

Closed-Form Maximum Likelihood Estimates of Nearest Neighbor Spatial Dependence

By

R. Kelley Pace
LREC Endowed Chair of Real Estate
Department of Finance
E.J. Ourso College of Business Administration
Louisiana State University
Baton Rouge, LA 70803-6308
(225)-388-6256
FAX: (225)-334-1227
kelley@spatial-statistics.com
kelley@pace.am

and

Dongya Zou
M.S. Candidate
Department of Computer Science
Louisiana State University
Baton Rouge, LA 70803
(225)-388-6238
FAX: (225)-334-1227

The authors gratefully acknowledge the research support they have received from Louisiana State University. In addition, the authors would like to thank Ronald Barry, Daniel Griffith, Carlos Slawson, the editor, and three anonymous reviewers for their valuable comments.

Forthcoming — *Geographical Analysis*, April 2000

“Closed-Form Maximum Likelihood Estimates of Nearest Neighbor Spatial Dependence” by R. Kelley Pace and Dongya Zou, *Geographical Analysis*, Volume 32, Number 2 (April 2000). Copyright 2000 by The Ohio State University. All rights reserved.

Abstract

Models of n^2 potential spatial dependencies among n observations spread irregularly over space seem unlikely to yield simple structure. However, the use of the nearest neighbor leads to a very parsimonious eigenstructure of the associated adjacency matrix which results in an extremely simple closed-form for the log-determinant. In turn, this leads to a closed form solution for the maximum likelihood estimates of the spatially autoregressive and mixed regressive spatially autoregressive models. With the closed-form solution, one can find the neighbors and compute maximum likelihood estimates for 100,000 observations in under one minute. The model has theoretical, pedagogical, diagnostic, modeling, and methodological applications. For example, the model could serve as a more enlightened null hypothesis for geographic data than spatial independence.

Keywords: spatial statistics, spatial autoregression, nearest neighbor, maximum likelihood, sparse matrices, log-determinants

Closed-Form Maximum Likelihood Estimates of Nearest Neighbor Spatial Dependence

I. Introduction

The task of modeling the n^2 potential dependencies among n spatially scattered observations has generated a substantial amount of intellectual activity. Naturally, many possible ways exist to specify the spatial interactions. Virtually all distance-based weighting methods such as variograms or spatial weight matrices assign some positive influence to the nearest observation, the nearest neighbor. In this sense, the nearest neighbor is common to most methods.

Given the ubiquitous nature of nearest neighbors, it seems worthwhile to examine a limiting model of spatial dependence, a model where observations depend only upon their nearest neighbor. Such special cases often help illustrate the more general case.

The nearest neighbor graph exhibits a number of important regularities. These regularities ensure the associated adjacency matrix (spatial weight matrix in this case) has a remarkable structure — the eigenvalues are all $-1, 0$, or 1 and the number of $1, -1$ pairs equals the number of pairs of symmetric elements in the adjacency matrix. This leads to a simple closed form for $\log|I - \alpha D|$, where α represents a spatial autoregressive parameter and D represents the n by n nearest neighbor adjacency matrix. Specifically, $\log|I - \alpha D| = K \log(1 - \alpha^2)$, where K represents the number of symmetric pairs of elements in the adjacency matrix. While such simple forms for the exact eigenvalues and log-determinants have been obtained before (*e.g.*, Ord (1975), Gasim (1988), Griffith and Sone (1995), Martin (1993)), such results pertain to various patterns based upon regular sets of

points or to very small numbers of points. The nearest neighbor eigenvalue results apply to arbitrary sets of points scattered across one, two, or more dimensions.

This simple form for the log-determinant leads to a closed-form solution for the spatially autoregressive and mixed regressive spatially autoregressive maximum likelihood estimators. In fact, we obtain a closed form not just for the log-determinant of the nearest neighbor adjacency matrix but also the log-determinant of any linear combination of the odd powers of this matrix. Finding the nearest neighbors and using them in estimation requires approximately $O(n\log(n))$ operations.

We illustrate the operation of the nearest neighbor model with an application to the data set analyzed by Pace and Barry (1997). By modeling just one neighbor, we obtain estimates whose log-likelihood is approximately halfway between those of maximum likelihood using zero neighbors (OLS) and maximum likelihood using four neighbors.

In addition, we conduct a Monte Carlo investigation of the sensitivity of the estimator to a simple form of zonal anisotropy. As expected, the anisotropy reduces the benefits from using the estimator. Nevertheless, the estimator still yields substantial improvements over the use of OLS even in the presence of this form of isotropy.

The simple form of the estimator can help clarify more complicated spatial dependence, aid in theoretical investigations, and serve as a useful pedagogical tool. In addition, the nearest neighbor maximum likelihood estimate of spatial dependence could have diagnostic value. Finally, it can provide a naïve model useful for gauging the

contribution of more complicated approaches while still keeping within the useful maximum likelihood framework.¹

Section II discusses the regularities in the nearest neighbor graph, section III uncovers the eigenvalues and log-determinant of the associated adjacency matrix, section IV shows how these lead to a closed-form solution to the maximum likelihood first order conditions, section V provides a more general model with a closed-form maximum likelihood solution, section VI applies the model to an actual data set, section VII examines the performance of the estimator in the presence of misspecified dependence and anisotropy, section VIII discusses the time required for computing the estimates, and section IX concludes with a discussion of the key results.

II. The Nearest Neighbor Graph

Given n points or nodes, we consider their nearest neighbor graph. We make the following assumption:

Assumption 1: Each node has an unique nearest neighbor (other than itself). This rules out ties among multiple nearest neighbors.

This Assumption eliminates the situation depicted in Diagram 1, where the arrows point to the nearest neighbor. Hence, nodes c and a cannot simultaneously be the nearest

¹ Other simple spatial estimators exist such as the coding estimator of Besag (1974). However, it becomes difficult to examine the null within a nested set of models, an almost essential requirement for a useful null hypothesis estimator.

neighbors of node b . Given Assumption 1, the results developed below apply only to irregularly spaced sets of nodes.²

$$a \leftarrow b \rightarrow c \quad (\text{Diagram 1})$$

Proposition 1: The nearest neighbor graph associated with the n nodes cannot have simple cycles such as illustrated in Diagram 2.

$$\begin{array}{ccccc} a & \rightarrow & b & \rightarrow & c \\ \uparrow & & & & \downarrow \\ y & \leftarrow & x & \leftarrow & \dots \end{array} \quad (\text{Diagram 2})$$

Proof. We use induction on the number of nodes. When the number is 3, we assume contrary to Proposition 1 that there is such cycle. Hence, we have the situation depicted in Diagram 3.

$$\begin{array}{ccc} a & \rightarrow & b \\ \nwarrow & & \swarrow \\ & c & \end{array} \quad (\text{Diagram 3})$$

This means that b is the nearest neighbor of a , c is the nearest neighbor of b , and a is the nearest neighbor of c . Or if we let $dist(x, y)$ represent the distance between

² Other techniques exist to deal with completely regular sets of points. If only some of the points have equal distances, the technique of randomly perturbing the points and computing the desired quantities across a number of iterations may have some value. This operation (termed “dithering”) has seen substantial usage in fields such as image processing.

two arbitrary points x and y , Diagram 3 is equivalent to the following inequalities presented in (1):

$$\begin{aligned}
& \text{dist}(a,b) < \text{dist}(a, j), \forall j \neq a,b \\
& \text{dist}(b,c) < \text{dist}(b, j), \forall j \neq b,c \\
& \text{dist}(c,a) < \text{dist}(c, j), \forall j \neq c,a.
\end{aligned} \tag{1}$$

In particular, (1) implies (2),

$$\begin{aligned}
(2.1) \quad & \text{dist}(a,b) < \text{dist}(a, c), \\
(2.2) \quad & \text{dist}(b,c) < \text{dist}(b, a), \\
(2.3) \quad & \text{dist}(c,a) < \text{dist}(c, b).
\end{aligned} \tag{2}$$

The system of inequalities in (2) and the symmetry of distance ($\text{dist}(b, a) = \text{dist}(a, b)$, $\text{dist}(c, b) = \text{dist}(b, c)$, $\text{dist}(c, a) = \text{dist}(a, c)$) implies (3),

$$\text{dist}(a,c) \stackrel{(2.3)}{<} \text{dist}(b,c) \stackrel{(2.2)}{<} \text{dist}(a,b) \stackrel{(2.1)}{<} \text{dist}(a,c) \tag{3}$$

which is impossible and proves by contradiction the 3 node case.

Suppose the proposition is true for any number of nodes less than m , then we prove it is also true for m nodes. From Diagram 2, we have the inequalities in (4).

$$\begin{aligned}
(4.1) \quad & \text{dist}(a,b) < \text{dist}(a, j), \forall j \neq a,b \\
(4.2) \quad & \text{dist}(b,c) < \text{dist}(b, j), \forall j \neq b,c \\
& \quad \quad \quad \vdots \\
(4.m-1) \quad & \text{dist}(x, y) < \text{dist}(x, j), \forall j \neq x, y \\
(4.m) \quad & \text{dist}(y, a) < \text{dist}(y, j), \forall j \neq y, a
\end{aligned} \tag{4}$$

Similarly, working from the last inequality in (4) back up to the first one, and using the symmetry of distance, gives us the inequalities in (5),

$$dist(a, y) \stackrel{(4.m)}{<} dist(x, y) \stackrel{(4.m-1)}{<} \cdots \stackrel{(4.3)}{<} dist(b, c) \stackrel{(4.2)}{<} dist(a, b) \stackrel{(4.1)}{<} dist(a, y) \quad (5)$$

which is impossible and by the contradiction proves the proposition. QED

Definition 1. Two subgraphs are said to be disconnected, if for any two nodes a and b , such that $a \in G_1, b \in G_2$, a is not connected with b , and b is not connected with a . Hence, there are no sequences of nodes c, d, \dots, e or u, v, \dots, w satisfying a set of connections such as illustrated in Diagram 4.

$$a \rightarrow c \rightarrow d \rightarrow \cdots \rightarrow e \rightarrow b \text{ or } b \rightarrow u \rightarrow v \rightarrow \cdots \rightarrow w \rightarrow a. \quad (\text{Diagram 4})$$

Proposition 2. Each subgraph has exactly one 2-node-cycle such as depicted in Diagram 5.

$$a \leftrightarrow b \quad (\text{Diagram 5})$$

Proof. Begin from any node a , whose nearest neighbor is b , and b whose nearest neighbor is c , and so on. This leads to a chain such as illustrated in Diagram 6 where the braces show a choice between two mutually exclusive potential neighbors:

$$a \rightarrow b \rightarrow \begin{cases} a \\ c \end{cases} \rightarrow \begin{cases} b \\ d \end{cases} \rightarrow \begin{cases} c \\ e \end{cases} \rightarrow \dots \rightarrow y \leftrightarrow z \quad (\text{Diagram 6})$$

In Diagram 6, a directed edge either (I) extends to a node not previously connected or (II) connects back to the previous node (the node and the previous node are nearest neighbors to each other). Proposition 1 allows no other possible connections. If case II occurs, this establishes the existence of the 2-node-cycle. If case I occurs, this process continues until encountering case II or until all nodes have been connected. If all nodes have been connected, the last node z must connect to some neighbor, which must be the previous node y by Proposition 1. This proves the existence of a 2-node-cycle.

Suppose we have two such cycles in a subgraph such as in Diagram 7:

$$g \leftrightarrow h \rightarrow k \rightarrow \dots \rightarrow a \rightarrow b \rightarrow \dots \rightarrow e \leftrightarrow f \quad (\text{Diagram 7})$$

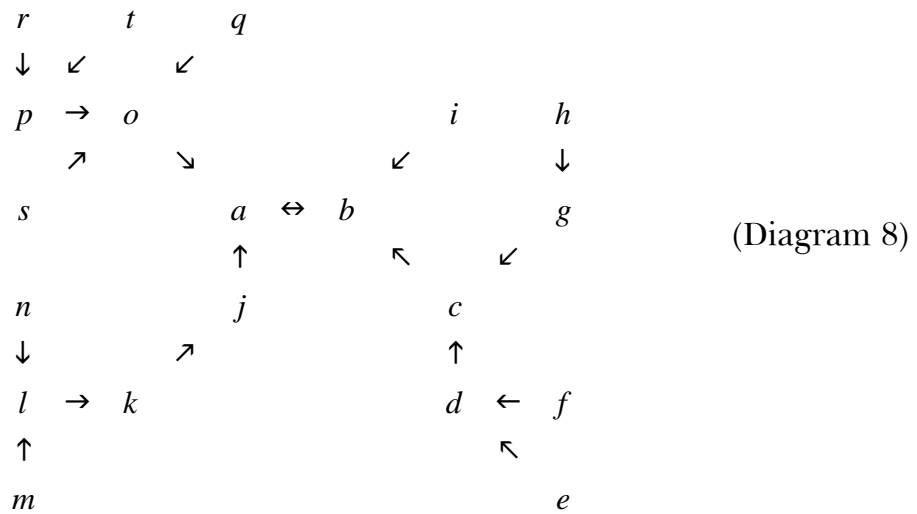
The part $g \leftarrow h \rightarrow k$ violates Assumption 1. This proves the uniqueness part of Proposition 2. QED

Proposition 3: Each subgraph is disconnected from the other subgraphs.

Proof. The proof is immediate in the case of one subgraph. If two subgraphs exist and if they were connected, by Definition 1 and Proposition 2 this would be the

situation discussed in Diagram 7. As shown in the proof of Proposition 2, this situation is not possible. Hence, the subgraphs must be disconnected. QED

As an additional illustration, Diagram 8 presents a more complicated individual subgraph. The diagram shows how multiple nodes can have an individual node as nearest neighbors but not *vice-versa*. Note, the existence of only one 2-node-cycle between nodes *a* and *b*.

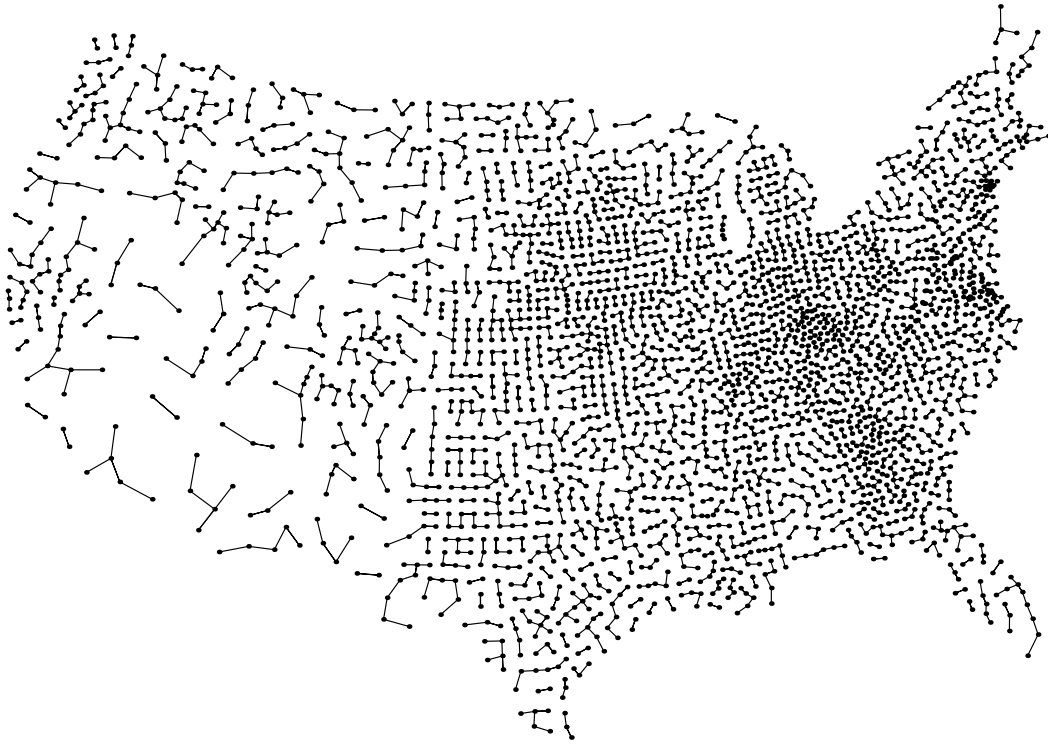


To give a better idea of the effects of the three propositions, we present a diagram of the United States counties and their nearest neighbor graph (without arrows) as depicted in Diagram 9.³ As expected from the Propositions stated above, all the subgraphs are disconnected, and each one has exactly one 2-node-cycle and no simple cycle (cycles

³ A Lambert conformal conic projection of the latitude and longitude of the county centroids yielded the *x, y* coordinates used in the construction of Diagram 9. The projection minimized the error along the 27.5 and 43.5 degree parallels.

containing three or more nodes). Intriguingly, the nearest neighbor graph leads to a potential mathematical description of regions.

Diagram 9 — US County Subgraphs



III. The Nearest Neighbor Adjacency Matrix and its Eigenvalues

Given the nearest neighbor graph discussed in section II, we now focus upon the associated adjacency matrix and its properties. We show the adjacency matrix has particularly simple eigenvalues and an extremely simple closed form determinant formula. Insofar as the matrix has no negative values, has only zeros on the main diagonal, and each row contains a single 1, this adjacency matrix is also a row-stochastic spatial weight matrix

and hence the log-determinant of this has immediate relevance in computing maximum likelihood estimates, a topic explored in section IV.

We begin by examining Proposition 4 which clarifies the structure of the adjacency matrix.

Proposition 4: The nearest neighbor graph possesses a block diagonal adjacency matrix representation. Furthermore, each block is a block triangular matrix itself.

Proof. Suppose the nearest neighbor directed graph has K subgraphs and construct the associated nearest neighbor adjacency matrix D for the n nodes according to the directed graph. Proposition 2 implies that the number of subgraphs in the directed graph equals the number of pairs of symmetric elements in the associated adjacency matrix D . Since these subgraphs are disconnected to each other by Proposition 3, we can label the nodes in such a way that the labels corresponding to one subgraph are contiguous. This way we get a block diagonal form of D as shown in (6).

$$D = \begin{bmatrix} D_1 & & & \\ & D_2 & & \\ & & \ddots & \\ & & & D_K \end{bmatrix} \quad (6)$$

Where each D_i corresponds to the i th subgraph ($i=1, 2, \dots, K$), and all the other elements in D that do not belong to the blocks are zeros, due to the fact that the subgraphs are disconnected. Moreover, when we label the nodes in each

subgraph, we can start with its 2-node-cycle so that each D_i has the following block triangular form as in (7).

$$D_i = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & \\ * & & B_{i2} \end{bmatrix} \quad (7)$$

Thus the proposition is proved. QED

Theorem 1: Given n nodes, whose nearest neighbor graph has K subgraphs.

Then their nearest neighbor matrix D has eigenvalues 1 and -1 each repeating K times, and 0 that repeats $n-2K$ times.

Proof. By Proposition 4, in order to find out the eigenvalues of D , we need only find out the eigenvalues of each block D_i , and take the union of them. Since all the subgraphs have the same structure, so do all the blocks of the adjacency matrix. So analyze D_1 as an example. Assume that D_1 's order is n_1 . By the proof of Proposition 4 we see that D_1 has the block triangular form in (8).

$$D_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & \\ * & & B_2 \end{bmatrix} = \begin{bmatrix} B_1 & 0 \\ * & B_2 \end{bmatrix} \quad (8)$$

Where $B_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, and B_2 is a n_1-2 by n_1-2 square matrix. Furthermore, the block triangular form of D_1 itself means it is reducible and hence the union of the eigenvalues of B_1 and B_2 constitute the eigenvalues of D_1 .

Note, that $tr(B_1)=0$ and $tr(B_1^2)=2$. Hence, the eigenvalues of B_1 satisfy (9).

$$\lambda_1 + \lambda_2 = 0 \text{ and } \lambda_1^2 + \lambda_2^2 = 2 \quad (9)$$

Solving these two equations leads to $\lambda_1 = 1, \lambda_2 = -1$.

Since the set of eigenvalues of D_1 is the union of that of B_1 and B_2 , thus all we have to show for the rest of the theorem is that B_2 has only one eigenvalue 0 that repeats n_1-2 times.

Following similar logic, $tr(B_2)=0$ and $tr(B_2^2)=0$. Proposition 2, which shows each subgraph has only one 2-node-cycle, implies $tr(B_2^2)=0$. The existence of an unique 2-node-cycle means D_1 has only two symmetric elements that are arranged in the submatrix B_1 . Therefore, all the elements in B_2 are asymmetric, and this implies $tr(B_2^2)=0$. Hence, the eigenvalues of B_2 satisfy (10).

$$\sum_{i=1}^{n_1-2} \lambda_i = 0 \text{ and } \sum_{i=1}^{n_1-2} \lambda_i^2 = 0 \quad (10)$$

Solving these two equations leads to $\lambda_i = 0$ for $i = 1, \dots, n_1 - 2$.

Hence, D_1 's only eigenvalues are 1, -1 and 0, with 0 repeating n_1-2 times. Similar results can be obtained for the rest of the $K-1$ blocks of D . Thus the theorem is proved. QED

Theorem 1 has many matrix function applications. The most obvious application leads to a closed-form solution for the determinant of functions involving the nearest neighbor adjacency matrix.

Application 1 of Theorem 1: Given any $-1 < \alpha < 1$, and any nearest neighbor matrix D with K symmetric pairs, the simple expression for the log-determinant appearing in (11) holds.

$$\log|I - \alpha D| = K \log(1 - \alpha^2) \quad (11)$$

Proof. Let $\{\lambda_i\}$ be the set of eigenvalues of A . From Theorem 1 we know that, matrix D has eigenvalues 1 and -1 each repeating K times, and all the other eigenvalues are 0.

$$\log|I - \alpha D| = \log \prod_i (1 - \alpha \lambda_i) = K \log(1 + \alpha) + K \log(1 - \alpha) = K \log(1 - \alpha^2) \quad (12)$$

Thus the application is proved. QED

Application 2 of Theorem 1: A weight matrix comprised of a linear combination of the odd powers of D leads to the same log-determinant as using the weight

matrix D for a given α . Hence, $\log|I - \alpha W| = K \log(1 - \alpha^2)$, where $W = \sum_{i=0}^m \omega_{1+2i} D^{1+2i}$

given $\sum_{i=0}^m \omega_{1+2i} = 1$ for a positive integer m .

Proof. By the definition of eigenvalues, $Dx = \lambda x$ where x is an eigenvector and λ is an eigenvalue. As well-known, $D^q x = \lambda^q x$ for a non-zero integer q (Strang (1976, p. 180)). Hence, one can show that $Wx = \left(\sum_{i=0}^m \omega_{1+2i} D^{1+2i} \right) x = \left(\sum_{i=0}^m \omega_{1+2i} \lambda^{1+2i} \right) x$. The eigenvalues of D are either $-1, 0$, or 1 . Odd-powers of these have the same value. The linear combination of these will also have the same value (*i.e.*, $\sum_{i=0}^m \omega_{1+2i} \lambda^{1+2i} = \lambda$). Hence, W has the same eigenvalues as D and hence the same log-determinant. Thus the application is proved. QED

Other applications of Theorem 1 are possible. For example, Bavaud (1998) discussed the possible importance of matrix functions of row-stochastic weight matrices. Some of these functions of the nearest neighbor weight matrix also have simple eigenvalues or determinants. As a limitation, attempts to create symmetry by adding the nearest neighbor adjacency matrix and its transpose destroy the simplifying structure imposed by using the nearest neighbor and do not yield simple eigenstructure.

IV. A Spatial Maximum Likelihood Estimator with Closed Form Solution

Computation of the spatial maximum likelihood estimators has traditionally required iterative, sometimes arduous, computations. The foregoing closed-form expression for the log-determinant enables the spatially autoregressive model to have a closed-form solution as well.

Let Y represent an n element vector of observations on the dependent variable, X represent an n by p matrix of observations on the independent variable (X may include

functions of the spatial weight matrix), let D represent the n by n nearest neighbor adjacency matrix as specified above, let β represent a p element vector of parameters associated with the independent variables, and let α represent a scalar autoregressive parameter. We fit the model $Y - \alpha DY = X\beta + \varepsilon$. The profile log-likelihood function is,

$$L(\alpha) = C + \log|I - \alpha D| - (\frac{n}{2})\log(SSE(\alpha)) \quad (13)$$

where C represents a scalar constant. Let $\beta_o = (X'X)^{-1}X'Y$, $\beta_d = (X'X)^{-1}X'DY$, $e_o = Y - X\beta_o$, $e_d = DY - X\beta_d$. Following Anselin (1988, p. 181) or Pace and Barry (1997, p. 235-236), we can write the sum-of-squared errors as a function of the autoregressive parameter as in (14)

$$SSE(\alpha) = e_o'e_o - 2\alpha e_o'e_d + \alpha^2 e_d'e_d \quad (14)$$

Substituting the expressions for $SSE(\alpha)$ from (14) and $\log|I - \alpha D|$ from (12) into the profile log-likelihood equation (13) yields (15),

$$L(\alpha) = C + K\log(1 - \alpha^2) - (\frac{n}{2})\log(e_o'e_o - 2\alpha e_o'e_d + \alpha^2 e_d'e_d) \quad (15)$$

where K represents the number of symmetric pairs in the adjacency matrix. We wish to find its maximum point with respect to the α parameter in $(-1,1)$. We assume $SSE > 0$, and given this we calculate the first derivative of L which appears in (16).

$$L'(\alpha) = \frac{-2\alpha K}{1 - \alpha^2} + \frac{n(e_d'e_o - \alpha e_d'e_d)}{SSE} \quad (16)$$

Setting $L'(\alpha) = 0$ and substituting the representation for SSE from (14), we obtain (17).

$$\alpha^3(2K - n)e_d'e_d + \alpha^2(n - 4K)e_d'e_o + \alpha(ne_d'e_d + 2Ke_d'e_o) - ne_d'e_o = 0 \quad (17)$$

We assume that not all points are symmetric (*i.e.*, $2K \neq n$). If $2K = n$, the cubic term vanishes and the first order condition in (17) becomes a simpler quadratic equation.

Assuming $2K \neq n$, we can rewrite the coefficients from (17) as in (18).

$$a_1 = \frac{(n-4K)e'_d e_o}{(2K-n)e'_d e_d}, \quad a_2 = \frac{ne'_d e_d + 2Ke'_o e_o}{(2K-n)e'_d e_d}, \quad a_3 = \frac{-ne'_d e_o}{(2K-n)e'_d e_d} \quad (18)$$

and rewrite the cubic equation from (17) using the coefficients from (18) as (19).

$$x^3 + a_1 x^2 + a_2 x + a_3 = 0 \quad (19)$$

We define in (20) the following constants used for solving the cubic equation.

$$Q = \frac{3a_2 - a_1^2}{9}, \quad R = \frac{9a_1 a_2 - 27a_3 - 2a_1^3}{54} \quad (20)$$

$$S = \sqrt[3]{R + \sqrt{Q^3 + R^2}}, \quad T = \sqrt[3]{R - \sqrt{Q^3 + R^2}}$$

Finally, the solutions appear in (21).

$$x_1 = S + T - \frac{1}{3}a_1$$

$$x_2 = -\frac{1}{2}(S + T) - \frac{1}{3}a_1 + \frac{1}{2}i\sqrt{3}(S - T) \quad (21)$$

$$x_3 = -\frac{1}{2}(S + T) - \frac{1}{3}a_1 - \frac{1}{2}i\sqrt{3}(S - T)$$

By the properties of cubic equations, at least one of these solutions is real.

Exactly one of the solutions from (21) falls in $(-1, 1)$. To see this, in (22) we take the second derivative of L with respect to α .

$$L''(\alpha) = \left(\frac{-1}{SSE} \right) \left[ne'_d e_d + \frac{2K(1+\alpha^2)SSE}{(1-\alpha^2)^2} + \left\{ \frac{n(e'_d e_o - 2\alpha e'_d e_d)}{SSE} \right\} \right] \quad (22)$$

The first and second terms within the brackets are unambiguously positive. The third term (the term in braces) is identical to the second term of the first order conditions in (16). As

the first term in the first order conditions is negative, it follows the only other term (the second term) must be non-negative at the optimum. The first two terms in brackets are positive and the third term is non-negative and thus the expression in brackets is positive. The positive expression within brackets is multiplied by a negative term in parentheses. Thus, $L''(\alpha) < 0$ for $\alpha \in (-1,1)$ and hence $L(\alpha)$ is strictly concave. Since $\lim_{|\alpha| \rightarrow 1} L(\alpha) = -\infty$, there exists at least one interior maxima. By the strict concavity of $L(\alpha)$, this maxima will be the maximum. Hence, a value of α from $(-1, 1)$ uniquely maximizes the profile log-likelihood.

V. A More General Spatial Maximum Likelihood Estimator

We consider a generalization of the model in section IV. We now find the closed-form solution for the autoregressive parameters associated with the odd powers of the nearest neighbor adjacency matrix. The odd powers of the nearest neighbor adjacency matrix have the desirable feature of zeros along the diagonals (unlike the even powers). Provided the linear combination of the odd powers have coefficients summing to unity, the resulting linear combination will have the same set of symmetric elements as the first-order nearest neighbor adjacency matrix, but will have different weights for the asymmetric elements. Because they share the same set of symmetric elements (same set of 2-node-cycles), they will have the same set of eigenvalues, as shown in Application 2 of Theorem 1. Hence, the linear combination provides a way of incorporating the effects of more distant observations without complicating the computation of the log-determinant. Thus, this generalization can improve the richness of the nearest neighbor model of spatial dependence at little additional computational cost.

Now we would like to consider the maximum likelihood function in (23),

$$L(\alpha_o, \omega_1, \omega_3, \dots, \omega_{2m+1}) = \log \left| I - \alpha_o \sum_{i=0}^m \omega_{2i+1} D^{2i+1} \right| - \frac{n}{2} \log(SSE(\alpha_o, \omega_1, \omega_3, \dots, \omega_{2m+1})) \quad (23)$$

where $-1 < \alpha_o < 1$, $\sum_{i=0}^m \omega_{2i+1} = 1$, $SSE(\alpha_o, \omega_1, \omega_3, \dots, \omega_{2m+1}) = v'(E'E)v$, $E' = (e_o, e_d, e_{d^3}, \dots, e_{d^{2m+1}})$,

and $v' = (1, -\alpha_o \omega_1, -\alpha_o \omega_3, \dots, -\alpha_o \omega_{2m+1})$. We assume E has full rank. Substituting the relation from (11) into (23) yields (24).

$$L(\alpha_o, \omega_1, \omega_3, \dots, \omega_{2m+1}) = K \log |1 - \alpha_o|^2 - \frac{n}{2} \log(v'(E'E)v) \quad (24)$$

To find the maximum of L we proceed in two stages. First, we maximize L with respect to $\omega_1, \omega_3, \dots, \omega_{2m+1}$ subject to $\sum_{i=0}^m \omega_{2i+1} = 1$ holding α_o constant where $-1 < \alpha_o < 1$. Holding α_o constant makes maximizing the log-likelihood equivalent to minimizing a quadratic form subject to linear equality constraints. We substitute the optimal $v^*(\alpha_o)$ into L to form the profile $L(\alpha_o)$. Second, we maximize the resulting profile log-likelihood $L(\alpha_o)$ with respect to α_o .

To begin the first stage of optimization, let (25)

$$\ell = v'(E'E)v + 2(r - Rv)\lambda \quad (25)$$

be the Lagrange function, where $r' = (1, -\alpha_o)$, $R = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots \\ 0 & 1 & 1 & 1 & \dots \end{pmatrix}$, and $Rv = r$. Taking first

derivatives yields (26).

$$\begin{aligned}\frac{\partial \ell}{\partial v} &= 2(E'E)v - 2R'\lambda = 0 \\ \frac{\partial \ell}{\partial \lambda} &= 2(r - Rv) = 0\end{aligned}\tag{26}$$

From the second equation, we get $r = Rv$; from this and the first one we get

$\lambda = (R(E'E)^{-1}R')^{-1}r$. Substituting this back into the first equation yields the optimal solution as a function of α_o in (27).

$$v^*(\alpha_o) = (E'E)^{-1}R'(R(E'E)^{-1}R')^{-1}r.\tag{27}$$

The matrix $E'E$ is a variance-covariance matrix and is positive definite (we assumed E had full rank). Hence, this solution should minimize the sum-of-squared errors subject to the constraints and thus maximize the log-likelihood subject to the equality constraint for a given value of α_o . Substituting the optimal $v^*(\alpha_o)$ into the log-likelihood in (24) yields the profile log-likelihood in (28).

$$L(\alpha_o) = K \log|1 - \alpha_o^2| - \frac{n}{2} \log(r'(R(E'E)^{-1}R')^{-1}r)\tag{28}$$

Let $G = (R(E'E)^{-1}R')^{-1}$, a 2 by 2 matrix. We can rewrite the profile log-likelihood in (28) as (29).

$$L(\alpha_o) = K \log|1 - \alpha_o^2| - \frac{n}{2} \log(G_{11} - 2\alpha_o G_{12} + \alpha_o^2 G_{22})\tag{29}$$

This is exactly the same problem as analyzed in section IV with G_{11} serving the same role as $e'_o e_o$, $G_{12} = G_{21}$ serving the same role as $e'_d e_o$, and G_{22} serving these same role as $e'_d e_d$.

Hence, we can use the solution developed there for the second stage of the optimization of the profile log-likelihood.

VI. Application of the Estimator to Geographical Data

To illustrate the effects of modeling spatial data with the nearest neighbor as opposed to multiple neighbors, we fitted the model and data introduced by Pace and Barry (1997). They examined voting behavior for the 1980 presidential election across 3,107 counties in the continental US using a mixed regressive spatially autoregressive estimator (Ord (1975), Anselin (1988)). Table 1 reproduces their OLS and maximum likelihood estimates using the four nearest neighbors ($\beta_{ols}, \beta_{ML-4}$) and augments these with the estimates using the nearest neighbor (β_{ML-1}) and the nearest neighbor with the addition of a third-order nearest neighbor ($\beta_{ML-1,3}$). We employed the Euclidean metric using unprojected latitude and longitude to determine the neighbors. Table 1 only shows the estimates for the independent variables common to all estimators. For brevity, it does not show the results for the numerous spatially lagged independent variables. We report likelihood ratio statistics (i.e., twice the difference between the restricted and unrestricted log-likelihoods) for the null hypothesis that a variable has no effect or in many cases that a variable and its spatial lag have no effect.

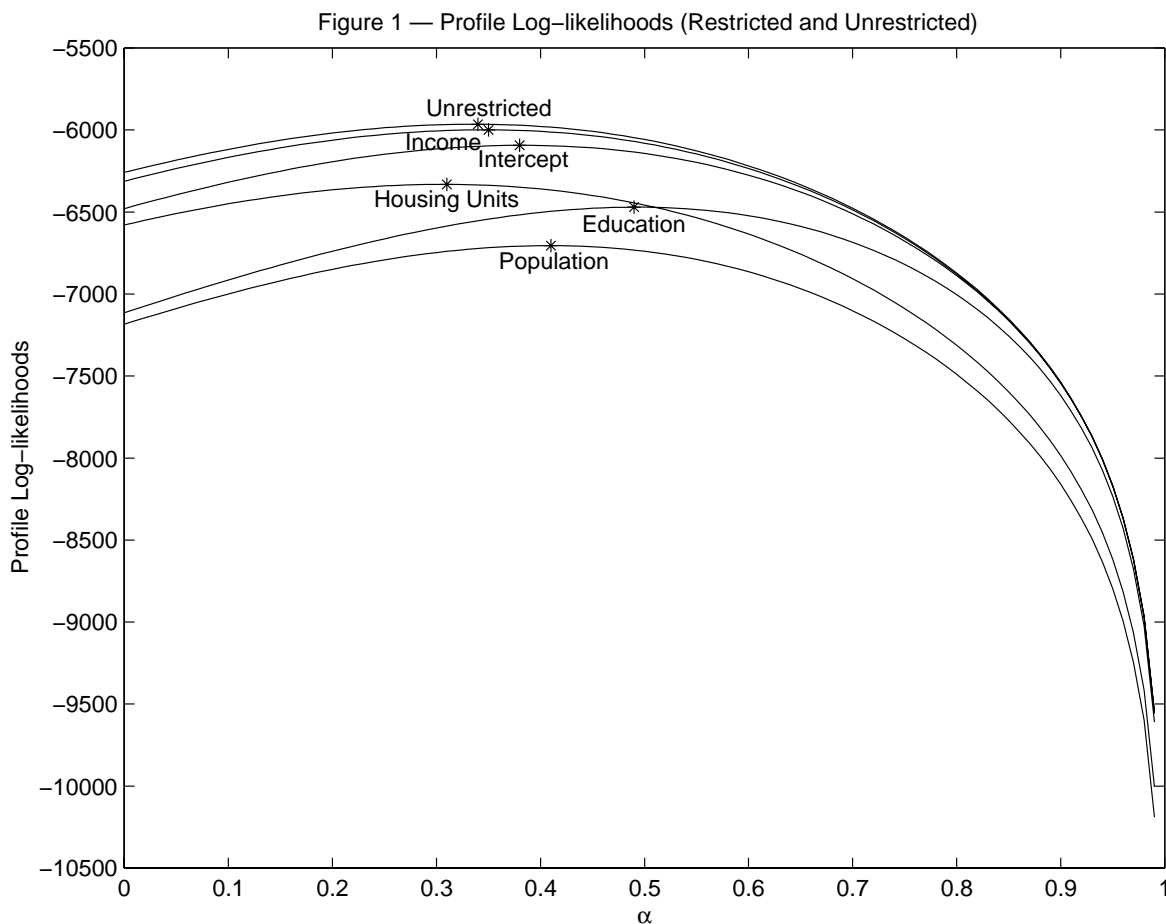
The third-order neighbor corresponds to using the nearest neighbor of the nearest neighbor of the nearest neighbor (i.e., D_1^3). For symmetric elements, the third-order neighbor adjacency matrix has the same entries as the first-order nearest neighbor adjacency matrix (in Diagram 8 the nearest neighbor to node b is node a , its nearest neighbor is node b , and finally its nearest neighbor is back to node a). For asymmetric elements, the third-order neighbor adjacency matrix elements are not the same as the first-order neighbor adjacency matrix elements (in Diagram 8 the nearest neighbor to node e is node d , its nearest neighbor is node c , and the third-order nearest neighbor is node b). For

graphs with many asymmetric nodes, the use of the third-order or higher odd-order adjacency matrices could capture some of the more complicated spatial dependence. See Anselin and Smirnov (1996) for more on higher-order weight matrices.

As one might expect, the nearest neighbor maximum likelihood estimates usually fell between the OLS and maximum likelihood using the four nearest neighbors. The log-likelihood was -6307.1 for OLS without lagged independent variables, -5964.6 for maximum likelihood with the nearest neighbor, -5954.9 for maximum likelihood with the nearest neighbor and the third-order neighbor, and -5678.8 for maximum likelihood with four nearest neighbors. As expected, maximum likelihood with four neighbors significantly outperformed the other specifications. However, the nearest neighbor maximum likelihood estimator clearly outperformed OLS. Its log-likelihood fell about halfway between that of OLS and the four nearest neighbor maximum likelihood estimators. The addition of the third-order neighbor increased the log-likelihood significantly, but did not materially change the estimates or the goodness-of-fit.

Figure 1 displays the profile log-likelihood (actually twice the log-likelihood) as a function of the autoregressive parameter α for both the unrestricted model and for the relevant restricted models. The restricted models include one without an intercept (1 restriction), one without the population over 18 years of age and its spatial lag (2 restrictions), one without the population with more than 12 years of education and its spatial lag (2 restrictions), one without the owner-occupied housing units and its spatial lag (2 restrictions), as well as one without aggregate income and its spatial lag (2 restrictions). For each profile log-likelihood curve, an asterisk denotes the α producing the maximum

log-likelihood for that model. As expected, all of the curves have clear concave shapes without any visible irregularities.



The profile log-likelihood curves provide a wealth of information on the importance of the variables, the autoregressive parameter, and their interactions. First, variables whose deletion (restriction to zero effect) results in a curve substantially below the unrestricted curve are clearly statistically significant. Under the null hypothesis of no effect, single restrictions (*i.e.*, no intercept) should have a likelihood ratio statistic (twice the difference between the restricted and unrestricted log-likelihood) of less than 4 at approximately the 5% significance level and of less than 6 at approximately the 1% significance level.

Similarly under the null hypothesis of no effect, dual restrictions (*e.g.*, no effect of income and spatially lagged income) should have a likelihood ratio statistic (twice the difference between the restricted and unrestricted log-likelihood) of less than 6 at approximately the 5% significance level and of less than 10 at approximately the 1% significance level.⁴ All of the variables in the model have likelihood ratio statistics exceeding these critical values. Second, the curvature of each curve conveys information concerning the estimation error associated with the autoregressive parameter. Third, the shift of the autoregressive parameter across the curves reveals the extent to which the autocorrelation adjustment can substitute for the deletion of a variable. For example, the autoregressive parameter rises from 0.34 to 0.49 in response to the deletion of the education variable and its spatial lag.

The optimal value of α for the nearest neighbor estimator was 0.3352 which contrasts with the optimal value for the four neighbor estimator of 0.6150. The optimal value for α_0 was 0.341 (ω_1, ω_3 was 0.733, 0.267), a small increase over the nearest neighbor result.

Median errors and other statistics show the nearest neighbor estimator results continue to fall between the results for OLS and the four neighbor maximum likelihood estimator.

Undoubtedly, other ways of specifying the model and the form of spatial dependence would result in superior performance to the simple four neighbor model used as a standard of comparison. In addition, other data and models may yield different degrees of improvement. To partially address this and to address the important problem of anisotropy, we conduct a Monte Carlo study in the following section.

⁴ The likelihood ratio statistic is asymptotically distributed as chi-squared with degrees-of-freedom equal to the number of equality restrictions (Cramer (1991, p. 39)).

VII. Simulated Performance With Misspecified Dependence and Anisotropy

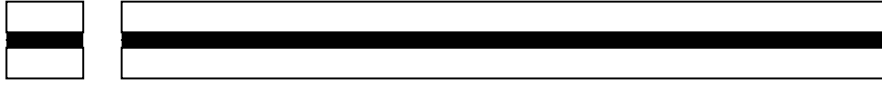
Real world data exhibit far more complicated spatial patterns than nearest neighbor spatial dependency. Actual spatial processes almost certainly involve multiple neighbors and may vary with direction or by zone (directional or zonal anisotropy). To better understand the roles of the various estimators, we devised a Monte Carlo experiment which employs random variables with multiple neighbor dependency and with a particularly simple form of zonal anisotropy.

We begin by examining the point distribution of the spatially correlated random variates. Points were uniformly distributed over the regions depicted in Diagram 10. We begin with the pair of boxes on the left separated by a heavy line. Each of these boxes has an aspect ratio (length/width) of 2. In contrast, the pair of boxes on the right have an aspect ratio of 10. As the aspect ratio becomes very large, the regions approach lines.

In the experiment, we will generate pairs of correlated random variables (with spatially random point distributions). One of each pair will be distributed in the top box and the other will be distributed in the bottom box. Each set of variates in the pair will be statistically independent of the other. Within each box the random variates are spatially correlated, but between boxes they are independent. Hence, this provides a simple way to simulate a form of zonal anisotropy.

The anisotropy in this case means that movements in one direction (along the longest dimension) can have different dependence than movements in the other direction (along the shortest dimension). One could imagine the heavy line represents a closed border, for

example, and economic activity on each side is correlated, but that the economic activities of each side are independent of the other side due to lack of trade.



(Diagram 10)

This situation presents a way of testing the performance of the nearest neighbor estimator in the presence of a form of anisotropy. Since estimator relies upon the nearest neighbor, selection of the neighbors disregarding the closed border will sometimes result in the selection of an across-the-border neighbor whose value is statistically independent of the dependent variable. Obviously, this should degrade the performance of the nearest neighbor estimator. As the aspect ratio rises (holding n constant), the chance of the nearest neighbor appearing across the border rises. Hence, one would expect a larger aspect ratio would hurt the performance of the nearest neighbor estimator. By the same argument, other estimators using a weight matrix based upon isotropic dependencies should also suffer a loss in performance.

Given the point distributions of the random variates, we simulated a mixed-regressive spatially autoregressive process where Y is an n vector of the correlated dependent variable, φ is the autoregressive parameter, W is an n by n Delaunay spatial weight matrix formed from the spatially random point distribution, X_2 is an n element $N(0,1)$ vector, $\beta = [0 \quad 1 \quad -0.5]$, and ε is an n element vector of $N(0,1)$ random errors.

$$Y = (I - \rho W)^{-1}([1]\beta_1 + X_2\beta_2 + WX_2\beta_3) + (I - \rho W)^{-1}\varepsilon \quad (30)$$

For each of the six values of the autoregressive parameter φ ($\varphi = [0.10 \ 0.25 \ 0.50 \ 0.75 \ 0.90 \ 0.95]$), we ran a pair of separate realizations of 5,000 observations each. We repeated this 100 times. Hence, we ran 100 pairs of realizations of the 5,000 element vector of correlated random variates for each of the six values of the autoregressive parameter.

To place the results in perspective, we compared the nearest neighbor estimates based upon the pooled observations to (1) estimates from using the correct model (Delaunay weights) for each one of the series comprising the pair, (2) estimates from using the Delaunay weights based upon the pooled observations, and (3) the OLS estimates based upon the pooled observations. Table 2 presents the results of the experiment. As an illustration of the accuracy of the simulation and estimation, the average estimated autoregressive parameter was on average within two places of the true autoregressive parameter.

The root mean-squared error of the nearest neighbor estimates relative to the root mean-squared error of the estimates based upon the true model appear in the column $RMSE_{ML-1} / RMSE_{True-Delaunay}$, the root mean-squared error of the maximum likelihood estimates using Delaunay weights for the pooled observations (ignoring the boundary) relative to the root mean-squared error of the estimates based upon the true model appear in the column $RMSE_{Iso-Delaunay} / RMSE_{True-Delaunay}$, while the root mean-squared error of the OLS estimates relative to the estimates based upon the true model appear in the column $RMSE_{OLS} / RMSE_{True-Delaunay}$.

The performance of the nearest neighbor (ML-1), the Delaunay model applied to the pooled observations, and OLS degrade as the degree of autocorrelation increases. However, the nearest neighbor performs relatively better than OLS for high levels of autocorrelation. For an aspect ratio of 2 and $\varphi = 0.95$ the closest neighbor estimator does 67% worse than the estimator based upon the true model but the OLS estimator performs 166% worse and the estimator using the Delaunay weights based upon the pooled observations performs only 2% worse.

The increase in the aspect ratio adversely affects the performance of all the estimators. However, OLS suffers relatively smaller losses compared to the nearest neighbor. For an aspect ratio of 200,000 and $\varphi = 0.95$ the closest neighbor estimator does 171% worse than the estimator based upon the true model, the OLS estimate performs 222% worse, and the estimator using the Delaunay weights based upon the pooled observations performs 126% worse. Interestingly, the anisotropy in this case produced a greater deleterious effect upon the more complicated model (Delaunay on the pooled observations) than it did upon the simpler nearest neighbor and *a fortiori* OLS. The point geometry (*e.g.*, aspect ratio) changes the nature of the spatial dependence and hence the performance of the estimators vary.

Note, the Delaunay W is symmetric and for this two-dimensional example will on average have six neighbors for each observation. Although the true process is symmetric, two-dimensional, and anisotropic, the nearest neighbor method outperforms OLS. For this experiment, an imperfect but limited model of spatial dependence can still outperform OLS, which assumes no dependence.

VIII. Computational Considerations

To find the nearest neighbors one usually employs a quadtree algorithm or a Delaunay triangularization algorithm. Optimal versions of these algorithms require $O(n\log(n))$ operations (Eppstein, Paterson, and Yao (1997)). The OLS computations used by the estimator require $O(nk^3)$ operations.

To examine this further, we generated some random points for the dependent variable, the independent variables ($k=21$), and their two-dimensional locational coordinates. Table 3 contains the timings. The timings coincide well with the predictions of $O(n\log(n))$ performance.

The nearest neighbor spatial autoregressions require almost no time for the sample sizes used in practice, require less than one minute for 100,000 observations, and under four minutes for 500,000 observations. We measured the times on a dual 500 megahertz Pentium III with one gigabyte of memory running Matlab 5.3 under Windows NT 4.0.

We used a Delaunay triangle based method for finding the nearest neighbor. The routines do not require the use of sparse matrices and hence could be implemented in other computer languages. The data and software is available at www.finance.lsu.edu/re.

IX. Conclusion

Even from a set of irregularly arranged points, nearest neighbor graphs display several important features which lead to a very simple eigenstructure of the associated adjacency matrix. In turn, this simple eigenstructure leads to an extremely simple closed-form for the log-determinant. Finally, this leads to a simple closed-form solution to the spatially

autoregressive or mixed regressive spatially autoregressive maximum likelihood first-order conditions.

The results for nearest neighbor dependency have a number of applications. First, its mathematical simplicity could aid theoretical investigations of more complicated situations where space plays a role. The nearest neighbor model combines certain aspects of spatial statistical and time series dependence. Some observations have the simultaneous dependence (*i.e.*, the 2-node-cycles) the hallmark of spatial statistics while other observations depend strictly upon the previous node in the graph. As demonstrated by the Monte Carlo experiment, the low cost of computation facilitates the execution of simulation experiments. Hence, one could investigate other estimators than the ones studied here. For example, one could easily compute the maximum likelihood SAR estimator. The asymmetry of the weights makes it difficult, however, to develop a traditional maximum likelihood CAR version of the nearest neighbor model (Besag (1974), Besag (1975)).

Spatial robustness is one area where the simplicity of the nearest neighbor estimator might play a role. Intuitively, the nearest neighbor model could suffer more severely from spatial outliers than more complicated models of spatial dependence. However, its simple structure might permit robustification without making the resulting estimator overly difficult to use. In fact, the nearest neighbor model might provide a means of detecting such spatial outliers.

Second, the easy computation of the estimates combined with the simplicity of the method may confer a pedagogical advantage to the nearest neighbor model of spatial

dependence. Since calculating the log-determinant only requires counting the number of symmetric elements in the weight matrix, students could compute simple examples by hand.

Third, since more general spatial dependence models usually include the nearest neighbor, focusing on the nearest neighbor has a certain generality with respect to the different approaches. As the nearest neighbor estimator should underfit the overall spatial dependence, it provides a conservative estimate of the spatial structure and of the possible improvement one could obtain with spatial methods. Both the empirical example and the Monte Carlo study provided herein corroborate this assertion. Indeed, underfitting may be preferable to overfitting. See Griffith (1996, p. 80)) for a discussion of the whether to underfit or overfit. Intuitively, underfitting may serve as a robust strategy in the presence of uncertain anisotropy. The Monte Carlo experiment conducted in this paper tends to support this view. In the experiment, anisotropy actually degraded the performance of the more complicated estimator (using Delaunay weights based upon all the points ignoring the boundary) relative to the simpler nearest neighbor estimator.

Overfitting may also play a role. A liberal estimate of improvement obtained by overfitting a spatial model (*e.g.*, placing many variants of spatially lagged dependent and independent variables on the right-hand-side of the equation and estimating by OLS) could approximately bound the possible performance gains obtainable through spatial estimation.

Fourth, the generality of nearest neighbor dependence and its ease of computation suggest the use of the nearest neighbor maximum likelihood estimate of spatial

dependence as a possible diagnostic. Again, it should underestimate the total degree of spatial dependence and serve as a useful foil to upwardly biased EGLS estimates of spatial dependence. In addition, decomposable or local spatial autocorrelation statistics have become quite popular.⁵ One could perform a similar decomposition (by subgraph) on the nearest neighbor estimator.

Finally, the nearest neighbor maximum likelihood estimator could provide a more enlightened null hypothesis than spatial independence. In many cases, such as with real estate pricing models, assuming the null hypothesis of spatial error independence is a “straw man,” since such data virtually always exhibit strong spatial dependence. Naïve models can serve as worthy adversaries, forcing more complicated models to justify their existence. For example, in the forecasting arena, many sophisticated models do not dominate their simpler counterparts in ex-sample prediction (Fildes and Makridakis (1995)). Possibly, the same phenomenon could occur with spatial models.

⁵ See Boots and Dufournaud (1994), Anselin (1995), Ord and Getis (1995), Lowell (1997), Tiefelsdorf and Boots (1997), and Sokal et al. (1998) for recent work on spatial diagnostics.

References

- Anselin, Luc, *Spatial Econometrics: Methods and Models*, Dordrecht: Kluwer Academic Publishers, 1988.
- Anselin, Luc, "Local Indicators of Spatial Association — LISA," *Geographical Analysis*, 27, (1995), p. 93-115.
- Anselin, Luc, and Oleg Smirnov, "Efficient Algorithms for Constructing Higher Order Spatial Lag Operators," *Journal of Regional Science*, 36, (1996), p. 67-89.
- Bavaud, Francois, "Models for Spatial Weights: A Systematic Look," *Geographical Analysis*, 30, (1998), p. 153-171.
- Besag, J. E., "Spatial Interaction and the Statistical Analysis of Lattice Systems," *Journal of the Royal Statistical Society, B*, 36, (1974), p. 192-225.
- Besag, J. E., "Statistical Analysis of Non-lattice Data," *The Statistician*, 24, (1975), p. 179-195.
- Boots, Barry and Christian Dufournaud, "A Programming Approach to Minimizing and Maximizing Spatial Autocorrelation Statistics," *Geographical Analysis*, 26, (1994), p. 54-66.
- Cramer, J. S., *Econometric Applications of Maximum Likelihood Methods*, New York: Cambridge University Press, 1991.
- Eppstein, D., M.S. Paterson, and F.F. Yao, "On Nearest-Neighbor Graphs," *Discrete and Computational Geometry*, 17, (1997), p. 263-282.

- Fildes, R. and S. Makridakis, "The Impact of Empirical Accuracy Studies on Time Series Analysis and Forecasting," *International Statistical Review*, 63, (1995), p. 289-308.
- Gasim, Ali Abdul, "First-Order Autoregressive Models: A Method of Obtaining Eigenvalues for Weighting Matrices," *Journal of Statistical Planning and Inference*, 18, (1988), p. 391-398.
- Griffith, Daniel, "Some Guidelines for Specifying the Geographical Weights Matrix Contained in Spatial Statistical Models," *Practical Handbook of Spatial Statistics*, Sandra Arlinghaus, Editor, Boca Raton: CRC Press, 1996, p. 65-82.
- Griffith, Daniel, and Akio Sone, "Trade-offs Associated with Normalizing Constant Computational Simplifications for Estimating Spatial Statistical Models," *Journal of Statistical Computation and Simulation*, 51, (1995), p. 165-183.
- Lowell, Kim, "Effect(s) of the "No-Same-Color-Touching" Constraint on the Join-Count Statistic: A Simulation Study," *Geographical Analysis*, 29, (1997), p. 339-353.
- Martin, R. J., "Approximations to the Determinant Term in Gaussian Maximum Likelihood Estimation of Some Spatial Models," *Communications in Statistics - Theory and Methods*, 22, (1993), p. 189-205.
- Ord, J. K., "Estimation Methods for Models of Spatial Interaction," *Journal of the American Statistical Association*, 70, (1975), p. 120-126.
- Ord, J. K. and Arthur Getis, "Local Spatial Autocorrelation Statistics: Distributional Issues and an Application," *Geographical Analysis*, 27, (1995), p. 286-306.

Pace, R. Kelley, and Ronald Barry, "Quick Computation of Regressions with a Spatially Autoregressive Dependent Variable," *Geographical Analysis*, 29, (1997), p. 232-247.

Sokal, Robert, Neal Oden, and Barbara Thomson, "Local Spatial Autocorrelation in a Biological Model," *Geographical Analysis*, 30, (1998), p. 331-354.

Strang, Gilbert, *Linear Algebra and Its Applications*, New York: Academic Press, 1976.

Tiefelsdorf, Michael, and Barry Boots, "A Note of the Extremities of Local Moran Iis and Their Impact on Global Moran I*," *Geographical Analysis*, 29, (1997), p. 248-257.

Table 1 — OLS, Nearest Neighbor, and Four Nearest Neighbors ML Estimates

	OLS	ML-1	ML-1,3	ML-4
Intercept	0.981	0.750	0.748	0.458
LR ($\beta_1 = 0$)	411.2	257.8	252.9	92.4
ln(Population > 18 years of Age)	-0.846	-0.749	-0.746	-0.717
LR ($\beta_2, \beta_6 = 0$)	1205.6	985.4	980.65	882.4
ln(Population with Education > 12 years)	0.517	0.290	0.284	0.191
LR ($\beta_3, \beta_7 = 0$)	954.4	595.1	591.0	257.1
ln(Owner Occupied Housing Units)	0.429	0.446	0.447	0.451
LR ($\beta_4, \beta_8 = 0$)	529.3	645.5	653.2	732.7
ln(Aggregate Income)	-0.144	-0.033	-0.032	0.027
LR ($\beta_5, \beta_9 = 0$)	52.4	55.9	57.0	27.0
Optimal α for ML-4				0.615
LR($\alpha=0$)				1256.5
Optimal α for ML-1		0.335		
LR($\alpha = 0$)		684.8		
Optimal α_o for ML-1,3			0.341	
LR($\alpha_o = 0$)			704.4	
Optimal ω_1 for ML-1,3			0.733	
Optimal ω_3 for ML-1,3			0.267	
R^2	0.524	0.642	0.645	0.712
Median $ e $	0.086	0.073	0.073	0.061
Log-likelihood	-6307.1	-5964.6	-5954.9	-5678.8
n	3107	3107	3107	3107
k	5	10	15	10

Table 2 — Closest Neighbor and OLS Performance Under Misspecification and Anisotropy

Aspect Ratio	α	$\tilde{\alpha}_{Iso-Delaunay}$	$RMSE_{Iso-Delaunay}$	$RMSE_{ML-1}$	$RMSE_{OLS}$
			$RMSE_{True-Delaunay}$	$RMSE_{True-Delaunay}$	$RMSE_{True-Delaunay}$
2	0.10	0.10	1.00	1.01	1.02
2	0.25	0.25	1.00	1.02	1.02
2	0.50	0.50	1.00	1.06	1.09
2	0.75	0.75	1.00	1.20	1.34
2	0.90	0.90	1.01	1.45	1.94
2	0.95	0.95	1.02	1.67	2.66
2,000	0.10	0.06	1.01	1.02	1.02
2,000	0.25	0.17	1.01	1.02	1.02
2,000	0.50	0.35	1.06	1.08	1.10
2,000	0.75	0.57	1.20	1.28	1.37
2,000	0.90	0.73	1.58	1.76	2.12
2,000	0.95	0.80	2.08	2.39	3.12
200,000	0.10	0.06	1.01	1.02	1.02
200,000	0.25	0.13	1.02	1.02	1.03
200,000	0.50	0.28	1.07	1.09	1.10
200,000	0.75	0.50	1.25	1.33	1.39
200,000	0.90	0.67	1.69	1.93	2.16
200,000	0.95	0.75	2.26	2.71	3.22

**Table 3 — Time in Seconds Needed for Locating Neighbors and
Computing Estimates versus n**

n	Neighbor Times	Estimate Times	Total Times
500	0.09	0.03	0.13
1,000	0.19	0.05	0.23
5,000	1.20	0.19	1.39
10,000	2.58	0.36	2.94
25,000	7.20	0.98	8.19
50,000	15.11	2.03	17.14
100,000	32.22	4.14	36.36
250,000	87.67	10.39	98.06
500,000	182.75	20.72	203.47