C.1 Spatial Econometric Models

James P. LeSage and *R. Kelley Pace*

C.1.1 Introduction

Spatial regression models allow us to account for dependence among observations, which often arises when observations are collected from points or regions located in space. The spatial sample of observations being analyzed could come from a number of sources. Examples of point-level observations would be individual homes, firms, or schools. Regional observations could reflect average regional household income, total employment or population levels, tax rates, and so on. Regions often have widely varying spatial scales (for example, European Union regions, countries, or administrative regions such as postal zones or census tracts).

Each observation is linked to a location which in the case of point-level samples could be latitude-longitude coordinates. For region-level observations we can rely on latitude-longitude coordinates of a point located within the region, perhaps a centroid point.

It is commonly observed that sample data collected for regions or points in space are not *independent*, but rather *positively spatially dependent*, which means that observations from one location tend to exhibit values similar to those from nearby locations.

The data generating process (DGP) that produced the sample data determines the type of spatial dependence. Of course, we never truly know the DGP, so alternative approaches to applied modeling situations have been advocated. One approach is to rely on flexible model specifications that can accommodate a wide range of different possible data generating processes. For example, LeSage and Pace (2009) advocate use of the spatial Durbin model (SDM), since it nests a number of other models as special cases.

A second approach would be to rely on economic or other types of theory to motivate the DGP. For example, Ertur and Koch (2007) use a theoretical model that posits physical and human capital externalities as well as technological interdependence between regions. They show that this leads to a reduced form growth

M.M. Fischer and A. Getis (eds.), *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*, DOI 10.1007/978-3-642-03647-7_18,

regression that should include an average of growth rates from neighboring regions as an explanatory variable in the model.

A third approach might be to rely on a purely econometric argument that favors use of particular models to protect against heterogeneity, omitted variables or other types of problems that arise in applied practice. For example, LeSage and Pace (2008) show that in the case of spatial interaction models of the type discussed in Chapter C.3, omitted variables or latent unobservable influences will lead to a model that includes a spatial lag of the dependent variable.

A fourth approach is to formally incorporate our uncertainty regarding the DGP into the estimation and inference procedure, which is illustrated in Chapter C.4. This involves drawing conclusions about the phenomena being modeled from a host of different model specifications, where each model is probabilistically weighted according to its consistency with the sample data evidence.

Conventional regression models commonly used to analyze cross-section and panel data assume that observations are independent of one another. In the case of spatial data samples where each observation represents a point or region located in space, this means that nearby regions are no more closely related than those more distant. A fundamental tenant of regional analysis is that regions located nearby tend to be more similar than those separated by great distances. This means that *positive spatial dependence* seems more plausible than *spatial independence* when analyzing regional data samples.

As an example, a conventional regression model that relates commuting times to work for region i to the number of persons in region i assumes that these commuting times are independent of those for persons located in a neighboring region j. Since it seems unlikely that regions i and j do not share parts of the road network, we would expect this assumption to be unrealistic. In addition to lack of realism, ignoring a violation of independence between observations can produce estimates that are biased and inconsistent. We pursue a demonstration of this in the sequel.

In our commuting time example, it may seem intuitively appealing to include an average of dependent variables observations from other nearby regions as a right-hand-side explanatory variable in the cross-sectional regression model. This could be formally implemented using a spatial indicator matrix that identifies neighboring observations in our sample. For example, in the case of regions located on a regular lattice we might specify that neighboring observations are the eight regions surrounding each region (ignoring the fact that regions on the edge have less than eight neighbors). This is sometimes referred to as Queen-based contiguity using an analogy to the board moves of the queen piece in the game of Chess. This would result in an extension of the regression model for observation i taking the form shown in Eq. (C.1.1), where the sample contains n observations.

$$y_i = \rho \sum_{j=1}^{n} W_{ij} y_j + \sum_{r=1}^{k} X_{ir} \beta_r + \varepsilon_i. \quad \text{(C.1.1)}$$

In Eq. (C.1.1), the dependent variable for observation i is y_i, the k explanatory variables are X_{ir}, $r = 1, \ldots, k$ with associated coefficients β_r, and the disturbance term is ε_i.[1] The n-by-n matrix $\boldsymbol{W}$ reflects the Queen's contiguity relations between the n regions and we use W_{ij} to denote the (i,j)th element. The matrix $\boldsymbol{W}$ is defined so that each element in row i of the matrix $\boldsymbol{W}$ contains values of zero for regions that are not neighbors to region i, and values of 1/8 for the eight contiguous neighbors to region i. By definition we do not allow region i to be a neighbor to itself, leading to the matrix $\boldsymbol{W}$ having zeros on the main diagonal. This leads to the product: $\sum_{j=1}^{n} W_{ij} y_i$ representing a scalar value equal to the average of values taken by the eight regions neighboring region i. The scalar ρ in model given by Eq. (C.1.1) is a parameter to be estimated that will determine the strength of the average (over all observations $i = 1, \ldots, n$) association between the dependent variable values for regions/observations and the average of those values for their neighbors.

There are of course numerous other ways to define the connectivity structure of the sample observations/regions embodied in the matrix $\boldsymbol{W}$, details of which are beyond the scope of this chapter. In cases involving irregular lattices or point observations these become a consideration in specifying a spatial regression model. For example, one could use some fixed number of nearest neighbors for the case of irregular lattices, a number of neighbors selected using a distance cut-off or some other contiguity definition such as Rook-based contiguity in lieu of 'Queen-based' contiguity described above. There is also flexibility in the way that weights are assigned to neighboring regions/observations. For example, weighting schemes based on the length of shared borders separating regions have been proposed as well as weights exhibiting distance decay (LeSage and Pace 2009, Chapter 4). Conventional wisdom is that the specification of the matrix $\boldsymbol{W}$ exerts a great deal of influence on estimates and inferences regarding the parameters of these models. However, LeSage and Pace (2009) argue that this is an incorrect conclusion that has arisen from invalid interpretation of parameters from these models, a subject that we take up later.

It should be clear that if the parameter $\rho = 0$, we have a conventional regression model: $y_i = \sum_{r=1}^{k} X_{ir} \beta_r + \varepsilon_i$, so a point of interest would be the statistical significance of the coefficient estimate for ρ.

We can write the model in Eq. (C.1.1) using matrix/vector notation as shown in Eq. (C.1.2), where $\boldsymbol{y}$ is an n-by-1 vector containing the dependent variable observations, $\boldsymbol{W}$ is our n-by-n spatial weight matrix that identifies the connectivity or

[1] Without loss of generality, one of the variable vectors X_r could represent an intercept vector of ones.

neighbor structure of the sample observations, $\boldsymbol{X}$ is the n-by-k matrix of explanatory variables which may include an intercept term. The n-by-1 vector $\boldsymbol{\varepsilon}$ represents zero mean, constant variance, zero covariance, normally distributed disturbances, for example, $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2 I_n)$, where we use I_n to denote an n-by-n identity matrix. The scalar parameter ρ and the k-by-1 vector $\boldsymbol{\beta}$ along with the scalar variance parameter σ^2 represent model parameters to be estimated. The associated DGP for this model which we label SAR is shown in Eq. (C.1.3), and the expected value or prediction from this model is shown in Eq. (C.1.4).

$$\boldsymbol{y} = \rho \boldsymbol{W}\, \boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{C.1.2}$$

$$\boldsymbol{y} = (\boldsymbol{I}_n - \rho \boldsymbol{W})^{-1} \boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{I}_n - \rho \boldsymbol{W})^{-1} \boldsymbol{\varepsilon} \tag{C.1.3}$$

$$E(\boldsymbol{y}) = (\boldsymbol{I}_n - \rho \boldsymbol{W})^{-1} \boldsymbol{X}\boldsymbol{\beta} \tag{C.1.4}$$

$$\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I}_n)\ . \tag{C.1.5}$$

The expectation follows from the assumption that elements of the matrix $\boldsymbol{W}$ are fixed/non-stochastic as are observations in the matrix $\boldsymbol{X}$. This results in $E\,[\boldsymbol{I}_n - \rho \boldsymbol{W})^{-1}\, \boldsymbol{\varepsilon}] = (\boldsymbol{I}_n - \rho \boldsymbol{W})^{-1}\, E\,[\boldsymbol{\varepsilon}] = 0$.

There are of course other ways we could envision spatial dependence arising as part of the DGP and these lead to other extensions of the conventional regression model. For example, it may be the case that dependence arises only in the disturbance process leading to the model in Eq. (C.1.6) (which we label SEM), associated DGP in Eq. (C.1.7), and expectation in Eq. (C.1.8).

$$\boldsymbol{y} = \boldsymbol{X}\, \boldsymbol{\beta} + \boldsymbol{u} \tag{C.1.6a}$$

$$\boldsymbol{u} = \rho \boldsymbol{W}\, \boldsymbol{u} + \boldsymbol{\varepsilon} \tag{C.1.6b}$$

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{I}_n - \rho \boldsymbol{W})^{-1} \boldsymbol{\varepsilon} \tag{C.1.7}$$

$$E(\boldsymbol{y}) = \boldsymbol{X}\boldsymbol{\beta} \tag{C.1.8}$$

$$\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I}_n). \tag{C.1.9}$$

Another elaboration of the basic model is one we label SDM shown in Eq. (C.1.10) with associated DGP in Eq. (C.1.11) and expectation in Eq. (C.1.12). In setting forth the SDM model we need to separate out the intercept term from the explanatory variables matrix $\boldsymbol{X}$ because $\boldsymbol{W}\boldsymbol{\iota}_n = \boldsymbol{\iota}_n$, where the n-by-1 intercept vector of ones is denoted by $\boldsymbol{\iota}_n$. This model includes spatial lags of the dependent variable is denoted by the matrix $\boldsymbol{W}\boldsymbol{y}$, and spatial lags of the explanatory variables denoted by the matrix product $\boldsymbol{W}\boldsymbol{X}$ in addition to the conventional explanatory variables $\boldsymbol{X}$. The matrix product $\boldsymbol{W}\boldsymbol{X}$ creates an average of explanatory variable values from neighboring regions which are added to the set of explanatory variables.

$$\boldsymbol{y} = \rho \boldsymbol{W}\boldsymbol{y} + \alpha \boldsymbol{\iota}_n + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{X}\theta + \boldsymbol{\varepsilon} \tag{C.1.10}$$

$$\boldsymbol{y} = (\boldsymbol{I}_n - \rho \boldsymbol{W})^{-1}\left(\alpha \boldsymbol{\iota}_n + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{X}\theta + \boldsymbol{\varepsilon}\right) \tag{C.1.11}$$

$$E(\boldsymbol{y}) = (\boldsymbol{I}_n - \rho \boldsymbol{W})^{-1}\left(\alpha \boldsymbol{\iota}_n + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{X}\theta\right) \tag{C.1.12}$$

$$\boldsymbol{\varepsilon} \sim \mathcal{N}(\sigma^2 \boldsymbol{I}_n)\,. \tag{C.1.13}$$

There are also models based on moving average spatial error processes, $\boldsymbol{u} = (\boldsymbol{I}_n - \rho \boldsymbol{W})\,\boldsymbol{\varepsilon}$ rather than the autoregressive spatial error process, $\boldsymbol{u} = (\boldsymbol{I}_n - \rho \boldsymbol{W})^{-1}\,\boldsymbol{\varepsilon}$ which we have described here (see LeSage and Pace 2009).

An important point to note is that the SEM model has an expectation equal to that from a conventional regression model where independence between the dependent variable observations is part of the maintained hypothesis. In large samples, point estimates for the parameters $\boldsymbol{\beta}$ from the SEM model and conventional regression will be the same, but in small samples there may be an efficiency gain from correctly modeling spatial dependence in the disturbance process. In contrast, the SAR and SDM models which are sometimes referred to as spatial lag models (because they contain terms $\boldsymbol{W}\boldsymbol{y}$ on the right-hand-side) produce expectations that differ from those of the conventional regression model. Use of least-squares regression methods to estimate the parameters of these models will result in biased and inconsistent estimates for the parameters $\boldsymbol{\beta}$ as well as ρ.

C.1.2 Estimation of spatial lag models

From the DGP associated with the SAR model, it should be clear that there is a Jacobian term involved in the transformation from $\boldsymbol{\varepsilon}$ to $\boldsymbol{y}$. The log-likelihood function for the SAR model takes the form in Eqs. (C.1.14) – (C.1.16) (see Ord 1975), where $\boldsymbol{\omega}$ is an n-by-1 vector containing eigenvalues of the matrix $\boldsymbol{W}$. If $\boldsymbol{\omega}$ contains only real eigenvalues, a positive definite variance-covariance matrix is ensured by conditions relating to the minimum and maximum eigenvalues of the matrix $\boldsymbol{W}$. LeSage and Pace (2009, Chapter 4) provide a discussion of situations involving complex eigenvalues that can arise for certain types of spatial weight matrices $\boldsymbol{W}$. Lee (2004) shows that maximum likelihood estimates are consistent for these models.

$$\ln L = -\frac{n}{2}\ln(\pi\sigma^2) + \ln|\boldsymbol{I}_n - \rho\boldsymbol{W}| - \frac{\boldsymbol{e}^{\mathrm{T}}\boldsymbol{e}}{2\sigma^2} \tag{C.1.14}$$

$$\boldsymbol{e} = \boldsymbol{y} - \rho\boldsymbol{W}\,\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} \tag{C.1.15}$$

$$\rho \in \left[\min(\boldsymbol{\omega})^{-1}, \max(\boldsymbol{\omega})^{-1}\right]. \tag{C.1.16}$$

A simple manipulation of the SAR model shown in Eq. (C.1.2): $\boldsymbol{y} - \rho\boldsymbol{W}\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ suggests that the log-likelihood in Eq. (C.1.14) can be concentrated with respect to the parameters $\boldsymbol{\beta}$ and σ^2. This is accomplished using: $\boldsymbol{\beta} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{I}_{\mathrm{n}} - \rho\,\boldsymbol{W})\boldsymbol{y}$ to replace this parameter vector in the full likelihood function. We also replace the parameter σ^2 with $\boldsymbol{e}^{\mathrm{T}}\boldsymbol{e} = (\boldsymbol{y} - \rho\,\boldsymbol{W}\,\boldsymbol{y} - \boldsymbol{X}\,\boldsymbol{\beta})^{\mathrm{T}}(\boldsymbol{y} - \rho\,\boldsymbol{W}\,\boldsymbol{y} - \boldsymbol{X}\,\boldsymbol{\beta})\,n^{-1}$, where $\boldsymbol{\beta}$ is as defined above. Concentrating the full likelihood in this fashion results in a univariate optimization problem over the parameter ρ. Since the parameter ρ has a well-defined range based on the eigenvalues of the matrix $\boldsymbol{W}$, this is a well-defined optimization problem. Given a maximum likelihood estimate for ρ, which we label ρ^*, we can use this estimate to recover maximum likelihood estimates for the parameters $\boldsymbol{\beta}^* = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{I}_n - \rho^*\boldsymbol{W})\,\boldsymbol{y}$, and $\hat{\sigma}^2 = \boldsymbol{e}^{\mathrm{T}}\boldsymbol{e} = (\boldsymbol{y} - \rho^*\boldsymbol{W}\,\boldsymbol{y} - \boldsymbol{X}\,\boldsymbol{\beta}^*)^{\mathrm{T}}(\boldsymbol{y} - \rho^*\boldsymbol{W}\,\boldsymbol{y} - \boldsymbol{X}\,\boldsymbol{\beta}^*)\,n^{-1}$.

Of course, similar likelihood functions exist for other spatial regression models such as the SEM, SDM and moving average processes. See LeSage and Pace (2009) for details regarding these and computationally efficient approaches to optimization. The most computationally challenging part of solving for maximum likelihood estimates using the concentrated log-likelihood function is evaluating the log-determinant for the n-by-n matrix: $\ln|\boldsymbol{I}_n - \rho\boldsymbol{W}|$, since the number of observations n can be large in spatial samples. There has been a great deal of research on computationally efficient ways to calculate this term. As a brief over-

view of the alternative approaches we note that Pace and Barry (1997) discuss use of sparse LU and Cholesky algorithms and set forth a vector expression for the concentrated log-likelihood as a gridded function of values taken by the parameter ρ involved in the univariate optimization problem. Barry and Pace (1999) describe an approach to producing a statistical estimate of this term along with confidence intervals for the estimate. There has been a great deal of literature on approximation approaches (see Pace and LeSage 2003, 2009b; Smirnov and Anselin 2009). In cases involving regular lattices and a repeating pattern of connectivity relations (a regular locational grid such as arises in satellite remote sensing) between the spatial units of observation, analytical formulas can be used to calculate the determinant (LeSage and Pace 2009).

An alternative to tackling what have been perceived as computational difficulties associated with maximum likelihood estimation is to rely on an estimation method that is not likelihood-based. Examples include the instrumental variables approach of Anselin (1988, pp.81-90), the instrumental variables/generalized moments estimator from Kelejian and Prucha (1998, 1999), or the maximum entropy method of Marsh and Mittelhammer (2004). These alternative methods suffer from a number of drawbacks. One is that they can produce dependence parameter estimates (ρ in our discussion) that fall outside the interval defined by the eigenvalue bounds arising from the matrix $\boldsymbol{W}$. In addition, inferential procedures for these methods can be sensitive to implementation issues such as the interaction between the choice of instruments and model specification, which are not always obvious to the practitioner.

There are alternative model specifications such as the matrix exponential spatial specification introduced by LeSage and Pace (2007) which they label MESS that can be estimated using maximum likelihood or Bayesian methods. This spatial regression model specification can be used in situations where the model DGP is that of the SAR or SDM to produce equivalent estimates and inferences. The MESS model eliminates the troublesome determinant term from the likelihood function, allowing rapid maximum likelihood and Bayesian estimation of these models for large spatial samples. LeSage and Pace (2007) provide a closed-form solution for estimates of this model. It is also possible to produce a closed-form solution for maximum likelihood estimates of the SAR, SDM and SEM models discussed here, a recent innovation introduced by LeSage and Pace (2009). These approaches greatly reduce the motivation for reliance on non likelihood-based methods which have been traditionally advocated as a work-around for the perceived computational difficulties of maximum likelihood estimation. These difficulties have been largely resolved with the recent advances described in LeSage and Pace (2009).

The bias of least-squares

As noted, one focus of inference is the magnitude and significance of the parameter ρ, since this distinguishes the SAR model from conventional regression and provides information regarding the strength of spatial dependence between dependent variable observations.

To contrast the maximum likelihood estimate ρ^* to that from least-squares which we label $\hat{\rho}$, consider the matrix expressions in Eq. (C.1.17).

$$\boldsymbol{y} = (\boldsymbol{W}\boldsymbol{y}\,\boldsymbol{X})\begin{pmatrix}\rho\\ \boldsymbol{\beta}\end{pmatrix} + \boldsymbol{\varepsilon} \tag{C.1.17}$$

$$\begin{pmatrix}\hat{\rho}\\ \hat{\boldsymbol{\beta}}\end{pmatrix} = \left[\begin{pmatrix}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{W}^{\mathrm{T}}\\ \boldsymbol{X}^{\mathrm{T}}\end{pmatrix}(\boldsymbol{W}\boldsymbol{y}\;\boldsymbol{X})\right]^{-1}\begin{pmatrix}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{W}^{\mathrm{T}}\\ \boldsymbol{X}^{\mathrm{T}}\end{pmatrix}\boldsymbol{y} = \left[\begin{pmatrix}\boldsymbol{y}\,\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}\,\boldsymbol{y} & \boldsymbol{y}^{\mathrm{T}}\,\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}\\ \boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\,\boldsymbol{y} & \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\end{pmatrix}\right]^{-1}\begin{pmatrix}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}\\ \boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}\end{pmatrix}. \tag{C.1.18}$$

If we assume zero covariance (or orthogonality) between $\boldsymbol{W}\,\boldsymbol{y}$ and $\boldsymbol{X}$, the inverse matrix in Eq. (C.1.18) becomes diagonal having a simple analytical inverse, leading to: $\hat{\rho} = (\boldsymbol{y}^{\mathrm{T}}\,\boldsymbol{W}^{\mathrm{T}}\,\boldsymbol{W}\,\boldsymbol{y})^{-1}\,\boldsymbol{y}^{\mathrm{T}}\,\boldsymbol{W}^{\mathrm{T}}\,\boldsymbol{y}$. Of course, for the case of non-zero covariance between $\boldsymbol{W}\,\boldsymbol{y}$ and $\boldsymbol{X}$ we could rely on a partitioned matrix inverse formulation to produce a similar, but more complicated result than the one we present here.

We can show that the least-squares estimate for the parameter ρ in this simple case of zero covariance is biased and inconsistent. This involved considering whether the definition of consistency: plim($\hat{\rho}$) = ρ, holds true.

$$\begin{aligned}\hat{\rho} &= (\boldsymbol{y}^{\mathrm{T}}\,\boldsymbol{W}^{\mathrm{T}}\,\boldsymbol{W}\,\boldsymbol{y})^{-1}\,\boldsymbol{y}^{\mathrm{T}}\,\boldsymbol{W}^{\mathrm{T}}\boldsymbol{y} = (\boldsymbol{y}^{\mathrm{T}}\,\boldsymbol{W}^{\mathrm{T}}\,\boldsymbol{W}\,\boldsymbol{y})^{-1}\,\boldsymbol{y}^{\mathrm{T}}\,\boldsymbol{W}^{\mathrm{T}}\,(\rho\,\boldsymbol{W}\,\boldsymbol{y} + \boldsymbol{X}\,\boldsymbol{\beta} + \boldsymbol{\varepsilon})\\ &= \rho + (\boldsymbol{y}^{\mathrm{T}}\,\boldsymbol{W}^{\mathrm{T}}\,\boldsymbol{W}\,\boldsymbol{y})^{-1}\,\boldsymbol{y}^{\mathrm{T}}\,\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}\,\boldsymbol{\beta} + (\boldsymbol{y}^{\mathrm{T}}\,\boldsymbol{W}^{\mathrm{T}}\,\boldsymbol{W}\,\boldsymbol{y})^{-1}\,\boldsymbol{y}^{\mathrm{T}}\,\boldsymbol{W}^{\mathrm{T}}\,\boldsymbol{\varepsilon}\\ &= \rho + (\boldsymbol{y}^{\mathrm{T}}\,\boldsymbol{W}^{\mathrm{T}}\,\boldsymbol{W}\,\boldsymbol{y})^{-1}\,\boldsymbol{y}^{\mathrm{T}}\,\boldsymbol{W}^{\mathrm{T}}\,\boldsymbol{\varepsilon}\end{aligned} \tag{C.1.19}$$

where the last equation follows from zero covariance, $\boldsymbol{y}^{\mathrm{T}}\,\boldsymbol{W}^{\mathrm{T}}\,\boldsymbol{X} = 0$. Now consider the probability limit (plim) of the expression: plim $(\boldsymbol{y}^{\mathrm{T}}\,\boldsymbol{W}^{\mathrm{T}}\,\boldsymbol{W}\,\boldsymbol{y})^{-1}\,\boldsymbol{y}^{\mathrm{T}}\,\boldsymbol{W}^{\mathrm{T}}\,\boldsymbol{\varepsilon}$.

The term: Q = plim $(1 / n)(\boldsymbol{y}^{\mathrm{T}}\,\boldsymbol{W}^{\mathrm{T}}\,\boldsymbol{W}\,\boldsymbol{y})^{-1}$ could obtain the status of a finite non-singular matrix with reasonable restrictions/assumptions made in typical applications. Specifically, we must view $\boldsymbol{W}$ as non-stochastic sample data information

and assume that as the sample size increases the number of non-zero elements in each row of the matrix $\boldsymbol{W}$ has a finite limit. In addition, the parameter ρ must obey the eigenvalue bounds to ensure bounded $\boldsymbol{y}$.

We turn attention to the term: $R = \text{plim}\ (1/n)\ \boldsymbol{y}^{\mathrm{T}}\ \boldsymbol{W}^{\mathrm{T}}\ \boldsymbol{\varepsilon}$. Using the model DGP: $\boldsymbol{y} = (\boldsymbol{I}_n - \rho \boldsymbol{W})^{-1}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})$, we find

$$R = \text{plim}\ (1/n)\ \boldsymbol{y}^{\mathrm{T}}\boldsymbol{W}^{\mathrm{T}}\ \boldsymbol{\varepsilon} \tag{C.1.20}$$

$$R = \text{plim}(1/n)[(\boldsymbol{I}_n - \rho \boldsymbol{W})^{-1}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})]^{\mathrm{T}}\ \boldsymbol{W}^{\mathrm{T}}\ \boldsymbol{\varepsilon} \tag{C.1.21}$$

$$R = \text{plim}\ (1/n)\ \boldsymbol{\varepsilon}^{\mathrm{T}}\ (\boldsymbol{I}_{\mathrm{n}} - \rho\ \boldsymbol{W}^{\mathrm{T}})^{-1}\ \boldsymbol{W}^{\mathrm{T}}\ \boldsymbol{\varepsilon} \tag{C.1.22}$$

$$\hat{\rho} = \rho + R. \tag{C.1.23}$$

It should be clear that the plim (the probability limit operator) of the quadratic form in the disturbances shown in Eq. (C.1.22), will not equal zero except in the trivial case where $\rho = 0$, or if the matrix $\boldsymbol{W}$ is strictly triangular. As noted, under the simplifying assumption that $\boldsymbol{Wy}$ and $\boldsymbol{X}$ are uncorrelated, the matrix inverse in Eq. (C.1.18) becomes diagonal having a simple analytical inverse, leading to: $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^{\mathrm{T}}\ \boldsymbol{X})^{-1}\ \boldsymbol{X}^{\mathrm{T}}\ \boldsymbol{y}$. It should be clear that a similar proof of inconsistency could be constructed for the least-squares estimate of this parameter vector. As already noted the maximum likelihood estimate should equal: $\boldsymbol{\beta}^*\ (\boldsymbol{X}^{\mathrm{T}}\ \boldsymbol{X})^{-1}\ \boldsymbol{X}^{\mathrm{T}}\ (\boldsymbol{I}_n - \rho^*\ \boldsymbol{W})\ \boldsymbol{y}$, which requires an unbiased estimate for ρ.

Pace and LeSage (2009a) discuss the biases of OLS when applied to spatially dependent data in more detail. In a richer setting spatial dependence in the explanatory variables as well as in the disturbances can further amplify the bias discussed here.

Bayesian estimation

An alternative to maximum likelihood estimation is Bayesian Markov Chain Monte Carlo (MCMC) estimation set forth in LeSage (1997) for the SAR model.[2] MCMC is based on the idea that a large sample from the Bayesian posterior distribution of our parameters can be used in place of an analytical Bayesian solution where this is difficult or impossible. We designate the posterior distribution using

[2] For an introduction to Bayesian methods in econometrics see Koop (2003).

$p(\boldsymbol{\theta} \mid D)$, where $\boldsymbol{\theta}$ represents the parameters $\rho, \boldsymbol{\beta}, \sigma^2$ and D the sample data. If the sample from $p(\boldsymbol{\theta} \mid D)$ were large enough, we could approximate the form of the posterior density using kernel density estimators or histograms, eliminating the need to know the precise analytical form of this complicated density. Simple statistics could also be used to construct means and variances based on the sampled values taken from the posterior.

The parameters $\boldsymbol{\beta}$ and σ^2 in the SAR model can be estimated by drawing sequentially from the conditional distributions of these two sets of parameters, a process known as Gibbs sampling because of its origins in image analysis, (Geman and Geman 1984). The conditional distributions for these sets of parameters take the form of a multivariate normal distribution (for $\boldsymbol{\beta}$) and inverse Gamma distribution (for σ^2). Gibbs sampling has also been labeled alternating conditional sampling, which seems a more accurate description of the procedure.

To illustrate how this works, assume for simplicity that we knew the true value for the parameter ρ. As already motivated in our discussion of concentrating the likelihood function, the parameter vector $\boldsymbol{\beta}$ can be expressed as: $\boldsymbol{\beta} = (\boldsymbol{X}^{\mathrm{T}} \boldsymbol{X})^{-1} \boldsymbol{X}^{\mathrm{T}} (\boldsymbol{I}_n - \rho \boldsymbol{W}) \boldsymbol{y}$, which is the mean of the normal conditional posterior distribution $\boldsymbol{\beta} \sim \mathcal{N}[(\boldsymbol{X}^{\mathrm{T}} \boldsymbol{X})^{-1} \boldsymbol{X}^{\mathrm{T}} (\boldsymbol{I}_n - \rho \boldsymbol{W}) \boldsymbol{y}, \sigma^2 (\boldsymbol{X}^{\mathrm{T}} \boldsymbol{X})^{-1}]$. We can use this mean expression in conjunction with the associated variance-covariance matrix: $\sigma^2 (\boldsymbol{X}^{\mathrm{T}} \boldsymbol{X})^{-1}$, to construct a multivariate normal draw for the k-by-1 parameter vector $\boldsymbol{\beta}$. We note that being able to condition on the parameter σ^2 (that is assume it is known) is what makes this calculation and multivariate normal draw simple. Similarly, the conditional posterior distribution for the parameter σ^2 takes the form of an inverse Gamma distribution that we denote $IG(a,b)$ with $a = n/2$, and $b = [(\boldsymbol{I}_n - \rho \boldsymbol{W}) \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}]^{\mathrm{T}} [(\boldsymbol{I}_n - \rho \boldsymbol{W}) \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}] /2$. Again, the fact that we can treat the parameter vector $\boldsymbol{\beta}$ as known makes the calculations required to produce this draw simple.

On each pass through the sequence of sampling from the two conditional distributions for $\boldsymbol{\beta}, \sigma^2$, we collect the parameter draws which are used to construct a joint posterior distribution for these model parameters. (We are ignoring the parameter ρ here, assuming it is known.) Gelfand and Smith (1990) demonstrate that sampling from the complete sequence of conditional distributions for all parameters in the model produces a set of estimates that converge in the limit to the true (joint) posterior distribution of the parameters. That is, despite the use of conditional distributions in our sampling scheme, a large sample of the draws can be used to produce valid posterior inferences regarding the joint posterior mean and moments of the parameters.

For the case of the SAR, SEM and SDM models, the conditional distribution for the spatial dependence parameter ρ does not take the form of a known distribution. However, LeSage (1997) describes an approach for sampling from the conditional distribution of this parameter using what has been labeled Metropolis-Hastings sampling, (Metropolis et al. 1953; Hastings 1970). This allows us to estimate spatial regression models using MCMC sampling which involves producing samples from the complete sequence of conditional distributions for the model parameters $\boldsymbol{\beta}, \sigma^2$ and ρ.

C.1.3 Estimates of parameter dispersion and inference

In addition to maximum likelihood or Bayesian estimates for the parameters ρ, $\boldsymbol{\beta}$ and σ^2, we are often interested in inference regarding these. Bayesian MCMC estimation leads to large samples of draws for the model parameters that can be used to construct measures of dispersion used in Bayesian inference. Maximum likelihood inference usually employs *likelihood ratio* (LR), *Lagrange multiplier* (LM), or *Wald* (W) tests. These are equivalent asymptotically, but can differ in small samples. The choice between these methods often comes down to computational convenience or personal preference.

Pace and Barry (1997) propose likelihood ratio tests for hypotheses such as the deletion of a single explanatory variable that exploit the computational advantages of being able to rapidly evaluate the likelihood. Pace and LeSage (2003) discuss use of *signed root deviance* statistics which can be used to transform likelihood ratio tests for single variable deletion to a form similar to *t*-tests.[3] The signed root deviance is the square root of the deviance statistic with a sign matching the sign of the coefficient estimates $\boldsymbol{\beta}$ (Chen and Jennrich 1996). These statistics behave similar to *t*-ratios for large samples, and can be used like a *t*-statistics for hypothesis testing.

Wald inference employs either an analytical or numerical version of the Hessian or the related information matrix to produce a variance-covariance matrix for the estimated parameters. This can be used to construct conventional regression *t*-statistics. An implementation issue is that constructing the analytical Hessian (or information matrix) involves computing the trace of a dense *n*-by-*n* matrix inverse $(\boldsymbol{I}_n - \rho \boldsymbol{W})^{-1}$. LeSage and Pace (2009) provide a number of alternative ways to rapidly approximate elements of the Hessian.

From a computational speed perspective the vector expressions from Pace and Barry (1997) for rapidly evaluating the log-likelihood function makes a purely numerical Hessian feasible for these models. However, there are some drawbacks to implementing this approach in software for general use, since practitioners often work with poorly scaled and multicollinear sample data. Such data can greatly degrade the accuracy of numerical estimates of the derivatives populating the Hessian. A second point is that univariate optimization takes place using the likelihood concentrated with respect to the parameters $\boldsymbol{\beta}$ and σ^2, so a numerical approximation to the full Hessian from the maximum likelihood estimation procedure requires additional work. LeSage and Pace (2009) show how a single computationally difficult term within the analytical Hessian can be replaced with a numerical approximation. This allows the remaining analytical terms to be employed, increasing the accuracy and overcoming scaling problems.

[3] Deviance is minus twice the log-likelihood ratio.

C.1.4 Interpreting parameter estimates

Simultaneous feedback is a feature of the spatial regression model that comes from dependence relations embodied in spatial lag terms such as $\boldsymbol{W}\boldsymbol{y}$. These lead to feedback effects from changes in explanatory variables in a region that neighbors *i*, say region *j*, that will impact the dependent variable for observation/region *i*. This can of course be a valuable feature of these models if we are interested in quantifying spatial spillover effects associated with the phenomena we are attempting to model.

To see how these feedback effects work, consider the data generating process associated with the SAR model, shown in Eq. (C.1.24), to which we have applied the well-known infinite series expansion in Eq. (C.1.25) to express the inverse.

$$\boldsymbol{y} = (\boldsymbol{I}_n - \rho\boldsymbol{W})^{-1}\boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{I}_n - \rho\boldsymbol{W})^{-1}\boldsymbol{\varepsilon} \tag{C.1.24}$$

$$(\boldsymbol{I}_n - \rho\boldsymbol{W})^{-1} = \boldsymbol{I}_n + \rho\boldsymbol{W} + \rho^2\boldsymbol{W}^2 + \rho^3\boldsymbol{W}^3 + \ \ldots \tag{C.1.25}$$

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \rho\boldsymbol{W}\boldsymbol{X}\boldsymbol{\beta} + \rho^2\boldsymbol{W}^2\boldsymbol{X}\boldsymbol{\beta} + \ldots + \boldsymbol{\varepsilon} + \rho\boldsymbol{W}\boldsymbol{\varepsilon} + \rho^2\boldsymbol{W}^2\boldsymbol{\varepsilon} + \rho^3\boldsymbol{W}^3\boldsymbol{\varepsilon} + \ldots \tag{C.1.26}$$

The model statement in Eq. (C.1.26) can be interpreted as indicating that the expected value of each observation y_i will depend on the mean value plus a linear combination of values taken by neighboring observations scaled by the dependence parameter ρ, ρ^2, ρ^3, ...

Consider powers of the row-stochastic spatial weight matrices $\boldsymbol{W}^2$, $\boldsymbol{W}^3$, … that appear in Eq. (C.1.26), where we assume that rows of the weight matrix $\boldsymbol{W}$ are constructed to represent *first-order* contiguous neighbors. The matrix $\boldsymbol{W}^2$ 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, $\boldsymbol{W}^2$ has positive elements on the diagonal. That is, higher-order spatial lags can lead to a connectivity relation for an observation *i* such that $\boldsymbol{W}^2\boldsymbol{X}\boldsymbol{\beta}$ and $\boldsymbol{W}^2\boldsymbol{\varepsilon}$ will extract observations from the vectors $\boldsymbol{X}\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}$ that point back to the observation *i* itself. This is in stark contrast with the conventional independence relation in ordinary least-squares regression 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.

Steady-state equilibrium interpretation

One might suppose that feedback effects would take time, but there is no explicit role for passage of time in our cross-sectional model. Instead, we view the cross-sectional sample of regions as the result of an equilibrium outcome or steady state of the regional process we are modeling. To elaborate on this point, consider a relationship where $\boldsymbol{y}$ represents regional income at time t, denoted by $\boldsymbol{y}_t$, and this depends on current period own-region characteristics $\boldsymbol{X}_t$ such as labor, human and physical capital and associated parameters $\boldsymbol{\beta}$, plus observed income levels of neighboring regions from the past period, $t-1$. This type of *space-time dependence* could be represented by a *space-time lag* variable $\boldsymbol{W}\boldsymbol{y}_{t-1}$, leading to the model in Eq. (C.1.27). It seems reasonable to assume that regional characteristics such as labor, human and physical capital change slowly over time, so we make the simplifying assumption that these do not change over time, that is we set $\boldsymbol{X}_t = \boldsymbol{X}$ in Eq. (C.1.27).[4]

$$\boldsymbol{y}_t = \rho \boldsymbol{W}\, \boldsymbol{y}_{t-1} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t \,. \tag{C.1.27}$$

Note that we can replace $\boldsymbol{y}_{t-1}$ on the right-hand-side of Eq. (C.1.27) with $\boldsymbol{y}_{t-1} = \rho \boldsymbol{W}\, \boldsymbol{y}_{t-2} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_{t-1}$, and continue this type of recursive substitution and in the limit with large t and q produce (LeSage and Pace 2009):

$$\begin{aligned} \lim_{q \to t} E(\boldsymbol{y}_t) &= \lim_{q \to t} E\left[\left(\boldsymbol{I}_n + \rho \boldsymbol{W} + \rho^2 \boldsymbol{W}^2 + \ldots + \rho^{q-1} \boldsymbol{W}^{q-1}\right) \boldsymbol{X}\boldsymbol{\beta} + \rho^q \boldsymbol{W}^q\, \boldsymbol{y}_{t-q} + \boldsymbol{u}\right] \\ &= (\boldsymbol{I}_n - \rho \boldsymbol{W})^{-1} \boldsymbol{X}\boldsymbol{\beta}. \end{aligned} \tag{C.1.28}$$

We conclude from this that the long-run expectation of the model in Eq. (C.1.27), can be interpreted as having a steady-state equilibrium that takes a form consistent with the data generating process for our cross-sectional SAR model. In other words, simultaneous feedback is a feature of the equilibrium steady-state for spatial regression models that include spatial lags of the dependent variable. In the context of our static cross-sectional SAR model where we treat the observed sample as reflecting a steady state equilibrium outcome, these feedback effects appear as instantaneous, but they should be interpreted as showing a movement to the next steady state.

[4] LeSage and Pace (2009) show that one can produce a similar result to that presented here if the explanatory variables $\boldsymbol{X}_t$ evolve over time in a number of ways.

Interpreting the parameters β

LeSage and Pace (2009) point out that interpretation of the parameter vector $\boldsymbol{\beta}$ in the SAR model is different from a conventional least squares interpretation. In least-squares the *r*th parameter, β_r, from the vector $\boldsymbol{\beta}$, is interpreted as representing the partial derivative of $\boldsymbol{y}$ with respect to a change in the *r*th explanatory variable from the matrix $\boldsymbol{X}$, which we write as X_r. In standard least-squares regression where the dependent variable vector contains *independent* observations, changes in observation *i* of the *r*th variable which we denote X_{ir} only influence observation y_i, whereas the SAR model allows this type of change to influence y_i as well as other observations y_j, where $j \neq i$. This type of impact arises due to the interdependence or connectivity between observations in the SAR model.

To see how this works, consider the SAR model expressed as shown in Eq. (C.1.29).

$$(\boldsymbol{I}_n - \rho \boldsymbol{W})\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{C.1.29}$$

$$\boldsymbol{y} = \sum_{r=1}^{k} S_r(\boldsymbol{W})\, X_r + V(\boldsymbol{W})\, \boldsymbol{\varepsilon} \tag{C.1.30}$$

$$S_r(\boldsymbol{W}) = V(\boldsymbol{W})(\boldsymbol{I}_n\, \beta_r) \tag{C.1.31}$$

$$V(\boldsymbol{W}) = (\boldsymbol{I}_n - \rho \boldsymbol{W})^{-1} = \boldsymbol{I}_n + \rho \boldsymbol{W} + \rho^2 \boldsymbol{W}^2 + \rho^3 \boldsymbol{W}^3 + \ldots \tag{C.1.32}$$

To illustrate the role of $S_r(\boldsymbol{W})$, consider the expansion of the data generating process in Eq. (C.1.30) as shown in Eq. (C.1.33).

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \sum_{r=1}^{k} \begin{pmatrix} S_r(W)_{11} & S_r(W)_{12} & \ldots & S_r(W)_{1n} \\ S_r(W)_{21} & S_r(W)_{22} & & \\ \vdots & \vdots & \ddots & \\ S_r(W)_{n1} & S_r(W)_{n2} & \ldots & S_r(W)_{nn} \end{pmatrix} \begin{pmatrix} X_{1r} \\ X_{2r} \\ \vdots \\ X_{nr} \end{pmatrix} + V(\boldsymbol{W})\boldsymbol{\iota}_n \alpha + V(\boldsymbol{W})\boldsymbol{\varepsilon}. \tag{C.1.33}$$

To make the role of $S_r(\boldsymbol{W})$ clear, consider the determination of a single dependent variable observation y_i shown in Eq. (C.1.34).

$$y_i = \sum_{r=1}^{k}\left[(S_r(W)_{i1}X_{1r} + S_r(W)_{i2}X_{2r} + \ldots + S_r(W)_{in}X_{nr}\right] + V(W)_i \ln\alpha + V(W)_i\boldsymbol{\varepsilon}. \tag{C.1.34}$$

It follows from Eq. (C.1.34) that the derivative of y_i with respect to X_{jr} takes the form shown in Eq. (C.1.35), where we use $S_r(W)_{ij}$ to represent the (i, j)th element from the matrix $S_r(\boldsymbol{W})$.

$$\frac{\partial y_i}{\partial X_{jr}} = S_r(W)_{ij}. \tag{C.1.35}$$

In contrast to the least-squares case, the derivative of y_i with respect to X_{ir} usually does not equal β_r, and the derivative of y_i with respect to X_{jr} for $j \neq i$ usually does not equal zero. Therefore, any change to an explanatory variable in a single region (observation) can affect the dependent variable in other regions (observations). This is of course a logical consequence of our simultaneous spatial dependence model. A change in the characteristics of neighboring regions can set in motion changes in the dependent variable that will impact the dependent variable in neighboring regions. These impacts will continue to diffuse through the system of regions.

Since the partial derivative impacts now take the form of a matrix, LeSage and Pace (2009) propose scalar summary measures for these impacts. These cumulate the impacts across all observations that arise from changes in all observations of the explanatory variables and then construct an average impact to simplify interpretation.

The scalar summary measures of impact are based on the idea that the *own derivative* for the ith region takes the form in Eq. (C.1.36), representing the ith diagonal element of the matrix $S_r(\boldsymbol{W})$, which we denote $S_r(W)_{ii}$.

$$\frac{\partial y_i}{\partial X_{ir}} = S_r(W)_{ii} \tag{C.1.36}$$

Of course, the cross-derivative would take the form shown in Eq. (C.1.35) for $i \neq j$, so we can construct scalars by averaging over elements of the matrix $S_r(\boldsymbol{W})$. Averaging over the main diagonal elements of the matrix produces a scalar summary that reflects own-derivatives while averaging over off-diagonal ele-

ments reflect cross-derivatives. The total impact arising from a change in explanatory variable X_r is reflected by all elements of the matrix $S_r(\boldsymbol{W})$. This can be decomposed into direct and indirect or spatial spillover impacts that sum to a total impact arising from a change (on average across all observations) in the variable X_r.

Formally, the LeSage and Pace (2009) definitions for the scalar summary measures of impact are

(a) *Average Direct Impact.* The impact of changes in the ith observation of X_r – which we denote X_{ir} – on y_i could be summarized by measuring the average of main diagonal elements $S_r(W)_{ii}$, from the matrix $S_r(\boldsymbol{W})$.
(b) *Average Total Impact.* The sum across the ith row of $S_r(\boldsymbol{W})$ represents the total impact on individual observation y_i resulting from changing the rth explanatory variable by the same amount across all n observations (for example, $X_r + \delta \boldsymbol{\iota}_n$ where δ is the scalar change). On the other hand, the sum across the ith column reflects the total impact on all y_i arising from changing the rth explanatory variable by an amount in the jth observation (for example, $X_{jr} + \delta$). Averaging either the sum of the row or column sums will produce the same number, which represents the total impact.
(c) *Average Indirect Impact.* This is by definition the difference between the total and direct impacts. This summary impact measure reflects what are commonly thought of as spatial spillovers, or impacts falling on regions other than the own-region.

LeSage and Pace (2009) point to an interpretative distinction between the average total impact summary measure that arises from averaging row-sums versus that from averaging columns-sums. Despite the equality of these two scalar summaries, the average of row-sums could be viewed as reflecting the (average) *Total Impact to an Observation*, whereas the average column-sums are more appropriately interpreted as the (average) *Total Impact from an Observation*.

To elaborate on the distinction between these two interpretative viewpoints, consider a modeling situation where interest centers on how a financial crisis in a single country/observation spills over to produce contagion in financial markets of other countries (Kelejian et al. 2006). This situation can be viewed as a change in the jth observation/country (for example, $X_{jr} + \delta$) impact on all countries y_i, $i = 1, \ldots, n$, or the (average) *Total Impact from an Observation*.

In contrast, if interest centers on how a rise in human capital levels across all regions by some amount will (on average) influence a single region's growth rate, then we are working with the (average) *Total Impact to an Observation* interpretative viewpoint (Dall'erba and LeGallo 2007).

It is easy to see that the numerical values of the summary measures for the two forms of average total impacts set forth above are equal, since the average of the column of row-sums $c_r = S_r(\boldsymbol{W})\boldsymbol{\iota}_n$, equal to $n^{-1}\,\boldsymbol{\iota}_n^{\mathrm{T}} c_r = n^{-1}\,\boldsymbol{\iota}_n^{\mathrm{T}} S_r(\boldsymbol{W})\,\boldsymbol{\iota}_n$. On the other

hand, the average of the row of column-sums $r_r = \boldsymbol{\iota}_n^{\mathrm{T}} S_r(\boldsymbol{W})$, equals $n^{-1} r_r \boldsymbol{\iota}_n$ which is also equal to $n^{-1} \boldsymbol{\iota}_n^{\mathrm{T}} S_r(\boldsymbol{W}) \boldsymbol{\iota}_n$.

The summary measure of total impacts, $n^{-1} \boldsymbol{\iota}_n^{\mathrm{T}} S_r(\boldsymbol{W}) \boldsymbol{\iota}_n$, for the SAR model take the simple form in Eq. (C.1.37) for a model that relies on a row-stochastic $\boldsymbol{W}$ matrix (where the row-sums equal one).

$$n^{-1} \boldsymbol{\iota}_n^{\mathrm{T}} S_r(\boldsymbol{W}) \boldsymbol{\iota}_n = n^{-1} \boldsymbol{\iota}_n^{\mathrm{T}} (\boldsymbol{I}_n - \rho \boldsymbol{W})^{-1} \beta_r \boldsymbol{\iota}_n = (1-\rho)^{-1} \beta_r. \tag{C.1.37}$$

One point to note is that even the average direct impact for this model does not equal the coefficient β_r as in the case of a conventional regression model. The difference between the coefficient estimate β_r and the scalar summary measure of average direct impact arises from the feedback loop reflecting how initial changes in y_i give rise to impacts on neighboring regions y_j which in turn pass through neighboring regions and feedback to region *i*. Of course, the magnitude of this type of feedback will depend on aspects of the spatial regression model used and the resulting parameter estimates. For example, the nature of the connectivity structure $\boldsymbol{W}$ used in the model and the magnitude of the parameter estimates for ρ and $\boldsymbol{\beta}$ both play a role in determining the impacts.

Finally, we should bear in mind the discussion in Section C.1.4, indicating that we should interpret these scalar summary measures of impact as reflecting how changes in the explanatory variables work through the simultaneous dependence system over time to culminate in a new steady state equilibrium. For example, if we find that a ten percent increase in regional levels of human capital give rise to a five percent direct impact on regional income growth and a ten percent indirect impact, we would conclude that these changes would be associated with regional income levels in the new steady-state equilibrium. In the context of our static cross-sectional model we cannot make informative statements about the time that will be required to reach this new equilibrium. Another point is that the indirect impacts will often exceed the direct impacts because the scalar summary measures cumulate impacts over all regions in the model. LeSage and Pace (2009) provide ways to decompose these cumulative impacts into those falling on first-order, second-order and higher-order neighboring regions. These decompositions result in the more intuitive situation where direct impacts exceed indirect impacts falling on first-order, second-order and higher-order neighbors. However, the cumulative impact scalar summary measures add up impacts falling on neighbors of all orders, which often results in indirect or spatial spillover impacts that exceed the direct impacts.

One applied illustration that uses these scalar summary impact estimates can be found in Chapter E.1. The application considers the direct, indirect and total impacts of changes in human capital on labor productivity levels in European Union regions. A number of other applications can be found in LeSage and Pace (2009) in a wide variety of applied contexts.

Inference regarding the impacts

For inference regarding the significance of these impacts, we need to determine their empirical or theoretical distribution. Since the impacts reflect a non-linear combination of the parameters ρ and $\boldsymbol{\beta}$ in the case of the SAR model, working with the theoretical distribution is not particularly convenient. Given the model estimates as well as associated variance-covariance matrix along with the knowledge that maximum likelihood estimates are (asymptotically) normally distributed, we can simulate the parameters ρ and $\boldsymbol{\beta}$. These empirically simulated magnitudes can be used in expressions for the scalar summary measures to produce an empirical distribution of the scalar impact measures.

For the case of Bayesian MCMC estimates we already have a sample of parameter draws for ρ and $\boldsymbol{\beta}$ which can be used in conjunction with the expressions for the scalar summary measures to produce a posterior distribution of the total, direct and indirect impact measures. Gelfand et al. (1990) show that this is a valid approach to derive the posterior distribution for non-linear combinations of model parameters.

For the case of the SAR model, this is relatively straightforward requiring that we need only evaluate the expression: $(1-\rho)^{-1}\beta_r$ to find the total impacts. Calculating the direct impacts requires that we work with the main diagonal of the matrix $(\boldsymbol{I}_n - \rho\boldsymbol{W})^{-1}$ for which LeSage and Pace (2009) provide computationally efficient methods. Recall that we would need to carry out these calculations thousands of times using the simulated parameter values or MCMC draws to determine the empirical measures of dispersion. These measures are used to determine the statistical significance of direct, indirect and total impacts associated with the various explanatory variables in the model, in a fashion similar to use of *t*-statistics in conventional regression models. In more complicated models such as the SDM, the scalar summary measures of impact take more complicated forms, but LeSage and Pace (2009) provide computationally efficient approaches for evaluating these expressions.

An applied illustration of a simulation approach to determining measures of dispersion for these scalar summary impact estimates can be found in Chapter E.1. Another illustration is given in LeSage and Fischer (2008) in the context of model averaging methods discussed in Chapter C.4.

Spatial heterogeneity, spatial dependence, and impacts

Many authors draw a distinction between models of spatial dependence and those of spatial heterogeneity. Typically, spatial dependence models estimate a parameter for each variable while spatial heterogeneity models effectively estimate an *n*-by-*n* matrix of parameters. The Casetti expansion method (Casetti 1997, see also Chapter C.6) and GWR (Fotheringham et al. 2002; see also Chapter C.5) exemplify this approach.

However, the distinction between models of spatial dependence and those of spatial heterogeneity is not as clear as it might initially appear. To motivate this discussion, consider the usual linear model (C.1.38) with the parameters written in matrix form in Eqs. (C.1.39) to (C.1.41).

$$E(\mathbf{y}) = X_1 \beta_1 + X_2 \beta_2 + \ldots + X_k \beta_k \tag{C.1.38}$$

$$E(\mathbf{y}) = \boldsymbol{\Theta}^{(1)} X_1 + \boldsymbol{\Theta}^{(2)} X_2 + \ldots + \boldsymbol{\Theta}^{(k)} X_k \tag{C.1.39}$$

$$B_{ii}^{(r)} = \beta_r \quad r = 1, \ldots, k, \quad i = 1, \ldots, n \tag{C.1.40}$$

$$\boldsymbol{\Theta}^{(r)} = \boldsymbol{B}^{(r)}. \tag{C.1.41}$$

Obviously, in the usual linear model the impact of changing the explanatory variable is the same across observations and a change in the explanatory variable for one observation does not affect the others.

What if we gave geometrically declining weights to the values of the parameters at the neighbors, including parameters at the neighbors of neighbors, and so forth as shown in Eq. (C.1.42). Given the formula for the infinite series expansion, this leads to Eq. (C.1.43). Interestingly, the matrix of parameters implied by this process equals the matrix of impacts ($S_r(\boldsymbol{W})$) discussed previously. As before, we can view the expected value of the dependent variable as a sum of the impacts from all the explanatory variables as in Eq. (C.1.44).

$$\boldsymbol{\Theta}^{(r)} = \boldsymbol{I}_n \boldsymbol{B}^{(r)} + \rho \boldsymbol{W} \boldsymbol{B}^{(r)} + \rho^2 \boldsymbol{W}^2 \boldsymbol{B}^{(r)} + \ldots \tag{C.1.42}$$

$$\boldsymbol{\Theta}^{(r)} = (\boldsymbol{I}_n - \rho \boldsymbol{W})^{-1} \boldsymbol{B}^{(r)} = S_r(\boldsymbol{W}) \tag{C.1.43}$$

$$E(\mathbf{y}) = S_1(\boldsymbol{W}) X_1 + S_2(\boldsymbol{W}) X_2 + \ldots + S_k(\boldsymbol{W}) X_k. \tag{C.1.44}$$

To summarize, spatial dependence involving a spatial lag of the dependent variable implies a form of spatial heterogeneity where the impacts measure the heterogeneity across observations. Error models, however, do not result in heterogenous impacts over space. Therefore, the traditional distinction between spatial

heterogeneity and spatial dependence is meaningful in the case of error models but misleading in the case of spatial autoregressive models.

C.1.5 Concluding remarks

Spatial autoregressive processes represent a parsimonious way to model spatial dependence between observations that often arises in regional economic research. We have shown how basic regression models can be augmented with spatial autoregressive processes to produce models that incorporate simultaneous feedback between regions located in space. It was also shown that conventional regression model estimates that ignore this feedback are biased and inconsistent.

Interpretation of estimates and inferences regarding the spatial connectivity relationships modeled require interpretation based on a steady-state equilibrium view. These models produce a situation where changes in the explanatory variables lead to a series of simultaneous feedbacks that ultimately result in a new steady-state equilibrium. Because we are working with cross-sectional sample data, these model adjustments appear as if they are simultaneous, but we argued that these models can be viewed as containing an implicit time dimension.

The availability of public domain software to implement estimation and inference for the models described here should make these methods widely accessible (Anselin 2006; Bivand 2002; LeSage 1999; Pace 2003).

References

Anselin L (1988) Spatial econometrics: Methods and models. Kluwer, Dordrecht

Anselin L (2006) GeoDa™ 0.9 User's guide, at geoda.uiuc.edu

Barry R, Pace RK (1999) A Monte Carlo estimator of the log determinant of large sparse matrices. Lin Algebra Appl 289(1-3):41-54

Bivand R (2002) Spatial econometrics functions in R: classes and methods. J Geogr Syst 4(4):405-421

Casetti E (1997) The expansion method, mathematical modeling, and spatial econometrics. Int Reg Sci Rev 20(1-2):9-33

Casetti E (2009) Expansion method, dependency, and modeling. In Fischer MM, Getis A (eds) Handbook of applied spatial analysis. Springer, Berlin, Heidelberg and New York, pp.487-505

Chen J, Jennrich R (1996) The signed root deviance profile and confidence intervals in maximum likelihood analysis. J Am Stat Assoc 91:993-998

Dall'erba S, LeGallo J (2007) Regional convergence and the impact of European structural funds over 1989-1999: a spatial econometric analysis. Papers in Reg Sci 87(2):219-244

Ertur C, Koch W (2007) Convergence, human capital and international spillovers. J Appl Econ 22(6):1033-1062

Fischer MM, Bartkowska M, Riedl A, Sardadvar S, Kunnert A (2009) The impact of human capital on regional labor productivity growth in Europe. In Fischer MM, Getis A

(eds) Handbook of applied spatial analysis. Springer, Berlin, Heidelberg and New York, pp.585-597
Fotheringham AS, Brunsdon C, Charlton M (2002) Geographically weighted regression: the analysis of spatially varying relationships. Wiley, New York, Chichester, Toronto and Brisbane
Gelfand AE, Smith AFM (1990) Sampling-based approaches to calculating marginal densities. J Am Stat Assoc 85:398-409
Gelfand AE, Hills SE, Racine-Poon A, Smith AFM (1990) Illustration of Bayesian inference in normal data models using Gibbs sampling. J Am Stat Assoc 85:972-985
Geman S, Geman D (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Transact Patt Anal Machine Intell 6(6):721-741
Hastings WK (1970) Monte Carlo sampling methods using Markov Chains and their applications. Biometrika 57(1):97-109
Kelejian H, Prucha IR (1998) A Generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. J Real Est Fin Econ 17(1):99-121.
Kelejian H, Prucha IR (1999) A generalized moments estimator for the autoregressive parameter in a spatial model. Int Econ Rev 40(2):509-533
Kelejian H, Tavlas HGS, Hondronyiannis G (2006) A spatial modeling approach to contagion among emerging economies. Open Econ5 Rev5 17(4/5):423-442
Koop G (2003) Bayesian econometrics. Wiley, West Sussex [UK]
Lee LF (2004) Asymptotic distributions of quasi-maximum likelihood estimators for spatial econometric models. Econometrica 72(6):1899-1926
LeSage JP (1997) Bayesian estimation of spatial autoregressive models. Int Reg Sci Rev 20(1-2):113-129
LeSage JP (1999) Spatial econometrics using MATLAB a manual for the spatial econometrics toolbox functions, available at www.spatial-econometrics.com
LeSage JP, Fischer MM (2008) Spatial growth regressions: model specification, estimation and interpretation. Spat Econ Anal 3(3):275-304
LeSage JP, Fischer MM (2009) Spatial econometric methods for modeling origin-destination flows. In Fischer MM, Getis A (eds) Handbook of applied spatial analysis. Springer, Berlin, Heidelberg and New York, pp.409-433
LeSage JP, Pace RK (2007) A matrix exponential spatial specification. J Econometrics 140(1):190-214
LeSage JP, Pace RK (2008) Spatial econometric modeling of origin-destination flows. J Reg Sci 48(5):941-967
LeSage JP, Pace RK (2009) Introduction to spatial econometrics. CRC Press (Taylor and Francis Group), Boca Raton [FL], London and New York
Marsh TL, Mittelhammer RC (2004) Generalized maximum entropy estimation of a first order spatial autoregressive model. In LeSage JP, Pace RK (eds) Advances in econometrics, volume 18. Elsevier, Oxford, pp.203-238
Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. J Chem Physics 21(6):1087-1092
Ord JK (1975) Estimation methods for models of spatial interaction. J Am Stat Assoc 70(349):120-126
Pace RK (2003) Matlab spatial statistics toolbox 2.0, available at www.spatial-statistics.com.
Pace RK, Barry B (1997) Quick computation of spatial autoregressive estimators. Geogr Anal 29(3):232-246
Pace RK, LeSage JP (2003) Likelihood dominance spatial inference. Geogr Anal 35(2):133-147

Pace RK, LeSage JP (2009a) Omitted variables biases of OLS and spatial lag models. In Páez A, LeGallo J, Buliung R, Dall'Erba S (eds) Progress in spatial analysis: theory and methods, and thematic applications. Springer, Berlin, Heidelberg and New York (forthcoming)

Pace K, LeSage JP (2009b) A sampling approach to estimating the log determinant used in spatial likelihood problems. J Geogr Syst (forthcoming)

Parent O, LeSage JP (2009) Spatial econometric model averaging. In Fischer MM, Getis A (eds) Handbook of applied spatial analysis. Springer, Berlin, Heidelberg and New York, pp.435-460

Smirnov O, Anselin L (2009) An O(N) parallel method of computing the Log-Jacobian of the variable transformation for models with spatial interaction on a lattice. Comput Stat Data Anal 53(8):2980-2988

Wheeler DC, Páez A (2009) Geographically weighted regression. 1er MM, Getis A (eds) Handbook of applied spatial analysis. Springer, Berlin, Heidelberg and New York, pp.465-486