. N. (1996): "Neighborhood Feedbacks, Endogenous Stratification, and Income Inequality," in *Dynamic Disequilibrium Modelling*, ed. by G. G. W. Barnett, and C. Hillinger, Cambridge. Cambridge University Press.

——— (2003): "Neighborhood Effects," in *Hanbook of Regional and Urban Economics*, ed. by J. F. Henderson, J. V. and Thisse. North Holland, Vol. 4, Economics.

Fernandez, R., and R. Rogerson (1997): "Keeping People Out: Income Distribution, Zoning, and the Quality of Public Education," *Quarterly Journal of Economics*, 38(1), 25–42.

Fielding, A. (1999): "Why Use Arbitrary Points Scores? Ordered Categories in Models of Educational Progress," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 162(3), 303–328.

Fingleton, B. (2008): "A Generalized Method of Moments Estimator for a Spatial Panel Model with an Endogenous Spatial Lag and Spatial Moving Average Errors'," *Spatial Economic Analysis*, 3, 27–44.

Fujita, M., and A. J. Krugman, P. and Venables (1999): *The Spatial Economy: Cities, Regions and International Trade*. MIT Press, Cambridge, MA.

Goldstein, H. (1986): "Multilevel Mixed Linear Model Analysis Using Iterative Generalised Least Squares," *Biometrika*, 73, 46–56.

——— (1989): "Restricted Unbiased Iterative Generalized Least-Squares Estimation," *Biometrika*, 76(3), 622–623.

Goldstein, H. (1998): "Multilevel Models for Analysing Social Data," *Encyclopedia of Social Research Methods*.

Goldstein, H., A. M. Jones, and N. Rice (2002): "Multilevel Models Where the Random Effects are Correlated with the Fixed Predictors," mimeo.

Goldstein, H., and P. Noden (2003): "Modelling Social Segregation," *Oxford Review of Education*, 29(2), 225–237.

Goldtsein, H., and J. Rasbash (1992): "Efficient Computational Procedures for the Estimation of Parameters in Multilevel Models Based on Iterative Least Squares," *Computational Statistics and Data Analysis*, 13, 63–71.

Graham, B. S. (2004): "Social Interactions and Excess Variance Contrasts: Identification, Estimation, and Inference with an Application on the Role of Peer Group Effects in Academic Achievement," mimeo, Department of Economics, Harvard University.

Greene, W. H. (2003): *Econometric Analysis*. Prentice Hall.

Greenhut, M., G. Norman, and C. Hung (1987): *The Economics of Imperfect Competition: A Spatial Approach*. Cambridge University Press.

Hausman, J., and W. Taylor (1981): "Panel Data and Unobservable Individual Effects," *Econometrica*, 49, 1377–1398.

Hausman, J. A. (1978): "Specification Tests in Econometrics," *Econometrica*, 46(6), 1251–71.

Hsiao, C., and M. H. Pesaran (2004): "Random Coefficient Panel Data Models," Cambridge Working Papers in Economics 0434.

Jones, A. (1994): "Health, Addiction, Social Interaction, and the Decision to Quit Smoking," *Journal of Health Economics*, 13, 93–110.

Kandori, M. (1992): "Social Norms and Community Enforcement," *Review of Economic Studies*, 59, 63–80.

Kapetanios, G., and M. H. Pesaran (2005): "Alternative Approaches to Estimation and Inference in Large Multifactor Panels: Small Sample Results with an Application to Modelling of Asset Results," in *The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis*, ed. by Phillips, and Tzavalis, Cambridge. Cambridge University Press.

Kapoor, M., H. H. Kelejian, and I. Prucha (2007): "Panel Data Models with Spatially Correlated Error Components," *Journal of Econometrics*, 140, 97–130.

Kelejian, H. H., and I. Prucha (1998): "A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances," *Journal of Real Estate Finance and Economics*, 17, 99–121.

Kelejian, H. H., and D. P. Robinson (1995): "Spatial Correlation: A Suggested Alternative to the Autoregressive Model," in *New Directions in Spatial Econometrics*, ed. by L. Anselin, and R. Florax, Berlin. Springer-Verlag, pp. 75-95.

Krosnick, J., and C. Judd (1982): "Transitions in Social Influence in Adolescence: Who Induces Cigarette Smoking," *Developmental Psychology*, 81, 359–368.

Langford, I. H., A. H. Leyland, J. Rasbash, and H. Goldstein (1999): "Multilevel Modelling of the Geographical Distributions of Diseases," *Applied Statistics*, 48, 253–268.

Lee, L. (2007): "Identification and Estimation of Econometric Models with Group Interactions, Contextual Factors and Fixed Effects," *Journal of Econometrics*, 140(2), 333–374.

Longford, N. T. (1993): *Random Coefficient Models*. Claredon Press, London.

Manski, C. F. (1993): "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, 60(3), 531–542.

——— (2000): "Economic Analysis of Social Interactions," National Bureau of Economic Research Working Paper 7580.

Moulton, B. R. (1986): "Random Group Effects and the Precision of Regression Estimates," *Journal of Econometrics*, 32, 385–397.

——— (1990): "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Unit," *The Review of Economics and Statistics*, 72(2), 334–338.

Muntaner, C., J. Lynch, and G. Oates (1999): "The Social Class Determinants of Income Inequality and Social Cohesion," *International Journal of Health Services*, 29, 699–732.

O'Campo, P. (2003): "Advancing Theory and Methods for Multilevel Models of Residential Neighborhoods and Health," *American Journal of Epidemiology*, 157(1), 9–13.

Pesaran, M. H. (2007): "A Simple Panel Unit Root Test in the Presence of Cross-Sectional Dependence," *Journal of Applied Econometrics*, 138(1), 265.

Pinkse, J., M. E. Slade, and C. Brett (2002): "Spatial Price Competition: a Semiparametric Approach," *Econometrica*, 70, 1111–1153.

Rabe-Hesketh, S., and A. Skrondal (2008): *Multilevel and Longitudinal Modeling Using Stata*. Chapman and Hall, CRC Press, Boca Raton, FL, 2 edn.

Ross, C., J. Reynolds, and K. Geis (2000): "The Contingent Meaning of Neighborhood Stability for Residents' Psychological Well-Being," *American Sociology Review*, 65, 581–597.

Shumow, L., D. Vandell, and J. Posner (1998): "Perceptions of Danger: A Psychological Mediator of Neighborhood Demographic Characteristics," *American Journal of Orthopsychiatry*, 68, 468–478.

Slade, M. E. (2005): "The Role of Economic Space in Decision Making," *ADRES Lecture*.

Snijders, T. A. B., and J. Berkhof (2008): "Diagnostic Checks for Multilevel Models," in *Handbook of Multilevel Analysis*, ed. by J. de Leeuw, and E. Meijer, pp. 141–175. Springer.

Tienda, M. (1991): "Poor People and Poor Places: Deciphering Neighborhood Effects on Poverty Outcomes," in *Macro-Micro Linkages in Sociology*, ed. by J. Huber, pp. 244–262, Newbury Park, CA. Sage Publications.

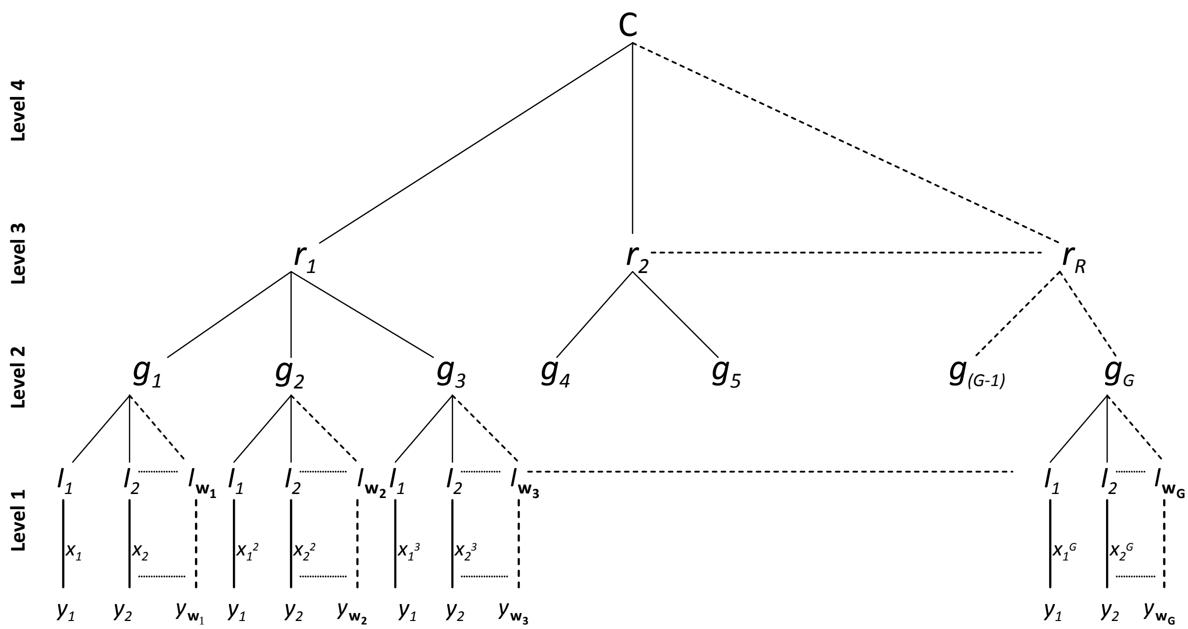

Figure 1: A Hierarchical Structure