

# STRATHCLYDE

DISCUSSION PAPERS IN ECONOMICS



WHERE IS THE ECONOMICS IN SPATIAL  
ECONOMETRICS?

BY

LUISA CORRADO AND BERNARD FINGLETON

No. 11-01

DEPARTMENT OF ECONOMICS  
UNIVERSITY OF STRATHCLYDE  
GLASGOW

# WHERE IS THE ECONOMICS IN SPATIAL ECONOMETRICS?

L. Corrado\* and B. Fingleton†

This Version: December 2010

## Abstract

Spatial econometrics has been criticized by some economists because some model specifications have been driven by data-analytic considerations rather than having a firm foundation in economic theory. In particular this applies to the so-called  $\mathbf{W}$  matrix, which is integral to the structure of endogenous and exogenous spatial lags, and to spatial error processes, and which are almost the *sine qua non* of spatial econometrics. Moreover it has been suggested that the significance of a spatially lagged dependent variable involving  $\mathbf{W}$  may be misleading, since it may be simply picking up the effects of omitted spatially dependent variables, incorrectly suggesting the existence of a spillover mechanism. In this paper we review the theoretical and empirical rationale for network dependence and spatial externalities as embodied in spatially lagged variables, arguing that failing to acknowledge their presence at least leads to biased inference, can be a cause of inconsistent estimation, and leads to an incorrect understanding of true causal processes.

Key-Words: Spatial econometrics, endogenous spatial lag, exogenous spatial lag, spatially dependent errors, network dependence, externalities, the  $\mathbf{W}$  matrix, panel data with spatial effects, multilevel models with spatial effects.

JEL code: C21,C31,R0

---

\*Marie Curie Fellow, Faculty of Economics, University of Cambridge, UK and Associate Professor in International Economics University of Rome Tor Vergata, Italy. E-mail: lc242@econ.cam.ac.uk. Luisa Corrado acknowledges the support of the Marie-Curie IE Fellowship 039326 and the Marie Curie Excellence Award 2007.

†Professor of Economics, University of Strathclyde, Sir William Duncan Building, 130 Rottenrow, Glasgow G4 0GE, UK: bf100@cam.ac.uk.

# 1 INTRODUCTION

The critique of spatial econometrics emanating from some economists is, we assert at the outset, based on imprecise and ill-informed perceptions of the sophistication and diversity of the work of the spatial econometrics and wider academic community. The argument is that standard spatial econometrics is typically applied in a mechanical fashion, variables are introduced simply because they are significant, without a priori rationale, spatial econometricians often work in isolation from urban economists and other regional scientists, and overall there is a lack of theoretical justification for variables that characterize spatial econometric models.<sup>1</sup> An engagement with the literature shows this to be a misrepresentation. There are numerous examples of single equation cross-sectional spatial econometric models, multiequation specifications, and panel models incorporating spatial effects, in which economic theory is fundamental to the specification of the reduced form, including specifications based on neoclassical growth theory (Ertur and Koch (2007), Fingleton and Lopez-Bazo (2006)), urban economics (Fingleton (2006), Barde (2010)) and on the wage equation from new economic geography (Fingleton (2003)). Almost invariably these specifications are elaborations of mainstream theory incorporating externalities in the form of spatial spillovers, being characterized by the presence of the trademark component of the spatial econometric model, namely the spatial lag, which can be considered to be the *sine qua non* of spatial econometrics. While the theory underlying these models is often exceptionally well established and well received, nevertheless it is also true that there are cases in which spatial econometric work has been too casual in its attempt to base model specifications on economic theory. Our first main contribution is to highlight this criticism. Often there is no attempt to make theory testing a focal point of the research, and too often we see an emphasis on diagnostics and empirical model validity as the most important criteria to be used to signify a good model, without any attempt to relate to real or theorized economic processes and mechanisms. Most significantly, when it comes to the spatial lag, which is based on the so-called  $\mathbf{W}$  matrix of spatial weights, many economists are skeptical, puzzled, or both. The values in the cells of  $\mathbf{W}$  comprise an explicit hypothesis about the strength of inter-location connection (typically towns, regions, or countries), and in many cases the matrix product of  $\mathbf{W}$  and endogenous variable  $\mathbf{Y}$ , namely the endogenous spatial lag  $\mathbf{WY}$ , often proves to be a highly significant variable. It has been suggested by the skeptics that the significance of explanatory variable  $\mathbf{WY}$  may be misleading, since it may be simply picking up the effects of omitted spatially dependent variables,  $\mathbf{WX}$ , incorrectly suggesting the existence of a spillover mechanism.

One way out of the  $\mathbf{W}$  matrix conundrum may appear to be via hierarchical modelling, which has had only very limited exposure in the spatial economics literature (exceptions being Smith and LeSage (2004), Parent and LeSage (2008)). Hierarchical models (also known as multilevel models) are becoming increasingly popular across the range of the social sciences, as researchers come to appreciate that observed outcomes depend on variables organized in a nested hierarchy.<sup>2</sup> In regional science and spatial economics, we can often envisage a hierarchy of effects from cities, regions containing cities, and countries containing regions. Failure to recognize these effects emanating from different hierarchical levels can lead to incorrect inference. The second major contribution of this paper is to point this out and to show that something equivalent to  $\mathbf{W}$  is very much a cognate part of a hierarchical approach, with an enhancement to hierarchical modelling coming by way of incorporating spatial effects via  $\mathbf{W}$ .

---

<sup>1</sup>The paper by Pinkse and Slade (2010) raises some additional difficult-to-resolve problems arising from a limited and somewhat atypical selection of the so-called ‘spatial econometrics’ literature.

<sup>2</sup>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).

We demonstrate that the inclusion of interdependence between groups in the form of spatial effects,  $\mathbf{WX}$ , has two main advantages: (i) it avoids the omitted variable problem that may afflict models with endogenous spatial lags and (ii) **it** introduces a source of exogenous variation that allows the identification of both endogenous and exogenous group effects.

## 2 W WITHIN SPATIAL ECONOMETRIC SPECIFICATIONS

The square matrix  $\mathbf{W}$  is of dimension  $N$ , where  $N$  is the number of nodes in a network, with the value in typical cell  $W_{jh}$  quantifying the hypothesized strength of interaction between nodes  $j$  and  $h$ . Here we stress that nodes need not necessarily be places in order to draw on the wider literature that provides additional support to the concept of network interaction. Typically all of the diagonal elements of  $\mathbf{W}$  are zero, and  $\mathbf{I} - \rho\mathbf{W}$  is non-singular. Also, following Kapoor, Kelejian, and Prucha (2007) and Kelejian and Prucha (1998),  $\mathbf{W}$  should be uniformly bounded in absolute value, meaning that a constant  $c$  exists such that  $\max_{1 \leq j \leq N} \sum_{h=1}^N |W_{jh}| \leq c < \infty$  and  $\max_{1 \leq h \leq N} \sum_{j=1}^N |W_{jh}| \leq c < \infty$  so as to produce asymptotic results required for consistent estimation.

In a single equation context, multiplying  $\mathbf{W}$  by a  $N \times 1$  vector (dependent variable)  $\mathbf{Y}$ , gives the endogenous spatial lag  $\mathbf{WY}$ , which is an integral component of numerous spatial econometric models. However given the  $N \times k$  matrix of variables  $\mathbf{X}$ , additional spatially lagged variables can be introduced, forming the columns of the matrix  $\mathbf{WX}$ . Also we can add a hypothesis that the errors may be spatially dependent. 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 (1988)), based on a matrix, say  $\mathbf{M}$ , which is  $N \times N$  with similar properties to  $\mathbf{W}$  and which may or may not be the same as  $\mathbf{W}$ , defining indirectly the spatial structure of the non-zero elements of the error variance-covariance matrix. Like  $\mathbf{W}$ , the  $\mathbf{M}$  matrix comprises non-negative values representing the *a priori* assumption about interaction strength between location pairs defined by specific rows and columns of  $\mathbf{M}$ , normally with zeros on the main diagonal. iii) common factor models originating from the time series literature (Hsiao and Pesaran (2008), Pesaran (2007), and Kapetanios and Pesaran (2005)) and iv) spatial error components models (Kelejian and Robinson (1995), Anselin and Moreno (2003)), which we consider subsequently.

Allowing endogenous and exogenous spatial lags, plus a spatial error process, and assuming nodes are linked by network dependence matrix  $\mathbf{W}$  for the lags and by  $\mathbf{M}$  for the error process, and assuming autoregressive processes, the typical single equation spatial econometric model specification, the SARAR(1,1) model, is

$$\mathbf{Y} = \rho\mathbf{WY} + \mathbf{X}\boldsymbol{\beta} + \mathbf{WX}\boldsymbol{\rho}_x + \boldsymbol{\varepsilon} \quad (1)$$

$$\boldsymbol{\varepsilon} = \lambda\mathbf{M}\boldsymbol{\varepsilon} + \boldsymbol{\xi} \quad (2)$$

$$\boldsymbol{\xi} \sim iid(0, \sigma^2\mathbf{I}) \quad (3)$$

In (1),  $\rho$  is a scalar coefficient,  $\boldsymbol{\beta}$  and  $\boldsymbol{\rho}_x$  are  $k \times 1$  vectors of coefficients, and  $\boldsymbol{\varepsilon}$  is an  $N \times 1$  vector

of disturbances. For the error process, we have the scalar  $\lambda$  and the  $N \times 1$  vector of innovations  $\boldsymbol{\xi}$  drawn from an *iid* distribution with variance  $\sigma^2$ .

We extend the scope of the spatial econometric models in two ways, first by introducing the time dimension, thus allowing panel data models with network dependence, and secondly by considering multilevel models (see Corrado and Fingleton (2010)).

Consider a simple random effects panel specification for time  $t = 1, \dots, T$  and for individual  $i = 1, \dots, N$  given by:

$$\begin{aligned} Y_{it} &= \beta_{0t} + \beta_1 X_{it} + \nu_{it} \\ \beta_{0t} &= \beta_0 + \mu_i \end{aligned} \quad (4)$$

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

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \mu_i + \nu_{it} \quad (5)$$

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

$$Y_{it} = \beta_0 + \rho W Y_{it} + \beta_1 X_{it} + \mu_i + \nu_{it} \quad (6)$$

and as an autoregressive error process:

$$\begin{aligned} Y_{it} &= \beta_0 + \rho W Y_{it} + \beta_1 X_{it} + e_{it} \\ e_{it} &= \lambda M e_{it} + \xi_{it} \end{aligned} \quad (7)$$

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

$$\mathbf{Y} = \rho(\mathbf{I}_T \otimes \mathbf{W})\mathbf{Y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \quad (8)$$

in which  $\mathbf{Y}$  is a  $TN \times 1$  vector of observations obtained by stacking  $Y_{it}$  for  $i = 1 \dots N$  and  $t = 1 \dots T$ ,  $\mathbf{X}$  is a  $TN \times k$  matrix of regressors and  $\boldsymbol{\beta}$  is a  $k \times 1$  vector of coefficients.

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

$$\mathbf{e} = (\mathbf{I}_{NT} - \lambda \mathbf{I}_T \otimes \mathbf{M})^{-1} \boldsymbol{\xi} \quad (9)$$

in which  $\boldsymbol{\xi}$  is an  $NT \times 1$  vector of innovations,  $\lambda$  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  $\mu$ , thus:

$$\boldsymbol{\mu} \sim iid(0, \sigma_\mu^2) \quad (10)$$

$$\boldsymbol{\nu} \sim iid(0, \sigma_\nu^2) \quad (11)$$

$$\boldsymbol{\xi} = (\boldsymbol{\nu}_T \otimes \mathbf{I}_N) \boldsymbol{\mu} + \boldsymbol{\nu} \quad (12)$$

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  $\sigma_1^2 = \sigma_\nu^2 + T\sigma_\mu^2$ . Note that this differs from the specifications given by Anselin (1988) 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.

### 3 PUTTING SOME ECONOMICS INTO $\mathbf{W}$

The suggestion that spatial econometrics may have been somewhat ‘mechanical’ in its application is undoubtedly true in instances where little care has been taken regarding the theoretical basis of the model specification. In this section we seek to show that in many applications of spatial econometrics, considerable attention has been given to specifying the matrices  $\mathbf{W}$  and  $\mathbf{M}$  in a rational manner that attempts to represent as far as possible real social and economic processes. Hence we argue that in many cases the presence of a spatial lag, say  $\mathbf{W}\mathbf{Y}$ , is necessary because it does reflect a true interaction and is not simply a surrogate for some omitted variables. In fact, it appears that an absence of detailed consideration for the structure of  $\mathbf{W}$ , or indeed rejection of an approach based on  $\mathbf{W}$ , has come from analysts who are not particularly interested in the spatial processes *per se*, but see spatial dependence as a something of a nuisance which, yes, has to be allowed for in modelling but which is not the focal point of their research. In this case spatial error dependence can be treated in very general terms, say by Spatial Heteroscedastic Autocorrelation Consistent estimation or by common factor approaches.

We first focus on the very basic form of  $\mathbf{W}$  matrix, which has close links with time series analysis. In fact, it is easy to show that an autoregressive process in time has an equivalent to the  $\mathbf{W}$  matrix, as demonstrated in Fingleton (2009).

Consider the  $\mathbf{W}^*$  matrix, in which  $W_{jh}^* = 1$  if location pair  $i$  and  $j$  are close to each other in space, and  $W_{jh}^* = 0$  otherwise. By close, we typically mean contiguous. For time series we have a comparable contiguity matrix, let us call it  $\mathbf{H}$ . To see the near equivalence of  $\mathbf{H}$  and  $\mathbf{W}^*$ , consider a data generating process for  $T$  periods,

$$\begin{aligned} Y(t) &= \alpha Y(t-1) + \varepsilon(t) \\ \varepsilon(1) &= Y(1) = 0 \\ \alpha &< 1, \quad \varepsilon \sim N(0, \sigma^2), \quad t = 2, \dots, T \end{aligned} \tag{13}$$

This generates a stationary time series. In matrix terms an entirely equivalent data generating process is given by

$$\mathbf{Y} = \alpha \mathbf{H}\mathbf{Y} + \boldsymbol{\varepsilon} \tag{14}$$

in which  $\mathbf{Y}$  is a  $T \times 1$  vector,  $\alpha$  is a scalar parameter,  $\boldsymbol{\varepsilon}$  is a  $T \times 1$  vector of disturbances and  $\mathbf{H}$  is a  $T \times T$  matrix with row  $t$  designating the same time point as column  $t$  and 1s indicating time proximity. In spatial series,  $\mathbf{W}^*$  is  $N \times N$ , where  $N$  is the number of places or nodes on the network. Below we show a typical  $\mathbf{W}^*$  matrix for  $N = 10$ , and its time series counterpart  $\mathbf{H}$ .

$$\mathbf{H} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \quad \mathbf{W}^* = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 \end{bmatrix}$$

Figures 1-3 show small hypothetical networks which can be represented as simple binary-valued  $\mathbf{W}^*$  matrices.

$$\mathbf{W}_A^* = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \mathbf{W}_B^* = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \quad \mathbf{W}_C^* = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Typically in practice one would scale the matrix  $\mathbf{W}^*$ , say by the maximum eigenvalue, since this clearly indicates ‘allowable’ values that avoid singularities. One way to do this is by

$$\mathbf{W} = \frac{\mathbf{W}^*}{\max(eig)} \quad (15)$$

in which  $\max(eig)$  denotes the maximum eigenvalue of  $\mathbf{W}^*$ , where  $\mathbf{W}^*$  is such that real eigenvalues can be obtained. Using this normalization, the maximum eigenvalue of  $\mathbf{W}$  is 1, and the continuous range for  $\rho$  in which  $(\mathbf{I} - \rho\mathbf{W})$  is non-singular is  $\frac{1}{\min(eig)} < \rho < 1$ . An alternative to this, due to Ord (1975), is

$$\mathbf{W} = \mathbf{D}^{-0.5}\mathbf{W}^*\mathbf{D}^{-0.5} \quad (16)$$

in which the diagonal matrix  $\mathbf{D}$  takes values equal to the row sums of  $\mathbf{W}^*$ . Most applications in spatial econometrics scale the individual rows (or columns) of  $\mathbf{W}^*$  by the row totals, so that the  $\mathbf{W}$  rows sum to 1. In both these alternatives the same conditions apply for non-singularity. With  $\mathbf{W}$  thus defined, the spatial data generating process is

$$\mathbf{Y} = \rho\mathbf{W}\mathbf{Y} + \boldsymbol{\varepsilon} \quad (17)$$

which is identical to the time series data generating process, except that  $\mathbf{W}$  replaces  $\mathbf{H}$  and  $\rho$  replaces  $\alpha$ . As  $\rho \rightarrow 1$ , we approach near unit root spatial autoregressive processes, with similar consequences as in non-stationary time series.

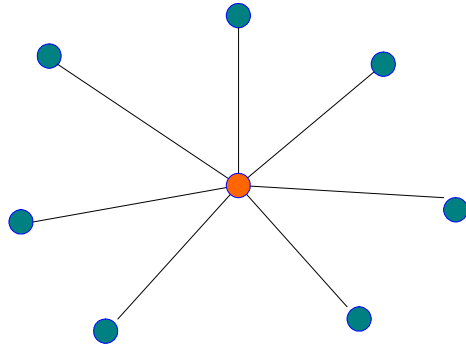


Figure 1: Star (A)

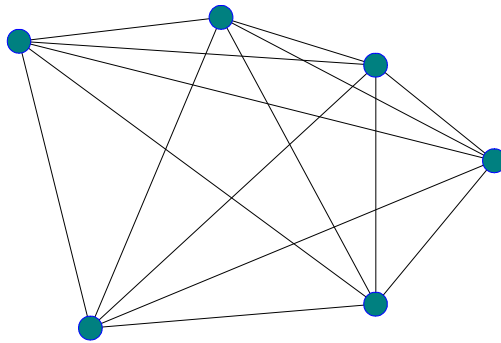


Figure 2: Complete Network (B)

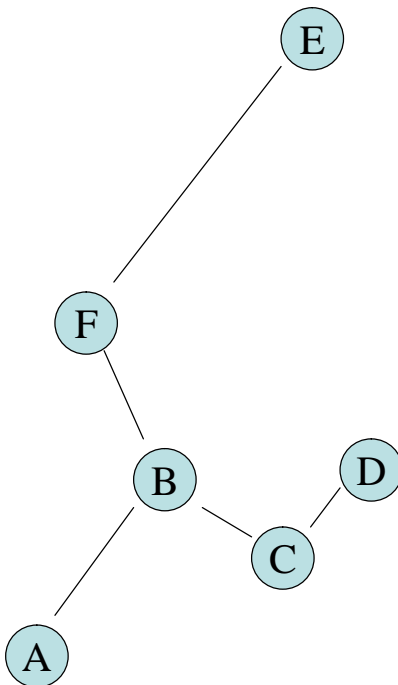


Figure 3: Nearest Neighbor (C)



While we have given some emphasis to the similarity of the DGP for autoregressive time series and spatial series, there are major dissimilarities on account of the multilateral nature of spatial interaction. In contrast, with time there are only two directions, forward and backwards. One of the contributions of Cliff and Ord (1973, 1981) was to extend the definition of the matrix  $\mathbf{W}$  to accommodate the distinctiveness of spatial processes. Spillovers have diverse origins, and therefore one would anticipate that the way to model them takes on various forms. For instance, they may be the outcome of network economics Goyal (2009), commuting,<sup>3</sup> migration, displaced demand and supply effects in the housing market, input-output linkages, competition and coordination between firms, localized information flows through social networks, strategic interaction between policy makers,<sup>4</sup> tax competition between local authorities, or even simply arbitrary boundaries causing spatial autocorrelation. We cannot deal with all of these cases, so in the following paragraphs we focus on a selection, commencing with the traditional distance-based unidimensional measures adopted in spatial econometrics (Anselin (1988)) 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.

Moving beyond crude measures of between-group spatial ‘distance’, such as the simple notions of proximity and contiguity, leads us to slightly more elaborate specifications which nonetheless are still based on the physical features of geographical units. For example Cliff and Ord (1973) combine distance and length of the common border between contiguous spatial units thus:

$$W_{jh} = (d_{jh})^a (\chi_{jh})^b \quad (18)$$

where  $d_{jh}$  denotes the distance between locations  $j$  and  $h$  and  $\chi_{jh}$  is the proportion of the boundary of  $j$  shared with  $h$  whereas  $a$  and  $b$  are parameters. More general distance measures include multidimensional indicator functions. For example, Bodson and Peters (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.:

$$W_{jh} = \sum_{j=1}^J p_j (a/1 + b \exp(-c_j d_{jh})) \quad (19)$$

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.

The above measures are less useful when the spatial interaction is determined by purely economic variables which may have little to do with spatial configuration of boundaries or geographical distance per se. This introduces the notion of economic distance, and developments in the conceptualization of economic distance have been surveyed in Greenhut, Norman, and Hung (1987). According to Fingleton and Le Gallo (2008) ‘the spillover between areas will not simply be a function of spatial propinquity, to the exclusion of other effects’ and ‘it is more realistic to base it on relative ‘economic distance’’. Big towns and cities are less remote than their geographical separation would imply, whereas very small locations are often isolated from one another’. Hence economic distance reflects the reduced

---

<sup>3</sup>For example, Holly, Pesaran, and Yamagata (2010) suggest a weighting matrix in a house price equation where connections between London and other UK regions are based on the inflow and outflow of commuters.

<sup>4</sup>Bhattacharjee and Holly (2006) find evidence of strategic spillovers across the members of the Bank of England’s Monetary Policy Committee in the way they vote on interest rate changes.

transaction costs associated with flows between geographically remote cities, which have better communications infrastructure, lower costs of information gathering and uncertainty, and similar economic and employment structures. Economic distance features in the work of 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.

Formulation of a  $\mathbf{W}$  matrix to reflect relative economic distance has been considered by, among others, Fingleton (2001), Fingleton (2008), LeSage and Pace (2008) and Fingleton and Le Gallo (2008). For example, consider the unstandardized matrix

$$\begin{aligned} W_{jh}^* &= Q_{h,0}^\alpha d_{j,h}^{-\gamma} & j \neq h \\ W_{jh}^* &= 0 & j = h \end{aligned} \quad (20)$$

in which  $Q_{h,0}$  is the level of output in economy  $h$  (at time 0) and  $d_{j,h}$  is a measure of geographical separation of locations  $j$  and  $h$ . There is no need to consider  $Q_{j,0}$  because it is nullified by the process of standardizing  $W^*$  by dividing by row totals to obtain  $\mathbf{W}$ . In the context to which this applies, the use of start of period values for  $Q$  excludes feedback from other model variables, thus ensuring the exogeneity of  $\mathbf{W}$ . The coefficients  $\alpha$  and  $\gamma$  reflect the weight attributed to  $Q_{h,0}$  and  $d_{j,h}$ , with  $\alpha = 0$  corresponding to a pure distance effect, and  $\gamma = 0$  corresponding to a pure economic size effect. These could be estimated alongside other model parameters, but because of the difficulty this would entail, it makes practical sense to assign values to these coefficients a priori.

Data are important in deciding  $\mathbf{W}$ , and since the spatial interaction we are attempting to model using  $\mathbf{W}$ , say as endogenous lag variable  $\mathbf{WY}$ , is typically in economics a spillover or externality, we conventionally look at where and to what extent spillovers are occurring. Typically, in the case of knowledge spillovers, the main sources have been input-output tables, patent concordances, innovation concordances, and proximity analysis. However this is sheer information, and we need to move closer to a theory of network emergence, dynamics and possible equilibrium conditions to have a more satisfying and coherent basis for  $\mathbf{W}$ .

In order to obtain a closer representation of the spatial interaction process in  $\mathbf{W}$  matrix construction choices, Anselin (2010) suggests greater focus on modelling agents involved in social and economic interaction. Looking back in this context, Patuelli, Reggiani, Gorman, Nijkamp, and Bade (2007) consider network interaction modelling with reference to earlier work on spatial interaction and discrete choice behavior such as Wilson (1967, 1970) a on entropy maximization and McFadden (1974, 1979) on the microeconomic basis of interaction models. Let us first consider  $\mathbf{W}$  as a representation of a network involving nodes (people or places) and links between nodes. These can be seen as dynamic evolving entities, and we can envisage network development to be a response to costs and benefits in being a node or a link on the network. Consequently some networks might be dynamic and ephemeral, some networks in a stable equilibrium, and some network slowly evolving. Following Goyal (2009), we envisage that ephemeral and dynamic networks occur when there are payoffs. This leads to a theory of network formation, thus ‘A game of network formation specifies a set of players, the link formation actions available to each player and the payoffs to each player from the networks that arise out of

individual linking decisions’, and ‘A network is said to be strategically stable or an equilibrium if there are no incentives for individual players (either acting alone or in groups) to form or delete links and thereby alter the network’ (Goyal (2009)). A quasi-stable network is akin to what is normally envisaged in the regional science literature, where typically a network will be a fixed or very slowly evolving one, as a consequence of major investment in transport infrastructure which define the inter-nodal links. Assume that network formation and evolution is a consequence of decisions by network providers (investors in infrastructures) on the one hand, and network users on the other. As a thought experiment, let us consider how a  $\mathbf{W}$  matrix might emerge and evolve. The network providers create, maintain or develop the network according to the profit generated, where profit equals revenue minus cost. The revenue comes from the number of network users and the prices they pay to use the network. The cost depends on the extent of the network (miles of railway to maintain for example) and may be divided between fixed costs and variable costs, which depend on network usage. The network users choose links on the network according to the level of utility they provide, with the choice of whether or not to use the network, and subsequently which network link to choose, modelled perhaps as a multilevel random utility model. With a poor network, which in a commuting sense might mean slow, unreliable and lengthy journeys, the level of utility will be low and users may prefer not to such use links. Consequently, usage and profits fall, although variable costs may also reduce. In such circumstances it seems that a poorly used network or link might fall into decay, although poor services may induce a reduction in prices, increase utility and usage, change profit levels, stimulate investment and revive the network. Evidently users and providers are involved in a strategic game, with a potential for equilibrium outcomes, and with dynamic changes to networks and user behavior a possibility. The emerging literature on endogenous network dynamics involving dynamic stochastic games of network formation could provide many insights into how the structure of the  $\mathbf{W}$  matrix can be placed on a more rational basis.

The potential for dynamic  $\mathbf{W}$  matrices poses some problems for estimation, given the assertion that  $\mathbf{W}$  is necessarily a fixed entity. While this may not be such an issue for cross-sectional approaches, where at a given snapshot in time this may be a reasonable approximation, with the extension of spatial econometrics to include panel data modeling it may be the case that  $\mathbf{W}$  is evolving, interacting with the regression variables. Such an endogenous interaction is implied by Anselin (2010), who remarks that ‘an endogenous spatial weights matrix would jointly determine who interacts (and why) and how that interaction affects the rest of the model. Much progress remains to be made...’. But we can have a dynamic  $W$  matrix as part of a simulation, with no consequence for estimation, as in Fingleton (2001).

## 4 DO WE NEED $\mathbf{W}$ ?

Let us now imagine that we have network dependence but we choose to ignore it. Assume that we have explanatory variables comprising exogenous  $\mathbf{X}$ s and an endogenous spatial lag  $\mathbf{W}\mathbf{Y}$ , so that the data generating process is

$$\begin{aligned}
 \mathbf{Y} &= \rho\mathbf{W}\mathbf{Y} + \mathbf{X}\mathbf{b} + \boldsymbol{\varepsilon} \\
 \mathbf{Y} &= (\mathbf{I} - \rho\mathbf{W})^{-1}(\mathbf{X}\mathbf{b} + \boldsymbol{\varepsilon}) \\
 \boldsymbol{\varepsilon} &\sim \mathbf{N}(\mathbf{0}, \sigma^2)
 \end{aligned}
 \tag{21}$$

and that  $X_{1,i} = 1$  for  $i = 1, \dots, N$ ,  $\mathbf{X}_2 = (\mathbf{I} - \rho_{x_1} \mathbf{W})^{-1} \mathbf{u}_1$ ,  $\mathbf{X}_3 = (\mathbf{I} - \rho_{x_2} \mathbf{W})^{-1} \mathbf{u}_2$ ,  $\mathbf{W}$  is a standardized  $N$  by  $N$  (Rook's case)<sup>5</sup> contiguity matrix,  $\mathbf{u}_1$  and  $\mathbf{u}_2$  are  $N$  by 1 vectors sampling from an  $N(0, 1)$  distribution, and  $\mathbf{I}$  is an  $N$  by 1 identity matrix. Vector  $\boldsymbol{\varepsilon}$  is  $N$  by 1 sampling from an  $N(0, \sigma^2 \mathbf{I})$  distribution. Also  $(\mathbf{I} - \rho \mathbf{W})$ ,  $(\mathbf{I} - \rho_{x_1} \mathbf{W})$  and  $(\mathbf{I} - \rho_{x_2} \mathbf{W})$  are non-singular,  $\rho = 0.25$ ,  $b_1 = 1$ ,  $b_2 = 8$ ,  $b_3 = 2$ ,  $\sigma^2 = 1$ , and  $N = 121$ .

Given these data, we generate  $\mathbf{Y}$  and estimate two models. One is the correctly specified model  $\mathbf{Y} = \rho \mathbf{W} \mathbf{Y} + \mathbf{X} \mathbf{b} + \boldsymbol{\varepsilon}$  estimated by maximum likelihood. The second is the (mis)specification  $\mathbf{Y} = \mathbf{X} \mathbf{b} + \boldsymbol{\varepsilon}$  estimated by OLS, which incorrectly assumes  $\rho = 0$ . Clearly, given spatial dependence in  $\mathbf{X}_2$  and  $\mathbf{X}_3$ , the OLS  $\mathbf{b}$  estimates will be biased, as is apparent from the 100 replications summarised in Table 1.

			ML		OLS	
	$b$	$\text{var}(b)$	mean $\hat{b}$	$\frac{\text{mean}\hat{b}-b}{\text{sqrt}(\text{var}(b))}$	mean $\hat{b}$	$\frac{\text{mean}\hat{b}-b}{\text{sqrt}(\text{var}(b))}$
$\rho_{x_1} = \rho_{x_2} = 0$						
$b_1$	1	0.0087	1.0121	0.1297	1.5522	5.9147
$b_2$	8	0.0085	7.9966	-0.0365	7.9995	-0.0049
$b_3$	2	0.0074	1.9948	-0.0600	2.0908	1.0536
$\rho_{x_1} = \rho_{x_2} = 0.5$						
$b_1$	1	0.0087	1.0019	0.0205	1.0729	0.7794
$b_2$	8	0.0081	8.0146	0.1630	8.8903	9.9187
$b_3$	2	0.0053	1.9948	-0.0711	2.3745	5.1567

Given that we often need  $\mathbf{W}$  to obtain unbiased estimates of  $\mathbf{b}$ , we also need it to obtain an unbiased measure of the true effect of a variable, which typically is not the same as  $\mathbf{b}$ , as emphasized by LeSage and Pace (2009). Given a SAR model of the form  $\mathbf{Y} = \rho \mathbf{W} \mathbf{Y} + \mathbf{X} \mathbf{b} + \boldsymbol{\varepsilon}$ , it is also the case that the interpretation of the effects on dependent variable  $\mathbf{Y}$  of a unit change in an exogenous variable  $\mathbf{X}_j$ , the derivative  $\frac{\partial \mathbf{Y}}{\partial \mathbf{X}_j}$  is not simply equal to the regression coefficient  $b_j$ . As pointed out by LeSage and Pace (2009), the true derivative also takes account of the spatial interdependencies and simultaneous feedback embodied in the model, leading to a total effect that differs somewhat (typically) from the regression coefficient estimate. It follows that

$$\frac{\partial \mathbf{Y}}{\partial \mathbf{X}_j} = (\mathbf{I} - \rho \mathbf{W})^{-1} \mathbf{I} b_j \quad (22)$$

in which  $\mathbf{I}$  is the  $N$  by  $N$  identity matrix and  $(\mathbf{I} - \rho \mathbf{W})^{-1} \mathbf{I} b_j$  is an asymmetric  $N$  by  $N$  matrix, so the derivative varies according to the cells of  $\mathbf{X}_j$  and  $\mathbf{Y}$  being considered. We can summarize these differentiated effects by their mean, which is

$$N^{-1} \sum_{ir} \frac{\partial \mathbf{Y}_i}{\partial \mathbf{X}_{rj}} = N^{-1} \boldsymbol{\iota}' (\mathbf{I} - \rho \mathbf{W})^{-1} \mathbf{I} b_j \boldsymbol{\iota} \quad (23)$$

In which  $\boldsymbol{\iota}$  is an  $N$  by 1 vector of 1s. This is the average total effect of a unit change in  $\mathbf{X}_j$ . Also we can partition the average total effect of a unit change in all cells of  $\mathbf{X}_j$  into a direct and an indirect

<sup>5</sup>Units are neighbors under Rook's criterion if they share the same borders.

component. The average direct effect of a unit change in  $\mathbf{X}_{rj}$  on  $\mathbf{Y}_r$  is given by the mean of the main diagonal of the matrix, hence

$$N^{-1} \sum_r^N \frac{\partial \mathbf{Y}_r}{\partial \mathbf{X}_{rj}} = N^{-1} \text{trace}[(\mathbf{I} - \rho \mathbf{W})^{-1} \mathbf{I} b_j] \quad (24)$$

This direct effect is somewhat different from  $b_j$  because it also allows for the fact that a change in  $\mathbf{X}_{rj}$  affects  $\mathbf{Y}_r$  which then affects  $\mathbf{Y}_s (s \neq r)$  and so on, cascading through all areas and coming back to produce an additional effect on  $\mathbf{Y}_r$ . The difference between the total effect and the direct effect is the average indirect effect of a variable. This is equal to the mean of the off-diagonal cells of the matrix  $(\mathbf{I} - \rho \mathbf{W})^{-1} \mathbf{I} b_j$ , hence

$$N^{-1} \sum_{r \neq s}^N \frac{\partial \mathbf{Y}_r}{\partial \mathbf{X}_{sj}} = N^{-1} (\iota' (\mathbf{I} - \rho \mathbf{W})^{-1} \mathbf{I} b_j \iota - \text{trace}[(\mathbf{I} - \rho \mathbf{W})^{-1} \mathbf{I} b_j]) \quad (25)$$

Table 2 gives the mean total effect of each of  $\mathbf{X}_2$  and  $\mathbf{X}_3$  for the small simulation with  $\rho_{x_1} = \rho_{x_2} = 0$ ,  $\rho_{x_1} = \rho_{x_2} = 0.5$  and with  $\rho_{x_1} = \rho_{x_2} = 0.95$ .

Table 2. Total, Direct and Indirect Effects (DGP with SAR Process).					
$\rho_{x_1} = \rho_{x_2} = 0$	OLS	ML	Total	Direct	Indirect
$b_1 = 1$	1.5795	0.9892	-	-	-
$b_2 = 8$	8.1616	8.0119	10.7411	8.1628	2.5784
$b_3 = 2$	2.0379	2.0109	2.6962	2.0488	0.6474
$\rho = 0.25$	-	0.2535	-	-	-
$\rho_{x_1} = \rho_{x_2} = 0.5$	OLS	ML	Total	Direct	Indirect
$b_1 = 1$	1.1855	0.9921	-	-	-
$b_2 = 8$	9.1596	7.9923	10.6748	8.1391	2.5357
$b_3 = 2$	2.3908	2.0036	2.6759	2.0404	0.6355
$\rho = 0.25$	-	0.2511	-	-	-
$\rho_{x_1} = \rho_{x_2} = 0.95$	OLS	ML	Total	Direct	Indirect
$b_1 = 1$	2.3287	1.0079	-	-	-
$b_2 = 8$	10.1432	8.0042	10.6645	8.1489	2.5156
$b_3 = 2$	2.4856	1.9998	2.6645	2.0359	0.6286
$\rho = 0.25$	-	0.2494	-	-	-

Consider next what happens if the true data generating process is

$$\begin{aligned} \mathbf{Y} &= \mathbf{X} \mathbf{b} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &\sim \mathbf{N}(\mathbf{0}, \sigma^2) \end{aligned} \quad (26)$$

and we (wrongly) fit the SAR model  $\mathbf{Y} = \rho \mathbf{W} \mathbf{Y} + \mathbf{X} \mathbf{b} + \boldsymbol{\varepsilon}$ . Will it be the case that the presence of  $\mathbf{W} \mathbf{Y}$  biases the  $\mathbf{b}$  estimates? This is a common criticism of spatial econometrics, that the significance

of the spatial lag is falsely interpreted as a true spatial spillover effect. Indeed too many spatial econometricians have been overenthusiastic in their adoption of the spatial lag without giving sufficient consideration to the theoretical rationale for the model specification. The consequences depend on the context. If the spatial lag  $\mathbf{WY}$  is simply an additional, unnecessary term in the model, then typically  $\rho \approx 0$ , which is of no significance. If however the SAR specification excludes  $\mathbf{X}_3 = ((\mathbf{I} - \rho_{x_2} \mathbf{W})^{-1} \mathbf{u}_2$ , then the outcome depends on whether  $\rho_{x_2} = 0$ . Most importantly, with  $\rho_{x_2} \neq 0$ , the inference that  $\rho \neq 0$  may be simply due to the fact that  $\rho_{x_2} \neq 0$ . On the other hand with  $\rho_{x_2} = 0$ , this means that  $\mathbf{X}_3$  is spatially random, and uncorrelated with the included variables, hence its absence does not affect the estimate obtained for  $\rho$ , in which case we should expect  $\rho \approx 0$ . Simply as an illustration, we generate data via  $\mathbf{Y} = \mathbf{X}\mathbf{b} + \boldsymbol{\varepsilon}$  reverting to  $\mathbf{X}_2 = ((\mathbf{I} - \rho_{x_1} \mathbf{W})^{-1} \mathbf{u}_1$  and  $\mathbf{X}_3 = ((\mathbf{I} - \rho_{x_2} \mathbf{W})^{-1} \mathbf{u}_2$ . With  $b_1 = 1$ ,  $b_2 = 8$ ,  $b_3 = 2$ ,  $\sigma^2 = 1$  and  $\rho_{x_1} = 0.5$ ,  $\rho_{x_2} = 0$ , our two estimating equations give the following mean estimates:

Table 3. Total, Direct and Indirect Effects (DGP without Spatial Effects)						
$\rho_{x_1} = 0.5$ $\rho_{x_2} = 0$	OLS	ML	Total	Direct	Indirect	
$b_1 = 1$	1.0041	1.1144	-	-	-	
$b_2 = 8$	8.0107	7.9017	8.0802	7.9035	0.1767	
$b_3 = 2$	2.0012	-	-	-	-	
$\rho = 0$	-	0.0219	-	-	-	
$\rho_{x_1} = 0.5$ $\rho_{x_2} = 0.95$	OLS	ML	Total	Direct	Indirect	
$b_1 = 1$	0.9910	0.6611	-	-	-	
$b_2 = 8$	8.0193	6.1322	12.6192	6.6749	5.9443	
$b_3 = 2$	1.9968	-	-	-	-	
$\rho = 0$	-	0.5137	-	-	-	

Incorrectly omitting  $\mathbf{X}_3$  when it is spatially dependent as a result of setting  $\rho_{x_2} = 0.95$  has a significant biasing effect on the ML estimates. In particular the spatial lag is evidently picking up the effect of the spatially dependent omitted variable. With regard to the  $t$  ratios for  $\rho$ , the mean of 100 replications is 7.85. Likewise the outcome is a biased estimate of  $b_2$ . Importantly, observe also that the total, direct and indirect effects of a variable will be incorrect when the specification wrongly includes the endogenous spatial lag. For example, in Table 3, the true effect of  $\mathbf{X}_2$  is given by  $b_2 = 8$ .

However it turns out that we just might be able to use  $\mathbf{W}$  to mitigate the bias arising from the omission of spatially dependent regressors, for it has been shown by LeSage and Pace (2008) and Pace and LeSage (2008) that fitting the so-called spatial Durbin model,

$$\mathbf{Y} = \rho \mathbf{WY} + \mathbf{Xb} + \mathbf{WX}\boldsymbol{\rho}_x + \boldsymbol{\varepsilon} \quad (27)$$

eliminates coefficient estimate bias, but this solution rests on the assumption that  $\mathbf{W}$  is the correct one for the omitted variable spatial autoregressive process. This is a topic which is explored and the analysis extended in Fingleton and Le Gallo (2009), who find via Monte Carlo simulation that when the omitted variable does not equate to a spatial autoregressive process in  $\mathbf{W}$ , an augmented spatial

Durbin specification, augmented to also includes an autoregressive error dependence process, produces biased estimates, but ones that are less biased than those obtained by ignoring the existence of omitted variables.

While we often need  $\mathbf{W}$ , sometimes its presence is unnecessary and can be misleading. A second note of caution also suggesting moderation of the emphasis by LeSage and Pace (2009) on  $\mathbf{W}$  leading to total, direct and indirect effects, as the proper interpretation of the impact of exogenous variables in the presence of a spatial lag, comes from the specification

$$\mathbf{Y} = \rho\mathbf{W}\mathbf{Y} + b_1(\mathbf{I} - \rho\mathbf{W})\mathbf{X}_1 + b_2\mathbf{X}_2 + \varepsilon \quad (28)$$

Following equation (22),  $\frac{\partial\mathbf{Y}}{\partial\mathbf{X}_2} = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{I}b_2$ , but equation (22) does not apply with regard to  $\mathbf{X}_1$ , and instead  $\frac{\partial\mathbf{Y}}{\partial\mathbf{X}_1} = b_1$ . This type of specification has been suggested by Fingleton (2003), Fingleton (2006) and Barde (2010) as the reduced form resulting from the existence of an ancillary SAR process.

In the following example, log labour efficiency ( $\ln \mathbf{A}$ ) is assumed to depend on local exogenous variables embodied in the  $N$  by  $k$  matrix  $\mathbf{X}$ , on log labour efficiency in ‘nearby’ areas ( $\mathbf{W} \ln \mathbf{A}$ ), and on random disturbances ( $\boldsymbol{\xi}$ ), hence  $\ln \mathbf{A} = \mathbf{X}\mathbf{b} + \rho\mathbf{W} \ln \mathbf{A} + \boldsymbol{\xi}$ ,  $\boldsymbol{\xi} \sim N(0, \boldsymbol{\Omega}^2)$ . It is convenient to specify this with the exogenous variables on the right hand side, hence  $\ln \mathbf{A} = (\mathbf{I} - \rho\mathbf{W})^{-1}(\mathbf{X}\mathbf{b} + \boldsymbol{\xi})$ . Starting from an explicit economic theory with microfoundations, they assume wages  $\mathbf{w}$  depend on employment density  $\mathbf{E}$  and labour efficiency  $\mathbf{A}$  in each area  $j$ ,  $j = 1, \dots, N$ , thus  $\ln w_j = k_1 + (\gamma - 1) \ln E_j + (\gamma - 1) \ln A_j$ . Substituting and rearranging obtains

$$\ln \mathbf{w} = \rho\mathbf{W} \ln \mathbf{w} + (\mathbf{I} - \rho\mathbf{W})k_1\boldsymbol{\iota} + (\gamma - 1)(\mathbf{I} - \rho\mathbf{W}) \ln \mathbf{E} + (\gamma - 1)\mathbf{X}\mathbf{b} + (\gamma - 1)\boldsymbol{\xi} \quad (29)$$

It is apparent that the partial derivative  $\frac{\partial \ln \mathbf{w}}{\partial \ln \mathbf{E}}$  is simply equal to  $(\gamma - 1)$ , so despite the existence of the spatial lag  $\mathbf{W} \ln \mathbf{w}$  in our model, there is no need to interpret the effect of this variable any differently from the normal interpretation.

This then leads us to the problem of inference and interpretation in a spatial econometric model that is driven by underlying economic theory, as is the case in Fingleton (2003), Fingleton (2006) and Barde (2010), compared with the inference and interpretation one would associate with a model specification that is governed entirely by empirical analysis. It is apparent that we could obtain misleading results if empirical analysis suggests a model specification that is contrary to the true specification.

Consider the following simple example, in which the DGP is the above model, but instead the following spatial Durbin specification is fitted to the data

$$\ln \mathbf{w} = \rho\mathbf{W} \ln \mathbf{w} + b_0\boldsymbol{\iota} + b_1 \ln \mathbf{E} + b_2\mathbf{W} \ln \mathbf{E} + \mathbf{X}\mathbf{c} + \mathbf{W}\mathbf{X}\mathbf{d} + \boldsymbol{\nu} \quad (30)$$

From this starting point, our best-fitting model would probably be a constrained version of this specification, but without knowledge of the underlying theory driving the DGP we would never consider the true specification among the set of optional models, and come to a false interpretation of the effects of the variables.

To illustrate this, consider the DGP based on an 11 by 11 lattice giving  $N = 121$  observations of variables  $\mathbf{w}$ ,  $\mathbf{E}$  and  $\mathbf{X}$ , with the 121 by 121  $\mathbf{W}$  matrix comprising a matrix of 1s and 0s according to the Rook’s contiguity criterion, subsequently standardize to row totals of 1. The values of  $\mathbf{E}$  and  $\mathbf{X}$  are respectively  $N$  by 1 and  $N$  by 2 matrices of pseudorandom numbers drawn from the standard uniform distribution on the open interval  $(0, 1)$  excluding 0s. Given  $k_1 = 1$ ,  $\gamma = 1.25$ ,  $\rho = 0.15$ ,  $c_1 = 8$ ,  $c_2 = 2$ , and  $(\gamma - 1)^2\boldsymbol{\Omega}^2 = (\gamma - 1)^2$ , we generate  $\ln \mathbf{w}$  via

$$\begin{aligned} \ln \mathbf{w} &= k_1 \boldsymbol{\iota} + (\gamma - 1) \ln \mathbf{E} + (\mathbf{I} - \rho \mathbf{W})^{-1} (\mathbf{X} \mathbf{c} + \boldsymbol{\Psi}) \\ \boldsymbol{\Psi} &\sim N(\mathbf{0}, (\gamma - 1)^2 \boldsymbol{\Omega}^2) \end{aligned} \quad (31)$$

Typical outcomes of this DGP and model fitting exercise are given in Table 4. We also eliminate the insignificant spatial lags (Model B).

	Model A			Model B		
	Coefficient	<i>t</i> Asymptotic	<i>z</i>	Coefficient	<i>t</i> Asymptotic	<i>z</i>
$b_0$	1.689	5.670	0.000	1.817	11.163	0.000
$b_1(\ln \mathbf{E})$	0.245	7.614	0.000	0.240	7.665	0.000
$b_2(\mathbf{W} \ln \mathbf{E})$	-0.022	-0.326	0.743	-	-	-
$c_1(x_1)$	7.993	633.3	0.000	7.999	846.2	0.000
$c_2(x_2)$	2.005	221.2	0.000	2.006	236.1	0.000
$d_1(\mathbf{W} x_1)$	-0.153	-0.709	0.477	-	-	-
$d_2(\mathbf{W} x_2)$	-0.036	-0.638	0.522	-	-	-
$\rho(\mathbf{W} \ln w)$	0.167	6.334	0.000	0.149	67.81	0.000
$R^2$	0.999			0.999		
$\bar{R}^2$	0.999			0.999		
$\sigma^2$	0.076			0.077		
$N$	121			121		
$ll$	25.14			24.88		

These model estimates closely approximate those assumed in the DGP, but they give a misleading indication of the true effect of the variables. Using the equations given above, we would infer that the average total effect of  $\ln \mathbf{E}$  is  $N^{-1} \boldsymbol{\iota}' (\mathbf{I} - \rho \mathbf{W})^{-1} \mathbf{I} b_1 \boldsymbol{\iota} = 0.3387$  compared to the true value of 0.25.

## 5 CHOOSING $\mathbf{W}$ EMPIRICALLY

Consider next the question of which matrix  $\mathbf{W}$  should be chosen given the obvious scope for numerous competing weights matrices. While theory will in the best practice cases drive the structure of  $\mathbf{W}$ , it nevertheless is true that there are a number of degrees of freedom in the exact  $\mathbf{W}$  specification, for example does one adopt a negative exponential decay or a power function for the distance term in equation 20. Harris, Moffat, and Kravtsova (2010) review some alternative approaches to constructing  $\mathbf{W}$ , such as trawling through data, perhaps designing  $\mathbf{W}$  around the residuals from a first stage regression, but this is atheoretical.

Burridge and Fingleton (2010) set out the history of the problem, commencing with Anselin (1986) who considers the simple model,

$$Y_i = \alpha + \rho \sum_{j=1}^N W_{ij} Y_j + u_i \quad (32)$$



where the  $N \times N$  weight matrix,  $\mathbf{W}$ , has three different forms,  $\mathbf{W}_A$ ,  $\mathbf{W}_B$  and  $\mathbf{W}_C$ . Taking  $\mathbf{W}_A$  to be the null hypothesis, Anselin considers  $J$ -type statistics in order to discriminate between these alternatives, obtained by augmenting (32) by additional explanatory variables equal to the fitted values from the model with weights  $\mathbf{W}_B$  or from the model with weights  $\mathbf{W}_C$ .

To further illustrate this, we follow Kelejian (2008) and Burridge and Fingleton (2010) and consider the more elaborate SARAR model, in which the choice of the  $\mathbf{W}$  matrix is accompanied by the question of which matrix  $\mathbf{M}$  to adopt, the latter defining the autoregressive spatial error dependence in the SARAR(1,1) model, thus

$$\begin{aligned}\mathbf{Y} &= \mathbf{X}_0\mathbf{b}_0 + \rho_0\mathbf{W}_0\mathbf{Y} + \mathbf{u}_0 \\ \mathbf{u}_0 &= \lambda_0\mathbf{M}_0\mathbf{u}_0 + \mathbf{v}_0.\end{aligned}\tag{33}$$

Here, the  $N \times k_0$  matrix of exogenous variables,  $\mathbf{X}_0$ , and the  $N \times 1$  vector for the dependent variable,  $\mathbf{Y}$ , are each measured without error, the two  $N \times N$  weight matrices,  $\mathbf{W}_0$  and  $\mathbf{M}_0$  are fixed *a priori*, and the unobserved shock vector,  $\mathbf{v}_0 \sim iid(0, \sigma_0^2\mathbf{I}_N)$  is independent of the exogenous regressors,  $\mathbf{X}_0$ . The parameters to be estimated are the slope coefficients,  $\mathbf{b}_0$ , the spatial lag and error coefficients,  $\rho_0$  and  $\lambda_0$ , and the variance,  $\sigma_0^2$ . The suffix 0 denotes that this specification is one of (at least) two competing non-nested hypotheses. Under the alternative, the data are generated by a similar structure, hence

$$\begin{aligned}\mathbf{Y} &= \mathbf{X}_1\mathbf{b}_1 + \rho_1\mathbf{W}_1\mathbf{Y} + \mathbf{u}_1 \\ \mathbf{u}_1 &= \lambda_1\mathbf{M}_1\mathbf{u}_1 + \mathbf{v}_1.\end{aligned}\tag{34}$$

Kelejian (2008) considers the tests of these competing models, extending the problem by allowing  $> 2$  non-nested alternatives. Among the hypotheses that can be tested, Burridge and Fingleton (2010) assume the explanatory variables,  $\mathbf{X}_0$  and  $\mathbf{X}_1$  are the same in the two models, but the spatial structures differ, so that  $\mathbf{W}_1 \neq \mathbf{W}_0$  and  $\mathbf{M}_1 \neq \mathbf{M}_0$ , but for simplicity they set  $\mathbf{M}_0 = \mathbf{W}_0$  and  $\mathbf{M}_1 = \mathbf{W}_1 \neq \mathbf{W}_0$ .

One alternative to the  $J$ -type statistics is to use an information criterion, thus avoiding several model comparisons, as suggested by Leenders (2002). However ‘.unfortunately different information criteria will in general lead to the selection of different models, so that the uniqueness of the chosen model relies on the investigator first selecting which criterion to adopt’ (Burridge and Fingleton (2010)).

Related approaches use Bayesian model averaging (LeSage and Parent (2007), LeSage and Fischer (2008)). While parameter uncertainty is well known, model uncertainty involving the unknown true structure of the  $\mathbf{W}$  matrix is less well explored. However, finding the true  $\mathbf{W}$  may involve searching through a very large number of competing specifications, which may include the true specification, rather than being decided on theoretical grounds.

## 6 ALTERNATIVES TO $\mathbf{W}$

It is evident that the specification of  $\mathbf{W}$  is fundamental and that for some the problems this poses, either from the perspective of logic, theory or empirics, poses a significant hurdle (McMillen (2010a)). Some have advocated rejecting the traditional a priori fixed  $\mathbf{W}$  matrix approach in favour of potentially less problematic ways of introducing spatial interaction in spatial econometric models (Harris, Moffat, and Kravtsova (2010), Folmer and Oud (2008)). Thus instead of constructing  $\mathbf{W}$ , they suggest directly entering variables in the regression model that proxy spillovers (Harris, Moffat, and Kravtsova (2010)). This is the approach adopted by Paci and Usai (2009). As Harris, Moffat, and Kravtsova (2010)

observe, ‘the essential difference between this and the standard approach using  $\mathbf{W}$  is that spillovers are not entered through the interaction between regions of the dependent or other (state) variables in the model, weighted by  $\mathbf{W}$ , but rather through constructing “stand alone” proxies for spatial spillovers’. However we argue that such an approach itself requires strong identifying assumptions and therefore possesses no real advantage compared to employing a  $\mathbf{W}$  matrix. In other words, this type of approach also typically involves some form of *a priori* variable selection and weighting usually in the form of a geographical proximity measure, and so does not really represent a complete departure from the  $\mathbf{W}$  matrix approach, and even if weighting can be avoided it introduces additional complexities of (arbitrary) variable definition.

With time series and panel data, we have more scope for a more refined approach. Seemingly unrelated regression (SUR) probably provides the most complete break from the  $\mathbf{W}$  matrix approach, because it estimates the error variance-covariance matrix on the basis of location-specific time series, thus, according to Anselin (1988), ‘the spatial dependence is not expressed in terms of a particular parameterized function, but left unspecified as a general covariance’, and, ‘in spatial econometrics, this model has been suggested as an alternative to the use of spatial weights’. Other time-series related alternatives in the form of vector autoregressions, which attempt to pick up spatial interaction via the presence of (any number of) lagged variables from ‘neighboring’ regions are an interesting alternative, but with a large number of regions ultimately collapse under the weight of a large number of parameters to estimate and interpret. However these problems may not always be fatal. One way out of this problem is via the Bayesian approach of LeSage and Krivelyova (1999), and Chang and Coulson (2001) successfully use a structural vector autoregression to model spatial spillovers. Another approach that has been advocated (Kelejian and Prucha (2007)) is spatial non-parametric heteroscedasticity and autocorrelation consistent estimation (SHAC). This gives consistent estimates of the error covariance matrix under rather general assumptions, allowing various patterns of correlation and heteroscedasticity, including spatial ARMA( $p, q$ ) errors. Kelejian and Prucha (2007) assume that the disturbance vector  $\mathbf{e}$  is:

$$\mathbf{e} = \mathbf{R}\boldsymbol{\xi}$$

where  $\mathbf{R}$  is a non-stochastic matrix with unknown elements. The asymptotic distribution of IV estimators implies that the variance-covariance matrix is

$$\boldsymbol{\Psi} = n^{-1}\tilde{\mathbf{z}}'\boldsymbol{\Sigma}\tilde{\mathbf{z}}$$

in which  $\boldsymbol{\Sigma} = \{\sigma_{ij}\}$  is the variance-covariance matrix of  $\mathbf{e}$  and  $\tilde{\mathbf{z}}$  is a full column rank matrix of instruments. The ( $r, s$ )th element is:

$$\Psi_{rs} = n^{-1} \sum_{i=1}^n \sum_{j=1}^n \tilde{z}_{ir} \tilde{z}_{js} \hat{e}_i \hat{e}_j K(d_{ij}/d_n)$$

where  $\hat{e}_i$  is the IV residual for observation  $i$ ,  $d_{ij}$  is the distance between locations  $i$  and  $j$ ,  $d_n$  is the bandwidth and  $K(\cdot)$  is a kernel function. Among the alternatives, we might opt for the Parzen kernel as given by Andrews (1991).

$$K(x) = \left\{ \begin{array}{ll} 1 - 6x^2 + 6|x|^3 & \text{for } 0 \leq |x| \leq \frac{1}{2} \\ 2(1 - |x|)^3 & \text{for } \frac{1}{2} < |x| \leq 1 \\ 0 & \text{otherwise} \end{array} \right\}$$

From this it is evident that this approach is not assumption-free. Alternative kernels, such as the Bartlett, Tukey-Hanning and Quadratic spectral kernels, each put different weights on the lagged covariances. Additionally, different bandwidth or lag truncation parameter options exist. This is also the case of semi-parametric approach proposed by McMillen (2010b), which uses a Cubic kernel transformation of the covariate,  $\mathbf{X}$ , as a function of geographical distance,  $\mathbf{f}(\mathbf{X}(\mathbf{d}))$ , in a relationship which suffers from missing variables and incorrect functional form. Smoothing over space adds a variable that is correlated with the omitted variable, which is also correlated with space, and so adds significant explanatory power to the model. In addition the use of the nonlinear spline function resolves the functional form misspecification. In practice these choices are essentially data driven, and the choices made drive the performance of both (S)HAC and semiparametric estimators. Evidently SHAC and semiparametric estimators provide no obvious advantage to the flexibility inherent in a data-driven, potentially asymmetric,  $\mathbf{W}$  matrix approach, which of course may be applied (perhaps using different  $\mathbf{W}$ s) not only to the error process, but also to endogenous and exogenous variables. Moreover, rather than neutralizing spatial dependence as a nuisance phenomenon, where spatial econometrics takes a lead is in its ability to identify and test theory relating to explicit spatial dependence mechanisms, as embodied in the parameterization of a  $\mathbf{W}$  matrix. Of course this demands a specific functional form, and values assigned to parameters, but careful data analysis may help to identify the most appropriate specification, in ways that parallel how optimal HAC estimators are obtained.

It is apparent that while alternatives have been advocated which attempt to model spatial dependence, ultimately their application also calls for some simplifying and operational assumptions, so that they do not, except for SUR modelling, represent a complete break from what is also required by the traditional  $\mathbf{W}$  matrix approach. However, there is one outstanding approach to modeling spatial effects which has had very little attention in the spatial econometrics literature, namely hierarchical modeling. At first glance this seems to also be a way to allow spatial effects without having to resort to a  $\mathbf{W}$  matrix as a convenient and practicable option. However, as we demonstrate for the first time, the  $\mathbf{W}$  matrix is also embodied within this approach.

## 7 HIERARCHICAL MODELS

Hierarchical models are becoming increasingly popular across the range of the social sciences, as researchers come to appreciate that observed outcomes depend on variables organized in a nested hierarchy. We see many applications of multilevel modelling in educational research where there exist a number of well defined groups organized within a hierarchical structure. 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.

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 neighbors 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 downward bias and to erroneous conclusions of statistical significance.

One way of incorporating the group effect in a multilevel framework is to evaluate the impact of higher level variables which measure one or more aspects of the composition of the group to which an 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 advantageously from the use of dummy variables which capture the net effect of several omitted 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 a 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 neighboring 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.

Figure 4 is a diagrammatic representation of a hierarchical structure, with  $r$  denoting the top level,  $g$  the second level, and  $I$  the individual level. There is a varying 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 regions ( $r$ ) each of which nests 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 varying 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 covariates in  $X_i$  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 by the use of dummy variables. However, a large number of levels result in a dramatic reduction in degrees of freedom. Second, a multilevel approach helps us to analyze the effect of heterogeneous groups in the small sample situation. In fact with unbalanced data, while OLS estimates of the coefficients give equal weights to each cluster, the preferred 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 individuals in each group (this is technically referred to as shrinkage). In most applications, shrinkage

is desirable so that clusters that provide little information have little influence in estimation. The hierarchical multilevel method, which is sample size dependent, seems to have a distinct edge over other methods in eliminating bias.

Compared to other approaches such as Clustered-Standard-Error OLS, Hierarchical Modelling (HLM) has some advantages: first, while CSE techniques treat the random variation as a simple nuisance, the objective of HLM is to estimate and decompose the total random variation in an individual component and a group component. Second, while CSE only adjusts standard errors for non-independence, HLM provides us with estimates of the variance components at each level and these affect point estimates also. 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)).

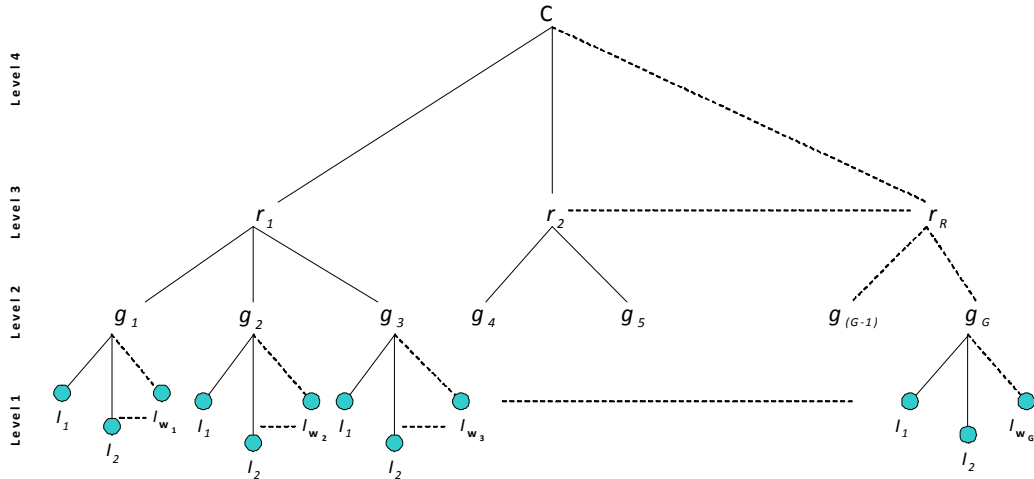


Figure 4: A Hierarchical Structure

## 8 MODELLING SPATIAL EFFECTS THROUGH THE ERROR TERM

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 ( $\mathbf{X}$ ) such as educational attainment, might vary if they reside in different areas. Part of the reason could be the effect of, say, fixed level two contextual factors ( $\mathbf{Z}$ ), and partly because of level two specific random effects  $\{u_j\}$ . However these fixed and random effects, while jointly accounting for heterogeneity across residential areas, are not spatially correlated, a topic we address subsequently. With this in mind, the specification

of a multilevel model is:

$$\mathbf{Y} = \boldsymbol{\beta}_0 + \mathbf{X}\boldsymbol{\beta}_1 + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \quad (35)$$

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. The additive error term  $\boldsymbol{\varepsilon}$  is composed of an idiosyncratic random error term  $e_{ij}$  for the  $i^{th}$  unit belonging to level  $j$  and a random effect  $u_j$  accounting for some level 2 heterogeneity. We make the following assumptions:

$$\begin{aligned} e_{ij}|X_{ij} &\sim N(0, \sigma_e) & \text{Cov}(e_{ij}, e_{i'j}) &= 0, \forall i \neq i' \\ u_j|X_{ij} &\sim N(0, \sigma_u) & \text{Cov}(u_j, e_{ij}) &= 0 \end{aligned} \quad (36)$$

If we let  $\sigma_e^2$  ( $\sigma_u^2$ ) denote the variance of  $e_{ij}$  ( $u_j$ ) such that for  $\text{cov}(e_{ij}, u_j) = 0$ , 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, the (equicorrelated) intra-class correlation is:

$$\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}. \quad (37)$$

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.

In order to accommodate spatial effects operating via the error term we can rewrite the composite error term as:

$$\varepsilon_{ij} = e_{ij} + \sum_{j \neq h} u_j W_{jh}, \quad (38)$$

where  $\mathbf{W} = \{W_{jh}\}$  allows us to specify the way neighboring 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:

$$\rho = \frac{\sigma_u^2}{\sum_{j \neq h} \sigma_u^2 W_{jh}^2 + \sigma_e^2}. \quad (39)$$

Trivially if  $\mathbf{W} = \mathbf{I}$ , where  $\mathbf{I}$  is the identity matrix, then we have the standard random effects model ignoring any between group effect.

If  $u_j$  are treated as fixed effects then we need to assume that  $\text{cov}(e_{ij}, u_j) = \sigma_{eu} = 0$ , 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 Generalized Least Squares (GLS) estimator would be biased and inconsistent. If  $u_j$  are random effects we also assume independence between these and the covariates such that  $\text{cov}(X_{ij}, u_j) = \sigma_{ux} = 0$  (Blundell and Windmeijer (1997)).

If we first rewrite (35) in compact form as:

$$\mathbf{Y} = \mathbf{J}\boldsymbol{\lambda} + \boldsymbol{\theta}_\varepsilon \boldsymbol{\varepsilon} \quad (40)$$

where  $\mathbf{J}$  is  $(N \times (k + q + 1))$  and  $\boldsymbol{\lambda}$  is  $((k + q + 1) \times 1)$  and  $\boldsymbol{\theta}_\varepsilon$  is the design matrix of the random parameters that is used in the estimation to derive the estimates for  $\hat{\sigma}_e^2$  and  $\hat{\sigma}_u^2$ . 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),<sup>6</sup> is based upon an Iterative Least Squares (IGLS) method that results in consistent and asymptotically efficient estimates of  $\boldsymbol{\lambda}$ .

<sup>6</sup>The method is currently implemented in the software MLwiN.

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 via RIGLS/REML 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/REML estimates don't.

## 9 LINKING SPATIAL MODELS TO HIERARCHICAL MODELS WITH INTERACTION EFFECTS

With the emergence of interaction-based models (Manski (2000), Brock and Durlauf (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 and Cho (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 and Durlauf (2001)). A typical example is the emergence of social networks where it is often observed that people belonging to the same group tend to behave similarly (Manski (2000)). The propensity that a person behaves in a certain way varies positively with the dominant behavior in the group (Bernheim (1994), Kandori (1992)).<sup>7</sup>

We consider how multilevel models can be linked to spatial models and generalize them to incorporate more general forms of network dependence involving individuals belonging to the same group. We start by assuming a specific form of spatial dependence where the dependent variable,  $\mathbf{Y}$ , depends on its spatial lag as in traditional spatial autoregressive (SAR) models:

$$\mathbf{Y} = \beta_0 + \rho \mathbf{W}\mathbf{Y} + \mathbf{X}\beta_1 + \mathbf{Z}\gamma + \mathbf{u} + \mathbf{e} \quad (41)$$

where  $\mathbf{W}$  is an  $N \times N$  matrix with  $G$  groups/areas each containing  $w_j$  individuals so that  $\sum_{j=1}^G w_j = N$  and  $\mathbf{Z}$  is a  $(N \times q)$  matrix of contextual variables defining group characteristics. Let's start by assuming that  $\mathbf{W}$  is a block matrix:

$$\begin{aligned} \mathbf{W} &= \text{Diag}(\mathbf{W}_1, \dots, \mathbf{W}_G) \\ \mathbf{W}_j &= \frac{1}{w_j} (\mathbf{1}_{w_j} \mathbf{1}'_{w_j}) \quad j = 1, \dots, G \end{aligned} \quad (42)$$

where  $\mathbf{1}_{w_j}$  is the  $w_j$ -dimensional column vector of ones. Elements in matrix  $\mathbf{W}$  indicate that individuals within a group are affected by the (average) behavior of other individuals residing in the same location, in other words by other members of the same group. Figure 5 depicts a situation resembling a complete network, as described in Section 3, where each unit  $I_{ij}$   $i = 1, \dots, w_j$   $j = 1, \dots, G$  interacts in the same way with all the other units in the group.

---

<sup>7</sup>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 and Fisher (1986); Krosnick and Judd (1982); Jones (1994)).

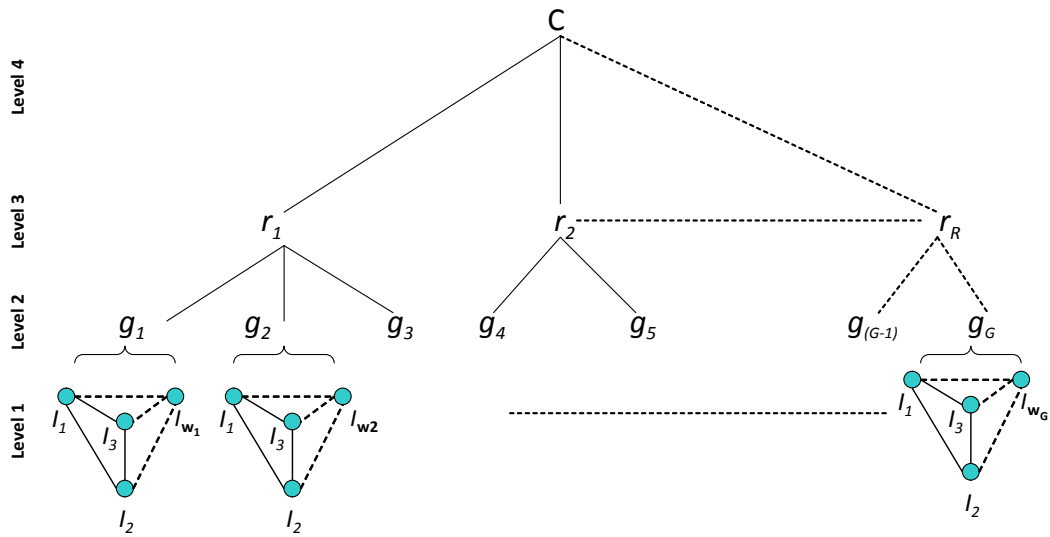


Figure 5: A Hierarchical Structure: Complete Network within Groups

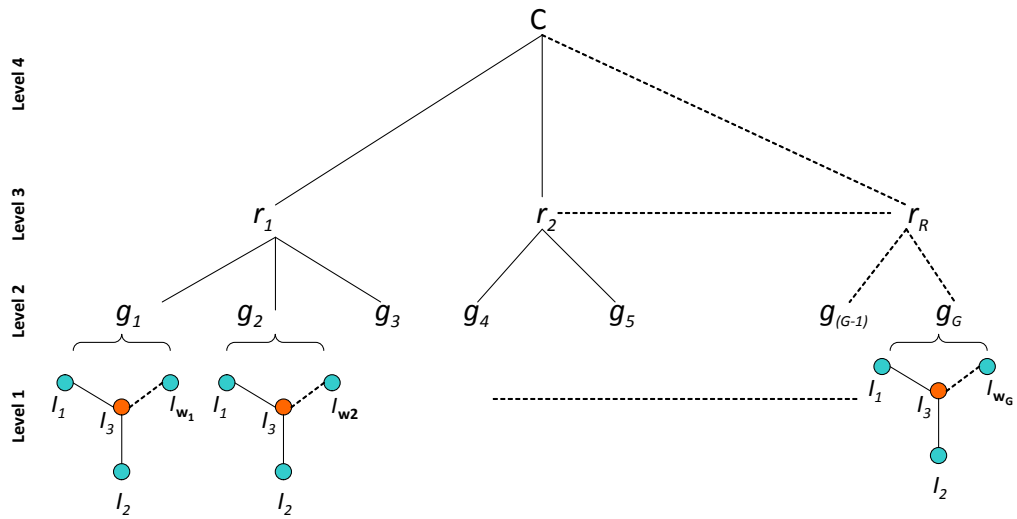


Figure 6: A Hierarchical Structure: 'Star' within Groups



We can envisage many different form of interactions among members of the same groups. Figure 6 illustrates a situation where individuals in the same group are affected by individual  $I_3$  without influencing each other. We can clearly generalize this analysis to other forms of network interactions, such as the nearest neighbor scenario introduced in Section 3 and described by the weight matrix  $\mathbf{W}^C$ , in which case the block diagonal matrix is:

$$\mathbf{W}^C = \text{Diag}(\mathbf{W}_1^C, \mathbf{W}_1^C, \dots, \mathbf{W}_G^C)$$

Finally we could also introduce different type of network structures across groups, further generalizing the structure of the block diagonal matrix  $\mathbf{W}^C$ .

To illustrate the implication of group dependence for spatial models we focus on the hierarchical complete network structure shown in Figure 5. The specification (41) together with the assumption of group dependence implied by (42) allow us to examine the particular case of a hierarchical model where, for ease of exposition, we assume for the moment that individuals (firms etc) within the same group are affected in the same way by all the other members of the group.

Focussing on within-group effects we therefore assume that  $\bar{Y}_j = \frac{1}{w_j} \sum_{i=1}^{w_j} Y_{ij}$  so that each individual 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. We will consider the implication of group spillovers in a hierarchical setting in the following section and show how introducing such externalities between groups is crucial for the identification of network interaction effects.

With this assumption, we can rewrite (41) as:

$$Y_{ij} = \beta_0 + \rho_1 \bar{Y}_j + \beta_1 X_{ij} + \gamma Z_j + e_{ij} + u_j \quad (43)$$

Following Manski (1993),  $\beta_1$  gives the effect of individual level characteristics  $X_{ij}$ ,  $\rho_1$  captures the strength of endogenous group effects  $\bar{Y}_j$ ,  $\gamma$  quantifies the exogenous or contextual effect  $Z_j$ ,  $u_j$  are random group effects and  $e_{ij}$  is an individual specific random component capturing other unmodelled sources of variation in  $Y_{ij}$ . The problem here is the identification of the parameters in the presence of what Manski calls reflection, something which we address below. Of course, identification only becomes a problem in linear-in-means models, and any nonlinearity, for instance as in an expanded spatial Durbin model (see Gibbons and Overman (2010)), automatically solves the problem.

As an example, consider workers within companies, with wages  $Y_{ij}$  dependent on individual worker attributes,  $X_{ij}$ , and on company level contextual effects  $Z_j$  (such as sector, company wages policy, level of research and development activity, investment etc.). In addition other unmeasured causes of individual wage variation are represented by random group (company) effects  $u_j$  and by individual random effects  $e_{ij}$ , and with  $\rho_1 \neq 0$  wages may also be endogenously determined so that a higher wage level achieved by one worker spills over (via  $\bar{Y}_j$ ) to other workers in the firm.

Taking group means of both sides of (43) and solving for  $\bar{Y}_j$  (assuming  $\rho_1 \neq 1$ ) results in the between group regression:

$$\bar{Y}_j = \frac{\beta_0}{1 - \rho_1} + \frac{\beta_1}{1 - \rho_1} \bar{X}_j + \frac{\gamma}{1 - \rho_1} Z_j + u'_j \quad (44)$$

where  $u'_j = \frac{\bar{e}_j + u_j}{1 - \rho_1}$ .<sup>8</sup> If  $\bar{X}_j = Z_j$  (reflection), putting together (44) in (43) and centering, we obtain:<sup>9</sup>

<sup>8</sup>See Snijders and Bosker (1999, page 53) for the derivation of the between group regression in a multilevel setting.

<sup>9</sup>Technical Appendix A with the proof of equation (45) is available at <http://sites.google.com/site/luisacor->

$$Y_{ij} = \frac{\beta_0}{1 - \rho_1} + \frac{\beta_1 + \gamma}{1 - \rho_1} \bar{X}_j + \beta_1 (X_{ij} - \bar{X}_j) + e_{ij} + u_j'' \quad (45)$$

with  $u_j'' = u_j + \rho_1 u_j'$ . This is an example of Manski's reflection problem: in situations where group average characteristics,  $\bar{X}_j$ , directly affect the individual outcome,  $Y_{ij}$ , the parameters in the structural model are not identified. In this case the number of coefficients in the reduced form (45) is not sufficient to identify the coefficients in the structural equation (43). We have four parameters to identify in the structural equation (43) but only three estimated coefficients in (45). 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 introducing mean peer characteristics,  $\bar{X}_j$ , as an effect on outcomes.<sup>10</sup>

In the following section we will show how the connection between spatial models and multilevel models with group interactions ( $\rho_1 \neq 0, \gamma \neq 0$ ), allows parameter identification using simple instruments for the endogenous effects,  $\bar{Y}_j$ .

## 9.1 Spatial Effects and Identification

Extending the work by Cohen-Cole (2006) to the area of modelling interaction effects in a multilevel setting (see Corrado (2009)) we can rewrite (43) in a way that takes into account possible interdependencies not only within groups but also across groups, where a group may be those living in a specific district or region. We assume intra-group effects  $\bar{Y}_j = \frac{1}{w_j} \sum_{i=1}^{w_j} Y_{ij}$  where  $\bar{Y}_j$  is an average of all  $w_j$  unit responses within group  $j$  and, importantly we now introduce out-group effects  $\bar{Y}_l$  which are equal to an average of the responses across all 'neighboring' 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 thus far the literature has assumed that all surrounding groups enter with the same weight, but in this paper we introduce the innovation of differential weighting. Similarly we introduce out-group contextual effects  $\bar{Z}_l$ , thus leading to the model:

$$Y_{ij} = \beta_0 + \rho_1 \bar{Y}_j + \rho_2 (\bar{Y}_j - \bar{Y}_l) + \beta_1 X_{ij} + \gamma (Z_j - \bar{Z}_l) + e_{ij} + u_j, \quad (46)$$

in which the outcome of individual  $i$  in group  $j$ ,  $Y_{ij}$ , depends on the average outcome of group  $j$ , where the implicit assumption is that unit  $i$  is affected equally by all other units in  $j$ .

We also assume that the individual outcome depends on the average outcome,  $\bar{Y}_l$ , and average contextual effects,  $\bar{Z}_l$ , of other groups '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 as  $Z_j - \bar{Z}_l$ . In order to identify all the relevant parameters in the model we consider the average relationship derived from (46):

$$\bar{Y}_j = \frac{1}{1 - \rho_1 - \rho_2} (\beta_0 + \beta_1 \bar{X}_j + \gamma (Z_j - \bar{Z}_l) - \rho_2 \bar{Y}_l) + u_j' \quad (47)$$

---

rado/publications.

<sup>10</sup>By allowing for the possibility that the conditional mean of group effect and the individual effect vary with group size, Graham (2008) also allows for the possibility that peer-group effects may differ according to class size, being stronger in bigger classes. A way to detect the presence of group effects in this case is to measure the excess between-group variance defined as the ratio of unconditional (scaled) between-group and within group variances  $EV = \frac{E[w_j (\bar{Y}_j - \mu_y)^2]}{E[(w_j - 1)^{-1} \sum_{i=1}^{w_j} (Y_{ij} - \bar{Y}_j)^2]}$ .

where  $u'_j = \frac{\bar{e}_j + u_j}{1 - \rho_1 - \rho_2}$ .<sup>11</sup>

Putting together (47) in (46) and assuming reflection so that  $Z_j = \bar{X}_j$  :

$$Y_{ij} = \frac{\beta_0}{1 - \rho_1 - \rho_2} + \beta_1 X_{ij} + \frac{\beta_1(\rho_1 + \rho_2) + \gamma}{1 - \rho_1 - \rho_2} \bar{X}_j - \frac{\rho_2}{1 - \rho_1 - \rho_2} \bar{Y}_l - \frac{\gamma}{1 - \rho_1 - \rho_2} \bar{Z}_l + e_{ij} + u''_j \quad (48)$$

where  $u''_j = u_j + (\rho_1 + \rho_2)u'_j$ . It is clear from (48) that we can identify all the parameters in the structural equation (46) if the number of level-one units (typically number of individual people) exceeds the number of groups and  $\frac{\partial Y_{ij}}{\partial Y_l} \neq 0$ ,  $\frac{\partial Y_{ij}}{\partial Z_l} \neq 0$  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 other surrounding groups. Hence one could use outer groups' behaviour as instruments to identify the endogenous effects,  $\bar{Y}_j$ . For example, in the wage equation one could use as instruments wage and company characteristics in neighbouring companies, in other sectors, or at different geographical/hierarchical levels.

One important issue is collinearity, but the collinearity between  $X_{ij}$  and  $\bar{X}_j$  can be nullified by centering the variables. Reparameterizing, we obtain an equivalent hierarchical model:<sup>12</sup>

$$Y_{ij} = \frac{\beta_0}{1 - \rho_1 - \rho_2} + \underbrace{(X_{ij} - \bar{X}_j)}_{\text{within-group exogenous}} \beta_1 + \underbrace{- \frac{\rho_2}{1 - \rho_1 - \rho_2} \bar{Y}_l}_{\text{out-group endogenous}} + \underbrace{\frac{\gamma + \beta_1}{1 - \rho_1 - \rho_2} \bar{X}_j}_{\text{between-group exogenous}} - \underbrace{\frac{\gamma}{1 - \rho_1 - \rho_2} \bar{Z}_l}_{\text{out-group contextual}} + e_{ij} + u''_j \quad (49)$$

So we find that if we have an HLM model specification that includes out-group effects  $\bar{Y}_l$  and  $\bar{Z}_l$ , then this facilitates identification (c.f. Manski, 1993) of the model parameters in equation (46) because we can use  $\bar{Y}_l$ ,  $\bar{Z}_l$  and  $\bar{X}_j$  as instruments for the endogenous variable  $\bar{Y}_j$ . In doing this we assume that our instruments are not weak, and are orthogonal to  $\mathbf{Y}$ .

## 9.2 Estimating Group Spillovers in Hierarchical Models

The specifications developed thus far in fact embody specific  $\mathbf{W}$  matrix assumptions (as described below). Given these, we could envisage a generalization of (49) in order to incorporate more general forms of group interaction effects:

$$\mathbf{Y} = (\mathbf{I} - (\rho_1 + \rho_2)\mathbf{W})^{-1} (\beta_0 - \rho_2 \mathbf{W}_l \mathbf{Y} + \mathbf{W}\mathbf{X}(\beta_1 + \gamma) - \mathbf{W}_l \mathbf{Z}\gamma) + (\mathbf{X} - \mathbf{W}\mathbf{X})\beta_1 + \mathbf{u}'' + \mathbf{e} \quad (50)$$

which resembles the spatial Durbin specification (30) in Section 4 where  $\mathbf{W}$  is the matrix which defines the within-group effects and  $\mathbf{W}_l$  the out-group effects. Hence, when estimating the simpler specification where  $\mathbf{W}_l \mathbf{Z} = \mathbf{W}_l \mathbf{Y} = 0$  not only we are unable to identify both endogenous and exogenous interaction effects, we also face an omitted variable problem. In this case unobserved (omitted) covariates would be correlated with the observed (included) covariates. Since the observed covariates also include the endogenous spatial lag parameter  $\rho_2$ , its significance would be misleading, since it would pick up the effects of the omitted variables  $\mathbf{W}_l \mathbf{Z}$  and  $\mathbf{W}_l \mathbf{Y}$ . This occurrence is particularly problematic when

<sup>11</sup>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  $\rho_1$  and  $\rho_2$ .

<sup>12</sup>Technical Appendix B with the proof of equation (49) is available at <http://sites.google.com/site/luisacor-rado/publications>.

estimating models with random effects since the maintained assumptions are that both the individual and the group error components should be uncorrelated with the regressors  $Cov(\mathbf{u}, \mathbf{X}) = 0$  and  $Cov(\mathbf{e}, \mathbf{X}) = 0$ . This assumption is of course the same as for random effects panel models, as would typically be tested by means of a Hausman test of parameter consistency. This involves comparison of fixed and random effects estimates, assuming the fixed effects are consistent by being independent of the idiosyncratic disturbances. In the hierarchical case omitting spatial effects for  $\mathbf{X}$  and  $\mathbf{Z}$  will likewise create inconsistent estimators since the error components will be correlated with the regressors.

While we do not have a test that is equivalent to the Hausman test, we do have a way of avoiding inconsistency resulting from omitted variables. This is by means of the inclusion of spatial effects  $\mathbf{W}_l \mathbf{Z}$  and  $\mathbf{W}_l \mathbf{Y}$ . In fact these play a dual role: (i) of avoiding the omitted variable problem that may afflict models with endogenous spatial lags and (ii) of introducing a source of exogenous variation that helps to identify both endogenous and exogenous group effects, which seems to be vital to most of the empirical work on network interactions. In the context of model (50) estimation of both the observed and unobserved components is achieved via RIGLS/REML, as mentioned in Section 9, given  $\mathbf{W}_l \mathbf{Y}$ ,  $\mathbf{W}_l \mathbf{Z}$  and  $\mathbf{W} \mathbf{X}$ .

We now consider a simulation exercise. The basis of this is the group set-up. Groups can be unequal in size, and the weights do not have to be equal as in equation (42), so that group members can carry differential weights, for example, for a 4 member group:

$$\mathbf{W}_1 = \begin{bmatrix} a & b & c & d \\ a & b & c & d \\ a & b & c & d \end{bmatrix} \quad (51)$$

in which  $a + b + c + d = 1$ . Thus  $\mathbf{W}_1$  can have a different number of members and different weights (also summing to 1 across rows), compared with  $\mathbf{W}_2, \dots, \mathbf{W}_G$  where  $G$  is the number of groups.

This leads to our square diagonal  $\mathbf{W}$  matrix, which is

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{W}_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \cdot & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdot & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdot & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{W}_G \end{bmatrix}$$

in which 0 stands for a sub-matrix of 0s of appropriate dimension.

The spatial lagging matrix  $\mathbf{W}_l$  identifies the location and strength of spillovers between groups, such as for example:

$$\mathbf{W}_l = \begin{bmatrix} 0 & \mathbf{W}_2 & 0 & 0 & \cdot & 0 & 0 \\ \mathbf{W}_1 & 0 & 0 & 0 & \cdot & 0 & 0 \\ 0 & 0 & 0 & \mathbf{W}_4 & \cdot & 0 & 0 \\ 0 & 0 & \mathbf{W}_3 & 0 & \cdot & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdot & 0 & \mathbf{W}_G \\ 0 & 0 & 0 & 0 & \cdot & \mathbf{W}_{G-1} & 0 \end{bmatrix}$$

Given this, we start by considering a generalization of the structural equation (46):

$$\mathbf{Y} = \beta_0 \mathbf{1} + \rho_1 \mathbf{W} \mathbf{Y} + \rho_2 (\mathbf{W} \mathbf{Y} - \mathbf{W}_l \mathbf{Y}) + \mathbf{X} \beta_1 + (\mathbf{Z} - \mathbf{W}_l \mathbf{Z}) \gamma + \mathbf{u} + \mathbf{e} \quad (52)$$

We can therefore simulate our data from the following process for  $\mathbf{Y}$ :<sup>13</sup>

$$\mathbf{Y} = (\mathbf{I} - (\rho_1 + \rho_2)\mathbf{W} + \rho_2\mathbf{W}_l)^{-1} (\beta_0\boldsymbol{\iota} + \mathbf{X}\boldsymbol{\beta}_1 + (\mathbf{Z} - \mathbf{W}_l\mathbf{Z})\boldsymbol{\gamma} + \mathbf{u} + \mathbf{e}) \quad (54)$$

We assume that we have 8 groups composed of six members each so that  $G = 8$ ,  $N = 6$  and the matrices of interactions within each group are assumed to take the form of (51) with values for row totals summing to one. Given the matrices  $\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_8$  we therefore define the block matrices of within and between group interactions  $\mathbf{W}$  and  $\mathbf{W}_l$ . The value of  $\mathbf{Y}$  is assumed to depend on an exogenous variable  $\mathbf{X}$ , on the contextual variable  $\mathbf{Z}$ , and on the random disturbances  $\mathbf{e} \sim N(0, \Omega_e^2)$ , and  $\mathbf{u} \sim N(0, \Omega_u^2)$  while  $\mathbf{I}$  is an identity matrix of dimension  $N$ . The values of the  $N$  by one vectors  $\mathbf{Z}$  and  $\mathbf{X}$  are set exogenously;  $\mathbf{e}$  and  $\mathbf{u}$  are  $N$  by 1 vectors of normally distributed pseudorandom numbers for the individual and the group random effects which are independent of the exogenous regressors. Given  $\beta_0 = 0.5$ ,  $\beta_1 = 0.6$ ,  $\gamma = 0.5$ ,  $\rho_1 = 0.5$  and  $\rho_2 = 0.9$  we generate  $\mathbf{Y}$  via (54). We ensure throughout that  $(\mathbf{I} - (\rho_1 + \rho_2)\mathbf{W} + \rho_2\mathbf{W}_l)$  is non-singular.<sup>14</sup>

	Model A		Model B	
	Coefficient	<i>s.e.</i>	Coefficient	<i>s.e.</i>
$\beta_0 = 0.5$	-0.045	0.275	-6.028	0.146
$\beta_1 = 0.6$	0.589	0.007	0.589	0.000
$\gamma = 0.5$	0.479	0.011	-0.114	0.065
$\rho_1 = 0.5$	0.507	0.010	0.761	0.000
$\rho_2 = 0.9$	0.873	0.021	-	-
$\sigma_u^2$	0.129	0.092	41.11	26.02
$\sigma_e^2$	0.008	0.001	0.008	0.001
$N$	48		48	
$ll$	13.05		1.388	

Having generated  $\mathbf{Y}$  using (54), we estimate two models. One is the correctly specified model (52), denoted model A in Table 5, and the second, model B, is the misspecified model which omits the spatially dependent variable  $\mathbf{W}_l\mathbf{Y}$ , since  $\rho_2 = 0$ , and the out group effect,  $\mathbf{W}_l\mathbf{Z}$ . Thus:

$$\mathbf{Y} = \beta_0\boldsymbol{\iota} + \rho_1\mathbf{W}\mathbf{Y} + \beta_1\mathbf{X} + \gamma\mathbf{Z} + \mathbf{u} + \mathbf{e} \quad (55)$$

The estimation via REML reported in Table 5 shows that when we incorrectly fit a DGP omitting the out-group effects  $\mathbf{W}_l\mathbf{Y}$  and  $\mathbf{W}_l\mathbf{Z}$ , this causes us to overstate the effect of the endogenous lag  $\mathbf{W}\mathbf{Y}$  as is apparent in the upwardly biased estimate of  $\rho_1$ . In addition the inflated value of the estimated random effects parameter  $\sigma_u^2$  might also be because it is also capturing an omitted variable, so that

<sup>13</sup>Note that the same result can also be obtained by simulating  $\mathbf{Y}$  from the reduced form (50):

$$\mathbf{Y} = (\mathbf{I} - (\rho_1 + \rho_2)\mathbf{W} + \rho_2\mathbf{W}_l)^{-1} (\beta_0\boldsymbol{\iota} + \mathbf{W}\mathbf{X}(\beta_1 + \gamma) - \mathbf{W}_l\mathbf{Z}\boldsymbol{\gamma}) + (\mathbf{X} - \mathbf{W}\mathbf{X})\boldsymbol{\beta}_1 + \mathbf{u}'' + \mathbf{e} \quad (53)$$

where  $\mathbf{u}'' = \mathbf{u} + ((\rho_1 + \rho_2)\mathbf{W} - \rho_2\mathbf{W}_l)\mathbf{u}'$  and  $\mathbf{u}' = (\mathbf{I} - (\rho_1 + \rho_2)\mathbf{W} + \rho_2\mathbf{W}_l)^{-1}(\mathbf{W}\mathbf{e} + \mathbf{u})$ .

<sup>14</sup>The code to generate the data for the estimations in Table 5 is available at <http://personal.strath.ac.uk/bernard.fingleton/>.

it is not a truly (unobserved) group effect. Therefore care must be taken in HLM estimation when assuming group independence.

## 10 CONCLUSIONS

We ask the question, ‘Where is the economics in spatial econometrics?’. Our answer is that economic theory is to be found underpinning many spatial econometric models, but it is evident that when it comes to the so-called  $\mathbf{W}$  matrix, the economic foundation of many models is at its weakest. Modeling spatial interaction in the economic context means in many cases modelling externalities and spillovers. These are elusive and difficult to pin down, which is probably why we have considerable difficulties in defining the structure of  $\mathbf{W}$  unambiguously. The difficulty of detecting and measuring spatial spillover phenomena was recognized by Krugman (1991), who famously remarked that knowledge flows ‘leave no paper trail by which they can be measured or tracked, and there is nothing to prevent the theorist from assuming anything about them that she likes’. We have called for a stronger more theoretical basis for  $\mathbf{W}$  to supplement the very significant atheoretical empirical foundations that dominate, something that might emerge from current work on games, network formation, dynamics and equilibria that is occurring within the social science, notably within the economics of networks. We have attempted to show that the concept of the  $\mathbf{W}$  matrix is however undeniably necessary in one form or another and is in any case almost inescapable. It first comes to our attention as a convenient, useful and succinct representation of spatial interaction, either in the form of endogenous or exogenous lagged variables, and/or as part of an explicit error process.

However we also find that the  $\mathbf{W}$  matrix has a pervasive presence extending far beyond cross-sectional models to multi-equation models and panel data models. The denial of  $\mathbf{W}$  does not completely eliminate it, merely suppresses it only to reappear as assumptions about distance and interaction in competing or complementary approaches, with a few exceptions, such as seemingly unrelated regression, vector autoregressions and vector error correction models applied to multivariate time series. Moreover, we have shown that it exists not only trivially in implied form in autoregressive time series processes, but is, we now discover, a cognate part of hierarchical models. We have considered in some detail the connection between hierarchical models and the standard spatial econometric specifications, because hierarchical models are almost completely absent from the spatial econometrics literature, and represent one major alternative way of capturing spatial effects, focussing on the multilevel aspects of causation that are a reality of many spatial processes. Our contribution in this regard is to show that  $\mathbf{W}$  appears also in a specific form as part of the structure of hierarchical models. This highlights the limitations of much of current hierarchical modelling, which should benefit from cross-fertilization from the spatial econometrics literature, with the prospect of a whole new research agenda embodying differential spatial dependence within and between groups in the multilevel context.

## References

- [1] Akerlof, George A. (1997). ‘Social Distance and Social Decisions,’ *Econometrica*, 65, pp. 1005–1028.
- [2] Andrews, Donald (1991). ‘Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation,’ *Econometrica*, 59, pp. 817–858.
- [3] Anselin, Luc (1986). ‘Non-nested Tests on the Weight Structure in Spatial Autoregressive Models,’ *Journal of Regional Science*, 26, pp. 267–284.

- [4] — (1988). *Spatial Econometrics: Methods and Models*. The Netherlands: Kluwer Academic Publishers.
- [5] — (1999). ‘The Future of Spatial Analysis in the Social Sciences,’ *Geographic Information Sciences*, 5, pp. 67–76.
- [6] — (2003). ‘Spatial Externalities,’ *International Regional Science Review*, 26, pp. 147–152.
- [7] — (2010). ‘Thirty Years of Spatial Econometrics,’ *Papers in Regional Science*, 89, pp. 3–25.
- [8] Anselin, Luc, and Wendy K. T. Cho (2002). ‘Spatial Effects and Ecological Inference,’ *Political Analysis*, 10, pp. 276–303.
- [9] Anselin, Luc, Julie Le Gallo and Hubert Jayet (2007). ‘Spatial Panel Econometrics,’ in L. Matyas and P. Sevestre (eds.), *The Econometrics of Panel Data, Fundamentals and Recent Developments in Theory and Practice*. Dordrecht: Kluwer, 3rd Edition, pp. 625–660.
- [10] Anselin, Luc and Rosina Moreno (2003). ‘Properties of Tests for Spatial Error Components,’ *Regional Science and Urban Economics*, 33, pp. 595–618.
- [11] Aslam, Aqib and Luisa 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.
- [12] Baltagi, Badi H. and Dong Li (2006). ‘Prediction in the Panel Data Model with Spatial Correlation: The Case of Liquor,’ *Spatial Economic Analysis*, 1, pp. 75–185.
- [13] Barde, Sylvain (2010). ‘Increasing Returns and the Spatial Structure of French Wages,’ *Spatial Economic Analysis*, 5, pp. 73–79.
- [14] Bauman, Karl, and Lynn Fisher (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, pp. 345–353.
- [15] Bénabou, Roland (1993). ‘Workings of a City: Location, Education, and Production,’ *Quarterly Journal of Economics*, CVIII, pp. 619–652.
- [16] Bernheim, Douglas (1994). ‘A Theory of Conformity,’ *Journal of Political Economy*, 102, pp. 841–877.
- [17] Bhattacharjee, Arnab and Sean Holly (2006). ‘Taking Personalities out of Monetary Policy Decision Making? Interactions, Heterogeneity and Committee Decisions in the Bank of England MPC,’ Centre for Dynamic Macroeconomic Analysis Working Paper Series 0612, School of Economics and Finance, University of St. Andrews.
- [18] Blundell, Richard and Frank Windmeijer (1997). ‘Correlated Cluster Effects and Simultaneity in Multilevel Models,’ *Health Economics*, 1, pp. 6–13.
- [19] Bodson, Paul and Dominique Peters (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*, 7, pp. 455–472.
- [20] Brock, William A. and Stephen N. Durlauf (2001). ‘Interactions-Based Models,’ in J. J. Heckman, and E. Leamer (eds), *Handbook of Econometrics*, Vol. 5. Amsterdam: Elsevier Science, pp. 3297–3380.
- [21] Brown, Bradford (1990). ‘Peer Groups and Peer Cultures,’ in S. Feldman, and G. Elliott (eds.), *At the Threshold: the Developing Adolescent*, Cambridge MA: Harvard University Press, pp. 171–196.

- [22] Brown, Bradford, Donna Clasen and Sue Eicher (1986). ‘Perceptions of Peer Pressure, Peer Conformity Dispositions and Self-Reported Behavior Among Adolescents,’ *Developmental Psychology*, 22, pp. 521–530.
- [23] Bryk, Anthony S. and Stephen W. Raudenbush (1992). *Hierarchical Linear Models for Social and Behavioral Research: Applications and Data Analysis Methods*. Newbury Park CA: Sage Publications.
- [24] Burridge, Peter and Bernard Fingleton (2010). ‘Bootstrap Inference in Spatial Econometrics : the J Test,’ *Spatial Economic Analysis*, 5, pp. 93–119.
- [25] Chang, Sheng-Wen and Edward Coulson (2001). ‘Sources of Sectoral Employment Fluctuations in Central Cities and Suburbs: Evidence from Four Eastern U.S. Cities,’ *Journal of Urban Economics*, 49, pp. 199–218.
- [26] Cliff, Andrew D. and John K. Ord (1973). *Spatial Autocorrelation*. London: Pion.
- [27] — (1981). *Spatial Processes: Models and Applications*. London: Pion.
- [28] Cohen-Cole, Ethan (2006). ‘Multiple Groups Identification in the Linear-in-Means Model,’ *Economics Letters*, 92, pp. 157–162.
- [29] Conley, Timothy G. (1999). ‘GMM Estimation with Cross-Sectional Dependence,’ *Journal of Econometrics*, 92, pp. 1–45.
- [30] Conley, Timothy G. and Ethan Ligon (2002). ‘Economic Distance, Spillovers and Cross Country Comparisons,’ *Journal of Economic Growth*, 7, pp. 157–187.
- [31] Conley, Timothy G. and Giorgio Topa (2002). ‘Socio-economic Distance and Spatial Patterns in Unemployment,’ *Journal of Applied Econometrics*, 17, pp. 303–327.
- [32] Corrado, Luisa (2009). ‘A Unified Solution to the Identification and Estimation of Social Interaction Effects,’ mimeo.
- [33] Corrado, Luisa and Bernard Fingleton (2010). ‘Multilevel Modelling with Spatial Effects,’ mimeo.
- [34] Cressie, Noel (2003). *Statistics for Spatial Data*. New York: Wiley.
- [35] Ertur, Cem and Wilfried Koch (2007). ‘Growth, Technological Interdependence and Spatial Externalities: Theory and Evidence,’ *Journal of Applied Econometrics*, 22, pp. 1033–1062.
- [36] Fingleton, Bernard (2001). ‘Equilibrium and Economic Growth: Spatial Econometric Models and Simulations,’ *Journal of Regional Science*, 41, pp. 117–147.
- [37] — (2003). ‘Increasing Returns: Evidence from Local Wage Rates in Great Britain,’ *Oxford Economic Papers*, 55, pp. 716–739.
- [38] — (2006). ‘The New Economic Geography versus Urban Economics: an Evaluation using Local Wage Rates in Great Britain,’ *Oxford Economic Papers*, 58, pp. 501–530.
- [39] — (2008). ‘A Generalized Method of Moments Estimator for a Spatial Model with Moving Average Errors, with Application to Real Estate Prices,’ *Empirical Economics*, 34, pp. 35–57.
- [40] — (2009). ‘Spatial Autoregression,’ *Geographical Analysis*, 41, pp. 385–391.
- [41] Fingleton, Bernard and Julie Le Gallo (2008). ‘Estimating Spatial Models with Endogenous Variables, a Spatial Lag and Spatially Dependent Disturbances: Finite Sample Properties,’ *Papers in Regional Science*, 87, pp. 319–339.



- [42] — (2009). ‘Endogeneity in a Spatial Context : Properties of Estimators,’ in A. Páez, J. Le Gallo, R. Buliung, and S. Dall’Erba. (eds.), *Progress in Spatial Analysis: Theory and Computation, and Thematic Applications, Advances in Spatial Sciences*. Berlin: Springer-Verlag, pp. 59–73.
- [43] Fingleton, Bernard and Enrique Lopez-Bazo (2006). ‘Empirical Growth Models with Spatial Effects,’ *Papers in Regional Science*, 85, pp. 177–198.
- [44] Folmer, Henk and Johan Oud (2008). ‘How to Get Rid of W: a Latent Variables Approach to Modelling Spatially Lagged Variables,’ *Environment and Planning A*, 40, pp. 2526–2538.
- [45] Gibbons, Stephen, and Henry Overman (2010). ‘Mostly Pointless Spatial Econometrics?,’ *Journal of Regional Science*, forthcoming.
- [46] Goldstein, Harvey (1986). ‘Multilevel Mixed Linear Model Analysis Using Iterative Generalised Least Squares,’ *Biometrika*, 73, pp. 46–56.
- [47] — (1989). ‘Restricted Unbiased Iterative Generalized Least-Squares Estimation,’ *Biometrika*, 76, pp. 622–623.
- [48] — (1998). ‘Multilevel Models for Analysing Social Data,’ in *Encyclopaedia of Social Research Methods*. Newbury Park CA: Sage Publications.
- [49] Goyal, Sanjeev (2009). *Connections: an Introduction to the Economics of Networks*. Princeton: Princeton University Press.
- [50] Graham, Bryan S. (2008). ‘Identifying Social Interactions through Conditional Variance Restrictions,’ *Econometrica*, 76, pp. 643–660.
- [51] Greenhut, Melvyn, Goerge Norman and Chao-Shun Hung (1987). *The Economics of Imperfect Competition: A Spatial Approach*. Cambridge: Cambridge University Press.
- [52] Harris, Richard, John Moffat and Victoria Kravtsova (2010). ‘In Search of W,’ mimeo.
- [53] Holly, Sean, M. Hashem Pesaran and Takashi Yamagata (2010). ‘Spatial and Temporal Diffusion of House Prices in the UK,’ *Journal of Urban Economics*, forthcoming.
- [54] Hsiao, Cheng and M. Hashem Pesaran (2008). ‘Random Coefficient Panel Data Models,’ in L. Matyas and P. Sevestre (eds.), *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*. Third Edition. Berlin: Springer Publishers, pp. 185–213.
- [55] Jones, Andrew (1994). ‘Health, Addiction, Social Interaction, and the Decision to Quit Smoking,’ *Journal of Health Economics*, 13, pp. 93–110.
- [56] Kandori, Michihiro (1992). ‘Social Norms and Community Enforcement,’ *Review of Economic Studies*, 59, pp. 63–80.
- [57] Kapetanios, George and M. Hashem Pesaran (2005). ‘Alternative Approaches to Estimation and Inference in Large Multifactor Panels: Small Sample Results with an Application to Modelling of Asset Returns,’ in G.D.A. Phillips, and E. Tzavalis (eds.), *The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis*. Cambridge: Cambridge University Press, pp. 239–281.
- [58] Kapoor, Mudit, Harry H. Kelejian, and Ingmar R. Prucha (2007). ‘Panel Data Models with Spatially Correlated Error Components,’ *Journal of Econometrics*, 140, pp. 97–130.

- [59] Kelejian, Harry H. (2008). ‘A Spatial J-test for Model Specification against a Single or a Set of Non-Nested Alternatives,’ *Letters in Spatial and Resource Sciences*, 1, pp. 3–11.
- [60] Kelejian, Harry H. and Ingmar R. 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, pp. 99–121.
- [61] — (2007). ‘HAC estimation in a Spatial Framework,’ *Journal of Econometrics*, 140, pp. 131–154.
- [62] Kelejian, Harry H. and Dennis P. Robinson (1995). ‘Spatial Correlation: A Suggested Alternative to the Autoregressive Model,’ in L. Anselin, and R. Florax (eds.), *New Directions in Spatial Econometrics*. Berlin: Springer-Verlag, pp. 75–95.
- [63] Krosnick, Jon A., and Charles M. Judd (1982). ‘Transitions in Social Influence in Adolescence: Who Induces Cigarette Smoking,’ *Developmental Psychology*, 81, pp. 359–368.
- [64] Krugman, Paul (1991). *Geography and Trade*. Cambridge, MA: MIT Press.
- [65] Leenders, Roger T. A. J. (2002). ‘Modeling Social Influence through Network Autocorrelation: Constructing the Weight Matrix,’ *Social Networks*, 24, pp. 21–47.
- [66] LeSage, James and Manfred Fischer (2008). ‘Spatial Growth Regressions: Model Specification, Estimation and Interpretation,’ *Spatial Economic Analysis*, 3, pp. 275–304.
- [67] LeSage, James and Kelley R. Pace (2008). ‘Spatial Econometric Modeling of Origin-Destination Flows,’ *Journal of Regional Science*, 48, pp. 941–967.
- [68] — (2009). *Introduction to Spatial Econometrics*. Boca Raton, FL: CRC Press.
- [69] LeSage, James and Olivier Parent (2007). ‘Bayesian Model Averaging for Spatial Econometric Models,’ *Geographical Analysis*, 39, pp. 241–267.
- [70] LeSage, James and Anna Krivelyova (1999). ‘A Spatial Prior for Bayesian Autoregressive Models,’ *Journal of Regional Science*, 39, pp. 297–317.
- [71] Longford, Nicholas T. (1993). *Random Coefficient Models*. London: Clarendon Press.
- [72] Manski, Charles F. (1993). ‘Identification of Endogenous Social Effects: The Reflection Problem,’ *Review of Economic Studies*, 60, pp. 531–542.
- [73] — (2000). ‘Economic Analysis of Social Interactions,’ *Journal of Economic Perspectives*, 3, pp. 115–136.
- [74] McFadden, Daniel (1974). ‘Conditional Logit Analysis of Qualitative Choice Behavior,’ in P. Zarembka (eds.), *Frontiers in Econometrics*. New York: Academic, pp. 105–142.
- [75] — (1979). ‘Qualitative Methods for Analysing Travel Behaviour of Individuals: some Recent Developments,’ in S. Hensher (eds.), *Behaviour Travel Modelling*. London: Croom Helm, pp. 279–318.
- [76] McMillen, Daniel P. (2010a). ‘Issues in Spatial Data Analysis,’ *Journal of Regional Science*, 50, pp. 119–141.
- [77] — (2010b). ‘Perspectives on Spatial Econometrics: Linear Smoothing with Structural Models,’ *Journal of Regional Science*, forthcoming.

- [78] Moulton, Brent R. (1986). ‘Random Group Effects and the Precision of Regression Estimates,’ *Journal of Econometrics*, 32, pp. 385–397.
- [79] — (1990). ‘An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Unit,’ *The Review of Economics and Statistics*, 72, pp. 334–338.
- [80] Ord, John K. (1975). ‘Estimation Methods for Models of Spatial Interaction,’ *Journal of the American Statistical Association*, 70, pp. 120–126.
- [81] Pace, Kelley R. and James LeSage (2008). ‘Omitted Variable Bias of OLS and Spatial Lag Models,’ in A. Páez, J. Le Gallo, R. Buliung, and S. Dall’Erba. (eds.), *Progress in Spatial Analysis: Theory and Methods and Thematic Applications*. Berlin: Springer, pp. 17–28.
- [82] Paci, Raffaele and Stefano Usai (2009). ‘Knowledge Flows across European Regions,’ *Annals of Regional Science*, 43, pp. 669–690.
- [83] Parent, Olivier and James LeSage (2008). ‘Using the Variance Structure of the Conditional Autoregressive Specification to Model Knowledge Spillovers,’ *Journal of Applied Econometrics*, 23, pp. 235–256.
- [84] Patuelli, Roberto, Aura Reggiani, Sean P. Gorman, Peter Nijkamp, and Franz-Josef Bade (2007). ‘Network Analysis of Commuting Flows: A Comparative Static Approach to German Data,’ *Networks and Spatial Economics*, 7, pp. 315–331.
- [85] Pesaran, M. Hashem (2007). ‘A Simple Panel Unit Root Test in the Presence of Cross-Sectional Dependence,’ *Journal of Applied Econometrics*, 138, pp. 2265–2312.
- [86] Pinkse, Joris and Margaret E. Slade (2010). ‘The Future of Spatial Econometrics,’ *Journal of Regional Science*, 50, pp. 103–1117.
- [87] Pinkse, Joris, Margaret E. Slade, and Craig Brett (2002). ‘Spatial Price Competition: a Semiparametric Approach,’ *Econometrica*, 70, pp. 1111–1153.
- [88] Slade, Margaret E. (2005). ‘The Role of Economic Space in Decision Making,’ ADRES Lecture, *Annales D’Économie et De Statistique*, 77, pp. 1–20.
- [89] Snijders, Tom and Roel Bosker (1999), *Multilevel Analysis: an Introduction to Basic and Advanced Multilevel Modeling*, London: Sage Publications.
- [90] Smith, Tony E. and James LeSage (2004). ‘A Bayesian Probit Model with Spatial Dependencies,’ in J. LeSage and R. Pace (eds.), *Advances in Econometrics, Volume 18: Spatial and Spatiotemporal Econometrics*. Oxford: Emerald Group Publishing, pp. 127–160.
- [91] Wilson, Alan G. (1967). ‘A Statistical Theory of Spatial Distribution Models,’ *Transportation Research*, 1, pp. 253–269.
- [92] — (1970). *Entropy in Urban and Regional Modelling*. London: Pion.