Estimation Methods for Models of Spatial Interaction

KEITH ORD*

Autoregressive models for spatial interaction have been proposed by several authors (Whittle [15] and Mead [11], for example). In the past, computational difficulties with the ML approach have led to the use of alternative estimators. In this article, a simplified computational scheme is given and extended to mixed regressive-autoregressive models. The ML estimator is compared with several alternatives.

1. INTRODUCTION

Suppose that observations on the random variable Y are recorded at each of n 'locations.' The purpose of this paper is to develop practical estimation methods for certain models which attempt to describe the interaction between (neighboring) locations for the variate. In this context, the locations may be either points (such as plants in a field) or areas (administrative counties in a state), and may form either a regular or an irregular pattern. The data refer to attributes of the particular location (height of a plant, population of a county) rather than defining the locations themselves (as would be the case in point pattern studies). The term "neighbor" has been used in previous studies to indicate strict physical contiguity. However, this is unnecessary and we shall formulate the model with reference to the variate Y_i, at location i, in terms of the set of locations $J(i) = \{i_1, i_2, \cdots, i_{m_i}\}$, whose variates interact with Y_i. $J(i)$ may include all locations other than i.

The models we consider follow the joint probability formulation of Whittle [15, 16]. We shall concentrate on the first order autoregressive model

$$Y_i = \alpha + \rho \sum w_{ij} Y_j + \epsilon_i \ , \quad (i = 1, \cdots, n) \ . \tag{1.1}$$

In Equation (1.1), as elsewhere, the sum is understood to be over $j \in J(i)$. Also, α and ρ are parameters, while the $\{w_{ij}\}$ are a set of nonnegative weights, to be defined, which represent the 'degree of possible interaction' of location j on location i (see Section 1.1). By convention, we always set $w_{ii} = 0$. The random disturbance terms, ϵ_i, are uncorrelated with zero means and equal variances. Usually, we shall take them to be normally distributed so that $\epsilon_i \sim IN(0, \sigma^2)$.

Taking expectations in (1.1) we see that

$$E(Y_i) = \alpha + \rho \sum w_{ij} E(Y_j) \ , \tag{1.2}$$

emphasizing the joint, or unconditional, nature of the model specification. This contrasts with the conditional model studied by Bartlett [2, 3] and Besag [4, 5] for which

$$E\{Y_i | Y_j = y_j, j \in J(i)\} = \alpha + \rho \sum w_{ij} y_j \ . \tag{1.3}$$

Equations (1.1) and (1.3) are equivalent only if

$$E\{\varepsilon_i | Y_j = y_j, j \in J(i)\} = 0 \ ; \tag{1.4}$$

that is, if ϵ_i and $\sum w_{ij} Y_j$ are uncorrelated. The nature of these restrictions may be indicated by an example.

Example. For the time series model, $w_{ij} = 1$ if $j = i - 1$ and $w_{ij} = 0$ elsewhere, so that equation (1.1) reduces to the familiar first order Markov scheme

$$Y_i = \alpha + \rho Y_{i-1} + \epsilon_i \ . \tag{1.5}$$

Here, the conditions (1.4) correspond to the familiar assumption that ϵ_i is uncorrelated with $Y_{i-1}, Y_{i-2}, \cdots$.

The time-dependent, one-sided nature of model (1.5) makes equation (1.4) a natural restriction, but there is no such ready motivation for a purely spatial model. If the conditional model is employed, estimation of the model may be carried out by splitting the locations into two mutually exclusive subsets, A and B, such that

$$E\{Y_i | Y_j = y_j, j \in J(i)\} = E\{Y_i | Y_j = y_j, j \in B\} \tag{1.6}$$

for all $i \in A$. Given this division, the parameters can be estimated by ordinary least squares (OLS). However, when the joint formulation is used, the failure to satisfy (1.4) implies that the OLS estimators are inconsistent (Whittle, [15]; see also Section 3.1). In the remainder of this article we develop viable alternative estimation procedures for the joint model.

1.1 The Choice of Weights

The set of weights specified in the example was a natural choice for a time series model, but the selection of weights is more difficult for spatial models. Thus, when the locations are equally spaced, we might set $w_{ij} > 0$

* Keith Ord is senior lecturer, Department of Statistics, University of Warwick, Coventry CV4 7AL, England. This research was carried out while the author was visiting Pennsylvania State University. The article is a revised version of a paper given at the American Statistical Association's annual meeting in Montreal, August 1972. The author wishes to thank Julian Besag of Liverpool University for several interesting discussions and for providing him with an advance copy of Dr. Besag's 1974 paper. He is also grateful to both referees for helpful comments and to Anne Kempson for typing the manuscript.

© Journal of the American Statistical Association
March 1975, Volume 70, Number 349
Theory and Methods Section

only when j is a near neighbor of i. For example, on a square lattice, each interior location has four immediate neighbors and we would use the set of weights $\{w_{ij} = \frac{1}{4}$, if j is an immediate neighbor of i; $w_{ij} = 0$, otherwise$\}$. Boundary locations may need separate treatment, depending on whether the boundary is natural or artificial, as in an experimental plot. More generally, for irregular areal patterns, Cliff and Ord [6] proposed the use of weights based upon the distances between locations (county centers) and lengths of common boundary. Mead [12] used a combination of distances, directions and item size in his study of plant competition.

To lend a natural interpretation to ρ, the scaling $\sum w_{ij} = 1$ may be used for each location, where the sums are over either i or j. This scaling implies that $\rho < 1$. The weights may be symmetric, but need not be so, while the suggested scaling will usually produce asymmetric weights.

2. ESTIMATION USING MAXIMUM LIKELIHOOD (ML)

2.1 Form of the Likelihood Function

The model given in (1.1) may be reformulated in matrix terms, taking $\alpha = 0$, as

$$\mathbf{Y} = \rho\mathbf{WY} + \boldsymbol{\varepsilon} \ ; \tag{2.1}$$

where $\mathbf{W}$ is the $(n \times n)$ matrix of weights and $\mathbf{Y}$, $\boldsymbol{\varepsilon}$ are $(n \times 1)$ vectors. The constant, α, is suppressed to simplify the initial exposition and will be restored, in the context of a mixed regressive-autoregressive model, in Section 4. From (2.1), we have $\boldsymbol{\varepsilon} = \mathbf{AY}$, where

$$\mathbf{A} = \mathbf{I} - \rho\mathbf{W} \ . \tag{2.2}$$

Given $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$, the log-likelihood function for ρ, σ^2, given that $\mathbf{Y} = \mathbf{y}$, is

$$\ell(\rho, \sigma^2) = -(n/2) \ln (2\pi\sigma^2) - (1/2\sigma^2)\mathbf{y}'\mathbf{A}'\mathbf{Ay} + \ln |\mathbf{A}| \ . \tag{2.3}$$

From (2.3), we have the ML estimators

$$\hat{\sigma}^2 = n^{-1}\mathbf{y}'\mathbf{A}'\mathbf{Ay} \tag{2.4}$$

and $\hat{\rho}$ as that value of ρ which maximizes (Mead)

$$\ell(\rho, \hat{\sigma}^2) = \text{const} - (n/2) \ln (\hat{\sigma}^2|\mathbf{A}|^{-2/n}) \ . \tag{2.5}$$

In this form, $|\mathbf{A}|$ is an nth order polynomial in ρ which must be evaluated afresh at each iteration. The explicit form of these polynomials is known for some regular configurations of sample points, but the evaluation of ρ becomes (computer) time consuming when n is not small and/or the lattice is irregular. To circumvent this difficulty, Whittle [15] developed a large sample approach based on spectral methods, with the considerable advantage that it can be applied to more general models than (1.1). However, many experimental situations involve small lattices for which the discounting of edge effects may not be negligible. Also, the approach is not easily extended to handle irregular lattices. It is evident that an alternative computational procedure is required, which we now outline.

2.2 A New Computational Procedure

The principal difficulty in determing $\hat{\rho}$ from (2.5) centers on the evaluation of

$$|\mathbf{A}| = |\mathbf{I} - \rho\mathbf{W}| \ . \tag{2.6}$$

However, if $\mathbf{W}$ has eigenvalues $\lambda_1, \cdots, \lambda_n$ it is well known that

$$|\lambda\mathbf{I} - \mathbf{W}| = \prod_{i=1}^{n} (\lambda - \lambda_i)$$

so that (2.7)

$$|\mathbf{A}| = \prod_{i=1}^{n} (1 - \rho\lambda_i) \ .$$

The advantage of (2.7) is that the $\{\lambda_i\}$ can be determined once and for all, so that $\hat{\rho}$ is the value of ρ which minimizes

$$\{\prod_{i=1}^{n} (1 - \rho\lambda_i)\}^{-2/n}(\mathbf{y}'\mathbf{y} - 2\rho\mathbf{y}'\mathbf{y}_L + \rho^2\mathbf{y}'_L\mathbf{y}_L) \tag{2.8}$$

where $\mathbf{y}_L = \mathbf{Wy}$.

For the results reported in Section 5, $\hat{\rho}$ was determined from (2.8) using a direct search procedure, although for more complex models a formal iterative procedure would be preferable (see Appendix A). The asymptotic variance-covariance matrix is given in Appendix B.

The evaluation of the $\{\lambda_i\}$ will usually be a computer job, and the time involved becomes large as n increases. However, the following points are worth noting:

a. All $\lambda_i = 0$ when $\mathbf{W}$ is upper or lower triangular, but with zeros on the main diagonal.
b. The largest eigenvalues are the most important, so that it may suffice to use only the first few values (most programs generate these in decreasing order).
c. For certain regular patterns, partial or complete evaluation may be possible (see Appendix C).
d. For large numbers of locations, it will often be possible to force $\mathbf{W}$ into a block diagonal form by the elimination of a small number of nonzero weights. Carried out with care, such a procedure considerably reduces the computational burden and will not affect the asymptotic properties of the estimator (see Appendix D).

3. OTHER ESTIMATION PROCEDURES

Although the results of Section 2 put the ML solution within computational reach, it seems worthwhile to review other approaches briefly and to comment on their performance.

3.1 The Least Squares Estimator of ρ

The LS estimator is the solution of the equation

$$0 = \mathbf{y}'\mathbf{Wy} - \rho\mathbf{y}'\mathbf{W}'\mathbf{Wy} = \mathbf{y}'(\mathbf{I} - \rho\mathbf{W}')\mathbf{Wy} = \mathbf{e}'\mathbf{Wy} \ , \tag{3.1}$$

where $\mathbf{e} = (\mathbf{I} - \rho\mathbf{W})\mathbf{y}$ represents the observed disturbance terms. Equation (3.1) is an estimating equation (Durbin, [10]). This will be an unbiased estimating equation only if

$$E(\boldsymbol{\varepsilon}'\mathbf{WY}) = 0 \tag{3.2}$$

implying that tr $\{\mathbf{W}'(\mathbf{I} - \rho\mathbf{W})^{-1}\} = 0$, where tr is the trace operator. Condition (3.2) is equivalent to condition

(1.4) considered previously. To ensure consistency, we require that

$$\operatorname*{plim}_{n\to\infty} \{\boldsymbol{\varepsilon}'\mathbf{WY}/\mathbf{Y}'\mathbf{W}'\mathbf{WY}\} = 0$$

or that (3.3)

$$\lim_{n\to\infty} [\operatorname{tr}\{\mathbf{W}'(\mathbf{I}-\rho\mathbf{W})^{-1}\}/ \operatorname{tr}\{(\mathbf{I}-\rho\mathbf{W}')^{-1}\mathbf{W}'\mathbf{W}(\mathbf{I}-\rho\mathbf{W})^{-1}\}] = 0 \ .$$

As noted earlier, condition (1.4) holds when $\mathbf{W}$ is upper or lower triangular, but not in general. Thus the LS estimator is inconsistent, as previously shown by Whittle [15]. To remove this lack of consistency, we searched for a matrix $\mathbf{G}$ such that

$$E(\boldsymbol{\varepsilon}'\mathbf{W}'\mathbf{GY}) = 0 \ . \tag{3.4}$$

It can be shown that $\mathbf{G}$ must be a function of ρ and the simplest choice is $\mathbf{G} = \mathbf{I} - \rho\mathbf{W} = \mathbf{A}$. The least squares estimator, $\tilde{\rho}$, is a solution of the quadratic in ρ,

$$\mathbf{y}'\mathbf{A}'\mathbf{WAy} = 0 \ . \tag{3.5}$$

Unfortunately, the efficiency of such estimators relative to the ML estimators declines drastically as ρ increases. The results will vary with $\mathbf{W}$, but the following example appears to be fairly typical.

Example. For the time series model given in equation (1.5), we find that

$$\lim_{n\to\infty} [\operatorname{var}\{\rho(\mathrm{ML})\}/\operatorname{var}(\tilde{\rho})] = (1-\rho^2)^2/(1+\rho^2) \ .$$

This declines from 0.89 at $\rho = 0.2$ to 0.30 at $\rho = 0.6$ and only 0.08 at $\rho = 0.8$.

3.2 The Pairs Estimator

An alternative procedure, suggested by Mead [12] for regular lattices, is to consider the interaction between pairs of locations, ignoring all other interactions. More generally, we could use an alternative weighting matrix $\mathbf{W}_1$, which possessed a simple structure for computational purposes. That is, we could work with the revised model

$$\mathbf{Y} = \rho\mathbf{W}_1\mathbf{Y} + \boldsymbol{\varepsilon}_1 \tag{3.6}$$

where

$$\boldsymbol{\varepsilon}_1 = \boldsymbol{\varepsilon} + \rho(\mathbf{W} - \mathbf{W}_1)\mathbf{Y} \ . \tag{3.7}$$

Unfortunately, this approach yields inconsistent estimators, as shown in Appendix D. Even when the lack of consistency is ignored, this approach seems to throw away a considerable amount of information. For example, if every location has equal nonzero weights for exactly k other locations, but only one of the k is used to form a pair, then the ratio of asymptotic variances is

$$\operatorname{var}\{\rho(\mathrm{ML})\}/\operatorname{var}\{\rho(\mathrm{pairs})\} = k^{-1} + O(\rho) \ . \tag{3.8}$$

4. REGRESSION MODELS

In addition to the adaptive model considered in Section 2, we usually wish to consider variations in the mean level. In general, we consider the mixed regressive-autoregressive model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \rho\mathbf{WY} + \boldsymbol{\varepsilon} \ , \tag{4.1}$$

where $\mathbf{X}$ is an $(n \times k)$ matrix with the first column consisting of ones, and $\boldsymbol{\beta}$ is a $(k \times 1)$ vector of parameters. Equation (4.1) may be rewritten as

$$\boldsymbol{\varepsilon} = (\mathbf{I} - \rho\mathbf{W})\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} \ . \tag{4.2}$$

For ρ known, the usual OLS estimators for $\boldsymbol{\beta}$ are best linear unbiased. For ρ unknown, the ML procedure leads to estimators of the same form, with the ML estimator $\hat{\rho}$ replacing ρ. To demonstrate this, we let $\mathbf{z} = (\mathbf{I} - \rho\mathbf{W})\mathbf{y}$ and solve the ML equations for $\boldsymbol{\beta}$ and ρ. The estimators for $\boldsymbol{\beta}$ and σ^2 are

$$\boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{z}} \tag{4.3}$$

$$\hat{\sigma}^2 = (1/n)\hat{\mathbf{z}}'\mathbf{M}\hat{\mathbf{z}} \tag{4.4}$$

where $\hat{\mathbf{z}} = (\mathbf{I} - \hat{\rho}\mathbf{W})\mathbf{y}$, and $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Substituting back into the likelihood function, $\hat{\rho}$ is that value of ρ which maximizes, as before,

$$\ell(\rho, \hat{\sigma}^2, \hat{\boldsymbol{\beta}}) = \text{const} - (n/2)\ln(\hat{\sigma}^2|\mathbf{A}|^{-2/n}) \ . \tag{2.5 bis}$$

For computational purposes, we use expression (2.8), except that the second bracket is replaced by

$$\mathbf{y}'\mathbf{My} - 2\rho\mathbf{y}'\mathbf{My}_L + \rho^2(\mathbf{y}_L)'\mathbf{My}_L \ . \tag{4.5}$$

If the weights are scaled so that $\sum_i w_{ij} = 1$ for all i, then the estimator for β_1 (corresponding to the constant term in the equation) has the intuitively appealing form $\hat{\beta}_1 = \bar{y}(1 - \hat{\rho})$.

The asymptotic variance-covariance matrix is given in Appendix B.

4.1 An Autoregressive Model for the Error Terms

An alternative to model (4.1), may be specified as follows:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U} \ , \tag{4.6}$$

and

$$\mathbf{U} = \rho\mathbf{WU} + \boldsymbol{\varepsilon} \tag{4.7}$$

where $\boldsymbol{\varepsilon}$ denotes a vector of random error terms with the same properties as in (2.1). For ρ known, we can eliminate $\mathbf{U}$ to obtain

$$\mathbf{Y} = \rho\mathbf{WY} + \mathbf{X}\boldsymbol{\beta} - \rho\mathbf{WX}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \ , \tag{4.8}$$

and, in this case, $\boldsymbol{\beta}$ is readily estimated by least squares. However, when ρ is unknown, even the use of constrained least squares produces inconsistent estimators. Cochrane and Orcutt [9] have devised an iterative estimation procedure for the time series model corresponding to (4.6) and (4.7), which can be adapted to handle the spatial model. The revised scheme is as follows:

1. Compute the OLS residuals from (4.6), $\tilde{\mathbf{u}}$ say.
2. Estimate ρ from $\tilde{\mathbf{u}} = \rho\mathbf{W}\tilde{\mathbf{u}} + \boldsymbol{\varepsilon}$, using the ML method outlined in Section 2.2. Call this estimate $\tilde{\rho}$.
3. Construct the new variables $\tilde{\mathbf{z}} = (\mathbf{I} - \tilde{\rho}\mathbf{W})\mathbf{y}$ and $\tilde{\mathbf{X}} = (\mathbf{I} - \tilde{\rho}\mathbf{W})\mathbf{X}$.
4. Apply OLS for $\tilde{\mathbf{z}}$ on $\tilde{\mathbf{X}}$, yielding new estimates $\tilde{\boldsymbol{\beta}}$.
5. Construct the new residuals $\tilde{\mathbf{u}} = \mathbf{Y} - \mathbf{X}\tilde{\boldsymbol{\beta}}$ and return to Step 2.

Repeat steps 2 through 5 until convergence is obtained. This procedure has been applied in only a limited number of cases, but convergence seems fairly rapid (see Section 5). For the time series case, Sargan [14] showed that convergence to a local minimum is guaranteed. This remains unproven for the spatial model. Also there is the possibility of several local minima, a situation which definitely arises for very small lattices.[1] This did not upset the analysis in Section 5.

5. ANALYSIS OF AN IRREGULAR LATTICE

O'Sullivan [13] explored the use of an index of arterial road accessibility (ARA) to decribe geographical patterns in Eire (Southern Ireland). The index was found as follows:

the index for county N is

$$A(i, N) = \sum_{j=1}^{m} d_{ij} \tag{5.1}$$

where m is the number of vertices in the network and d_{ij} is the shortest distance in miles (by road) between vertices i and j.

Thus, a high value of the index implies that the city is relatively inaccessible. The road network was taken as all T class roads in Eire and n represented the 31 towns of population 5,000 or over at the time of the 1961 census. Index values for the 26 counties were taken corresponding to the largest town in the county (if it was one of the 31), or by interpolation from neighboring towns otherwise. Criticisms of this index are possible (see, e.g., [8, Ch. 6]), but it does appear to describe the broad features of road accessibility to produce-markets satisfactorily.

O'Sullivan considered several features of the Irish economy and tried to relate these to his ARA index. In particular, he considered the relationship between

Y = percentage of gross agricultural output of each county consumed by itself (output in value terms)

and

X = ARA. The data are given in Table 1.

It is argued that remote areas will tend to be more self-sufficient so that Y and X should be positively correlated. This hypothesis is supported by O'Sullivan's regression analysis (see Table 2).

However, before accepting such a model the pattern of residuals should be examined. Cliff and Ord [7] have given a test for spatial autocorrelation among the residuals. This test considers an error model of the form of (4.7) and compares H_0: $\rho = 0$ with the general alternative H_1: $\rho \neq 0$. The test requires the specification of the weighting matrix **W**. To allow for the irregular

1. Data and Regression Residuals for the Example

County	Pct. of own produce consumed		ARA		Residuals from ML model
	y	y_L	x	x_L	$(y - \hat{y})$
Carlow	8.6	9.0	3664	3676	−1.74
Cavan	15.0	15.7	5000	4464	−2.35
Clare	19.0	16.6	4321	4326	3.55
Cork	9.0	13.7	4118	4543	−3.77
Donegal	27.0	23.1	7500	5000	−2.71
Dublin	9.4	9.1	3078	3733	1.39
Galway	21.9	22.4	4537	5006	4.02
Kerry	17.0	10.3	5140	4116	1.15
Kildare	9.0	10.5	3200	3681	−0.29
Kilkenny	8.0	9.0	3708	3711	−2.44
Laoghis	10.3	11.2	3455	3813	−0.07
Leitrim	23.1	19.7	5000	5106	4.98
Limerick	11.4	12.7	4018	4220	−1.11
Longford	19.0	18.3	4250	4348	3.01
Louth	10.1	10.0	3948	4152	−0.91
Mayo	30.0	22.2	6815	5131	3.65
Meath	8.7	12.3	4008	3915	−4.17
Monaghan	13.0	13.5	4500	4695	−0.85
Offaly	14.3	12.1	4108	3685	0.70
Roscommon	23.0	22.1	4500	5060	5.52
Sligo	22.0	25.9	5997	5611	−2.12
Tipperary	9.0	11.4	3926	3918	−3.09
Waterford	8.0	8.7	3691	3907	−1.84
Westmeath	16.0	15.5	3872	4179	2.56
Wexford	8.6	8.9	3940	3658	−2.79
Wicklow	10.2	8.9	3600	3456	−0.27

nature of the county pattern, Cliff and Ord [6] took

$$w_{ij} \propto q_i(j)/d_{ij}$$

where

$q_i(j)$ is the proportion of the interior boundary of County i which is in contact with County j;
d_{ij} is the distance between the (geometric) centers of Counties i and j.

These weights were then scaled so that $\sum_j w_{ij} = 1$ for all i.

The full set of weights appears in Cliff and Ord [7, Appendix 2] the eigenvalues of **W** appear in Appendix 3 of the same work.

Given the matrix **W**, the test showed significant positive spatial autocorrelation among the residuals (see Table 3), so it was decided to fit the extended model given by (4.6) and (4.7). The preliminary test of significance is justified in terms of the potentially simpler estimation procedure, as in time series studies.

To assess the importance of using the correct estimation procedure the model was reformulated as

$$\mathbf{Y} = \alpha + \rho \mathbf{Y}_L + \beta \mathbf{x} - \gamma \mathbf{x}_L + \boldsymbol{\varepsilon} \tag{5.2}$$

where $\gamma = \rho\beta$ should be imposed as a constraint. The following alternative estimators were evaluated:

(1) least squares, with $\rho = 0$;
(2) least squares, after setting $\rho = 1$ (taking first differences);
(3) least squares for the full model, ignoring the constraint;
(4) constrained least squares for the full model, subject to $\gamma = \rho\beta$;
(5) maximum likelihood, as in Section 4.1.

[1] Howard Ross-Parker (University of Reading, England) has demonstrated this empirically in an unpublished paper.

2. Results of Analysis of O'Sullivan's Data

Model and method of estimation	Coefficients with standard errors in brackets (all multiplied by 1,000) const	x	x_L	y_L	R^2	VP [a]
Original (OLS)	−8.49	5.27 (0.70)	—	—	0.700	0.700
First differences (OLS)	−0.04	2.45 (0.67)	−2.45 (0.67)	1.0	0.356	0.880
Full model (OLS)	0.63	2.45 (0.68)	−2.65 (2.13)	1.018 (0.233)	0.882	0.882
Full model[b] (constrained LS)	−0.04	2.60 (0.67)	−2.65 (1.06)	1.018 (0.114)	0.881	0.881
Full model[b] (ML)	−1.21	3.87 (0.66)	−1.95 (0.70)	0.505 (0.160)	—	0.815

[a] VP denotes the proportion of variance accounted for by the model; that is error variance/var (y).
[b] The standard errors were computed from the asymptotic formulas. For the ML formulae, see (B.3) in Appendix B.

Schemes (1) and (2) yield consistent and unbiased, but inefficient, estimators for α and β, but an arbitrary value for ρ. Schemes (3) and (4) yield consistent estimators for α and β, but the estimators for ρ are inconsistent. Only the ML scheme yields fully efficient, and consistent estimators for all parameters. The results listed in Table 2 show marked differences between the estimates, and indicate the need for the use of the ML approach. Computationally, the iterative scheme performed well as the successive values for $\tilde{\rho}$ were 0.433, 0.496, 0.5037, 0.5046, 0.5047, the last being correct to four decimal places. The value of the likelihood function changed very little after the second cycle. The standard errors were computed from the asymptotic variance-covariance matrix given in Appendix B.

The distribution of the test statistic is changed when the model of Section 4.1 is used, so that the value of the statistic is computed only as a rough guide. However, there is some evidence (see Table 3) that the model is still not completely satisfactory, although there is a substantial improvement relative to the original formulation. The coefficient for X is reduced by about a quarter, which suggests that the effect of road accessibility is somewhat less important than the original model suggested. The positive value of ρ may be interpreted as an indication of similarity between neighboring counties, probably because of climatic and historical effects on the choice of crops grown.

3. Tests for Spatial Autocorrelation Among the Regression Residuals

Model	Value of I [a]	Value of standard deviate
(1)	0.436	3.6[b]
(2)	0.387	−2.5[b]
(3)	0.392	−2.5[c]
(4)	0.408	−2.6[c]
(5)	0.239	2.0[c]

[a] The statistic $I = \mathbf{e}'\mathbf{W}\mathbf{e}/\mathbf{e}'\mathbf{e}$, and $E(I|\rho = 0) = -(n-1)^{-1} = -0.04$ for this example.
[b] Computed using the expression for var (I) given by Cliff and Ord [8].
[c] Computed using the set of n! random permutations of the $\{e_i\}$; This procedure ignores both the spatial dependence among the regressor variables and the autoregressive nature of the model, but the values are given as a guide in the absence of a correct procedure.

APPENDIX A: THE NEWTON-RAPHSON PROCEDURE FOR THE DETERMINATION OF $\hat{\rho}$

The ML estimator is that value of ρ which minimizes

$$f(\rho) = -\frac{2}{n}\sum_{i=1}^{n} \ln(1-\rho\lambda_i) + \ln(s^2) \tag{A.1}$$

where

$$s^2 \equiv s^2(\rho) = \mathbf{y}'\mathbf{y} - 2\rho\mathbf{y}'\mathbf{y}_L + \rho^2(\mathbf{y}_L)'\mathbf{y}_L$$

and $f(\rho)$ is the logarithm of expression (2.8). From (A.1), the derivatives of $f(\rho)$ are

$$f_\rho(\rho) = \frac{2}{n}\sum_{i=1}^{n} \lambda_i/(1-\rho\lambda_i) + 2(\rho(\mathbf{y}_L)'\mathbf{y}_L - \mathbf{y}'\mathbf{y}_L)/s^2 \tag{A.2}$$

and

$$f_{\rho\rho}(\rho) = \frac{2}{n}\sum_{i=1}^{n} (\lambda_i)^2/(1-\rho\lambda_i)^2 + 2(\mathbf{y}_L)'\mathbf{y}_L/s^2 - 4(\rho(\mathbf{y}_L)'\mathbf{y}_L - \mathbf{y}'\mathbf{y}_L)^2/s^4 \ . \tag{A.3}$$

Then $\hat{\rho}$ may be determined iteratively from the usual expression

$$\rho_{r+1} = \rho_r - f_\rho(\rho_r)/f_{\rho\rho}(\rho_r) \ . \tag{A.4}$$

A useful starting value is $\rho_0 = \mathbf{y}'\mathbf{y}_L/\mathbf{y}'\mathbf{y}$. The alternative $\rho_0 = \mathbf{y}'\mathbf{y}_L/(\mathbf{y}_L)'\mathbf{y}_L$ is not recommended since this value may exceed $1/\lambda_{\max}$, and lie outside the region of feasible values for $\hat{\rho}$.

APPENDIX B: THE VARIANCE OF THE MAXIMUM LIKELIHOOD ESTIMATOR

From the procedure outlined in Appendix A, it can be seen that $(\lambda_{\max})^{-1} > \hat{\rho} > (\lambda_{\min})^{-1}$, since $\lambda_{\min}$ is always negative. Thus, the ML solution yields an interior point in the range of feasible values for ρ, and the large sample variances may be determined from the second derivatives of the log-likelihood function.

The asymptotic variance-covariance matrix, for $\omega = \sigma^2$ and ρ in that order, is

$$\mathbf{V}(\omega, \rho) = \omega^2 \begin{bmatrix} n/2 & E(\boldsymbol{\varepsilon}'\mathbf{Y}_L) \\ & \omega E((\mathbf{Y}_L)'\mathbf{Y}_L) - \alpha\omega^2 \end{bmatrix}^{-1} \tag{B.1}$$

where $\alpha = \partial^2 \ln |\mathbf{A}|/\partial\rho^2 = \sum_{i=1}^{n} (\lambda_i)^2/(1-\rho\lambda_i)^2$. If we set $\mathbf{B} = \mathbf{A}^{-1}\mathbf{W}$, then $E(\boldsymbol{\varepsilon}'\mathbf{Y}_L) = \omega \operatorname{tr}(\mathbf{B})$ and $E((\mathbf{Y}_L)'\mathbf{Y}_L) = \omega \operatorname{tr}(\mathbf{B}'\mathbf{B})$.

For the mixed regressive-autoregressive model of Section 4, the corresponding expression is

$$\mathbf{V}(\omega, \rho, \boldsymbol{\beta}) = \omega^2 \begin{bmatrix} n/2 & E(\boldsymbol{\varepsilon}'\mathbf{Y}_L) & \mathbf{0}' \\ & \omega E((\mathbf{Y}_L)'\mathbf{Y}_L) - \alpha\omega^2 & \omega \mathbf{X}'E(\mathbf{Y}_L) \\ & & \omega \mathbf{X}'\mathbf{X} \end{bmatrix}^{-1} \quad \text{(B.2)}$$

where

$$E(\mathbf{Y}_L) = \mathbf{BX}\boldsymbol{\beta}, \quad E(\boldsymbol{\varepsilon}'\mathbf{Y}_L) = \omega \operatorname{tr}(\mathbf{B})$$

and

$$E((\mathbf{Y}_L)'\mathbf{Y}_L) = \omega \operatorname{tr}(\mathbf{B}'\mathbf{B}) + \{E(\mathbf{Y}_L)\}'\{E(\mathbf{Y}_L)\} .$$

Finally, for the model with autoregressive error terms given in Section 4.1,

$$\mathbf{V}(\omega, \rho, \boldsymbol{\beta}) = \omega^2 \begin{bmatrix} n/2 & E(\boldsymbol{\varepsilon}'\mathbf{Z}_L) & \mathbf{0}' \\ & \omega E((\mathbf{Z}_L)'\mathbf{Z}_L) - \alpha\omega^2 & \mathbf{0}' \\ & & \omega \mathbf{X}'\mathbf{A}'\mathbf{AX} \end{bmatrix}^{-1} \quad \text{(B.3)}$$

where $\mathbf{Z}_L = \mathbf{W}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$, $E(\boldsymbol{\varepsilon}'\mathbf{Z}_L) = \omega \operatorname{tr}(\mathbf{B})$ and $E((\mathbf{Z}_L)'\mathbf{Z}_L) = \omega \operatorname{tr}(\mathbf{B}'\mathbf{B})$.

The expressions (B.1)–(B.3) have been written in expectation form to indicate the most straightforward method of computing an approximate $\mathbf{V}$ from the sample data.

APPENDIX C: THE EVALUATION OF THE EIGENVALUES OF W

For irregularly spaced points, analytic simplification of the equation for the eigenvalues is usually impossible. Existing computers will handle symmetric matrices of order up to 50, depending on the size of the machine. If the number of data points is in excess of this, it may be possible to force $\mathbf{W}$ into a block structure, so that the eigenvalues are evaluated separately for each block. This may involve the elimination of some nonzero weights, and could lead to more than one eigenvalue equal to one. However, the total bias imparted to the likelihood estimator will be slight, given a carefully chosen "blocking" procedure.

Programs for symmetric matrices are more readily available than those for asymmetric forms, but in some cases $\mathbf{W}$ can be "symmetrized." Thus, suppose that the symmetric (unstandardized) matrix $\mathbf{W}_1$ with row sums $\mathbf{w} = \mathbf{W}_1\mathbf{I}$ is converted to the weighting matrix $\mathbf{W} = \mathbf{W}_1\mathbf{D}$ where $\mathbf{D} = \operatorname{diag}(\mathbf{w}^{-1})$. That is, $\mathbf{W}$ is scaled to have row sums equal to one. Then the matrices $\mathbf{W}$ and $\mathbf{W}^* = \mathbf{D}^{\frac{1}{2}}\mathbf{W}_1\mathbf{D}^{\frac{1}{2}}$ have identical eigenvalues, while $\mathbf{W}^*$ is symmetric. Obviously the same results apply for scaling column sums. It is clear that when $\mathbf{W}$ is thus constructed, all the eigenvalues must be real.

When the lattice has a regular grid structure, some analytic progress may be possible. Some results are now given.

Case 1. Whenever $\mathbf{W}$ is upper or lower triangular, all the eigenvalues are zero.

Case 2. The bilateral model $y_i = \frac{1}{2}\rho(y_{i-1} + y_{i+1}) + \epsilon_i$, for $i \neq 1$ or n has weights $w_{ij} = \frac{1}{2}$ for $j = i - 1,\ i + 1,\ 1 \leq j \leq n$, and $w_{ij} = 0$, otherwise. This special form of weighting matrix is denoted by $\frac{1}{2}\mathbf{C}$, for future reference. For a set of n cells, let $d_n = |\alpha\mathbf{I} - \mathbf{C}|$. Then it is readily established that

$$d_n = \alpha d_{n-1} - d_{n-2}, \quad n = 1, 2, \cdots ; \quad \text{(C.1)}$$

where $d_0 = 1$, $d_{-1} = 0$. However, (C.1) is the recurrence relation for the Chebyshev polynomials of the first kind for $-2 \leq \alpha \leq 2$, based on the generating function $(1 - \alpha^2/4)^{\frac{1}{2}}$; see Abramowitz and Stegun [1, Ch. 22]. The zeros of the nth degree polynomial are

$$\alpha_m = 2 \cos\{\pi m/(n + 1)\}, \quad m = 1, \cdots, n .$$

Hence the eigenvalues for $\mathbf{W}$ are just

$$\lambda_m = \alpha_m/2, \quad m = 1, \cdots, n .$$

Case 3. The rectangular lattice with p rows and q columns ($n = pq$) has a model of the form

$$y_{rs} = \tfrac{1}{4}\rho\{y_{r-1,s} + y_{r+1,s} + y_{r,s-1} + y_{r,s+1}\} + \epsilon_{rs} ,$$

with appropriate adjustments for boundary cells. Suppose that the cells are numbered as in the diagram below:

$$\begin{array}{cccccc} 1 & 2 & 3 & 4 & \cdots & q \\ \mathrm{x} & \mathrm{x} & \mathrm{x} & \mathrm{x} & & \mathrm{x} \\ q+1 & q+2 & & & & 2q \\ \mathrm{x} & \mathrm{x} & & & & \mathrm{x} \\ \vdots & & & & & \\ pq-q+1 & pq-q+2 & & & & pq \\ \mathrm{x} & \mathrm{x} & & & \cdots & \mathrm{x} \end{array}$$

The weighting matrix can be written as

$$\mathbf{W} = \frac{1}{4}\begin{bmatrix} \mathbf{C} & \mathbf{I} & & \\ \mathbf{I} & \mathbf{C} & \mathbf{I} & \\ & \ddots & \ddots & \ddots \\ & & \mathbf{I} & \mathbf{C} \end{bmatrix}$$

where $\mathbf{C}$ denotes a $(q \times q)$ matrix of the kind specified previously, and $\mathbf{I}$ is the identity matrix of order q. There are p submatrices on the leading diagonal of $\mathbf{W}$.

Using the same approach as for Case 2, we find that the pq eigenvalues required are

$$\begin{aligned} \lambda_{rs} = \alpha_{rs}/4 &= \tfrac{1}{4}(\theta_r + \mu_s) \\ &= \tfrac{1}{2}[\cos\{r\pi/(p + 1)\} + \cos\{s\pi/(q + 1)\}] \end{aligned} \quad \text{(C.2)}$$

for $r = 1, \cdots, p;\ s = 1, \cdots, q$.

Some progress is also possible with special forms of hexagonal lattice.

APPENDIX D: INCONSISTENCY OF THE PAIRS ESTIMATOR

We adopt the notation of Section 3.2 and consider the estimator $\tilde{\rho}_1$ derived from (A.1) in Appendix A, but based on $\mathbf{W}_1$ rather than $\mathbf{W}$. The estimate corresponds to the solution of the equation ${}_1f_\rho(\tilde{\rho}_1) = 0$ in (A.2), where the subscript denotes the use of $\mathbf{W}_1$. The estimator will be consistent if the iterative procedure described in Appendix A converges to the true value of ρ as $n \to \infty$. That is, if

$$\operatorname*{plim}_{n\to\infty} [\{n_1 f_\rho(\rho)\}/\{n_1 f_{\rho\rho}(\rho)\}] = 0 . \quad \text{(D.1)}$$

To evaluate this expression, we first quote three standard results:

$R1$: For any matrix $\mathbf{G}$, the probability limit of $\mathbf{Y}'\mathbf{GY}$ is $\{\operatorname{tr} \mathbf{G}(\mathbf{A}'\mathbf{A})^{-1}\}$, since $\mathbf{Y} = \mathbf{A}\boldsymbol{\varepsilon}$.
$R2$: By definition $\operatorname{tr}(\mathbf{W}) = \sum_{i=1}^{n} \lambda_i = 0$, since $w_{ii} = 0$ for all i.
$R3$: $\operatorname{tr}(\mathbf{W}^j) = \sum_{i=1}^{n} (\lambda_i)^j$.

Using $R1$, the plim of the numerator of expression (D.1) is

$$2 \sum_{i=1}^{n} \lambda_{1i}(1 - \rho\lambda_{1i})^{-1} - 2n \operatorname{tr}\{(\mathbf{W}_1)'\mathbf{A}_1(\mathbf{A}'\mathbf{A})^{-1}\}/\operatorname{tr}\{(\mathbf{A}_1)'\mathbf{A}_1(\mathbf{A}'\mathbf{A})^{-1}\} . \quad \text{(D.2)}$$

This can be written as the series expansion for ρ,

$$-2\rho \operatorname{tr}\{((\mathbf{W}_1)' + \mathbf{W}_1)(\mathbf{W} - \mathbf{W}_1)\} + O(\rho^2) . \quad \text{(D.3)}$$

When $\mathbf{W}_1 = \mathbf{W}$, (D.2) reduces to $2 \operatorname{tr}(\mathbf{B}) - 2 \operatorname{tr}(\mathbf{B}) \equiv 0$. A similar procedure for the denominator of (D.1) yields the final form for (D.1), as a series in ρ,

$$-\rho \operatorname{tr}\{((\mathbf{W}_1)' + \mathbf{W}_1)(\mathbf{W} - \mathbf{W}_1)\} \operatorname{tr}\{((\mathbf{W}_1)' + \mathbf{W}_1)\mathbf{W}_1\} + O(\rho^2) . \quad \text{(D.4)}$$

From (D.4), it can be seen that the estimator will only be consistent if either

(i) ρ is $O(n^{-\delta})$, $\delta > 0$ or
(ii) the coefficients of terms in (D.4) are $O(n^{-\delta})$, $\delta > 0$.

Condition (ii) requires that $\mathbf{W}_1$ be sufficiently close to $\mathbf{W}$.

[*Received May 1973. Revised July 1974.*]

REFERENCES

[1] Abramowitz, M. and Stegun, I.A., *Handbook of Mathematical Functions*, Dover: New York, 1965.

[2] Bartlett, M.S., *An Introduction to Stochastic Processes*, 2nd ed., London: Cambridge University Press, 1966.

[3] ———, "Physical Nearest Neighbour Models and Non-linear Time Series," *Journal of Applied Probability*, 8, No. 2 (1971), 222–32.

[4] Besag, J.E., "Nearest Neighbour Systems and the Auto-Logistic Model for Binary Data," *Journal of the Royal Statistical Society*, Ser. B, 34, No. 1 (1972), 75–83.

[5] ———, "Spatial Interaction and the Statistical Analysis of Lattice Systems," *Journal of the Royal Statistical Society*, Ser. B, 36, No. 2 (1974), to appear.

[6] Cliff, A.D. and Ord, J.K., "The Problem of Spatial Autocorrelation," in A. J. Scott, ed., *Studies in Regional Science*, London: Pion, 1969, 25–55.

[7] ——— and Ord, J.K., "Testing for Spatial Autocorrelation Among Regression Residuals," *Geographical Analysis*, 4 (July 1972), 267–84.

[8] ——— and Ord, J.K., *Spatial Autocorrelation*, London: Pion, 1973.

[9] Cochrane, D. and Orcutt, G.H., "Applications of Least Square Regressions to Relationships Containing Autocorrelated Error Terms," *Journal of the American Statistical Association*, 44 (March 1949), 32–61.

[10] Durbin, J., "Estimation of Parameters in Time-Series Regression Models," *Journal of the Royal Statistical Society*, Ser. B, 22, No. 1 (1960), 139–53.

[11] Mead, R., "A Mathematical Model for the Estimation of Interplant Competition," *Biometrics*, 23 (June 1967), 189–205.

[12] ———, "Models for Interplant Competition in Irregularly Spaced Populations," in G.P. Patil, E.C. Pielou and W.E. Waters, eds., *Statistical Ecology*, Vol. 2, University Park: Pennsylvania State University Press, 1971, 13–30.

[13] O'Sullivan, P.M., *Transportation Networks and the Irish Economy*, L.S.E. Geographical Papers No. 4, London: Weidenfeld and Nicholson, 1969.

[14] Sargan, J.D., "Wages and Prices in the U.K.: A Study in Econometric Methodology," in P.E. Hart *et al.*, eds., *Econometric Analysis for National Economic Planning*, London: Butterworth, 1964, 25–63.

[15] Whittle, P., "On Stationary Processes in the Plane," *Biometrika*, 41 (Parts 3 and 4, 1954), 434–49.

[16] ———, "Stochastic Processes in Several Dimensions," *Bulletin of the International Statistical Institute*, 40 (Bk. I, 1963), 974–94.