Semiparametric Maximum Likelihood Estimates of Spatial Dependence

R. Kelley Pace, James P. LeSage

R. Kelley Pace
and James P. LeSage

Semiparametric Maximum Likelihood Estimates of Spatial Dependence

We semiparametrically model spatial dependence via a combination of simpler weight matrices (termed spatial basis matrices) and fit this model via maximum likelihood. Estimation of the model relies on the intuition that bounds to the log-determinant term in the log-likelihood can provide penalties to overfitting both the level and pattern of spatial dependence. By relying on symmetric and doubly stochastic spatial basis matrices that reflect different weight specifications assigned to neighboring observations, we are able to derive a mathematical expression for bounds on the log-determinant term that appears in the likelihood function. These bounds can be conveniently calculated allowing us to solve for maximum likelihood estimates at the bounds using a simple optimization over two quadratic forms that involve small matrices. An intuitively pleasing aspect of our approach is that the objective function for the bounded log-likelihoods contains one quadratic form equal to the sum-of-squared errors measuring the quality of fit, and another quadratic form reflecting a penalty to overfitting spatial dependence. We apply our semiparametric estimation method to a housing model using 57,647 U.S. census tracts.

The computational exigencies of spatial maximum likelihood estimation have led researchers to concentrate on specifying spatial dependence as a function of a small number of parameters (often just one). However, in many estimation problems the performance displays more sensitivity to the spatial specification than to the traditional independent variable specification (Bell and Bockstael 2000). Tools to flexibly specify spatial dependence and visually examine the impact of alternative choices could help practitioners choose superior models.

We address this issue using an overparameterized model of spatial dependence that assumes a monotonic decline in spatial dependence with the order of the neigh-

boring observations. Estimation of the model relies on the intuition that bounds to the log-determinant term in the log-likelihood can provide penalties to overfitting both the level and pattern of spatial dependence. In our approach, the overall weight matrix results from a convex combination of *spatial basis matrices*. The term spatial basis matrices arises by way of analogy with *B*-splines, that employ linear combinations of basis vectors to produce a vector with a flexible form [see Härdle (1990) for more information on splines]. Similarly, linear combinations of spatial basis matrices can produce a spatial weight matrix with a flexible form. Each spatial basis matrix reflects a different weight specification assigned to neighboring observations. The spatial basis matrices used in the linear combination are symmetric and doubly stochastic (each row and column sums to 1), resulting in an overall spatial weight matrix that is both symmetric and doubly stochastic. Symmetric and doubly stochastic weight matrices have a number of appealing properties that facilitate semiparametric modeling of spatial dependence.

We derive a mathematical expression for bounds on the log-determinant term that appears in the likelihood function for this symmetric doubly stochastic weight matrix. Conveniently, the bounds can be expressed as a function of a weighted sum of the traces of pairwise products among the spatial basis matrices. From a computational standpoint, calculating these traces does not require forming the pairwise products explicitly, so solving for maximum likelihood estimates at the bounds becomes a simple optimization exercise involving two quadratic forms based on small matrices (for example, 23 by 23). An intuitively pleasing aspect of our approach is that the objective functions (bounded log-likelihoods) contain one quadratic form equal to the sum-of-squared errors that measures the quality of fit, and another quadratic form that reflects a penalty to overfitting spatial dependence. Overfitting spatial dependence can arise from overfitting the level of spatial dependence (holding the pattern constant) as well as fitting an overly flexible pattern of spatial dependence (holding the level of dependence constant). Again, this bears resemblance to the type of objective functions used in fitting some forms of splines in nonparametric regression.

The ability to observe semiparametric estimates of the weights given to neighboring observations at both the lower- and upper-bounded log-likelihood provides a feel for the nature and extent of possible spatial dependence that may prove useful in estimation and inference, or as an exploratory tool. We label our approach the *doubly spatial model* to emphasize the double bounding of the log-likelihood and the use of doubly stochastic spatial weight matrices.

In addition to the advantages noted above, inference can be simplified as well, while remaining in the traditional likelihood framework. Given upper- and lower-bounded log-likelihoods for a model and any restricted submodel, one can carry out likelihood ratio tests that take into account the range of the Jacobian term. In some cases, one can reject or fail to reject a null hypothesis with the same validity as in cases where the true log-likelihood is known. For example, suppose the unrestricted lower-bound log-likelihood is -1000 and the unrestricted upper-bound log-likelihood is -900. In addition, assume that imposing one restriction causes the restricted lower-bound log-likelihood to fall to -1300 while the restricted upper-bound log-likelihood falls to -1200. This difference of 200 between the lower bound of the unrestricted log-likelihood and upper bound of the restricted log-likelihood implies that the difference between the true unrestricted and true restricted log-likelihoods must equal or exceed 200. Such a difference certainly exceeds the threshold required for significance of a single hypothesis.

A notable aspect of our approach is that we can carry out inference that considers the range of the Jacobian term in the problem, without explicit computation of the Jacobian term, which greatly accelerates computation of estimates. To demonstrate

scalability of the technique, we semiparametrically estimate the spatial dependence for a housing model using 57,647 observations.

Section 1 develops log-determinant bounds as a function of individual basis matrices and sets up the optimization problem. A small data set involving 506 observations is analyzed in section 2 to examine our approach in a setting where true likelihood function values are known. We show that weights based on the bounds enclose the true weights. Section 2 also provides an application of our semiparametric estimates to a housing model using all continental U.S. census tracts. Section 3 concludes with the key results.

1. THE DOUBLY SPATIAL MODEL

Spatial weight matrices have a number of characteristics that affect both their statistical and numerical properties. Section 1.1 presents a doubly stochastic, symmetric spatial weight matrix composed of primitive spatial basis matrices. This sets up the log-determinant bounds discussed in section 1.2. In turn, the bounds lead to bounded log-likelihoods presented in section 1.3. Section 1.4 discusses how bounded log-likelihoods can lead to the same inferences as would arise from using the exact log-likelihood via likelihood dominance arguments.

1.1 Spatial Weight Matrix Properties

A weight matrix specifies the dependence among n observations and thus its specification and properties greatly affect spatial model computation, estimation, as well as inference.

DEFINITION 1. Let $N^{(1)}, N^{(2)}, \ldots, N^{(m)}$ represent a sequence of m individual nearest neighbor weight matrices such that $N_{ij}^{(k)} = 1$ if observation j is the k-th nearest neighbor to observation i ($i \neq j$) and 0 otherwise ($i, j = 1, \ldots, n$). Let

$$B^{(q)} = S^{(q)} \left[\sum_{k=1}^{m} w_k^{(q)} (N^{(k)} + N^{(k)'}) \right] S^{(q)} \tag{1}$$

represent the qth spatial basis matrix where $w_1^{(q)}, w_2^{(q)}, \ldots, w_m^{(q)}$ represent the weights given to the first, second, and up to the mth neighbor and $S^{(q)}$ is a diagonal n by n matrix that insures that $B^{(q)}$ is doubly stochastic (that is, $B^{(q)} \iota = \iota$ and $\iota' B^{(q)} = \iota'$ where ι is a column vector of ones). Furthermore, assume that

$$w_1^{(q)} + w_2^{(q)} + \ldots + w_m^{(q)} = 1 \, ;$$

$$w_1^{(q)} \geq w_2^{(q)} \geq \ldots \geq w_m^{(q)} \geq 0 \quad q = 1, \ldots, p \, ;$$

$$S_{ii}^{(q)} > 0, \ S_{ij}^{(q)} = 0, (i \neq j) \, .$$

Define the overall spatial weight matrix B as a combination of the p primitive spatial basis matrices:

$$B = a_1 B^{(1)} + a_2 B^{(2)} + \ldots + a_p B^{(p)} \tag{2}$$

where:

$$a_1 + a_2 + ... + a_p = 1 \,;$$

$$a_1, a_2, ..., a_p \geq 0 \,;$$

$$a = [a_1 \quad a_2 \quad ... \quad a_p]' \,.$$

Finally, let T represent a p by p matrix containing all possible traces of pairwise multiplications of the basis matrices. Specifically, $T_{uv} = tr(B^{(u)}B^{(v)})$ for $u,v = 1,...,p$.

The diagonal elements of each spatial basis matrix $B^{(p)}$ equal 0 to preclude observations from directly influencing their own predictions in the statistical model. By the definition of a symmetric doubly stochastic matrix, $B^{(q)}\iota = \iota$ and $\iota'B^{(q)} = \iota'$, where ι is a column vector of ones. Since $B^{(q)}$ comprises a limited number of very sparse individual neighbor matrices, it is also sparse. Note, the weights w_i used in constructing $B^{(q)}$ stay the same or decline in value with the order of neighbors, giving $B^{(q)}$ a monotonic character. In summary, each of the p spatial basis matrices $B^{(q)}$ is a sparse, doubly stochastic, symmetric, non-negative, n by n matrix with a trace equal to 0. The overall spatial weight matrix B inherits these properties as well. See Bavaud (1998), Griffith and Lagona (1998), Tiefelsdorf, Griffith, and Boots (1999) for recent discussions concerning spatial weight matrices.

A numerical example may help clarify construction of B. Let $B_{ex}^{(1)}$ represent a spatial basis matrix with equal weight given to the four nearest neighbors.

$$B_{ex}^{(1)} = \begin{bmatrix} 0 & 0.2500 & 0.2500 & 0.2500 & 0.2500 \\ 0.2500 & 0 & 0.2500 & 0.2500 & 0.2500 \\ 0.2500 & 0.2500 & 0 & 0.2500 & 0.2500 \\ 0.2500 & 0.2500 & 0.2500 & 0 & 0.2500 \\ 0.2500 & 0.2500 & 0.2500 & 0.2500 & 0 \end{bmatrix}.$$

Let $B_{ex}^{(2)}$ represent a spatial basis matrix where different weights are given to the four nearest neighbors.

$$B_{ex}^{(2)} = \begin{bmatrix} 0 & 0.4999 & 0.3105 & 0.1273 & 0.0623 \\ 0.4999 & 0 & 0.2414 & 0.1394 & 0.1194 \\ 0.3105 & 0.2414 & 0 & 0.1816 & 0.2666 \\ 0.1273 & 0.1394 & 0.1816 & 0 & 0.5517 \\ 0.0623 & 0.1194 & 0.2666 & 0.5517 & 0 \end{bmatrix}.$$

If a_{ex1} and a_{ex2} equal 0.5, we can define $B_{ex} = 0.5B_{ex}^{(1)} + 0.5B_{ex}^{(2)}$.

$$B_{ex} = \begin{bmatrix} 0 & 0.3749 & 0.2803 & 0.1887 & 0.1562 \\ 0.3749 & 0 & 0.2457 & 0.1947 & 0.1847 \\ 0.2803 & 0.2457 & 0 & 0.2158 & 0.2583 \\ 0.1887 & 0.1947 & 0.2158 & 0 & 0.4009 \\ 0.1562 & 0.1847 & 0.2583 & 0.4009 & 0 \end{bmatrix}.$$

Note, the individual spatial basis matrices $B_{ex}^{(1)}$ and $B_{ex}^{(2)}$ are doubly stochastic and symmetric, as is the overall example spatial weight matrix B_{ex}. The functions to produce these doubly stochastic scalings as well as to estimate various spatial models reside at www.spatial-statistics.com or www.spatial-econometrics.com.

Turning attention to the character of the weights underlying the spatial basis matrices, Figure 1 shows a plot of the weights $w_1^{(q)} \geq w_2^{(q)} \geq \ldots \geq w_m^{(q)}$ associated with $B^{(q)}$, $q = 1,\ldots,22$. Each of the continuous lines in Figure 1 depicts weights for one of the twenty-two spatial basis matrices. For example, the line extending away from 0.25 represents weights for four nearest neighbors receiving 0.25 each (similar to $B_{ex}^{(1)}$). The fifth nearest neighbor has a weight of 0 which accounts for the segment connecting 0.25 at neighbor 4 to 0 at neighbor 5. Some of the weights in the figure exhibit a linear decline with the order of neighbor, while others take on constant values over a range of neighbors. Obviously, a number of shapes can be produced by assigning different weights to the basis matrices. In implementing the doubly spatial model, users are, of course, free to select weights other than those employed here.

Reliance on doubly stochastic matrices is not inconsistent with the usual approaches to defining spatial weight structures because any symmetric matrix with enough positive entries in each row and column can be transformed to doubly stochastic form. This is accomplished using iterative scaling of each row and column by a diagonal matrix containing the inverse of the square root of the row sums (which equal the column sums). The use of too few neighbors can, however, lead to a matrix for which no scaling will yield a doubly stochastic matrix. This might be a problem in situations where, for example, observations are located along a major transportation route and only one or two neighbors are used. For the case of a single neighbor, the weight matrix is asymmetric so the techniques described here do not apply. For com-
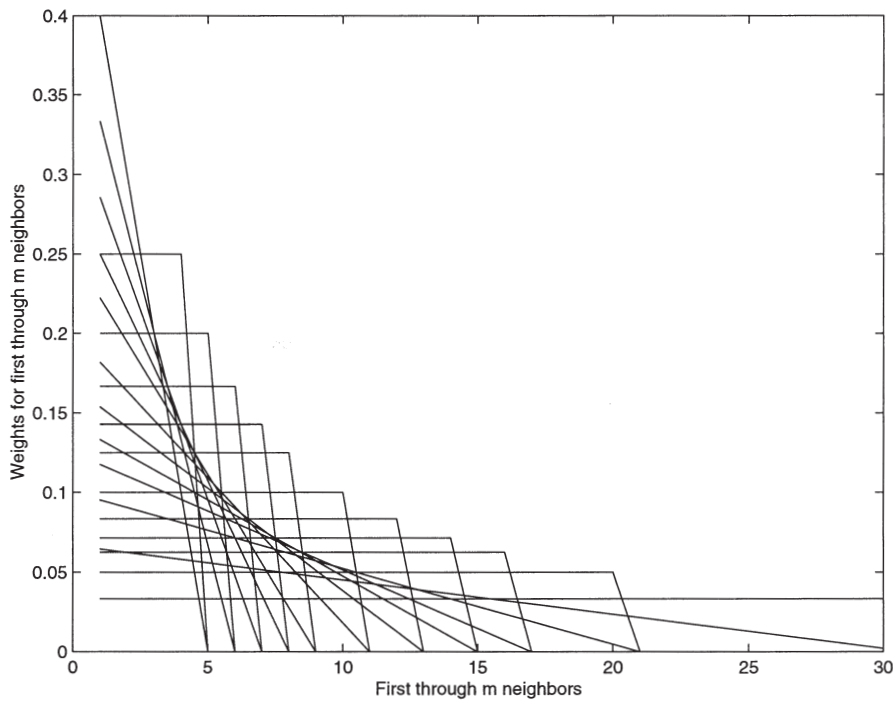

Fig. 1. Basis Matrix Weights

putationally efficient solutions to single neighbor problems based on a closed-form solution, see Pace and Zou (2000). The case of two symmetric neighbors is also unlikely to permit doubly stochastic scaling, but problems involving a few neighbors (for example, edge sites, roads, and peninsulas) can take advantage of sparsity using the methods described by Pace and Barry (1997) to produce rapid log-determinant computations.

We note that problems encountered with doubly stochastic scaling for small numbers of neighbors provided the motivation for use of the basis matrix approach described here. The most sparse basis matrix used here contains four neighbors, but we have been able to form doubly stochastic spatial weight matrices for typical contiguity matrices constructed using Delaunay triangles. It appears that the benefits of doubly stochastic scaling can be combined with the interpretive benefits associated with the usual contiguity matrix. See Bapat and Raghavan (1997, pp. 261–63) for more on doubly stochastic scaling algorithms.

Finally, combining different matrices to model dependence has been employed in statistics generally and in spatial statistics specifically (for example, Streitberg 1979; Cressie 1993, p. 94). Discontinuities and problems of ensuring positive definiteness sometimes impede these attempts. The doubly stochastic scaling automatically ensures positive definiteness of $(I-\alpha B)$. The identical scaling of all the matrices, the quadratic log-determinant bounds, and the use of a grid to set the overall level of dependence appear to have made this easier than the general problem.

1.2 Log-Determinant Bounds

The log-determinant of the spatial weight matrix plays a key role in maximum likelihood spatial estimation. Here, we derive easily computable lower and upper bounds to the log-determinant.

PROPOSITION 1. *For a symmetric nonnegative matrix D with eigenvalues $\lambda_{\min} \geq -1$, $\lambda_{\max} \leq 1$ and $1/\lambda_{\min} < \alpha < 1$, and $tr(D) = 0$,*

$$[\alpha + \ln(1-\alpha)]tr(D^2) \leq \ln|I - \alpha D| \leq -(1/2)\alpha^2 tr(D^2). \tag{3}$$

PROOF. Via a Taylor series expansion:

$$\ln|I-\alpha D| = -\sum_{i=1}^{\infty} tr(D^i)(\alpha^i/i)$$

and $-\sum_{i=1}^{\infty} tr(D^i)(\alpha^i/i) \leq -(1/2)\alpha^2 tr(D^2)$, since $-(1/2)\alpha\ tr(D^2)$ equals the first two terms of the series (given $tr(D) = 0$) while $tr(D^i) \geq 0$ for $i = 3,\dots,\infty$, since D is nonnegative. Hence, $\ln|I-\alpha D| \leq -(1/2)\alpha^2 tr(D^2)$ and thus the upper bound is proved.

For the lower bound, symmetry of D implies real eigenvalues λ such that $|\lambda \leq 1|$ for $i = 1,\dots, n$. Even powers have all nonnegative eigenvalues since $\lambda_i^{2j} \geq 0$ for positive integer j. Recall the trace equals the sum of the real eigenvalues, $tr(D^{2j}) = \sum_{i=1}^{n}\lambda_i^{2j}$ and for positive integer r,

$$tr(D^{2j+r}) = \sum_{i=1}^{n}\lambda_j^{2j}(\lambda_i^r)$$

since the maximum value for λ_i^r is 1 for any positive integer r and since $\lambda_i^{2j} \geq 0$, $tr(D^{2j}) \geq tr(D^{2j+r})$. In particular, $tr(D^2) \geq tr(D^{2+r})$. Since $tr(D) = 0$, $\ln|I-\alpha D| = -\sum_{i=2}^{\infty} tr(D^i)(\alpha^i/i)$. However, $-\sum_{i=2}^{\infty} tr(D^i)(\alpha^i/i) \geq -tr(D^2)\sum_{i=2}^{\infty}(\alpha^i/i)$. Moreover, $-\sum_{i=2}^{\infty}(\alpha^i/i) = (\alpha - \sum_{i=1}^{\infty}(\alpha^i/i)) = (\alpha + \ln(1 - \alpha))$. Hence, $\ln|I-\alpha D| \geq tr(D^2)((\alpha + \ln(1 - \alpha))$ and thus the lower bound is proved. QED.

COROLLARY 1. $(\alpha + \ln(1 - \alpha))a'Ta \leq \ln|I - \alpha B| \leq -(1/2)\alpha^2 a'Ta$.

Since $tr(B^2) = a'Ta$ and B meets the conditions of Proposition 1, the Corollary is proved. QED.

In terms of implementation, $tr(B^{(u)}B^{(v)}) = \sum_{i=1}^{n}\sum_{j=1}^{n}(B^{(u)} \odot B^{(v)'})$ where $\odot$ denotes elementwise or Hadamard multiplication. This requires a maximum of $O(2mn)$ operations (there is a maximum of $2m$ elements in each row of $B^{(u)}$, there are n rows, and sparse matrix methods avoid unnecessary multiplications of the zero elements). As a result, computation of $tr(B^2)$ requires $O(mnp(p + 1)))$ operations, since the cyclical redundancy of the trace requires only $p(p + 1)/2$ basis matrix traces. Note also that given T, computation of $tr(B^2)$ for another set of weights requires only $O(p^2)$ operations.

While we mainly dwell upon the utility of (3) for more complicated combinations of spatial weight matrices, one could employ the bounds to estimate conventional single weight matrix models. An anonymous reviewer estimated the Cressie (1993, pp. 386–89) SIDS data using the exact log-determinant, two versions of the approximation proposed by Griffith and Sone (1995), and an approximation based upon the average of the lower and upper log-determinant bounds from (3). The reviewer obtained estimates of 0.4958, 0.5041, 0.4982, and 0.4792 using these approaches. Insofar as the crude approximation based upon the average of the lower and upper log-determinant bounds from (3) comes reasonably close to the exact estimate, this suggests a potential exists for other simple approximations such as the ones proposed by Griffith and Sone (1995), Martin (1993), as well as Barry and Pace (1999). Whatever approximation used, (3) can provide bounds on the accuracy of the approximation. Note, the bound computations (and also using an average of the bounds as an approximation) for a single weight matrix requires fewer operations than these other, typically more accurate, approximations. In fact, the use of these bounds and an approximation comprised of an average of the bounds can lead to a closed-form solution for the mixed regressive spatially autoregressive estimates at the bounds as well as the approximation. Specifically, the bounds lead to a cubic first-order condition and the approximation composed of the average of the bounds leads to a quartic first-order condition and these all have closed-form solutions. The speed of these solutions could help in spatial data mining applications. Finally, one could use some of these simple approximations and bounds to estimate spatial autoregressions using commercial statistical packages (Griffith 1988).

1.3 The Spatial Autoregressive Profile Log-likelihood

The spatial autoregressive model in equation (4) provides a very useful way of combining independent variables and spatial information to predict the dependent variable.

$$y = X\beta + \alpha By + \varepsilon$$
$$(y - \alpha a_1 B^{(1)}y - \alpha a_2 B^{(2)} - \dots - \alpha a_p B^{(p)}y) = X\beta + \varepsilon \qquad (4)$$

where y represents the nx1 vector of observations on the dependent variable, X represents n observations on k variables, and α represents a scalar autoregressive parameter. Let $M = I - X(X'X)^{-1}X'$, an idempotent matrix. One can write the residuals $e(\omega)$ as the product of an n by $(p + 1)$ matrix R containing the residuals from regressing y and $B^{(q)}y$, for $q = 1,\dots,p$ onto the columns of X and a $p + 1$ by 1 vector, $\omega = [1 - \alpha a']'$, comprised of a 1 followed by complement of the scalar α times the p element column vector a.

$$R = \begin{bmatrix} My & MB^{(1)}y & MB^{(2)}y & \dots & MB^{(p)}y \end{bmatrix}$$

$$e(\omega) = R \begin{bmatrix} 1 \\ -\alpha a \end{bmatrix} = R\omega$$

One can form the sum-of-squared errors easily as in (5),

$$e(\omega)'e(\omega) = \omega'(R'R)\omega = \omega'Q\omega, \tag{5}$$

where Q is a $(p + 1)$ by $(p + 1)$ matrix. The profile log-likelihood becomes

$$L(\omega) = c + \ln|I - \alpha B| - (n / 2)\ln(\omega'Q\omega) \tag{6}$$

where c in (6) represents a constant (Anselin 1988, p. 182). Let A denote the matrix T augmented with a leading row and column of zeros as shown in (7). Note that $\omega'A\omega = \alpha^2 a'Ta$, so the term $\omega'A\omega$ combines the overall level of spatial dependence embodied in α as well as the pattern of spatial dependence found in a.

$$A = \begin{pmatrix} 0 & 0 \\ 0 & T \end{pmatrix}. \tag{7}$$

For normal iid errors, we can write the upper- and lower-bounded profile log-likelihoods, that we designate $L^u(\omega)$ and $L^\ell(\omega)$ as in (8):

$$L^u(\omega) = c - (1 / 2)\omega'A\omega - (n / 2)\ln(\omega'Q\omega);$$

$$L^\ell(\omega) = c - (\alpha + \ln(1 - \alpha))\alpha^{-2}\omega'A\omega - (n / 2)\ln(\omega'Q\omega). \tag{8}$$

Note, the term $\omega'A\omega$ contains the sum of squared weights as part of the overall penalty and thus reduces the variance in the estimated weights. As noted earlier, this resembles the penalty approach used with smoothing splines. The lower-bound likelihood can be interpreted as a penalized likelihood since it has a larger penalty than the actual log-determinant. This will result in less variance among the estimated weights than the weights associated with the actual log-determinant. If the log-determinant term did not exist in the likelihood, one might have invented it anyway to smooth the estimated weights.

A number of ways exist to model the independent variables. We examine both a nonspatial specification for X commonly known as the spatial autoregressive model in (9) and a fully mixed model shown in (10).

$$X_{AR} = \begin{bmatrix} U & \iota \end{bmatrix} \tag{9}$$

$$X_{mixed} = \begin{bmatrix} U & B^{(1)}U & \dots & B^{(p)}U & \iota \end{bmatrix} \tag{10}$$

Note that estimation of the doubly spatial model based on X_{mixed} nests a spatial error autoregression since X_{mixed} spans $(I - \alpha B)[U \ \iota]$.

1.4 Likelihood Dominance and Inference for the Doubly Spatial Model

The double-bounded log-likelihoods enable inference in many cases without computing the actual log-likelihood. Suppose J restrictions are imposed upon the model. The conventional likelihood ratio test would compute the deviance (twice the difference in the log-likelihoods) and this would asymptotically follow a $\chi^2(J)$ distribution (Fan, Hung, and Wong 2000). Let $L_R(\omega)$ represent the restricted log-likelihood and $L_U(\omega)$ the unrestricted log-likelihood; then $2(L_U(\omega) - L_R(\omega)) \sim \chi^2(J)$. By construction, $L_U^u(\omega) \geq L_U(\omega) \geq L_U^\ell(\omega)$ and $L_R^u(\omega) \geq L_R(\omega) \geq L_R^\ell(\omega)$ where the superscripts u and l denote upper and lower bounds. Hence, if $2(L_U^\ell(\omega) - L_R^u(\omega))$ exceeds the selected critical value of the chi-squared distribution with J degrees of freedom, this implies that the deviance based on the actual log-likelihoods, $2(L_U(\omega) - L_R(\omega))$, would also exceed this selected critical value.

From this we see that in some cases one would accept the same inference based on the bounded log-likelihoods as with the actual log-likelihoods. This resembles the likelihood dominance criteria discussed by Pollack and Wales (1991). Note also that this condition is likely to be met in a spatial model that contains a large number of observations with explanatory variables that are highly significant.

2. EMPIRICAL PERFORMANCE OF THE DOUBLY SPATIAL MODEL

The utility of the techniques presented in section 1 depend on the quality of estimates and inferences in an applied setting involving spatial data sets. Section 2.1 applies the techniques to a data set containing 506 observations on pollution and housing values from Harrison and Rubinfeld (1978). Use of this small spatial data set allows us to contrast the solutions obtained using our bounded method with exact results based on maximizing the log-likelihood. Section 2.2 applies the techniques to a data set containing 57,647 observations reflecting a problem that is over 1,000 times larger than considered in section 2.1. This illustrates the feasibility of our approach for large data sets.

2.1 Pollution Data

Harrison and Rubinfeld (1978) investigated the use of housing data to estimate the demand for clean air. They illustrated some of their ideas by employing data from the Boston SMSA. These data represent a collection of 506 observations (one observation per census tract) on levels of nitrogen oxides (*NOX*), average number of rooms (*RM*), proportion of structures built before 1940 (*AGE*), black population proportion (*B*), lower-status population proportion (*LSTAT*), crime rate (*CRIM*), proportion of area zoned with large lots (*ZN*), proportion of nonretail business area (*INDUS*), property tax rate (*TAX*), pupil-teacher ratio (*PTRATIO*), location contiguous to the Charles River (*CHAS*), weighted distances to the employment centers (*DIS*), and an index of accessibility (*RAD*).

A number of researchers have used these data to illustrate various statistical points. For example, Krasker, Kuh, and Welch (1983), Subramanian and Carson (1988), Breiman and Friedman (1985), Lange and Ryan (1989), and Breiman et al. (1993) have used the data to examine robust estimation, normality of residuals, and nonparametric and semiparametric estimation. Gilley and Pace (1996) augmented these data with the corresponding spatial coordinates and corrected some mistakes in the data, and we use this version of the data for our analysis.

We selected these data since (1) they have substantial spatial dependencies; (2) the data are well known; and (3) the number of observations is small enough to permit estimation of the weights given to each spatial basis matrix using the exact log-determinant. Table 1 contains the estimates from OLS, from the lower- bounded log-likelihood, the exact log-likelihood, and the upper-bounded log-likelihood. The maximum

likelihood estimates used the autoregressive model for the dependent variable. As one would hope, the results from the exact approach fall between the bounded estimates. Figure 2 displays the weights given to the different numbers of neighbors at the bounds and for the exact log-likelihood. Qualitatively, all estimation approaches produced the same shape. Note, estimation results at the bounds would enable a researcher to reject independence in favor of spatial dependence, since the lower bound of the unrestricted

TABLE 1

Bounded and Exact Estimation Results for the Harrison and Rubinfeld Data Using an Autoregressive Model

| Variables	OLS	Lower-Bound Maximum Likelihood	Exact Maximum Likelihood Lower-Bound Weights	Upper-Bound Maximum Likelihood
CRIM	−0.0118	−0.0072	−0.0069	−0.0067
ZN	0.0001	0.0004	0.0004	0.0004
INDUS	0.0002	0.0009	0.0010	0.0011
CHAS	0.0921	0.0146	0.0114	0.0077
NOX^2	−0.6372	−0.2524	−0.2346	−0.2161
RM^2	0.0063	0.0072	0.0072	0.0073
AGE	0.0001	−0.0004	−0.0005	−0.0005
LDIS	−0.1978	−0.1643	−0.1616	−0.1599
LRAD	0.0896	0.0592	0.0574	0.0560
TAX	−0.0004	−0.0003	−0.0003	−0.0003
PTRATIO	−0.0296	−0.0092	−0.0083	−0.0073
B	0.0004	0.0003	0.0003	0.0003
LSTAT	−37.4895	−23.3274	−22.4674	−21.7632
α	0	0.5150	0.5400	0.5650
Log-likelihood	−700.35	−584.4019	−580.1425	−576.7777
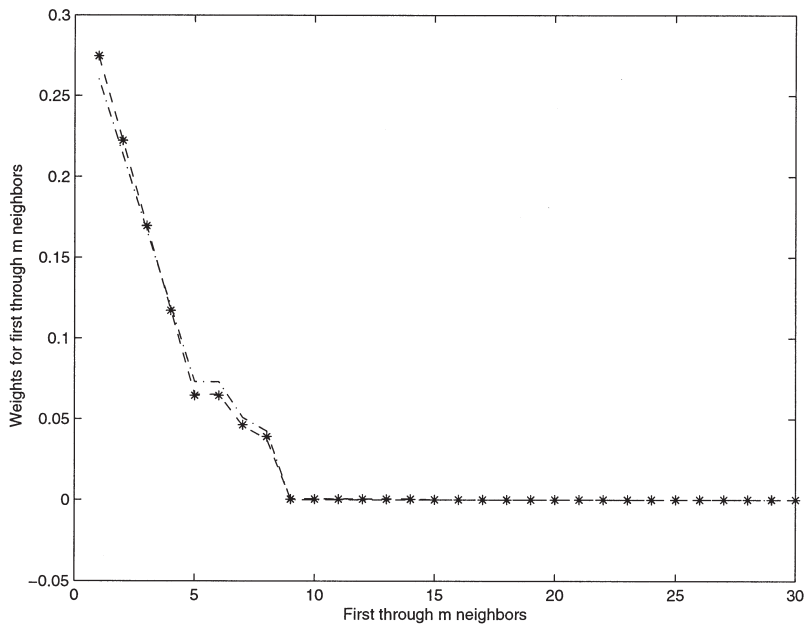
FIG. 2. Optimal Weights at the Exact (*), Upper- (- -), and Lower-Bound (-.) Likelihoods for HR Data, AR Model

log-likelihood equals -584.4, which is significantly different from the exact restricted log-likelihood equal to -700.4 under independence.[1]

2.2 U.S. Census Tract Data

To provide a more challenging example where the computational benefits associated with the doubly spatial model become clear, we collected 57,647 observations using all census tracts in the continental United States from the 1990 Census. Observations on the median price of housing (*Price*), median per capita income (*Income*), median year built, population (*Pop*), the tract's land area (*Area*), as well as the latitude and longitude of the centroid for each tract were obtained from the census information. The variable *Age* equals 1990 minus the median year built and this was strictly positive.

Table 2 contains estimation results from the autoregressive model based on 5 independent variables and Table 3 presents results based on the mixed model that includes 93 independent variables (5 independent variables, plus 22 basis matrices multiplied by the 4 nonconstant independent variables). Comparing the nonspatial

TABLE 2

Restricted and Unrestricted Bounded Log-Likelihoods for the U.S. Tract Data Mixed Model (Spatially lagged independent variable estimates not shown)

Variables	OLS	Restricted Log-Likelihood	Lower Bound	Restricted Log-Likelihood	Upper Bound	Restricted Log-Likelihood
Land Area	-0.0025	$-264,834$	-0.0003	$-227,572$	0.0003	$-224,099$
Pop	0.0253	$-261,261$	0.0200	$-227,373$	0.0189	$-224,098$
Income	0.6628	$-282,537$	0.6471	$-244,844$	0.6400	$-239,264$
Age	-0.1338	$-260,785$	-0.1337	$-228,754$	-0.1326	$-225,448$
α			0.8850	$-260,176$	0.9850	$-260,176$
k	93		93		93	
$p+1$	0		23		23	
Unrestricted Log-Likelihood		$-260,176$		$-227,218$		$-223,995$

TABLE 3

Restricted and Unrestricted Bounded Log-Likelihoods for the U.S. Tract Data Autoregressive Model

Variables	OLS	Restricted Log-Likelihood	Lower Bound	Restricted Log-Likelihood	Upper Bound	Restricted Log-Likelihood
Land Area	-0.0850	$-271,195$	-0.0177	$-232,756$	-0.0082	$-230,914$
Pop	0.1146	$-267,184$	0.0195	$-232,526$	0.0142	$-230,803$
Income	1.0837	$-288,183$	0.4840	$-246,472$	0.4358	$-240,730$
Age	-0.1269	$-267,093$	-0.0763	$-233,199$	-0.0700	$-231,408$
α	0		0.7750	$-266,505$	0.8150	$-266,505$
k	5		5		5	
$p+1$	0		23		23	
Unrestricted Log-Likelihood		$-266,505$		$-232,456$		$-230,764$

1. It took 13,630 seconds to determine the optimal weights using the exact log-determinant across a grid of 100 values of α while it took 5.2 seconds to determine the optimal bounded weights over the same grid using a 600 Mhz Pentium III running NT 4.0. This 5.2 seconds included 2.8 seconds to form the doubly stochastic matrices, 1.8 seconds to compute the traces, and 0.56 seconds to find optimal weights. Optimal weights were determined using Fortran 90 code while all other operations were carried out with Matlab 6.0 code. This small problem illustrates that choosing optimal weights based on the exact log-likelihood is computationally tedious. The ratio between the time required for an exact solution and that needed to compute bounds would increase with the number of observations.

least-squares estimates based on 5 nonspatial independent variables to a least-squares model with 93 spatial and nonspatial independent variables resulted in an increase in the log-likelihood from −266,505 to −260,176. In contrast, the use of the 22 spatial basis matrices plus the parameter α in addition to the 5 independent variables resulted in an increase in the log-likelihood from −266,505 to a point within the interval [−232,456, −230,764]. Adding the extra 88 spatial independent variables shifted this interval to [−227,218, −223,995]. Because these intervals are not close to overlapping, one can reject the autoregressive model in favor of the mixed model using a likelihood dominance argument.

The number of added variables (that is, 88) in going from the autoregressive to the mixed model may seem excessive. However, each parameter has over 600 observations (that is, 57647/93 > 93). Nevertheless, one could add additional penalties to the likelihood function to enforce more parsimony. For example, one could simply increase the weight given to the log-determinant bounds to yield a more parsimonious solution (penalized likelihood approach). A second approach would be to add other model selection penalties of the type embodied in the Akaike information criterion (AIC) or Schwarz criterion (see Judge et al. 1985, p. 244) that would also lead to more parsimonious models.

Figures 3 and 4 show the estimated weight curves for the mixed and autoregressive models. Both curves show relatively smooth declines that appear almost linear near the origin and display a geometric decline away from the origin. Figure 5 displays the bounded profile log-likelihoods for the mixed model, that also exhibit a smooth shape.

In determining the optimal weights, we used one hundred values of α ranging from 0 to 1 in 0.01 increments to implement our estimation method, and computed the optimal weights at each one of these values of α. In addition, a random starting
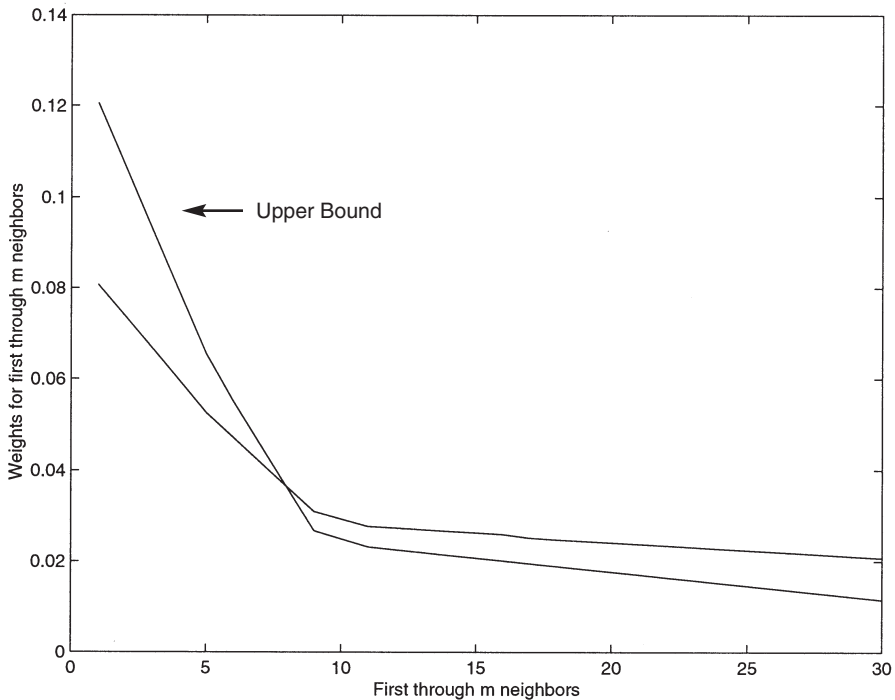

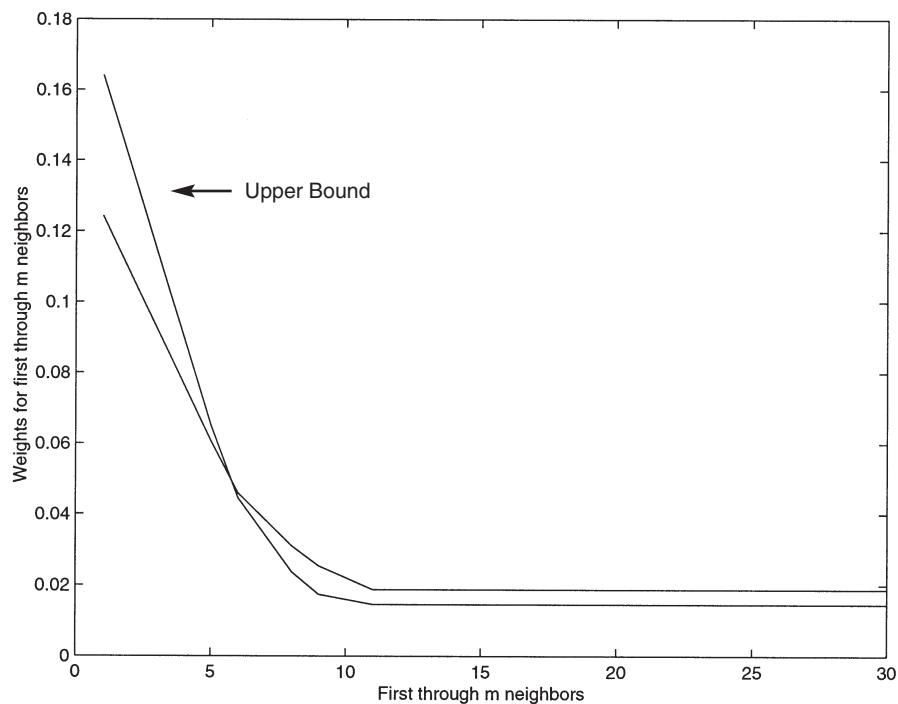Fɪɢ. 3. Optimal Weights at the Upper- and Lower-Bound Likelihoods for Tract Data, Mixed Model

FIG. 4. Optimal Weights at the Upper- and Lower-Bound Likelihoods for Tract Data, AR Model
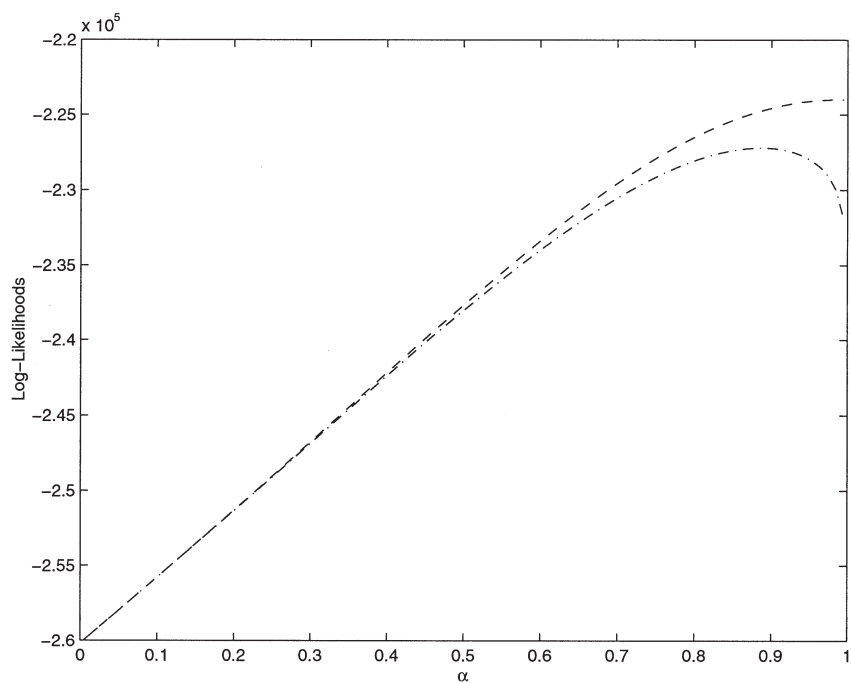

FIG. 5. α versus Upper- (- -) and Lower-Bound (- ·) Likelihoods for Tract Data, AR Model

set of weights was used for each value of α. Note, problems with multiple optima would appear in this curve as dips or downward jumps in the curve, which would alert the practitioner to potential problems. Also, if for some reason the optimization procedure failed at a particular value of α, this would cause the "optimal" log-likelihood to lie below its true value at this value of α. The overall procedure would then pick a level of α on either side of the value where optimization failed, unless the optimization failed for a compact range of values that actually contained the optimal value of α. Near the optimum the derivative of the objective function is near zero, so the value of the objective function at the neighbors would be similar. This means that choosing the neighboring value rather than the optimal value of α would not result in a substantial loss for a relatively fine grid of α values. As a result, we have a robust approach to optimization that simultaneously provides an informative profile likelihood curve.[2]

3. CONCLUSION

Large spatial data sets have become increasingly common and the computational challenges posed by these data make straightforward likelihood-based estimation difficult. Classical and Bayesian schools of inference rely on the likelihood, so abandoning the likelihood paradigm in favor of problem specific solutions should not be undertaken lightly. By deriving simple upper and lower bounds to the actual likelihood, one can simplify computations greatly while retaining many of the benefits of likelihood estimation. In addition, the bounds have an intuitively appealing interpretation as penalty terms analogous to past work in the area of semiparametric smoothing splines.

To illustrate our method, we estimated a very flexible functional form of spatial dependence constructed using a convex combination of spatial basis matrices. Because the overall errors on a convex combination of parts of the dependent variable are just the convex combination of the errors on each part, the overall sum-of-squared errors is a quadratic form involving a small matrix. Further, the size of this matrix does not rise with the number of observations. Similarly, the penalty term is a quadratic form based upon a small matrix of traces containing pairwise basis matrix products. Again, the size of this small matrix does not rise with the number of observations, preserving the tractability of our approach as the sample size increases.

We initially illustrated the method using 506 observations on the Boston housing market. In this setting, weights obtained via the doubly spatial model using bounds enclosed weights based on the true log-likelihood. Using a likelihood dominance argument, estimated bounds allowed us to reject spatial independence in favor of the spatial autoregressive model, without knowledge of the actual likelihood.

As a more computationally challenging illustration, we estimated the spatial dependence among neighbors using a sample of 57,647 U.S. Census tracts that contained complete information on housing characteristics. It took under 8.5 minutes to estimate the semiparametric model that embodied both nonparametric dependence as well as parametric regression. In our application, the estimated dependence showed an almost linear decline for nearby neighbors and a more geometric decline for higher-order neighbors. In support of our approach, the estimated dependence at both lower and upper bounds demonstrated this same pattern of decay in spatial influence. Using a likelihood dominance argument, estimated bounds allowed us to re-

---

2. In terms of computation, the time required to locate neighbors needed to form the spatial basis matrices was 122.2 seconds. It took 380.9 seconds to form twenty-two doubly stochastic spatial basis matrices, 123.7 seconds to compute the trace matrix, and 0.41 seconds to compute estimates. In total, less than 8.5 minutes were required for the entire procedure. Additional performance improvements are possible since formulation of the twenty-two doubly stochastic spatial basis matrices and trace matrix lend themselves to parallel processing.

ject a spatial autoregressive model in favor of a mixed model, without knowledge of the actual likelihood. A fortiori, we are also able to reject independent errors as well.

In conclusion, modest modifications to numerical aspects of likelihood estimation can transform the computationally difficult problem of estimating the pattern and level of spatial dependence into a relatively simple problem, for which estimators exist. Such estimators can provide a means of visualizing spatial dependence through less tinted glasses than current methods. In addition to its exploratory role, the doubly spatial model may also be used to reject some hypotheses of interest via likelihood dominance arguments.

LITERATURE CITED

Anselin, Luc (1988). *Spatial Econometrics: Methods and Models*. Dorddrecht: Kluwer Academic Publishers.

Bapat, R. B., and T.E.S. Raghavan (1997). *Nonnegative Matrices and Applications*. Cambridge: Cambridge University Press.

Barry, Ronald, and R. Kelley Pace (1999). "A Monte Carlo Estimator of the Log Determinant of Large Sparse Matrices." *Linear Algebra and its Applications* 289, 41–54.

Bavaud, Francois (1998). "Models for Spatial Weights: A Systematic Look." *Geographical Analysis* 50, 155–71.

Bell, Kathleen P., and Nancy E. Bockstael (2000). "Applying the Generalized-Moments Estimation Approach to Spatial Problems Involving Microlevel Data." *Review of Economics and Statistics* 87, 72–82.

Breiman, Leo, and Jerome Friedman (1985). "Estimating Optimal Transformations for Multiple Regression and Correlation." *Journal of the American Statistical Association* 80, 580–619.

Breiman Leo, Jerome Friedman, R. Olshen, and C. J. Stone (1993). *Classification and Regression Trees*. New York: Chapman and Hall.

Cressie, Noel (1993). *Statistics for Spatial Data.* New York: Wiley.

Fan, Jianqing, Hui-Nien Hung, and Wing-Hung Wong (2000). "Geometric Understanding of Likelihood Ratio Statistics." *Journal of the American Statistical Association* 95, 836–41.

Gilley, O. W., and R. Kelley Pace (1996). "On the Harrison and Rubinfeld Data." *Journal of Environmental Economics and Management* 31, 403–5.

Griffith, Daniel (1988). "Estimating Spatial Autoregressive Model Parameters with Commercial Statistical Packages." *Geographical Analysis* 20, 176–86.

Griffith, Daniel, and F. Lagona (1998). "On the Quality of Likelihood-based Estimators in Spatial Autoregressive Models When the Data Neighborhood Structure Is Misspecified." *Journal of Statistical Planning and Inference* 69, 153–74.

Griffith, Daniel, and Akio Sone (1995). "Trade-offs Associated with Normalizing Constant Computational Simplifications for Estimating Spatial Statistical Models." *Journal of Statistical Computation and Simulation* 51, 165–83.

Härdle, Wolfgang (1990). *Applied Nonparametric Regression*. Cambridge: Cambridge University Press.

Harrison, D., and D. L. Rubinfeld (1978). "Hedonic Prices and the Demand for Clean Air." *Journal of Environmental Economics and Management* 5, 19–40.

Judge, George, W. E. Griffiths, R. Carter Hill, Helmut Lütkepohl, and Tsoung-Chao Lee (1985). *The Theory and Practice of Econometrics*. New York: Wiley.

Krasker, William S., Edwin Kuh, and Roy E. Welch (1983). "Estimation for Dirty Data and Flawed Models." *Handbook of Econometrics.* Amsterdam: North-Holland, 651–98.

Lange, Nicholas, and Louise Ryan (1989). "Assessing Normality in Random Effects Models." *Annals of Statistics* 17, 624–42.

Martin, R. (1993). "Approximation to the Determinant in Gaussian Maximum Likelihood Estimation of Some Spatial Models." *Communications in Statistics: Theory and Methods* 22, 189–205.

Pace, R. Kelley, and Ronald Barry (1997). "Quick Computation of Regressions with a Spatially Autoregressive Dependent Variable." *Geographical Analysis* 29, 232–47.

Pace, R. Kelley, and Dongya Zou (2000). "Closed-Form Maximum Likelihood Estimates of Nearest Neighbor Spatial Dependence." *Geographical Analysis* 32, 154–72.

Pollack, R.A., and T. J. Wales (1991). "The Likelihood Dominance Criterion: A New Approach to Model Selection." *Journal of Econometrics* 47, 227–42.

Streitberg, B. (1979). "Multivariate Models of Dependent Spatial Data." *Exploratory and Explanatory Statistical Analysis of Spatial Data*, edited by C.P.A Bartels and R.H. Ketellaper, pp. 139–77. Martinus Nijhoff.

Subramanian, Shankar, and Richard T. Carson (1988). "Robust Regression in the Presence of Heteroskedasticity." JAI Press, *Advances in Econometrics* 7, 85–138.

Tiefelsdorf, M., Daniel Griffith, and Barry Boots (1999). "A Variance Stabilizing Scheme for Spatial Link Matrices." *Environment and Planning A* 31, 165–80.