MONASH University

Australia

Department of Econometrics

and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

An Anisotropic Model For Spatial Processes

Minfeng Deng

March 2006

Working Paper 7/06

An Anisotropic Model For Spatial Processes

Minfeng Deng

Department of Econometrics and Business Statistics,
Monash University, Melbourne, Victoria, 3168, Australia.
Email: gary.deng@buseco.monash.edu.au

Abstract

One of the key assumptions in spatial econometric modelling is that the spatial process is isotropic, which means that direction is irrelevant in the specification of the spatial structure. On one hand, this assumption largely reduces the complexity of the spatial models and facilitates estimation and interpretation; on the other hand, it appears rather restrictive and hard to justify in many empirical applications. In this paper a very general anisotropic spatial model, which allows for a high level of flexibility in the spatial structure, is proposed. This new model can be estimated using maximum likelihood and its asymptotic properties are well understood. When the model is applied to the well-known 1970 Boston housing prices data, it significantly outperforms the isotropic spatial lag model. It also provides interesting additional insights into the price determination process in the properties market.

Keywords: anisotropy, spatial econometrics, maximum likelihoods estimation, housing prices.

JEL CLASSIFICATION: C21, R15m R31

1. Introduction

The most important aspect of spatial econometric modelling is the incorporation of a dependence structure between cross-sectional observations. With N spatial units, potentially there are up to $\left(N^2 - N\right)$ unique spatial relationships. The problem of incidental parameters arises, and, unless the phenomenon of interest can be observed for the same set of cross-sectional units over a large number of time periods, it is impossible to separately identify these $\left(N^2 - N\right)$ spatial relationships.

In most empirical applications rather restrictive assumptions are made with regards to the extent and form of the spatial dependence structure. One of the key assumptions is that of isotropy, which requires that spatial effects are of equal strength in all directions between all spatially contiguous observations. For example, the well-known spatial lag model (Anselin 1988):

$$y = \rho W y + X\beta + \varepsilon \tag{1}$$

where the error term ε is assumed to be $N\left(0, \sigma^2 I\right)$, assumes isotropy. The spatial weights matrix W is an $\left(N \times N\right)$ matrix that identifies spatial relationships between observations. The most commonly employed spatial weights matrices, as proposed by Cliff and Ord (1973), are either of a binary contiguity design or of a distance decay design, and do not take direction into account. The spatial parameter ρ is a scalar that is multiplied to each and every spatial neighbour, making it impossible to allow for varying degrees of spatial influence to originate from different spatial neighbours.

While the assumption of isotropy is a standard practice in spatial econometric modelling, it seems overly restrictive in many empirical applications. Just as in our personal relationships not every acquaintance is a best friend, it is unlikely that every spatial neighbour is equally important. It appears more reasonable to allow spatial effects to vary between different spatial units and in different directions.

Over the years, some attempts have been made to incorporate more information and greater flexibility into the spatial structure. For example, Dacey (1968) suggests weights that combine distance, size and length of border. Bodson and Peeters (1975) introduce a general accessibility weight that utilizes a logistic function to combine the influences of several channels of communication between regions (such as roads and railways). Cliff and Ord (1981) propose spatial weights that combine distance measures and common borders. Besner (2002) proposes the so-called SARS (spatial autoregressive model with similarity measures) model that explicitly incorporates similarity measures in socio-economic variables in the construction of the spatial weights. Getis and Aldstadt (2004) propose a model where the spatial weights are constructed using the G_i^* local statistic of Ord and Getis (1995). But the above-mentioned models all contain unknown parameters in the spatial weights matrix. Estimation of these parameters often takes place at a step separate from the regression model of interest. As pointed out in Anselin (1988), "...it could potentially lead to the inference of spurious relationships, since the validity of estimates is pre-conditioned by the extent to which the spatial structure is correctly reflected in the weights." They also present great difficulties in the interpretation and significance testing of the parameters present in the spatial weights matrix, since their estimation is separate from the regression of interest. On the other hand, if the parameters were to be estimated along with the rest of the parameters in the regression model, since traditionally the weights matrix is row-standardized, at each iteration of the optimization process the spatial weights matrix must be re-standardized. For large samples this could become a computationally very heavy task. Given the difficulties involved in the parameterization of the spatial weights matrix, empirical application of these models is limited. Gillen, Thibodeau, and Wachter (2001) examine anisotropic spatial autocorrelation in the residuals of a hedonic house-price model of the suburban Philadelphia by comparing the empirical semivariograms for the North-South and East-West directions. Although they find evidence of anisotropy in the semivariograms, they fail to provide a framework in which this information can be used to further improve the efficiency of the estimators. Also, in a situation where anisotropic spatial effect is expressed in a substantive form, that is, not as a pattern in the residuals but as spatially lagged dependent variables on the right hand side, their anisotropic semivariogram approach becomes unusable.

Generally speaking, a process is said to be anisotropic when its behaviour exhibits directional dependence. It is fair to say that thus far anisotropy has not been systematically built into spatial econometric modelling and as a result the explanatory power of the existing spatial models might have been limited. In this paper an anisotropic model for spatial processes is proposed. Moreover, anisotropy is generalized beyond a physical sense. As will be demonstrated shortly, in the new model spatial influence is allowed to vary in strength based on not only the usual North/South or East/West directions but also on the relative socio-economic characteristics of the spatial pairs.

This paper is organised as follows: In section 2 the new anisotropic model is presented. In section 3 estimation method based on the maximum likelihood principle is discussed, and the asymptotic properties of these estimators are derived in full. In section 4 an empirical study on the well-known 1970 Boston housing prices data is conducted, where it is demonstrated that the anisotropic model significantly outperforms the isotropic spatial lag model. Section 5 concludes the paper.

2. Formulation Of An Anisotropic Model

It is well-understood that spatial autocorrelation can be presented either as a spatial lag model or as a spatial error model. A spatial lag model represents substantiative dependence between spatial units, while a spatial error model is often used as a way of dealing with missing variables that are also spatially correlated. In this paper, anisotropic forms of both models will be considered. Discussion will start from the spatial lag model. The following is a fully expanded version of the spatial lag model:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \cdots \\ y_N \end{bmatrix} = \begin{bmatrix} 0 & \rho w_{12} & \rho w_{13} & \cdots & \rho w_{1N} \\ \rho w_{21} & 0 & \rho w_{23} & \cdots & \rho w_{2N} \\ \rho w_{31} & \rho w_{32} & 0 & \cdots & \rho w_{3N} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho w_{N1} & \rho w_{N2} & \rho w_{N3} & \cdots & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \cdots \\ y_N \end{bmatrix} + \begin{bmatrix} X_1'\beta \\ X_2'\beta \\ X_3'\beta \\ \cdots \\ X_N'\beta \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \cdots \\ \varepsilon_N \end{bmatrix} \quad (2)$$

where w_{ij} is the spatial weight given to the jth spatial neighbour of the ith spatial unit. The most commonly used spatial weights matrices are the binary contiguity matrix and the distance decay matrix, and they do not take direction into account. Prior to row-standardization these weights matrices are symmetric, which reflects the inherent isotropic assumption. Row-standardization normally makes these matrices asymmetric since generally speaking not all spatial units have the same number of spatial neighbours. But this asymmetry is a result of row-standardization and does not bring extra information on the asymmetry of the spatial effects into the model. One also notices that in equation (2) the same scalar spatial parameter ρ is being multiplied into each and every one of the spatial neighbours, implying that all spatial neighbours are given equal amount of importance, further confirming the isotropic nature of the model.

In many empirical studies the assumption of isotropy seems highly restrictive. For example, in the study of crime patterns, it is reasonable to expect neighbourhoods that suffer from severe poverty to "export" crimes to other neighbourhoods, but not the other way around. In the study of water-born diseases, it is reasonable to expect the diffusion of the disease to closely follow the flow of the water system. In hedonic house-price studies neighbours with "strong" characteristics, such as high crime rates, are expected to exhibit greater influences. Therefore, it appears more reasonable to allow the strength of spatial effects to differ depending on the relative characteristics of the two spatial neighbours.

In a spatial lag model the spatial structure is jointly expressed by the spatially lagged dependent variables Wy and the spatial autoregressive coefficient ρ. Intuitively the incorporation of anisotropy in such a model must involve changes in either one or both of these factors. As mentioned earlier many difficulties exist in the parameterization of the spatial weights matrix, it is therefore proposed that anisotropy should be introduced through the spatial autoregressive coefficient ρ.

Let $\psi_{ij}^{'}$ be a vector of variables that capture the relative characteristics between the ith observation and its jth spatial neighbour, which are thought to be important in generating the anisotropic spatial influences. Let $\psi_{q,ij}$ be the qth element of vector $\psi_{ij}^{'}$, it could be a binary variable that indicates whether the jth neighbour is to the North of the ith spatial unit, or a variable that captures the similarity in poverty level between the two spatial regions, or a binary variable that indicates whether the jth neighbour is located on the upstream of the ith spatial unit. Replacing the scalar spatial parameter ρ in equation (2) with a function $f(.)$ of $\psi_{ij}^{'}$ for all observations, the model becomes:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \dots \\ y_N \end{bmatrix} = \begin{bmatrix} 0 & f(\psi_{12}^{'} | \theta) w_{12} & f(\psi_{13}^{'} | \theta) w_{13} & \dots & f(\psi_{1N}^{'} | \theta) w_{1N} \\ f(\psi_{21}^{'} | \theta) w_{21} & 0 & f(\psi_{23}^{'} | \theta) w_{23} & \dots & f(\psi_{2N}^{'} | \theta) w_{2N} \\ f(\psi_{31}^{'} | \theta) w_{31} & f(\psi_{32}^{'} | \theta) w_{32} & 0 & \dots & f(\psi_{3N}^{'} | \theta) w_{3N} \\ \dots & \dots & \dots & \dots & \dots \\ f(\psi_{N1}^{'} | \theta) w_{N1} & f(\psi_{N2}^{'} | \theta) w_{N2} & f(\psi_{N3}^{'} | \theta) w_{N3} & \dots & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \dots \\ y_N \end{bmatrix} + \begin{bmatrix} X_1^{'} \\ X_2^{'} \\ X_3^{'} \\ \dots \\ X_N^{'} \end{bmatrix} \beta + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \dots \\ \varepsilon_N \end{bmatrix} \quad (3)$$

where $\theta = \{\theta_0, \theta_1, ..., \theta_Q\}^{'}$ is a vector of parameters associated with the function $f(.)$ and Q is the total number of variables included in vector $\psi_{ij}^{'}$.

The above model is referred to as the *Anisotropic Spatial Lag Model*, which can be written more succinctly as:

$$y_i = \sum_{j=1}^{N} f(\psi_{ij}^{'} | \theta) w_{ij} y_j + X_i^{'} \beta + \varepsilon_i \quad \forall i = 1,...,N \quad (4)$$

And in matrix notation:

$$y = [F(\psi | \theta) \bullet W] y + X\beta + \varepsilon \quad (5)$$

where W is the usual spatial weights matrix of choice. $F(\psi | \theta)$ is an *(N x N)* matrix, where its *i,j*th element is $f(\psi_{ij}^{'} | \theta)$. Since every $f(\psi_{ij}^{'} | \theta)$ is specific to one particular pair of spatial observations i and j, $[F(\psi \mid \theta) \bullet W]$ is a Hadamard product between $F(\psi \mid \theta)$ and the spatial weights matrix W.

Let $f(.)$ be a linear function of $\psi_{ij}^{'}$ and let $\psi_{ij}^{'}$ contain a constant term. Let θ_0 be the parameter in $f(\psi_{ij}^{'} \mid \theta)$ that is associated with this constant term. If the remaining variables in $\psi_{ij}^{'}$ are jointly insignificant, i.e., $\theta_q = 0 \;\forall\, q \neq 0$, then the anisotropic spatial lag model collapses to the usual spatial lag model. Thus, a spatial lag model is a special case of the more general anisotropic spatial lag model. Estimates of $\theta_q \;\forall\, q \neq 0$ can provide valuable information on what determines the anisotropy in the spatial structure.

If anisotropy manifests itself in the residuals, as in the case where a missing variable (such as the presence of a river or a highway) contains directional spatial patterns, the following *Anisotropic Spatial Error Model* should be used instead:

$$y = X\beta + \varepsilon \qquad \text{where } \varepsilon = [F(\psi \mid \theta) \bullet W]\varepsilon + u \tag{6}$$

where u is assumed to be $N(0, \sigma^2 I)$ and $[F(\psi \mid \theta) \bullet W]$ is defined in exactly the same way as before.

3. Methodology

3.1 Estimation

Maximum likelihood estimation of spatial models is first outlined in Ord (1975) and has since become the most popular estimation method in spatial econometric modelling. For the anisotropic spatial lag model, the log likelihood function is:

$$\ln L = -\left(\frac{N}{2}\right)\ln(2\pi) - \left(\frac{N}{2}\right)\ln(\sigma^2) + \ln\left| I - [F(\psi \mid \theta) \bullet W] \right|$$
$$-\left(\frac{1}{2\sigma^2}\right)\{y - [F(\psi \mid \theta) \bullet W]y - X\beta\}^T \{y - [F(\psi \mid \theta) \bullet W]y - X\beta\} \tag{7}$$

which can be expressed more succinctly as

$$\ln L = -\left(\frac{N}{2}\right)\ln(2\pi) - \left(\frac{N}{2}\right)\ln(\sigma^2) + \ln|A| - \left(\frac{1}{2\sigma^2}\right)\{Ay - X\beta\}^T \{Ay - X\beta\} \tag{8}$$

where:

$$A = \{ I - [F(\psi \mid \theta) \bullet W] \} \tag{9}$$

The first partial derivatives can be obtained as:

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}\{Ay - X\beta\}^T \{Ay - X\beta\} \tag{10}$$

$$\frac{\partial \ln L}{\partial \beta} = \frac{1}{\sigma^2} X^T \{Ay - X\beta\} \tag{11}$$

$$\frac{\partial \ln L}{\partial \theta_q} = tr\left[A^{-1}\left(-F_q \bullet W\right)\right] + \frac{1}{\sigma^2}\left[\left(F_q \bullet W\right)y\right]^T \{Ay - X\beta\} \qquad \forall q = 0,1,\ldots,Q \tag{12}$$

where F_q is an $(N \times N)$ matrix with its *i,j*th element as:

$$F_{q_{ij}} = \frac{\partial f(\psi_{ij}^{'} \mid \theta)}{\partial \theta_q} \tag{13}$$

The above system of highly nonlinear equations will not have analytical solutions and numerical methods are needed to solve for the maximum likelihood estimates. Estimation is considerably simplified when β and σ^2 are concentrated out of the log likelihood. Notice that, conditional on some θ the maximum likelihood estimates of β and σ^2, denoted as β_{ML} and ρ_{ML}, can be expressed as:

$$\beta_{ML} = \left(X^T X\right)^{-1} X^T \left(Ay\right) \tag{14}$$

$$\sigma_{ML}^2 = \varepsilon_{ML}^T \varepsilon_{ML} / N \tag{15}$$

where $\varepsilon_{ML} = Ay - X\beta_{ML}$. The above expressions can be substituted into the original log likelihood function to obtain a concentrated log likelihood function, where maximum likelihood estimation involves only the $Q+1$ parameters in θ.

For the anisotropic spatial error model, the log likelihood function is:

$$\begin{aligned}\ln L = &-\left(\frac{N}{2}\right)\ln(2\pi) - \left(\frac{N}{2}\right)\ln(\sigma^2) + \ln\left| \; I - \left[F(\psi \mid \theta) \bullet W\right] \; \right| \\ &-\left(\frac{1}{2\sigma^2}\right)\{y - X\beta\}^T \{I - \left[F(\psi \mid \theta) \bullet W\right]\}^T \{I - \left[F(\psi \mid \theta) \bullet W\right]\}\{y - X\beta\}\end{aligned} \tag{16}$$

which can be written more succinctly as:

$$\ln L = -\left(\frac{N}{2}\right)\ln(2\pi) - \left(\frac{N}{2}\right)\ln(\sigma^2) + \ln|A| - \left(\frac{1}{2\sigma^2}\right)\{y - X\beta\}^T A^T A\{y - X\beta\} \tag{17}$$

where A is defined as before. The first partial derivatives can be obtained as:

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}\{y - X\beta\}^T A^T A\{y - X\beta\} \tag{18}$$

$$\frac{\partial \ln L}{\partial \beta} = \frac{1}{\sigma^2} X^T A^T A\{y - X\beta\} \tag{19}$$

$$\frac{\partial \ln L}{\partial \theta_q} = tr\left[A^{-1}\left(-F_q \bullet W\right)\right] + \frac{1}{\sigma^2}\{y - X\beta\}A^T\left(F_q \bullet W\right)\{y - X\beta\} \ \forall q = 0,1,...,Q \tag{20}$$

The above system of highly nonlinear equations will not have analytical solutions and numerical methods are needed to solve for the maximum likelihood estimates. Once again estimation can be considerably simplified when both β and σ^2 are concentrated out of the log likelihood function. Notice that the maximum likelihood estimates β_{ML} and ρ_{ML} can be expressed as:

$$\beta_{ML} = \left[(AX)^T(AX)\right]^{-1}(AX)^T(Ay) \tag{21}$$

$$\sigma^2_{ML} = \left(A\varepsilon_{ML}\right)^T\left(A\varepsilon_{ML}\right)/N \tag{22}$$

where $\varepsilon_{ML} = y - X\beta_{ML}$. The above expressions can be substituted into the original log likelihood function to obtain a concentrated log likelihood function, which involves fewer parameters and is easier to solve.

3.2 Asymptotic Properties

Formal conditions for consistency of maximum likelihood estimators with dependent observations have been derived in the works of Bates and White (1985), and Heijmans and Magnus (1986a, 1986b, 1986c), and their implications in spatial modelling discussed in Anselin (1988). Generally speaking these conditions are similar to those required of the non-spatial models.

One additional requirement unique to spatial models is that the structure of spatial interaction expressed jointly by the spatial parameter and the spatial weights matrix should not be explosive. In a spatial model where the spatial weights matrix is row-standardized, this requirement is satisfied when ρ is bounded between $(1/\omega_{min}) < \rho < (1/\omega_{max})$, as discussed in Anselin (1980, 1982, 1988), where ω_{min} and ω_{max} are the smallest and largest eigenvalues of the spatial weights matrix. This bound is derived from the decomposition of the log of the Jacobian term, which is shown in Ord (1975) to be:

$$\ln|I - \rho W| = \sum_{i=1}^{N} \ln(1 - \rho\omega_i) \tag{23}$$

where ω_i is the ith eigenvalue of the spatial weights matrix W. For row-standardized W, $\omega_{max} = +1$ and $\omega_{min} > -1$, thus the lower bound on ρ could be less than -1, while the upper bound is strictly +1. So long as ρ does not exceed these bounds, the log of the Jacobian term is well-defined.

Establishing similar bounds in an anisotropic spatial model is more complicated, since the spatial parameter is not a fixed scalar but a function of a number of parameters and it varies in value according to the relative characteristics of the spatial pair. However, an intuitive and perhaps conservative bound will be proposed here. Recall that in an anisotropic model the Jacobian term is:

$$\ln\left|I - [F(\psi \mid \theta) \bullet W]\right| \tag{24}$$

where $F(\psi \mid \theta)$ is an *(N x N)* matrix, and its *i,j*th element is $f(\psi_{ij}^{'} \mid \theta)$. When $f(\psi_{ij}^{'} \mid \theta)$ is equal to a fixed constant for all spatial pairs (i, j), the model collapses to a normal spatial model and the usual bound $\left(\frac{1}{\omega_{\min}}, \frac{1}{\omega_{\max}}\right)$ apply. When $f(\psi_{ij}^{'} \mid \theta)$ is not a fixed constant, however, it is reasonable to expect that if the most extreme values of $f(\psi_{ij}^{'} \mid \theta)$ fall within $\left(\frac{1}{\omega_{\min}}, \frac{1}{\omega_{\max}}\right)$, then the rest of the $f(\psi_{ij}^{'} \mid \theta)$'s would also fall within the bound and the log of the Jacobian term is well-defined. In particular, assume that $f(\psi_{ij}^{'} \mid \theta)$ takes on a linear functional form of:

$$f(\psi_{ij}^{'} \mid \theta) = \theta_0 + \theta_1\psi_1 + \theta_2\psi_2 + \ldots + \theta_Q\psi_Q \tag{25}$$

The extreme values of $f(\psi_{ij}^{'} \mid \theta)$ must exist at the corners of the *Q* dimensional space defined by the extreme values of $\psi_1, \psi_2, \ldots, \psi_Q$, which results in 2^Q corner points. The proposed bound is such that these corner points all fall within $\left(\frac{1}{\omega_{\min}}, \frac{1}{\omega_{\max}}\right)$. This bound is likely to be sufficient to guarantee a well-behaved log Jacobian term. From a practical point of view, if the optimization routine fails to converge, the spatial structure might be ill-defined.

Assuming that all regularity conditions are satisfied, maximum likelihood estimators are consistent and asymptotically normally distributed, with an asymptotic variance-covariance matrix given by the inverse of the information matrix evaluated at the maximum likelihood estimates. Following Anselin (1988), the individual elements of the information matrix for the anisotropic spatial lag model can be derived as:

$$I_{\sigma^2\sigma^2} = \frac{N}{2\sigma^4} \tag{26}$$

$$I_{\beta\beta} = \frac{1}{\sigma^2} X^T X \tag{27}$$

$$\begin{aligned} I_{\theta_q\theta_p} &= tr\left[A^{-1}\left(F_q \bullet W\right)A^{-1}\left(F_p \bullet W\right)\right] \\ &+ \left(\frac{1}{\sigma^2}\right)\left\{\left[\left(F_q \bullet W\right)\left(A^{-1}X\beta\right)\right]^T\left[\left(F_p \bullet W\right)\left(A^{-1}X\beta\right)\right]\right\} \\ &+ tr\left\{\left[\left(F_q \bullet W\right)A^{-1}\right]^T\left[\left(F_p \bullet W\right)A^{-1}\right]\right\} \quad \forall q, p = 0,1,\ldots,Q \end{aligned} \tag{28}$$

$$I_{\sigma^2\beta`} = 0 \tag{29}$$

$$I_{\sigma^2\theta_q} = \left(\frac{1}{\sigma^2}\right) tr\left\{\left[\left(F_q \bullet W\right)A^{-1}\right]^T\right\} \quad \forall q = 0,1,\ldots,Q \tag{30}$$

$$I_{\beta\theta_q} = \frac{1}{\sigma^2} X^T\left[\left(F_q \bullet W\right)A^{-1}\right]X\beta \quad \forall q = 0,1,\ldots,Q \tag{31}$$

where F_q is as defined before and for simplicity $f(\psi_{ij}^{'} \mid \theta)$ is assumed to be linear in θ. And for the anisotropic spatial error model the information matrix can be derived as:

$$I_{\sigma^2\sigma^2} = \frac{N}{2\sigma^4} \tag{32}$$

$$I_{\beta\beta} = \frac{1}{\sigma^2} X^T A^T A X \tag{33}$$

$$\begin{aligned} I_{\theta_q\theta_p} &= tr\left[A^{-1}\left(F_p \bullet W\right)A^{-1}\left(F_q \bullet W\right)\right] \\ &+ tr\left\{\left[\left(F_q \bullet W\right)A^{-1}\right]^T\left[\left(F_p \bullet W\right)A^{-1}\right]\right\} \quad \forall q, p = 0,1,\ldots,Q \end{aligned} \tag{34}$$

$$I_{\sigma^2\beta`} = 0 \tag{35}$$

$$I_{\sigma^2\theta_q} = \left(\frac{1}{\sigma^2}\right) tr\left\{\left[\left(F_q \bullet W\right)A^{-1}\right]^T\right\} \quad \forall q = 0,1,\ldots,Q \tag{36}$$

$$I_{\beta\theta_q} = 0 \quad \forall q = 0,1,\ldots,Q \tag{37}$$

Once maximum likelihood estimates have been obtained, they can be substituted into the above formulas to obtain the estimated information matrix.

4. A case study: Boston housing prices

4.1 The data

The data set used in this empirical study is the well-known housing prices data for the city of Boston in 1970, originally used by Harrison and Rubinfeld (1978) to investigate the effect of air pollution on housing prices. In Pace and Gilley (1997), spatial coordinates are obtained for each of the spatial observations and minor corrections are made for a few recording errors. The data is obtained from UIUC-ACE Spatial Analysis Laboratory. The data comprises of 506 census tracts in the Boston Standard Metropolitan Statistical Area (SMSA) in 1970. The dependent variable is the log of the median value (in 1,000 USD) of the owner-occupied homes in each of the census tracts. 13 explanatory variables are included in the model, which are summarized in table 1.

In terms of the construction of the spatial weights matrix, a triangulation approach is taken. A Delaunay triangulation is first performed, which automatically yields a number for closest neighbours for every spatial point. The spatial weights matrix is then constructed based on the number of spatial neighbours computed for each spatial point and based on distance decay. The main advantage of this approach is that, while it ensures that every spatial point has at least one neighbour, it does not put any restrictions on the total number of neighbours for each spatial point.

4.2 The spatial lag model

Before proceeding to the results from an anisotropic model, it is instructive to present the simple spatial lag model as the baseline model. Table 2 gives the maximum likelihood estimates from a spatial lag model.

ρ_0 is positive and highly significant, indicating that strong positive spatial autocorrelation is present in the 1970 Boston housing prices. Looking at the explanatory variables, *CRIM, ZN, INDUS, NOX, RM, AGE, DIS, TAX, PTRATIO*, and *LSTAT* are found to be significant in determining housing prices. Higher per capita crime rate (*CRIM*), higher pollution (*NOX*), longer travel distance to work (*DIS*), higher property tax rates (*TAX*), higher pupil-teacher ratio (*PTRATIO*), and higher percentage of low status population (*LSTAT*) are all expected to lead to lower housing prices. On the other hand, larger average residential zoning area (*ZN*), larger average number of rooms (*RM*), greater accessibility to highways (*RAD*) and greater proportion of African American population (*B*) are all expected to lead to higher housing prices. Overall these estimates are consistent with one's expectations,

4.3 The anisotropic spatial model

When an investor evaluates her property, she is likely to take into account the property prices of neighbouring areas. But it is unlikely that all neighbouring housing prices are considered as equally important. A reasonable assumption is that she might pay more attention to neighbourhoods with “strong” social characteristics, such as high crime rates or high pollution levels. Looking at the list of explanatory variables, the following four variables: *CRIM* (crime rate), *NOX* (pollution level), *RAD* (easy access to highways), and *LSTAT* (concentration of low status population), appear to attract the most attention.

Furthermore, it also appears reasonable to assume that it is not the absolute level of these variables per se that generates the anisotropic effects. Rather, it is the relative levels of these variables between spatial neighbours that are of great importance. A likely scenario is that neighbours with higher levels of *CRIM*, *NOX*, *RAD*, and *LSTAT* would attract greater attention than those with lower levels of these factors.

The possibility of directional asymmetry in the spatial effects can be incorporated into the anisotropic spatial lag model through the use of directional dummies. Over the space defined by the qth explanatory variable, a directional dummy variable for the ith observation and its jth spatial neighbour is defined as:

$$D_{qij} = \begin{cases} 1 & if \quad X_{qj} > X_{qi} \\ 0 & \quad otherwise \end{cases} \tag{38}$$

For example, let X_{qi} correspond to the crime rate of the ith observation. D_{qij} would only record a value of 1 if the jth neighbour of the ith observation has a crime rate higher than that of the ith observation. It is then possible to define D_q as the qth directional matrix, which is of dimension $(N \times N)$ with the i,jth element being D_{qij}. In this case study, there are four such spatial directional matrices, namely D_{CRIM}, D_{NOX}, D_{RAD}, and D_{LSTAT}.

If $f(\psi'_{ij} \mid \theta)$ takes on a linear form, the anisotropic spatial lag model becomes:

$$y = \left[\left(\rho_0 W_1 + \sum_q \rho_q D_q \right) \bullet W \right] y + X\beta + \varepsilon \qquad (39)$$

where $q = \{CRIM, NOX, RAD, LSTAT\}$, and W_1 is a binary contiguity matrix that is not row-standardised, and it acts as a constant term in the linear function $F(\psi \mid \theta)$. More specifically, for the *i*th observation:

$$y_i = \sum_{j=1}^{N} \left(\rho_0 W_{1ij} + \rho_C D_{CRIM,ij} + \rho_N D_{NOX,ij} + \rho_R D_{RAD,ij} + \rho_L D_{LSTAT,ij} \right) W_{ij} y_j + X_i^{`} \beta + \varepsilon_i \qquad (40)$$

If the parameters ρ_C, ρ_N, ρ_R, and ρ_L are all equal to zero, the model collapses to the usual spatial lag model. But if their true values are not equal to zero, the spatial lag model is overly restrictive and yields inconsistent estimates. Table 3 gives the maximum likelihood estimates from the anisotropic spatial lag model.

The first thing that comes to one's attention is that the log likelihood experiences a substantial improvement from 239.0220 (of the spatial lag model) to 254.126 (of the anisotropic spatial model). A likelihood ratio test on the joint significance of the directional matrices D_{CRIM}, D_{NOX}, D_{RAD}, and D_{LSTAT} can be conducted. The LR test statistic is:

$$2(LnLU - \ln LR) = 2[254.126 - 239.0220] = 30.208.$$

Under the null hypothesis that the four directional dummies are jointly insignificant, the *LR* statistic is asymptotically X^2 distributed with four degrees of freedom. At 1% significance level its critical value is 13.277. Clearly, the test statistic exceeds this critical value. Hence the null hypothesis that the spatial effects are isotropic can be rejected. Although the true form of the anisotropic spatial structure may never be known, at least we can say with great confidence that the assumption of isotropy is overly restrictive.

Looking at the estimated spatial parameters associated with the anisotropic directional dummies, it is found that D_{CRIM}, D_{NOX}, and D_{LSTAT}, which correspond to the directional dummies for CRIM (crime), NOX (pollution), and LSTAT (proportion of low status population) respectively, are all positive and significant, and D_{NOX} is slightly less significant than the other two.

Firstly, ρ_C is 0.01301066 and is significant, which suggests that the housing price of the *i*th observation is influenced to a greater extent by its high crime neighbours than by its low crime neighbours. Also notice that the estimated marginal effect of the *CRIM* (crime) variable is negative (-0.006358373) and significant, which suggests that higher crime rate leads to significantly lower housing prices. The combination of a significantly positive directional matrix D_{CRIM} and a significantly negative regressor *CRIM* suggests that if the *i*th observation had a high crime neighbour, its housing price would be influenced to a greater extent by this high crime neighbour than by its other neighbours and it could be significantly lower than it would have been had there not been a high crime neighbour.

Similarly, a positive (0.01494643) and significant ρ_N suggests that the *i*th observation's housing price is influenced to a greater extent by its high pollution neighbours. Also notice that the estimated marginal effect of the NOX (pollution) variable is negative (-0.174673747) and significant, which suggests that higher pollution leads to significantly lower housing prices. The combination of a significantly positive directional dummy D_{NOX} and a significantly negative regressor

NOX^2 suggests that if the *i*th observation had a high pollution neighbour, its housing price would be influenced more strongly by this high pollution neighbour and it could be significantly lower than it otherwise would have been.

Finally, the combination of a positive (0.04057186) and significantly ρ_L and a negative (-0.011451756) and significant estimated marginal effect of the *LSTAT* (proportion of low status population) tells a similar story. If the *i*th observation had a neighbour with high proportion of low status population, its housing price would be influenced more strongly by this neighbour and could be significantly lower than it otherwise would have been.

These results suggest that in 1970 real estate investors in Boston pay significantly more attention to neighbours that have higher crime rates, higher pollution levels, and higher proportion of low status population. These results are consistent with one's expectations, since crime, pollution, and low income population can all be considered as mobile and it is reasonable for investors to pay extra attention to neighbours with high levels of these factors. Also notice that the *LSTAT* variable seems to generate the most significant anisotropic effect out of all four directional dummies. This suggests that out of all the factors that could have lead to asymmetric spatial influences, in 1970 people in Boston worry the most about the concentration of low status population in certain areas.

Finally, it is interesting to look at the remaining variable *RAD* (access to highway). As a regressor, it is found to be positive (0.009704452) and significant. But as a possible generator of anisotropy, it is found to be negative (-0.01014937) and insignificant. In fact, it fails to be significant by only a relatively small margin, which suggests that there is some evidence that *RAD* plays a part in generating anisotropy, though the evidence is not conclusive. The interesting point to note here is that, when a variable appears as a regressor it is used to explain the underlying level of the dependent variable, but as a part of the anisotropic spatial design it is used to explain the dependence structure between spatially connected neighbourhoods. In an anisotropic model the same variable can be used to explain very different aspects of the underlying data generating process and there is no reason why it should be found to be significant in both parts of the model.

5. Conclusion

In this paper, a general anisotropic spatial model is proposed. The functional form of the anisotropic design is highly flexible, and it can be shown that the spatial lag model is a special case nested in this general model. Estimation of such a model is straightforward and the asymptotic properties of the maximum likelihood estimators are well understood. A case study on 1970 Boston housing prices data demonstrates the significant gain in explanatory power when spatial effects are no longer restricted to be isotropic. Also the new model provides additional insights into the workings of the properties market in Boston in 1978. In particular it is found that extra attention is being paid to safety and discrimination against low status population groups seems to be present. Such insights are not possible under isotropic spatial designs.

Table 1: List of explanatory variables in the Boston housing dataset.

Variable Name	Description
CRIM	Per capita crime rate by town.
ZN	Proportion of residential land zoned for lots over 25000 sq.ft per town.
INDUS	Proportions of non-retail business acres per town.
CHAS	1 if tract borders Charles River; 0 otherwise.
NOX^2	Sq nitric oxides concentration (parts per 10 mil) per town.
RM^2	Sq average number of rooms per dwelling.
AGE	Proportions of owner-occupied units built prior to 1940.
DIS	Weighted distances to five Boston employment centres.
RAD	Index of accessibility to radial highways per town.
TAX	Full-value property tax rate per USD 10,000 per town.
PTRATIO	Pupil-teacher ratios per town.
B	$= 1000(Bk - 0.63)^2$, where *Bk* is the proportion of African Americans by town.
LSTAT	Percentage of low status population (population without high school education or are classified as labourers).

Table 2: Maximum likelihood estimates for the spatial lag model.

Spatial Lag Model		
	Estimate	Significance
Log Likelihood	239.0220	
constant	1.7999838024 (10.68362698)	***
CRIM	−0.0067013268 (−6.513428082)	***
ZN	0.0009968185 (2.343677693)	***
INDUS	0.0028692194 (1.510196633)	*
CHAS	0.0183905955 (0.685795062)	
NOX^2	−0.2220620161 (−2.468511634)	***
RM^2	0.0086685499 (8.613266493)	***
AGE	−0.0001937162 (−0.478369062)	
DIS	−0.0322098681 (−5.202552573)	***
RAD	0.0106851834 (5.183081367)	***
TAX	−0.0005098445 (−4.363829302)	***
PTRATIO	−0.0143529634 (−3.378954851)	***
B	0.0003153001 (3.765003975)	***
LSTAT	−0.0168697912 (−10.0629736)	***
σ^2	0.02154431 (1.58E[illegible]01)	***
ρ	0.51342409 (17.47427973)	***

Where: ***: significant at 1% level; **: at 5% level; * at 10% level.

Table 3: Maximum likelihood estimates for the anisotropic spatial lag model.

Anisotropic Spatial Lag Model		
	Estimate	Significance
Log Likelihood	254.126	
constant	1.407385444 (3.199051059)	***
CRIM	-0.006358373 (-6.232422899)	***
ZN	0.000588673 (1.280702195)	
INDUS	0.002471482 (1.355296487)	*
CHAS	0.020261572 (0.782717134)	
NOX^2	-0.174673747 (-1.920734427)	**
RM^2	0.008101809 (3.151117219)	***
AGE	-0.000132529 (-0.338706977)	
DIS	-0.024074631 (-3.853275397)	***
RAD	0.009704452 (4.857887982)	***
TAX	-0.000456523 (-4.035116794)	***
PTRATIO	-0.01328587 (-3.258582928)	***
B	0.000310635 (3.775817121)	***
LSTAT	-0.011451756 (-6.021345128)	***
α	0.01979092 (15.697761)	***
γ_0	0.57485994 (19.53072717)	***
γ_C	0.01301066 (1.83190089)	**
γ_N	0.01494643 (1.506389911)	*
γ_R	-0.01014937 (-1.251093329)	
γ_L	0.04057186 (4.434069025)	***

Where: ***: significant at 1% level; **: at 5% level; * at 10% level.

Literature cited

Anselin, L. (1980). “Estimation Methods for Spatial Autoregressive Structures: A Study in Spatial Econometrics.” *Regional Science Dissertation and Monograph Series 8, Program in Urban and Regional Studies*, Cornell University, Ithaca, NY.

Anselin, L. (1982). “A Note on Small Sample Properties of Estimators in a First Order Spatial Autoregressive Model.” *Environment and Planning A*, Vol 14, p. 1023 – 1030.

Anselin, L. (1988). *Spatial Econometrics: methods and Models*. Dorecht: Kluwer.
Bates, C. & H. White (1985). “A Unified Theory of Consistent Estimation for Parametric Models.” *Econometric Theory 1*, p.151-178.

Besner, C. (2002). “A Spatial Autoregressive Specification with a Comparable Sales Weighting Scheme.” *Journal of Real Estate Research*, Vol.24, p.193-212.

Bodson P., Peeters D., (1975). “Estimations of the coefficients in a linear regression in the presence of spatial autocorrelation: an application to a Belgian labour-demand function.” *Environment and Planning A*, n°7, p. 455-72.

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

Dacey, M. F. (1965). “A Review on Measures of Contiguity for Two and k-Color Maps*.” Technical Report No.2, Spatial Diffusion Study*, Department of Geography, Evanston: Northwestern University.

Getis, A., Aldstadt, J. (2003). “Constructing the Spatial Weights Matrix Using A Local Statistic.” *Geographical Analysis*, Vol 36, No. 2.

Harrison, D. and Rubinfeld, D. L. (1978). “Hedonic Housing Prices and the Demand for Clean Air.” *Journal of Environmental Economics and Management*, Vol.5, p.81-102.

Heijmans, R. & J. Magnus (1986a). “On the First-Order Efficiency and Asymptotic Normality of maximum Likelihood Estimators Obtained from Dependent Observations.” *Statistica Neerlandica 40*, p.169-188.

Heijmans, R. & J. Magnus (1986b). “Consistent Maximum-Likelihood Estimation with Dependent Observations: The General (Non-Normal) Case and the Normal Case.” *Journal of Econometrics 32*, p.253-285.

Heijmans, R. & J. Magnus (1986c). “Asymptotic Normality of Maximum Likelihood Estimators Obtained from Normally Distributed but Dependent Observations.” *Econometric Theory 2*, p.374-412.

Smirnov, O. Anselin, L. (2001). “Fast Maximum Likelihood Estimation of Very Large Spatial Autoregressive Models: a Characteristic Polynomial Approach.” *Computational Statistics and Data Analysis*, Vol 35, p. 301 – 319.

Ord, J. K. (1975). “Estimation methods for models of spatial interaction.” *Journal of the American Statistical Association 70,* p.120-126.