Chapter 1

Spatial Models: Basic Issues*

In this chapter we provide the reader with several basic notions that relate to the use of spatial econometric models. In Section 1.1 we give a definition of the first law of Geography, which is a basic principle behind spatial models. In many applications, space often relates to geographical distance. However, other representations of spaces can be considered. Some illustrative examples are given below. Section 1.2 introduces the concept of a neighbor and presents one of the most important tools in spatial econometrics, namely the spatial weighting matrix. Formal definitions of some typology of a weighting matrix are given in Section 1.3. Finally, the last section relates to spatial weighting matrices that are usually employed in computer studies. Nevertheless, some of the criteria used in Monte Carlo studies are also valid, *mutatis mutandis*, for real world applications.

1.1 ILLUSTRATIONS INVOLVING SPATIAL INTERACTIONS

Spatial models account for the role that space plays in determining many of the variables that economists and other social scientists (such as geographers, planners, and regional scientists, etc.) deal with. On an intuitive level, the typology of a model is based on a simple principle, namely the further apart in space are the observational units, the weaker the connections between them. In many applications space is related to geographic distances, but it can also relate to differences in products, markets, political systems, city size, and a host of other "spaces." The following examples should clarify this[1]:

1. Many empirical models try to explain the level of police expenditures per capita in given areas (e.g., counties or states).[2] Good candidates to explain such

∗ Basic texts in spatial analysis and econometrics are Cliff and Ord (1973, 1981), Anselin (1988), and Cressie (1993). More recent texts and compilations are Anselin and Florax (1995b), Anselin et al. (2004), Anselin and Rey (2014), LeSage and Pace (2004, 2009), Elhorst (2014), and Arbia (2006, 2014). See also a nice overview of spatial models, and the development of spatial econometrics by Anselin (2002, 2009) and Anselin and Bera (1998).

1. The interesting paper by Brueckner (2003) contains additional motivations for spatial models.

2. See, for example, Ajilore and Smith (2011); Di Tella and Schargrodsky (2004); Levitt (1997); Kovandzic and Sloan (2002), and Marvell and Moody (1996) among others.

Spatial Econometrics. http://dx.doi.org/10.1016/B978-0-12-813387-3.00001-9

a variable would be, among others, crime and unemployment rates, average levels of education and income, and the proportion of housing units that are rented in each of the considered areas. However, with respect to each given area, it might also be of interest to consider the levels of police expenditures per capita in areas surrounding that given area. That is, one might expect that the higher the expenditures in these surrounding areas, the higher the level of police expenditures in the given area. The reason for this is that, *ceteris paribus*, if neighboring areas have high levels of police protection, and a given area does not, that given area might be a magnet for criminals and so be more at risk for crime. Clearly, the further away an area is from the one considered, the less important its characteristics might be to the given area.[3] In this example distance might simply relate to geographical proximity.[4]

2. Similar spillover issues would arise if one were to consider the education budget per capita in a given city as they relate to those of other cities. In this case one would think of competitive issues between cities as they try to attract firms as well as higher income people for tax purposes. In this case the distance between cities might relate to the size of their population rather than their geographic distance. For example, New York might seek to compete with Los Angeles, rather than with College Park, Maryland even though College Park, Maryland is a lot closer to New York than is Los Angeles.[5]

3. As a third set of examples, consider the problem of explaining foreign exchange rates between countries. Clearly, the exchange rate between countries A and B will influence the rate between countries A and C. These relationships may be especially strong during times of crisis. For example, if there is a run on the currency of country A due to its "poor" macro conditions, then speculators may decide to flee from the currency of countries which are geographically close but, perhaps more importantly, have similar macro conditions. This type of spillover between countries in times of crisis is often referred to as contagion. In such cases the connectedness, or "closeness", between countries which may induce contagion may involve more than one characteristic, e.g., geographic proximity, similar credit conditions, the extent of public debt, trade shares, etc.[6]

3. This statement is, generally, referred to as the Tobler's first law of geography.

4. An example of a crime model along the lines described here was given by Kelejian and Robinson (1993). Many other studies involve geographical distance. Shroder (1995) analyzes whether or not states tend to "export" their welfare recipients by setting the level of welfare payments lower than their surrounding areas. Other studies involving geographical distance relate to housing prices, Case (1991), Fingleton (2008b), regional growth rates, LeSage and Fischer (2008), Piras et al. (2012), Dall'Erba et al. (2008), Arbia et al. (2008), Abreu et al. (2005); total factor productivity, Angeriz et al. (2009), labor productivity, Fischer et al. (2009).

5. A study along these lines was given by Vigil (1998).

6. A study involving contagion in foreign markets is Hondroyiannis et al. (2009).

Clearly, there are many other cases involving spatial "spillovers" between various types of "units" and their characteristics. Some evident examples relate to air and water pollution issues, local tax rates and migration, GDP fluctuations between countries, and the extent of foreign trade between countries. A somewhat less evident case relates to various characteristics associated with government quality, such as freedom of speech, of the press, etc. For example, citizens in a given country may be influenced in their demands on their government by political conditions observed in nearby countries!

1.2 CONCEPT OF A NEIGHBOR AND THE WEIGHTING MATRIX

In spatial models the spatial interactions described above are typically accounted for by what has been termed the "weighting (or weights) matrix", which is a concept that was introduced by Moran (1948). The (i, j)th element of this matrix, say w_{ij}, describes the "closeness" between unit i and unit j in terms of a distance measure. If $w_{ij} \neq 0$, unit j is said to be a neighbor of unit i; on the other hand, if $w_{ij} = 0$, unit j is not a neighbor of unit i. Units that are viewed as neighbors to a given unit interact with that given unit in a meaningful way. This interaction could relate to spillovers, externalities, copy-cat policies, geographic proximity issues, industrial structure, similarity of markets, sharing of infrastructure, welfare benefits, banking regulations, tax and reelection issues, just to mention a few.

The description above indicates that the weighting matrix is a matrix that selects neighbors for each unit and indicates the "importance" of each neighbor. For example, suppose we have N observations on a dependent variable, say $y' = (y_1, ..., y_N)$, where y_i is the value on the dependent variable corresponding to the ith unit, e.g., y_i might be the crime rate relating to the ith unit, which might be a region. Suppose the neighbors of unit i are units 1, 2, and 3. Then the ith row of the $N \times N$ weighting matrix, W, will only have the nonzero elements: w_{i1}, w_{i2}, and w_{i3}. If $\sum_{j=1}^{N} w_{ij} = 1$ for all i, the weighting matrix is said to be row normalized. Because a unit is not viewed as its own neighbor $w_{ii} = 0$, $i = 1, ..., N$.

As an example of use, let $W_{i.}$ be the ith row of W and let x_i be a scalar variable in unit i which is used to explain y_i. Then a model such as

$$\begin{aligned} y_i &= b_0 + b_1 x_i + b_2 W_{i.} X + \varepsilon_i, \qquad (1.2.1) \\ X' &= (x_1, ..., x_N), \end{aligned}$$

where ε_i is the corresponding error term, suggests that y_i depends upon x_i (a within unit effect), as well as $\sum_{j=1}^{N} w_{ij} x_j$, which is a weighted sum of the regressor in neighboring units. If W is specified to be row normalized then y_i depends on x_i and a weighted average of this regressor corresponding to neighboring units. Clearly, the simplest weighted average is the uniform, e.g., if y_i has 5 neighbors, then the nonzero weights in the ith row of W are all 1/5. Other weighting schemes will be considered below.

1.3 SOME DIFFERENT WAYS TO SPECIFY SPATIAL WEIGHTING MATRICES

Rewriting the above relation in scalar terms, i.e.,

$$y_i = b_0 + b_1 x_i + b_2 \Sigma_{j=1}^{N} w_{ij} x_j + \varepsilon_i, \tag{1.3.1}$$

one can clearly see that w_{ij} is part of the effect that x_j has on y_i, namely $b_2 w_{ij}$. If unit j is not a neighbor to unit i (i.e., $w_{ij} = 0$), y_i would not be effected by x_j. If unit j is a neighbor of unit i, the effect of x_j on y_i would, in part, depend upon w_{ij} which, in turn, depends upon the distance measure, or some measure of connectedness, between unit i and unit j.

There are various ways researchers have specified w_{ij}. An incomplete list of possible specifications is given below:

(A) Let N_i be the number of neighbors that unit i has. Then, if j is a neighbor to i, researchers often take $w_{ij} = 1/N_i$. In this case W would be a row normalized weighting matrix. Suppose that W is a 10×10 matrix, and that the first unit has three neighbors corresponding to units 2, 5, and 8. Then, the first row of the spatial weighting matrix will be

$$W_{1.} = [0, 1/3, 0, 0, 1/3, 0, 0, 1/3, 0, 0].$$

(B) Let j be a neighbor to i, and let d_{ij} be a distance measure between i and j. Then, one typically wants $w_{ij} \geq 0$, and also wants w_{ij} to be negatively related to d_{ij}, that is, the more distant j and i are, the smaller w_{ij} should be, see, e.g., Fig. 1.3.1 below. Some researchers take $w_{ij} = 1/d_{ij}$. In many cases d_{ij} is taken to be the geographical distance between i and j. Also, in many cases a row normalized version is specified as

$$w_{ij} = \frac{1/d_{ij}}{\Sigma_{r=1}^{N}(1/d_{ir})}, \quad i = 1, ..., N. \tag{1.3.2}$$

(C) A specification need not relate to geographical distance. As an example, let us define INC_r as the income per capita in cross-sectional unit $r = 1, ..., N$; then one specification of w_{ij} would be

$$w_{ij} = |INC_i - INC_j|^{-1}. \tag{1.3.3}$$

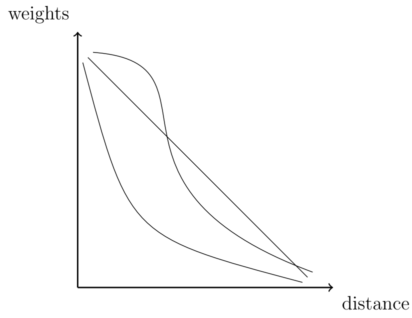

FIGURE 1.3.1 Possible patterns of weights and distance

Although this particular specification is interesting, it also presents a disadvantage. In particular, w_{ij} is not necessarily *bounded* because INC_i and INC_j could be arbitrarily close. As we will see in the next chapter, this is a serious shortcoming because many formal results in spatial econometrics assume that the elements of a weighting matrix are bounded.

Given that income is nonnegative, one possible improvement over (1.3.3) is

$$w_{ij} = 1 - \frac{|INC_i - INC_j|}{INC_i + INC_j}. \tag{1.3.4}$$

The reason for the denominator in (1.3.4) is that elements of a weighting matrix are always taken to be $w_{ij} \geq 0$, and in the absence of the denominator, w_{ij} in (1.3.4) could be negative. At first sight, one might think that w_{ij} in (1.3.4) has an evident shortcoming in that its value is bounded by 1 which occurs when $INC_i = INC_j$. However, this is not a real issue since the maximum value of w_{ij} can be viewed simply as a scaling issue which is accounted for by the coefficient of the term multiplying the spatial weights. For example, one could define a variant of w_{ij} in (1.3.4) as

$$w_{ij} = \alpha \left[1 - \frac{|INC_i - INC_j|}{INC_i + INC_j} \right] \tag{1.3.4A}$$

where α is a preselected positive constant whose value can be very large (or small). If large, the upper limit of w_{ij} would be α. However, in practice one need not be concerned with preselecting the value of a constant such as α. For example, in a model such (1.3.1) if w_{ij} is taken to be as in (1.3.4A), but with α not specified, then that model could be rewritten as

$$y_i = b_0 + b_1 x_i + b_3 \Sigma_{j=1}^{N} \bar{w}_{ij} x_j + \varepsilon_i,$$
$$b_3 = b_2 \alpha,$$

where

$$\bar{w}_{ij} = \left[1 - \frac{|INC_i - INC_j|}{INC_i + INC_j}\right].$$

In this case, the weighting matrix would be defined in terms of $\bar{w}_{ij}$ and the researcher would estimate b_0, b_1, and b_3. The constant term α is not identified, and so cannot be estimated. Although there is no benefit in doing so, if a value were assigned to α, the estimated value of b_2 would be $\hat{b}_2$, where $\hat{b}_2 = \hat{b}_3/\alpha$. Generalizing, the coefficient of the term multiplying the weights, such as α, can always be viewed as accounting for the scale factor of the weights. An exception is the case in which the weighting matrix is row normalized.[7]

Another possible variant of (1.3.3) is

$$w_{ij} = F[-|INC_i - INC_j|] \tag{1.3.5}$$

where $F \geq 0$ is a cumulative distribution function – as one example, the CDF of the normal. In this case $0 < w_{ij} < 1$.

Other economic variables whose differences could signify distance between neighboring units i and j are:

(a) the average level of education
(b) the proportion of housing units that are rental units
(c) ethnic group composition differences
(d) trade shares
(e) the proportion of people in a given area that are registered in a particular political party, e.g., as democrats, republicans, etc.

Finally, in studies involving states, countries, or other geographical areas, researchers often define $w_{ij} = 1$ if area i borders area j at some point, and $w_{ij} = 0$, otherwise. Row normalized versions of such a weighting matrix are typically considered. In the examples above, each unit typically would not be specified to have **all** other units as neighboring units.

In all of the examples we gave, w_{ij} is specified in terms of a single distance measure. In some cases units may differ in important ways by more than one measure. As an example, suppose one were studying the determinants of the volatilities of the GDPs of countries. In a spatial framework, one might want the elements of the weighting matrix to involve more than one comparison of variables between countries. For example, one might want w_{ij} to involve comparisons of unemployment rates, GDP growth rates, inflation rates, etc. Then one might specify w_{ij} in terms of the Euclidean distance between the important

7. In this case we leave it to the reader to demonstrate that a scale factor such as α would cancel if the weighting matrix were row normalized.

variables of the countries such as

$$w_{ij} = 1/(d_{ij}), \tag{1.3.6}$$
$$d_{ij} = [(z_{i1} - z_{j1})^2 + ... + (z_{ir} - z_{jr})^2]^{1/2}$$

where z_{iq} is the q-th "relevant" variable for country i, $q = 1, ..., r$.

As is, the measure in (1.3.6) has three potential problems. The first relates to the units of measurement for the variables involved. The second relates to the relative importance of each variable in the distance measure. The third relates to the possibility that w_{ij} is unbounded.

For instance, suppose there are only two variables in the comparison and they are the unemployment rate and income per capita. Unless care is taken with units of measurement, in this case d_{ij} in (1.3.6) could be completely determined by income comparisons because the unemployment rates are bounded by zero and one, and the measure of income per capita (say in dollars) could lead to much large numbers. The evident solution is that the researcher should use units of measurement such that all the variables determining the distance measure are within bounds of each other, e.g., as one example, they are all within, say, [0, 10].

Consider now the issue of relative importance. It is unlikely that the variables determining d_{ij} in (1.3.6) are all equally important, and so one might want to weight these variables. As an example, instead of (1.3.6), one might **think** of specifying d_{ij} as

$$d_{ij} = [a_1(z_{i1} - z_{j1})^2 + ... + a_r(z_{ir} - z_{jr})^2]^{1/2} \tag{1.3.7}$$

and estimate the parameters $a_1, ..., a_r$ along with the other model parameters. As our discussion in a later chapter will indicate, the specification in (1.3.7) will lead to serious estimation problems. Therefore, in such cases we suggest that instead of defining a single distance measure with numerous weights, such as that in (1.3.7), the researcher should define r distance measures each of which relates to a single variable, such as

$$d_{ij}^q = (z_{iq} - z_{jq})^2, \quad q = 1, ..., r, \tag{1.3.8}$$

and correspondingly, r weighting matrices whose elements are

$$w_{ij}^q = 1/d_{ij}^q, \quad q = 1, ..., r. \tag{1.3.9}$$

Using evident notation, in this case the extended form of the basic model in (1.3.1) would be

$$y_i = b_0 + b_1 X_i + a_1 W_i^1 X + a_2 W_i^2 X + ... + a_r W_i^r X + \varepsilon_i. \tag{1.3.10}$$

$$\begin{bmatrix} NN & NN & NN & NN & NN \\ NN & N & R & N & NN \\ NN & R & i & R & NN \\ NN & N & R & N & NN \\ NN & NN & NN & NN & NN \end{bmatrix}$$

FIGURE 1.4.1 Some neighbor types in computer studies

Finally if, for the chosen variables, the resulting elements of the weighting matrix w_{ij}^q, are not bounded, solutions such as those described in (1.3.4) or (1.3.5) should be considered.

1.4 TYPICAL WEIGHTING MATRICES IN COMPUTER STUDIES

Economists often consider computer studies relating to the properties of spatial models. These studies are typically based on data that are generated using Monte Carlo methods.

Unlike in standard econometric applications, the sample size is often taken as the square of an integer and the "geometry" underlying the sample is often taken in terms of a regular grid (i.e., checkerboard of squares). In many of these studies the elements of the weighting matrices are often determined in such a way that they relate to the game of chess. For instance, in a sample of size 25 one might have a scenario as described in Fig. 1.4.1.

There are 25 squares in Fig. 1.4.1 and each square corresponds to a unit in the computer study. Reading across the first row, the units would be $1, ..., 5$; the second row starts with unit 6, etc. For the ith unit, the squares labeled N and R denote "close" neighbors; the units NN are neighbors that are further away, etc. For unit i, the close neighbors would be units 7, 8, 9, 12, 14, 17, 18, and 19. The pattern of neighbors of unit i described by R is called a rook; this corresponds to some of the movements a rook can make in the game of chess. The pattern described by N and R combined is called a queen for similar reasons. The pattern described by NN is called a double queen.

Another configuration for computer studies is the one based on the Euclidean distance between the units in a checker board of squares based on the "(x, y)" coordinates of the units as described in Fig. 1.4.2 below. For example, in the figure the origin is taken with coordinates $(0, 0)$. Again, the number of the units in the top most row are units, $1, 2, ..., 5$; row 2 begins with unit 6, etc.

In this framework, the distance between unit 12, whose coordinates are $(1, 2)$, and unit 9, whose coordinates are $(3, 3)$ is

$$\begin{aligned} d_{12,9} &= [(1-3)^2 + (2-3)^2]^{1/2} \\ &= 2.236. \end{aligned} \tag{1.4.1}$$

$$\begin{bmatrix} 0,4 & 1,4 & 2,4 & 3,4 & 4,4 \\ 0,3 & 1,3 & 2,3 & 3,3 & 4,3 \\ 0,2 & 1,2 & 2,2 & 3,2 & 4,2 \\ 0,1 & 1,1 & 2,1 & 3,1 & 4,1 \\ 0,0 & 1,0 & 2,0 & 3,0 & 4,0 \end{bmatrix}$$

FIGURE 1.4.2 Neighbors types based on Euclidean distance

In some cases, a configuration such as that in Fig. 1.4.2 is used to compute a spatial weighting matrix which is based on the k-nearest neighbors criteria. That is, corresponding to each unit, the Euclidean distance from all the other units is calculated and sorted in an increasing order. The neighbors for each unit are then taken to be the nearest k of those units.

Still another method used in computer studies is the "k-ahead and k-behind" criterion in a circular world introduced by Kelejian and Prucha (1999). In this framework, each unit is assumed to have k neighbors which are ahead of it in the order of the sample, and k units which are behind it. The number k is typically chosen to be small relative to the sample size. In this scenario, each unit has $2k$ neighbors. Weighting matrices which are built on this framework are typically row normalized, and all of the nonzero elements in the matrix are $1/(2k)$. The meaning of "circular world" is best explained in terms of an illustration.

Suppose the sample size is 9, and a value of k is taken to be 3. In this case each unit would have $2 \times 3 = 6$ neighbors and the weighting matrix would be 9×9. Since the diagonal elements of a weighting matrix are always taken to be zero, the first row of that matrix, which relates to the first unit, would be

$$W_{1.} = [0, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, 0, 0, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}]. \tag{1.4.2}$$

The three neighbors that are ahead of the first unit are units 2, 3, and 4. Because of the circular world condition, the three neighbors that are behind the first unit are units 9, 8, and 7. In a similar light, the second row of the weighting matrix would be

$$W_{2.} = [\frac{1}{6}, 0, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, 0, 0, \frac{1}{6}, \frac{1}{6}]. \tag{1.4.3}$$

In (1.4.3), the three units that are behind the second unit are units $1, 9$, and 8.

Many of the available software packages that deal with spatial econometric models (such as R, Stata, GeoDa/GeoDaSpace/PySAL, and Matlab) have the capability of generating weighting matrices according to the previous criteria. However, the examples presented in the book will be based on R statistical software (see Appendix B).

SUGGESTED PROBLEMS

1. Demonstrate that if w_{ij} were defined as in (1.3.4A), the scale factor α would cancel if the weighting matrix is row normalized.
2. In (1.3.5) suppose F is the CDF of the normal $(0, \sigma^2)$ distribution. If σ^2 is not known, would the model fit into the framework of (1.2.1)?
3. Suppose data are available on the annual percentage changes in the GDPs of N countries. Suggest a model which might explain these percentage changes.
4. For a general sample size, say N, which corresponds to a checkerboard of squares, what is the minimum number of neighbors a unit can have if the weighting matrix is based on a queen pattern?
5. In the k-ahead and k-behind circular world model, suppose $N = 10$ and $k = 2$. Specify the third row of the 10×10 weighting matrix.