6. Spatial Regression Models for Areal Data Analysis

The primary models of interest for areal data analysis are regression models. In the same way that *geo-regression models* were used to study relations among continuous-data attributes of selected point locations (such as the California rainfall example), the present *spatial regression models* are designed to study relations among attributes of areal units (such as the English Mortality example in Section 1.3 above). The key difference is of course the underlying spatial structure of this data. In the case of geo-regression, the fundamental spatial assumption was in terms *covariance stationarity*, which together with multi-normality, enabled the full distribution of spatial residuals to be modeled by mean of variograms and their associated covariograms. In the present case, this stationarity assumption is replaced by spatial autogressive hypotheses that are based on specific choices of spatial weights matrices, as developed in Section 5. Here we start with the most fundamental spatial autogressive hypothesis in terms of regression residuals themselves.

6.1 The Spatial Errors Model (SEM)

The most direct analogue to geo-regression is the spatial regression already developed in Section 3 above. In particular, if we start with the regression model in (3.1) above, i.e.,

(6.1.1) $$Y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + u_i \;\;,\;\; i = 1,..,n$$

and postulate that dependencies among the regression residuals (errors), u_i, at each areal unit i are governed by the spatial autoregressive model in (3.5) and (3.6), i.e., by

(6.1.2) $$u_i = \rho \sum_j w_{ij} u_j + \varepsilon_i \;\;,\;\; \varepsilon_i \sim N(0,\sigma^2) \;\;,\;\; i = 1,..,n$$

for some choice of spatial weights matrix, $W = (w_{ij} : i, j = 1,..,n)$ [with $diag(W) = 0$] then the resulting model summarized in matrix form by (3.2) and (3.9) as:

(6.1.3) $$Y = X\beta + u \;\;,\;\; u = \rho W u + \varepsilon \;\;,\;\; \varepsilon \sim N(0, \sigma^2 I_n)$$

is now designated as the *Spatial Errors Model* (also denoted as the *SE-model* or simply *SEM*).[1]

As mentioned above, this constitutes the most direct application of the spatial autogressive model in Section 3. In essence it is hypothesized here that all spatial dependencies are among the unobserved *errors* in the model (and hence the name, SEM). In the case of the English Mortality data for example, it is clear that while the Jarman index includes many socio-economic and demographic factors influencing rates of myocardial infarctions, there are surely other factors involved. Moreover, since many of

[1] See footnote 3 below for further discussion of this terminology.

these excluded factors will exhibit spatial dependencies, such dependencies will necessarily be reflected by the corresponding residual errors, u, in (6.1.3).

Before considering other types of autoregressive dependencies, it is of interest to reformulate this model as an instance of the General Linear Regression Model. First, if for notational convenience, we now let

$$B_\rho = I_n - \rho W \tag{6.1.4}$$

then by expression (3.2.5) above, we may solve for *u* in terms of ε as follows:

$$u = (I_n - \rho W)^{-1}\varepsilon = B_\rho^{-1}\varepsilon, \quad \varepsilon \sim N(0, \sigma^2 I_n) \tag{6.1.5}$$

Thus by the *Invariance Theorem* for multi-normal distributions, it follows at once from the multi-normality of ε that *u* is also multi-normal with covariance given by [2]

$$\text{cov}(u) = \text{cov}(B_\rho^{-1}\varepsilon) = B_\rho^{-1}\text{cov}(\varepsilon)(B_\rho^{-1})' \tag{6.1.6}$$

$$= B_\rho^{-1}(\sigma^2 I_n)(B_\rho^{-1})' = \sigma^2 B_\rho^{-1}(B_\rho')^{-1} = \sigma^2(B_\rho' B_\rho)^{-1} = \sigma^2 V_\rho$$

where the *spatial covariance structure*, V_ρ, is given by[3]

$$V_\rho = (B_\rho' B_\rho)^{-1} \tag{6.1.7}$$

This in turn implies that (6.1.3) can be rewritten as

$$\boxed{Y = X\beta + u, \quad u \sim N(0, \sigma^2 V_\rho)} \tag{6.1.8}$$

which is seen to be an instance of the *General Linear Regression Model* in expression (7.1.8) of Part II, where in this case the matrix *C* is replaced by V_ρ in (6.1.7). This will allow us to apply some of the GLS methods in Section 7.1.1 of Part II to *SE*-models.

Finally, there is a third equivalent way of writing this *SE*-model which is also useful for analysis. If we simply substitute (6.1.5) directly into (6.1.3) and eliminate u altogether, then this same model can be written as

$$\boxed{Y = X\beta + B_\rho^{-1}\varepsilon \;\;, \;\; \varepsilon \sim N(0, \sigma^2 I_n)} \tag{6.1.9}$$

Since all simultaneous relations, $u = \rho W u + \varepsilon$, have been eliminated, expression (6.1.9) is usually called the *reduced form* of (6.3).

---

[2] Here we have used the matrix identities, $(A')^{-1} = (A^{-1})'$, and, $A^{-1}B^{-1} = (BA)^{-1}$, which are established, respectively, in expressions (A3.1.20) and (A3.1.18) of the Appendix.

[3] This terminology is motivated by the fact that all *spatial* aspects of covariance (6.1.6) are defined by V_ρ.

6.2 The Spatial Lag Model (SLM)

An alternative linear model based on the spatial autoregressive model is obtained by assuming that these autoregressive relations are among the *dependent variables* themselves. If we again assume that the underlying spatial relations among areal units are representable by a spatial weights matrix, $W = (w_{ij} : i, j = 1,..,n)$ [with $diag(W) = 0$], then the simplest way to write such a model in terms of W is by modifying expression (6.1.1) as follows,

(6.2.1) $$Y_i = \beta_0 + \rho \sum_h w_{ih} Y_h + \sum_{j=1}^k \beta_j x_{ij} + \varepsilon_i \ , \ i = 1,..,n$$

where again $\varepsilon_i \sim N(0,\sigma^2)$, $i = 1,..,n$. Here the autoregressive term, $\rho \Sigma_{h=1}^n w_{ih} Y_h$, reflects possible dependencies of Y_i on values, Y_h , in other areal units. A standard example of (6.2.1) is in terms of *housing prices*. If the relevant areal units are say city blocks within a metropolitan area, and if Y_i is interpreted as the average price (per square foot) of housing on block i , then in addition to other housing attributes $(x_{ij} : j = 1,..,k)$, of block i , such prices may well be influenced by prices in surrounding blocks. So the relevant autoregressive relations here are among the housing prices, Y , and not the spatial residuals, ε . Such relations are typically called *spatial lag* relations, which motivates the name *spatial lag model* (*SLM*).[4]

6.2.1 Simultaneity Structure

Before analyzing this model in detail, it is important to emphasize one fundamental difference between (6.1.1) and (6.2.1). Since the residuals are here assumed to be independent,[5] one might at first glance conclude that (6.2.1) is nothing more than an OLS model with an added term, $\rho(\Sigma_{h=1}^n w_{ih} Y_h)$, where the unknown spatial dependency parameter, ρ , is simply the relevant "beta coefficient". But the key points to notice are that (*i*) the Y_h values are *random variables*, and moreover that (*ii*) they appear on *both* sides of the equation system (6.2.1), i.e., that Y_i will also appear in equations for Y_h , whenever $w_{hi} > 0$. Thus, in the same way that "opinions" $(u_1,..,u_n)$ among households in Figure 3.1 involved simultaneities, the housing prices $(Y_1,..,Y_n)$ in the present illustration also involve simultaneities. So this is *not* simply another term in an OLS model.

---

[4] At this point, it should be emphasized that (much like "variograms" versus "semivariograms" in the Kriging models of Part II), there is no general agreement regarding the names of various spatial regression models. For example, while we have reserved the term *Spatial Autogressive Model* (*SAR*) for the basic residual process in expression (3.9) above, this term is used by LeSage and Pace (2009) for the spatial lag model (SLM). Our present terminology follows that of the open-source software, GEODA, (to be discussed later) and has the advantage of clarifying *where* the basic spatial autoregressive model is being applied, i.e., to the *error terms* in SEM and to the *dependent variable* in SLM.

[5] Relaxations of this assumption will be considered in the "combined model" of Section 6.3.1 below.

This can be seen more clearly by formalizing this model in matrix terms and solving for its reduced form. By employing the same notation as in (6.1.3), the *Spatial Lag Model* (*SL-model* or simply *SLM*) can be written as

(6.2.2) $$Y = \rho WY + X\beta + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 I_n)$$

As a parallel to (6.2.1), we can rewrite this model by grouping Y terms in (6.2.2) as follows:

(6.2.3) $$Y - \rho WY = X\beta + \varepsilon \;\Rightarrow\; (I_n - \rho W)Y = X\beta + \varepsilon$$

$$\Rightarrow\; B_\rho Y = X\beta + \varepsilon$$

$$\Rightarrow\; Y = B_\rho^{-1} X\beta + B_\rho^{-1}\varepsilon$$

which then yields the corresponding *reduced form* of the *SL*-model:

(6.2.4) $$Y = B_\rho^{-1} X\beta + B_\rho^{-1}\varepsilon \quad , \quad \varepsilon \sim N(0, \sigma^2 I_n)$$

In this reduced form, it should now be clear that the spatial lag term, ρWY, in (6.2.2) is not simply another "regression term".

Finally, one can also view this model as an instance of the General Linear Regression Model, though the correspondence is not as simple as that of *SEM*. In particular, if we now treat the spatial dependency parameter, ρ, as a *known* quantity, or more properly, if we *condition* (6.2.4) on a given value of ρ, then [in a manner similar to the Cholesky transformation in expression (7.1.16) of Part II] we can treat

(6.2.5) $$X_\rho = B_\rho^{-1} X$$

as a transformed data set, and again use (6.1.5) through (6.1.7) to write (6.2.4) as

(6.2.6) $$Y = X_\rho \beta + u \quad , \quad u \sim N(0, \sigma^2 V_\rho)$$

with spatial covariance structure, V_ρ, again given by (6.1.7). The key difference here is that ρ is no longer simply an unknown parameter in the covariance matrix, V_ρ, but now also appears in X_ρ. So while (6.2.6) does permit the GLS methods in Section 7.1.1 in

Part II to also be applied to *SL*-models, these applications are somewhat more restrictive than for *SE*-models.

6.2.2 Interpretation of Beta Coefficients

One final difference between *SE*-models and *SL*-models that needs to be emphasized is the interpretation of the standard beta coefficients, β, in (6.1.8) versus (6.2.4) [or equivalently, (6.1.9) versus (6.2.6)]. Recall that one of the appealing features of OLS is the simple interpretation of beta coefficients. For example, consider an OLS version of the housing price example above, namely

(6.2.7) $$Y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \varepsilon_i \ , \quad i = 1,..,n$$

with $\varepsilon_i \sim N(0,\sigma^2)$, $i = 1,..,n$. If say x_{i1} denotes the *average age* of housing on block i (as a surrogate for structural quality), then one would expect that β_1 is negative. In particular since,

(6.2.8) $$E(Y_i \mid x_{i1},..,x_{ik}) = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} \ , \ i = 1,..,n$$

the value of β_1 should indicate the *expected decrease in mean housing prices on block* i *resulting from a one-year increase in the average age of houses on block* i. More generally, these marginal changes can be expressed as partial derivatives of the form:

(6.2.9) $$\frac{\partial}{\partial x_{ji}} E(Y_i \mid x_{i1},..,x_{ij},..,x_{ik}) = \beta_j \ , \ \ i = 1,..,n, \ j = 1,..,k$$

and are seen to be precisely the corresponding β_j coefficient for variable x_j.

Of course this OLS model ignores spatial dependencies between blocks. So if (6.2.7) is reformulated as an SE-model to account for such dependencies, say of the form in (6.1.8):

(6.2.10) $$Y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + u_i \ , \quad (u_1,..,u_n) \sim N(0,V_\rho)$$

then since $E(u_i) = 0$, $i = 1,..,n$, it follows that (6.2.8) and (6.2.9) continue to hold. Thus, while certain types of spatial dependencies have been accounted for, the interpretation of betas (such as β_1 above) continues to hold.

However, if the major spatial dependencies are among these price levels themselves, so that an *SL*-model is more appropriate, then the situation is far more complex. This can be seen by observing from the reduced form in (6.2.4), together with the "ripple" decomposition of $(I_n - \rho W)^{-1}$ in expression (3.3.26) above that [6]

[6] Here it is implicitly assumed that the convergence condition, $|\rho| < 1/\lambda_w$, holds for ρ and *W*.

(6.2.11) $$E(Y\,|\,X) = B_\rho^{-1}X\beta = (I_n - \rho W)^{-1}X\beta = (I_n + \rho W + \rho^2 W^2 + \cdots)X\beta$$

$$= X\beta + \rho WX\beta + \rho^2 W^2 X\beta + \cdots$$

So the partial derivative in (6.2.9) cannot even be defined without specifying *all* attributes on *all* blocks. Moreover, while (6.2.8) implies that there are no interaction effects between blocks, i.e., that the partial derivatives of $E(Y_i\,|\,x_{1i},..,x_{ji},..,x_{ki})$ with respect to housing attributes on any other block are identically zero, this is no longer true in (6.2.11). For example, if the age of housing on block i is increased, then this not only has a direct effect on expected mean prices in block i, but also has *indirect effects* on prices in all other blocks. Moreover, such indirect effects in turn affect prices in i. So this spatial ripple effect leads to complex interdependencies that must be taken into account when interpreting each beta coefficient. These effects can be summarized by analyzing (6.2.11) in more detail. To do so, we now employ the following notation. For any $n\times m$ matrix, $A = (a_{ij} : i = 1,..,n,\ j = 1,..,m)$, let $A(i,j) = a_{ij}$ denote the $(ij)^{th}$ element of *A*, and let $A(\cdot,j)$ denote the j^{th} column of *A*. In these terms, (6.2.11) can be decomposed as follows:

(6.2.12) $$E(Y\,|\,X) = B_\rho^{-1}X\beta = B_\rho^{-1}\sum\nolimits_{j=1}^k \beta_j X(\cdot,j) = \sum\nolimits_{j=1}^k \beta_j [B_\rho^{-1}X(\cdot,j)]$$

$$= \sum\nolimits_{j=1}^k \beta_j \left[\sum\nolimits_{h=1}^n X(h,j)B_\rho^{-1}(\cdot,h)\right] = \sum\nolimits_{j=1}^k \beta_j \left[\sum\nolimits_{h=1}^n x_{hj}B_\rho^{-1}(\cdot,h)\right]$$

$$= \sum\nolimits_{h=1}^n \sum\nolimits_{j=1}^k x_{hj}\beta_j B_\rho^{-1}(\cdot,h)$$

so that each i^{th} row of $E(Y\,|\,X)$ can be written as

(6.2.13) $$E(Y_i\,|\,X) = \sum\nolimits_{h=1}^n \sum\nolimits_{j=1}^k x_{hj}\beta_j B_\rho^{-1}(i,h)$$

In terms of this decomposition, it now follows that the desired partial derivatives can be obtained directly. First, as a parallel to (6.2.9) we see that

(6.2.14) $$\frac{\partial}{\partial x_{ij}} E(Y_i\,|\,X) = \beta_j B_\rho^{-1}(i,i)$$

So this marginal effect depends not just on β_j but also on the i^{th} diagonal element of B_ρ^{-1}, which has the more explicit form

(6.2.15) $$B_\rho^{-1}(i,i) = 1 + \rho W(i,i) + \rho^2 W^2(i,i) + \cdots$$

$$= 1 + \rho^2 W^2(i,i) + \cdots$$

where the last line follows from the zero-diagonal assumption on W. But since $\rho^2 W^2(i,i)$ together with all higher order effects are positive, it is clear that the effect of each β_j is being inflated by these spatial effects, as described informally above. Moreover it is also clear from (6.2.13) that expected mean prices in i are affected by housing attribute changes in other blocks. In particular, for attribute j in block h, it now follows that

$$\frac{\partial}{\partial x_{hj}} E(Y_i \mid X) = \beta_j B_\rho^{-1}(i,h) \tag{6.2.16}$$

Total effects on $E(Y_i \mid X)$ of attributes in the same areal unit i are designated as *direct effects* by LeSage and Pace (2009, Section 2.7.1), and similarly, the total effects of attributes in different areal units are designated as *indirect effects*. For further analysis of these effects see LeSage and Pace (2009).

6.3 Other Spatial Regression Models

While there are many variations on the *SE*-model and *SL*-model above, we focus only on those that are of particular interest for our purposes.

6.3.1 The Combined Model

When developing the *SL*-model above, a question that naturally arises is why all unobserved factors should be treated as spatially independent. Clearly it is possible to have spatial autoregressive dependencies both among the Y variables and the residuals, ε. If we now distinguish between these by letting M and λ denote the *spatial weights matrix* and *spatial dependency parameter* for the spatial-error component, then one may combine these two models as follows,[7]

$$Y = \rho WY + X\beta + u \ , \ u = \lambda Mu + \varepsilon \ , \ \varepsilon \sim N(0, \sigma^2 I_n) \tag{6.3.1}$$

with corresponding *reduced form* given by

$$(I_n - \rho W)Y = X\beta \ (I_n - \lambda M)^{-1}\varepsilon \tag{6.3.2}$$

$$\Rightarrow \quad Y = (I_n - \rho W)^{-1} X\beta + (I_n - \rho W)^{-1}(I_n - \lambda M)^{-1}\varepsilon$$

However, our primary interest in this model will be to construct comparative tests of *SEM* versus *SLM* as instances of the same model structure. Hence we shall focus on the special case with $M = W$,

[7] This model has been designated by Kelejian and Prucha (2010) as the SARAR(1,1) model, standing for Spatial Autoregressive Model with Autoregressive disturbances of order (1,1).

(6.3.3) $$Y = \rho WY + X\beta + u \;\;,\;\; u = \lambda Wu + \varepsilon \;\;,\;\; \varepsilon \sim N(0, \sigma^2 I_n) \;\;,$$

which we now designate as the *combined model*, with corresponding reduced form:

(6.3.4) $$Y = (I_n - \rho W)^{-1} X\beta + (I_n - \rho W)^{-1}(I_n - \lambda W)^{-1}\varepsilon$$

So for any given spatial weights matrix, *W*, the corresponding *SE*-model (*SL*-model) is seen to be the special case of (6.3.3) with $\rho = 0$ ($\lambda = 0$).

One additional point worth noting here is that while this combined model is mathematically well defined, and can in principle be used to obtain joint estimates of both ρ and λ, these joint estimates are in practice often very unstable. In particular, since both ρ and λ serve as dependency parameters for the *same* matrix, *W*, they in fact play very similar rolls in (6.3.4). But, as will be seen in Section 10.4 below, this instability will turn out to have little effect on the usefulness of this model for comparing *SEM* and *SLM*.

6.3.2 The Spatial Durbin Model

A second model that will prove useful for our comparisons of SEM and SLM can again be motivated by the housing price example above. In particular, if housing prices, Y_i, in block group i are influenced by housing prices in neighboring block groups, then it is not unreasonable to suppose that they may be influenced by other housing attributes in these block groups. If so, then a natural extension of the *SL*-model in (6.2.1) would be to include these spatial effects as additional terms, i.e.,

(6.3.5) $$Y_i = \beta_0 + \rho \sum\nolimits_{h \neq i} w_{ih} Y_h + \sum\nolimits_{j=1}^{k} \beta_j x_{ij} + \sum\nolimits_{h=1}^{n} w_{ih} \left(\sum\nolimits_{j=1}^{k} \alpha_j x_{hj} \right) + \varepsilon_i \;\;,\;\; i = 1,..,n$$

Following Anselin (1988) this extended model is designated as the *Spatial Durbin Model* (also *SD-model* or simply *SDM*). This *SD*-model can be written in matrix form by letting $\alpha = (\alpha_1,..,\alpha_k)'$. However, one important additional difference is that (as in all previous models) the matrix, *X*, is defined to include the intercept term in (6.3.5). So here it is convenient to introduce the more specific notation,

(6.3.6) $$X = [1_n, X_v] \quad \text{and} \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_v \end{pmatrix}$$

where both $X_v [= (x_1,..,x_k)]$ and β_v now refer explicitly to the explanatory *variables*. With this additional notation, (6.3.5) can be written in matrix form as follows: [8]

[8] It is of interest to note here that in many ways it seems more natural to use *X* for the *x* variables, and to employ separate notation for the intercept. But while some authors have chosen to do so, including LeSage and Pace (2009) [compare (6.3.7) above with their expression (2.34)], the linear-model notation ($Y = X\beta + \varepsilon$) is so standard that it seems awkward at this point to attempt to introduce new conventions.

(6.3.7) $$Y = \rho WY + \beta_0 1_n + X_v\beta_v + WX_v\alpha + \varepsilon, \;\; \varepsilon \sim N(0,\sigma^2 I_n)$$

As pointed out by LeSage and Pace (2009, Sections 2.2, 6.1) this model is also useful for capturing omitted explanatory variables that may be correlated with the *x* variables. In this sense, it may serve to make the *SL*-model somewhat more robust. However, as developed more fully in Section 10.3 below, our main interest in this model is that it provides an alternative method for comparing *SLM* and *SEM*.

6.3.3 The Conditional Autoregressive (CAR) Model

There is one additional spatial regression model that should be mentioned in view of its wide application in the literature. While this model is conceptually similar to the *SE*-model, it involves a fundamentally different approach from a statistical viewpoint. In terms of our housing price example, rather than modeling the joint distribution of all housing prices $(Y_1,..,Y_n)$ among block groups, this approach focuses on the *conditional distributions* of each housing price, Y_i , given all the others. The advantage of this approach is that it avoids all of the *simultaneity* issues that we have thus far encountered. In particular, since all univariate conditional distributions derivable from a multi-normal distribution are themselves normal, this approach starts off by assuming only that the conditional distribution of each price, Y_i, given any values $(y_h : h \neq i)$ of all other prices $(Y_h : h \neq i)$, is normally distributed. So these distributions are completely determined by their conditional means and variances. To construct these moments, we start by rewriting the reduced *SE*-model in (6.1.9) as follows:

(6.3.8) $$Y = X\beta + B_\rho^{-1}\varepsilon \;\Rightarrow\; Y - X\beta = B_\rho^{-1}\varepsilon \;\Rightarrow\; B_\rho(Y - X\beta) = \varepsilon$$

$$\Rightarrow\; (I_n - \rho W)(Y - X\beta) = \varepsilon$$

$$\Rightarrow\; Y - X\beta - \rho W(Y - X\beta) = \varepsilon$$

$$\Rightarrow\; Y = X\beta + \rho W(Y - X\beta) + \varepsilon$$

But if we now denote the i^{th} row of W by $w_i' = (w_{i1},..,w_{in})$, then the i^{th} line of this relation can be written as,

(6.3.9) $$Y_i = x_i'\beta + \rho w_i'(Y - X\beta) + \varepsilon_i \;=\; x_i'\beta + \rho \sum\nolimits_{h\neq i} w_{ih}(Y_h - x_h'\beta) + \varepsilon_i$$

where the last equality follows from the assumption that $w_{ii} = 0$. This suggests that if we if we now *condition* Y_i on given values $(y_h : h \neq i)$ of $(Y_h : h \neq i)$, then the natural conditional model of Y_i to consider it the following:

(6.3.10) $$Y_i \mid (y_h : h \neq i) = x_i'\beta + \rho \sum_{h \neq i} w_{ih}(y_h - x_h'\beta) + \varepsilon_i \quad , \quad i = 1,..,n$$

where again $\varepsilon_i \sim N(0,\sigma^2),\ i = 1,..,n$. In this form, it is now immediate that

(6.3.11) $$E[Y_i \mid (y_h : h \neq i)] = x_i'\beta + \rho \sum_{h \neq i} w_{ih}(y_h - x_h'\beta),\ \ i = 1,..,n$$

Moreover, since $Y_i \mid (y_h : h \neq i)$ in (6.3.11) is simply a constant plus ε_i it also follows that $Y_i \mid (y_h : h \neq i)$ must be *normally distribibuted* with the same variance as ε_i, i.e.,

(6.3.12) $$\text{var}[Y_i \mid (y_h : h \neq i)] = \sigma^2 \quad , \quad i = 1,..,n$$

Such conditional models are usually designated as *Conditional Autoregressive* (*CAR*) models. The advantages of such conditional formulations are most evident in *Bayesian* spatial models, where standard "Gibbs sampling" procedures for parameter estimation require only the specification of all conditional distributions. However, such Bayesian models are beyond the scope of this NOTEBOOK. [For an excellent discussion of CAR models in a Bayesian context, see Banerjee, Carlin and Gelfand (2004, Section 3.3).]

Thus our present analysis will focus on the *Spatial Errors Model* (*SEM*) and the *Spatial Lag Model* (*SLM*), which are by far the most commonly used spatial regression models. In the next section, we shall develop the basic methods for estimating the parameters of these models. This will be followed in Section 8 with a development of the standard regression diagnostics for these models.