Estimating Linear Models with Spatially Distributed Data
Author(s): Patrick Doreian
Source: *Sociological Methodology*, Vol. 12 (1981), pp. 359-388
Published by: Wiley
Stable URL: http://www.jstor.org/stable/270747

10

ESTIMATING LINEAR MODELS WITH SPATIALLY DISTRIBUTED DATA

Patrick Doreian
UNIVERSITY OF PITTSBURGH

Sociologists study a wide variety of social, political, and economic phenomena. Many of these phenomena—for example, urbanization, political mobilization, economic development, diffusion of innovations—take place in and are distributed across geographical space. It is reasonable, therefore, to argue that sociologists are interested, indeed have long been interested, in social phenomena distributed in geographical space. Yet, in the main, our theoretical frameworks and data-analytic capabilities do not include the geography of

I am grateful to Norman P. Hummon, who designed and implemented REPOMAT, a matrix algebra package I have used for estimation throughout this chapter, and to Philip Sidel, of the Social Science Computer Research Institute at the University of Pittsburgh, for programming the computation of the spatial autocorrelation statistics. Comments of the anonymous reviewers led to considerable improvements in the chapter and are appreciated.

social phenomena.[1] As a result, the geographical characteristics of social phenomena are overlooked, especially when data analyses are performed. Thus there is a large lacuna in our theoretical frameworks and methodological apparatus. This chapter represents an initial attempt at filling that lacuna. I should make clear that my objective in this chapter is not to claim that geographical space *must* be included; rather, it is to claim that *when* it is appropriate to include geographical space, a variety of conceptual and methodological issues need to be addressed. The following pages outline some of these issues and discuss estimation strategies that are appropriate for linear equations where the data are spatially distributed.

Some examples of the data structure considered in this study are appropriate.[2] Frisbie and Poston (1975), working in the human ecology tradition, analyzed the relationships between sustenance organization and population change. Using data for all nonmetropolitan counties of the 48 contiguous states of the United States, they present regression equations linking population changes to components of sustenance organization and also to other social characteristics that have been hypothesized to account for population change. Inverarity (1976) used a linear model in which the final dependent variables were lynchings and electoral support, the exogenous variables were racial composition, urbanization, and religious homogeneity, and there was an unmeasured endogenous variable of mechanical solidarity. The data were from the (then) 59 parishes of Louisiana. Ragin (1977), using British county data, regressed Conservative and Labour votes on a variety of measures of class composition and two regional dummies. Chirot and Ragin (1975) analyzed the Romanian peasant rebellion of 1907 by using data for counties and multiple-regression methods.

These examples were taken from the *American Sociological Review*. Moving elsewhere, we can find the classic papers of Matthews and Prothro (1963a, 1963b). In the first of these papers, they studied the relationship between social and economic factors and

[1]A stronger argument can be made: Not only is geographical space frequently excluded; so too is consideration of the physical environment (Dunlap and Catton, 1979). The two omissions are not unrelated. While entry into the debate over the "environmental sociology" paradigm would take us too far from our objectives, it can be remarked that the issues discussed here are germane to the empirical study of society-environment interactions.

[2]These examples are not included in a critical vein since the issue of the relevance of geographical space has not been decided.

black voter registration; in the second they extended their analysis to consider political factors. The central workhorse was again multiple regression, and the data were for 997 Southern counties. Salamon and Van Evera (1973) examined three competing explanations of political participation using data for Mississippi's 29 black-majority counties and regression methods. Kernall (1973), in an extended critique of Salamon and Van Evera's thesis, used the same methodology with data for all of Mississippi's counties. Schoenberger and Segal (1971) examined the (linear) relation between voting support for Wallace in 1968 and a variety of socioeconomic characteristics for 77 Southern congressional districts. Wasserman and Segal (1973) performed essentially the same analysis, only they used data for counties and split the South into the Deep South and the marginal South. Capecchi and Galli (1969) presented a linear causal model of voting determinants in Italy using data for 88 territorial units comprising all but four of the Italian provinces. Cox (1969), also employing a linear causal model, analyzed voting participation and the Conservative vote with data for the parliamentary constituencies of London.

In a quite different vein, Aigner and Heins (1967) used regression equations to account for variations in income equality in the 50 states and Washington D.C.; the exogenous variables were a variety of social, demographic, economic, and political variables. Mitchell (1969) adapted econometric procedures to analyze the Huk rebellion in the Philippines, linking insurgent control to a variety of cultural and economic factors; the data were for 57 municipalities in four provinces. Doreian and Hummon (1976) reanalyzed these data with the same objectives as Mitchell.

The list of examples, while long, is far from being exhaustive. These are simply examples of a particular type of data structure; in all cases, the data are defined for areal units and these units together comprise a region.[3] This data structure has prompted the use of multiple-regression analysis to estimate a linear relation. This coupling of a particular data structure and the use of linear structural equations is the focus of this chapter.

[3]There are partial exceptions to this statement. Salamon and Van Evera (1973) did not use all the counties of Mississippi; however, Kernall (1973) did. Matthews and Prothro (1963a, 1963b) did exclude some Southern counties, and Frisbie and Poston (1975) excluded the metropolitan counties. Nevertheless, the general point still stands. Further, the issues and procedures discussed here apply even if there are holes, so to speak, in a region.

The essence of the data structure arises through aggregation. (However, this chapter is *not* about the "aggregation problem.") Whenever data are aggregated to represent areal units, which are usually defined administratively or politically, it is likely that the geography of a social phenomenon has been retained, albeit in an implicit way. Except for the examples represented by Mitchell (1969) and Doreian and Hummon (1976), all the examples cited here ignore geographical space. This is not necessarily meant as a negative appraisal. Whether geographical space should or should not be explicitly included is an issue that hinges on theoretical and methodological considerations. It is these considerations, especially the latter, to which our attention is now directed.

REPRESENTING GEOGRAPHICAL SPACE

The representation of geographical space is not obvious as it involves choices based on substantive concerns and technical constraints. Sociologists can benefit from the efforts of geographers, regional economists, and mathematical ecologists (see Cliff and Ord, 1973; Pielou, 1969) who grapple with representing geographical space. There are two broad strategies: using measurements of distances between geographical locations and partitioning a region into areas.[4] The focus of this chapter is on the latter. In this strategy a region is partitioned into areas and data are recorded for each area. Geographical space is then represented by a $(N \times N)$ matrix where there are N areas in the region. Doreian and Hummon (1976, pp. 117–125) provide a general discussion of this representation of geographical space, and part of their conceptualization is used here.

The overriding reason for deciding on a specific representation is substantive. Generally, an explanation of some phenomenon is sought where the value of a variable of interest in a given area is (systematically) related to the values of that variable in some other areas. That is, observations are presumed to be interdependent rather than independent and this interdependency is presumed to be geographically based. The interdependency, or connectedness in geographical space, determines the representation of geographical

[4]It is possible, of course, to use distances between geographical areas. For example, the distances between area centroids or salient points (such as administrative capitals) can be used.

space. Consider the example of the Huk insurgency (Mitchell, 1969; Doreian and Hummon, 1976). Rebel control (or governmental control) of an area has immediate consequences for adjacent areas. If one side of the conflict can move weaponry and troops into an area, that area becomes a base for attempting to control adjacent areas. Adjacency is then the key geographical characteristic. More generally, accessibility of one area from another may be the key geographical characteristic with adjacency simply being a special case of accessibility. The properties of a transportation network could be used to determine accessibility with respect to each pair of areas that make up a region. Collective violence, as represented by the Romanian peasant rebellion of 1907 (Chirot and Ragin, 1975) or lynchings in the South (Inverarity, 1976), is likely to spread spatially, especially in an era before mass communication. Black enfranchisement in the United States is another process that can be seen to operate in terms of adjacency.[5] More generally, diffusion can be viewed as a spatial process operative over spatial connections between areas.

The first step of spatial representation is deciding which spatial property is to be represented. Even with that choice made, there are further options. Consider adjacency as a spatial property to be represented. Suppose a region R can be partitioned into N mutually exclusive areas. Louisiana is partitioned into 64 parishes, the contiguous United States into 48 states, and so on. Let $\mathbf{S} = [s_{ij}]$ be an $(N \times N)$ matrix, where s_{ij} is 1 if area i is adjacent to area j and 0 otherwise. Throughout, the s_{ii} are taken to be zero. The adjacency characteristics of R are completely specified in terms of $\mathbf{S}$. The entries of $\mathbf{S}$ are either 0 or 1. Such a binary matrix can be converted into a set of weights in the following fashion. Let s_i be the row sum for the ith row of $\mathbf{S}$. Then a matrix, $\mathbf{W} = [w_{ij}]$, can be constructed where $w_{ij} = s_{ij}/s_i$. The entries of $\mathbf{W}$ lie between 0 and 1 (inclusive, although $w_{ij} = 1$ is only possible for a pair of mutually adjacent but elsewise disconnected areas) and are proportions based on the number of other areas adjacent to a specific area. Adjacency could be operationalized slightly differently (Mitchell, 1969). Let b_{ij} be the length of common border between area i and area j. Then define $w_{ij} = b_{ij}/b_{ii}$, where b_{ii} is the total perimeter of area i.

[5]A much more general notion of adjacency could be used such that areas are adjacent if they are sufficiently similar with respect to relevant social, political, or economic characteristics. (See Cliff and Ord, 1970.)

It is clear that in any empirical situation the choice of a matrix **W** to represent geographical space is not obvious. There is, literally, an infinity of possible representations.[6] This infinite number has led some critics, for example, Arora and Brown (1977), to argue that the specification of **W** should be abandoned altogether. Such a judgment is premature. Some representations will be more compelling and soundly based than others. The choice of a representation can be made where the substantive concerns of the investigator will be paramount. A wide range of choice means that care should be exercised in the choice made, not that the investigator should refrain from making choices altogether.

For the following exposition, I assume that geographical space can be represented by a matrix of weights **W**. The elements of **W** are nonnegative and bounded by 0 and 1.

LINEAR EQUATIONS WITH SPATIALLY DISTRIBUTED DATA

Let **Y** be a vector of observations on an endogenous variable; let **X** be a matrix of observations on a set of exogenous variables (including a column of 1's for the intercept term); let $\boldsymbol{\beta}$ be a vector of parameters; and let ϵ be the disturbance term. Then, in abstract terms, the model estimated in most of the preceding examples is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \epsilon \tag{1}$$

where

$$E[\epsilon] = \mathbf{0} \qquad E[\epsilon\epsilon'] = \sigma^2\mathbf{I} \tag{2}$$

and ϵ is multivariate normal. Such is the conventional population regression function. As such, the spatial structure is ignored. The areal units are treated as units of analysis in the conventional sense, and nothing further is done concerning space. However, the spatial structure of the region can be incorporated in a variety of ways. Alternatives have been outlined by Ord (1975), and maximum-likelihood estimation (MLE) procedures have been proposed for each of these. The

[6]Further, a whole variety of distance decay functions could be specified for the elements of **W** (which would introduce the distance approach into this areal partitioning approach).

technical nature and terse presentation of Ord's procedures render them inaccessible to most social scientists, which leads to the need to make them more widely available. Thus this chapter is largely an exegesis of Ord's procedures, presenting derivations of the properties of those procedures and their sociological applications. The chapter is not solely an exegesis, however, for Ord's procedures can be improved and there are cases where much simpler procedures may suffice.

To express the idea of a spatial effect via adjacency,[7] a very simple notion is that Y, for a particular area, is related to, or a function of, the values of Y in adjacent areas. More precisely, Y is related to a weighted combination of values of Y in adjacent areas. If the weights are given in the matrix $\mathbf{W}$, then $\mathbf{Y}$ is a function of $\mathbf{WY}$.

The simplest linear model specifies

$$\mathbf{Y} = \rho\mathbf{WY} + \epsilon \tag{3}$$

where ϵ is as specified in (2). This is a *pure endogenous* model: Only the values of $\mathbf{Y}$ in adjacent areas determine the value of $\mathbf{Y}$ in a given area. Alternatively, we can describe this as a spatially autocovarying model. While such a simple formulation is likely to be of limited sociological utility, extended discussion of this model is warranted, as it underlies the models that do include exogenous variables in their specification. From (3),

$$\epsilon = (\mathbf{I} - \rho\mathbf{W})\mathbf{Y} = \mathbf{AY} \tag{4}$$

where $\mathbf{A} = (\mathbf{I} - \rho\mathbf{W})$. The joint likelihood of the ϵ_i is given by Mead (1967):

$$L(\epsilon) = (1/\sigma^2 2\pi)^{N/2} \exp\left(-\epsilon'\epsilon/2\sigma^2\right) \tag{5}$$

As Mead observes, however, it is the Y_i that are observed and not the ϵ_i. Thus it is the joint likelihood of the Y_i that needs to be maximized and not the function given in Equation (5). From (4) and (5) we have as the joint likelihood function, given $\mathbf{Y} = \mathbf{y}$:

$$L(\mathbf{y}) = |\mathbf{A}|\,(1/\sigma^2 2\pi)^{N/2} \exp\left[-(\mathbf{Ay})'(\mathbf{Ay})/2\sigma^2\right] \tag{6}$$

where $|\mathbf{A}|$ is the Jacobian of the transformation from the ϵ to the $\mathbf{y}$. It is

[7]Adjacency is but one spatial characteristic that can be represented as a matrix $\mathbf{W}$. To avoid cumbersome phrases like "interdependency due to a relevant spatial characteristic," adjacency is used simply as an exemplar.

easier to work with the log-likelihood function,[8] which is obtained from Equation (6) as

$$\begin{aligned} l(\mathbf{y}) &= -(N/2)\ln(2\pi\sigma^2) - [1/(2\sigma^2)]\mathbf{y}'\mathbf{A}'\mathbf{A}\mathbf{y} + \ln|\mathbf{A}| \\ &= \text{const} - (N/2)\ln\sigma^2 - (1/2\sigma^2)\mathbf{y}'\mathbf{A}'\mathbf{A}\mathbf{y} + \ln|\mathbf{A}| \end{aligned} \tag{7}$$

The parameters of the model requiring estimation are ρ and σ^2, so $l(y)$ has to be maximized with respect to these.[9] To simplify notation slightly, $\omega = \sigma^2$ and (7) is rewritten as

$$l(\mathbf{y}) = \text{const} - (N/2)\ln\omega - (1/2\omega)\,\mathbf{y}'\mathbf{A}'\mathbf{A}\mathbf{y} + \ln|\mathbf{A}| \tag{8}$$

We first consider ω and then ρ. Minimizing $l(y)$ with respect to ω is straightforward (Mead, 1967). Differentiating $l(y)$ with respect to ω gives

$$\partial l(\mathbf{y})/\partial\omega = -\,(N/2\omega) + (1/2\omega^2)\,(\mathbf{y}'\mathbf{A}'\mathbf{A}\mathbf{y}) \tag{9}$$

Setting this derivative to zero and solving gives

$$\hat{\omega} = \hat{\sigma}^2 = \mathbf{y}'\mathbf{A}'\mathbf{A}\mathbf{y}/N \tag{10}$$

as the estimator of σ^2. As yet, $\hat{\omega}$ is not known, since it depends upon ρ because $\mathbf{A} = \mathbf{I} - \rho\mathbf{W}$. From (10) and (8), however, $\hat{\rho}$ is the value of ρ that maximizes

$$l(\mathbf{y}) = l(\mathbf{y}, \rho; \hat{\omega}) = \text{const} - (N/2)\ln\hat{\omega} + \ln|\mathbf{A}| \tag{11}$$

Computationally this is burdensome, since determining $\hat{\rho}$ rests on the evaluation of $|\mathbf{A}| = |\mathbf{I} - \rho\mathbf{W}|$. Ord, using a simple result in a clever way, reduced the burden of computation and obtained an easier way of estimating ρ. Let $\mathbf{W}$ have $\lambda_1, \ldots, \lambda_N$ as its eigenvalues. Then, by definition of the characteristic equation,

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

Further, the determinant of a matrix is equal to the product of its eigenvalues. If $f(\mathbf{W})$ is an algebraic polynomial in $\mathbf{W}$ (Lancaster,

[8]If θ is a parameter being estimated, then, under general conditions, $\partial/\partial\theta$, $(\ln L) = (\partial L/\partial\theta)/L$. Thus L and $\ln L$ will have maxima together, and L is always greater than zero (Kendall and Stuart, 1967, pp. 35–36).

[9]Given $\mathbf{Y} = \mathbf{y}$, $l(\mathbf{y})$ is a function of the parameters to be estimated. I could have written $l(\rho, \sigma^2)$ instead of $l(\mathbf{y})$ to emphasize this.

1971, p. 289), then the eigenvalues of $f(\mathbf{W})$ are $f(\lambda_i)$. Thus the eigenvalues of $\mathbf{I} - \rho\mathbf{W}$ are $\{1 - \rho\lambda_i\}$ and

$$|\mathbf{A}| = \prod_{i=1}^{N} (1 - \rho\lambda_i) \tag{12}$$

The λ_i can be determined once and for all and, from (11), $\hat{\rho}$ is the value of ρ that minimizes

$$-(2/N) \sum_{i=1}^{N} \ln (1 - \rho\lambda_i) + \ln [\mathbf{y}'\mathbf{y} - 2\rho\mathbf{y}'\mathbf{W}\mathbf{y} + \rho^2(\mathbf{W}\mathbf{y})'\mathbf{W}\mathbf{y}] \tag{13}$$

where $\hat{\omega}$ has been substituted from (10). The value of $\hat{\rho}$ that minimizes (13) can be found by a direct-search procedure, and such a procedure is used throughout this chapter.[10] With $\hat{\rho}$ found in this fashion, $\hat{\omega}$ can be obtained from (10). With these estimates established, it is necessary to be able to estimate the variance-covariance matrix of the estimates. In general, the asymptotic variance-covariance matrix is given by $\mathbf{V}$, where

$$\mathbf{V}^{-1} = -E\,[\partial^2 l/\partial\theta_r\partial\theta_s] \tag{14}$$

and where θ_r and θ_s are parameters being estimated (Kendall and Stuart, 1967, p. 55). In the simple model being considered here, there are only two parameters: θ_r is ω and θ_s is ρ. The derivation of this information matrix is given in Appendix A. The asymptotic variance-covariance matrix is

$$\mathbf{V}(\omega, \rho) = \omega^2 \begin{bmatrix} N/2 & \omega \operatorname{tr}(\mathbf{B}) \\ \omega \operatorname{tr}(\mathbf{B}) & \omega^2[\operatorname{tr}(\mathbf{B}'\mathbf{B}) - \alpha] \end{bmatrix}^{-1} \tag{15}$$

As an example of the pure endogenous model, consider the data assembled by Mitchell (1969) on the Huk rebellion in the Philippines. The vector $\mathbf{y}$ is the percentage of barrios in a municipality under Huk control. Using a direct-search procedure, $\hat{\rho} = 0.83$. Use of this

[10]Ord suggests that for more complex models a formal iterative procedure may be preferable. The direct-search procedure is rather cumbersome. It amounts to evaluating the value of the function given in (13) for each value of ρ in a given range and selecting the value of ρ for which the function takes a minimum value. Large increments in ρ can be used to get the minimum value roughly; then smaller increments in ρ can be used in that restricted region.

estimate,[11] together with Equation (10), gives $\hat{\sigma}^2 = 211.4$. Thus the parameters of the (simple) model have been estimated. Using Equation (15), and computing the matrices there, gives

$$\mathbf{V}(\hat{\omega}, \hat{\rho}) = \begin{bmatrix} 1710 & -0.836 \\ -0.836 & 0.005 \end{bmatrix}$$

Straightforwardly, the standard error of $\hat{\rho}$ is 0.07 and the standard error of $\hat{\sigma}$ is 41.36. Thus $\hat{\rho}$ is clearly significantly different from zero and a spatial process could be said to operate if the specification of Equation (3) is accepted. The measure of fit, FIT, is 0.69.[12]

In most social science contexts, however, the pure endogenous model is of limited utility. Exogenous variables must be included in relational specifications, and this leads to

$$\mathbf{y} = \rho \mathbf{W} \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \epsilon \tag{16}$$

where ϵ is specified as before. Equation (16) is referred to as the *mixed endogenous-exogenous* model or, in Ord's terms, the regressive-autoregressive model. The joint likelihood function for the ϵ_i is given, as before, by Equation (5). For this specification, however, $\epsilon = \mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}$ and the joint likelihood function for the y_i is given by

$$L(\mathbf{y}) = |A|(1/2\pi\sigma^2)^{N/2} \exp\{-(1/2\sigma^2)[\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}]'[\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}]\} \tag{17}$$

Equation (17) points out the error in the establishment of the estimation procedure of Doreian and Hummon (1976, p. 138), as they omitted $|\mathbf{A}|$, the Jacobian of the transformation from ϵ to $\mathbf{y}$. The

[11]Actually, the estimation and inference statistics were computed by using the results of the following section. Rather than estimate the model of (3) directly, the estimation was done for $\mathbf{Y} = \alpha + \rho \mathbf{W} \mathbf{Y} + \epsilon$, where α is an intercept term. For simplicity, the pure endogenous model is discussed here without the intercept term. If α is significantly different from zero, then $\hat{\rho}$ will be biased; if $\hat{\alpha} > 0$, then $\hat{\rho}$ will be biased upwards; and if $\hat{\alpha} < 0$, then $\hat{\rho}$ will be biased downward. In the estimation, $\hat{\alpha} = 2.67$ with a standard error of 2.22; in this case, therefore, omission of α would not have been problematic. Indeed, estimation with α omitted leads to only small modifications of the estimates.

[12]The measure of fit (FIT) is the square of the correlation between y and the fitted value, $\hat{y}$. For OLS this is equivalent to the coefficient of determination (see Johnston, 1963, p. 58). However, R^2 does not have meaning here due to the interdependence of the observations. For this reason, the notation of R^2 has been avoided and the measure of fit used here cannot be interpreted as the proportion of variance explained. One hopes that a more adequate measure of the goodness of fit of these spatial models will be developed.

log-likelihood function is given by

$$l(y) = \text{const} - (N/2)\ln\omega - (1/2\omega)$$
$$(\mathbf{y}'\mathbf{A}'\mathbf{A}\mathbf{y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{A}\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}) + \ln|\mathbf{A}| \quad (18)$$

which has to be minimized with respect to ω, ρ, and $\boldsymbol{\beta}$. As before, we start by minimizing $l(y)$ with respect to ω:

$$\partial l/\partial\omega = -(N/2\omega)$$
$$+ (1/2\omega^2)(\mathbf{y}'\mathbf{A}'\mathbf{A}\mathbf{y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{A}\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}) \quad (19)$$

Setting (19) to zero gives

$$\hat{\omega} = \hat{\sigma}^2 = (1/N)(\mathbf{y}'\mathbf{A}'\mathbf{A}\mathbf{y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{A}\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}) \quad (20)$$

Of course, $\hat{\omega}$ is numerically unknown since $\hat{\rho}$ and $\hat{\beta}$ have not been determined. Define $\mathbf{z} = (\mathbf{I} - \rho\mathbf{W})\mathbf{y} = \mathbf{A}\mathbf{y}$. Then $\epsilon = \mathbf{z} - \mathbf{X}\boldsymbol{\beta}$ and $\epsilon'\epsilon = \mathbf{z}'\mathbf{z} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{z} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$, which begins to look like the situation of ordinary least squares (OLS) of $\mathbf{z}$ on $\mathbf{X}$. With this change, we can write

$$l(\mathbf{y}) = \text{const} - (N/2)\ln\omega$$
$$- (1/2\omega)(\mathbf{z}'\mathbf{z} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{z} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}) + \ln|\mathbf{A}| \quad (21)$$

Differentiating (21) with respect to $\boldsymbol{\beta}$ gives

$$\partial l/\partial\beta = -(1/2\omega)[-2\mathbf{X}'\mathbf{z} + 2(\mathbf{X}'\mathbf{X})\boldsymbol{\beta}] \quad (22)$$

Setting this to zero gives

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{z} \quad (23)$$

as the estimator of $\boldsymbol{\beta}$, which, if ρ were known, would be given by OLS of $\mathbf{z}$ on $\mathbf{X}$. Substituting (23) into (20) gives

$$\begin{aligned}\hat{\omega} &= (1/N)[\mathbf{z}'\mathbf{z} - 2\mathbf{z}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{z} + \mathbf{z}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{z}] \\ &= (1/N)\mathbf{z}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{z} \quad (24) \\ &= (1/N)\mathbf{z}'\mathbf{M}\mathbf{z}\end{aligned}$$

where $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is a symmetric, idempotent matrix. While (24) and (23) give estimation equations for ω and $\boldsymbol{\beta}$, ρ is, as yet, unknown. As before, it has to be found by a direct-search procedure. Equation (13) is replaced by another expression to be minimized that takes into account the exogenous variables. As before, $\hat{\rho}$ maximizes

$$l(\mathbf{y}) = l(\mathbf{y}:\quad \rho, \hat{\omega}, \hat{\beta}) = \text{const} - (N/2)\ln\hat{\omega} + \ln|\mathbf{A}|$$

Using the simplified expression for $\ln|\mathbf{A}|$, ρ minimizes

$$-(2/N)\sum_{i=1}^{N}\ln(1-\rho\lambda_i) + \ln\hat{\omega}$$

But

$$\begin{aligned}\hat{\omega} &= (1/N)\mathbf{z}'\mathbf{M}\mathbf{z} \\ &= (1/N)\mathbf{y}'\mathbf{A}'\mathbf{M}\mathbf{A}\mathbf{y} \\ &= (1/N)\mathbf{y}'(\mathbf{I}-\rho\mathbf{W})'\mathbf{M}(\mathbf{I}-\rho\mathbf{W})\mathbf{y} \\ &= (1/N)[\mathbf{y}'\mathbf{M}\mathbf{y} - 2\rho\mathbf{y}'\mathbf{M}\mathbf{W}\mathbf{y} + \rho^2(\mathbf{W}\mathbf{y})'\mathbf{M}\mathbf{W}\mathbf{y}]\end{aligned}$$

Thus $\hat{\rho}$ minimizes

$$-(2/N)\sum_{i=1}^{N}\ln(1-\rho\lambda_i) + \ln[\mathbf{y}'\mathbf{M}\mathbf{y} - 2\rho\mathbf{y}'\mathbf{M}\mathbf{W}\mathbf{y} + \rho^2(\mathbf{W}\mathbf{y})'\mathbf{M}\mathbf{W}\mathbf{y}] \tag{25}$$

and this is again done by a direct-search procedure. From Appendix A,

$$\mathbf{V}(\hat{\omega}, \hat{\rho}, \hat{\boldsymbol{\beta}}) = \omega^2\begin{bmatrix} N/2 & \omega\,\mathrm{tr}\,(\mathbf{B}) & \mathbf{0}' \\ \omega\,\mathrm{tr}\,(\mathbf{B}) & \omega^2\,\mathrm{tr}\,(\mathbf{B}'\mathbf{B}) + \omega\boldsymbol{\beta}'\mathbf{X}'\mathbf{B}'\mathbf{B}\mathbf{X}\boldsymbol{\beta} - \alpha\omega^2 & \omega\mathbf{X}'\mathbf{B}'\mathbf{X}\boldsymbol{\beta} \\ \mathbf{0} & \omega\mathbf{X}'\mathbf{B}\mathbf{X}\boldsymbol{\beta} & \omega\mathbf{X}'\mathbf{X}\end{bmatrix}^{-1} \tag{26}$$

is the asymptotic variance-covariance matrix for the estimators of the parameters of the mixed endogenous-exogenous model specified in (16). Before providing examples of these estimation procedures, I take a brief digression into the topic of spatial autocorrelation.

In the specification of the classic linear regression model, we have $E[\epsilon\epsilon'] = \sigma^2\mathbf{I}$, which indicates that the disturbance term is homoskedastic and not autocorrelated. If some off-diagonal elements of the variance-covariance matrix are nonzero, we have autocorrelation. The realm in which this has been most extensively discussed is time-series analysis. In that context a disturbance term is autocorrelated if $E[\epsilon_t\epsilon_{t-t_1}] \neq 0$ for $t > t_1$. A variety of tests have been used to detect the presence of (temporal) autocorrelation; of these, the

Durban–Watson statistic is the most widely used. Once the autocorrelation is detected, its form can be diagnosed and estimation strategies can be devised that take into account the autocorrelation that has been empirically diagnosed. (See, for example, Hibbs, 1974.) Of course, any variable of direct interest (endogenous or exogenous) in a model may be autocorrelated. Temporal interdependencies amount to temporal autocorrelation and, by the same token, spatial interdependencies amount to spatial autocorrelation. However, spatial autocorrelation is far more than a simple spatial analog of temporal autocorrelation. Cliff and Ord, in a series of publications culminating in their book (1973), have dealt extensively with this problem by reviewing earlier efforts, providing an exhaustive account of measures of spatial autocorrelation, and applying these measures to a variety of empirical situations.

Given a spatially distributed variable $\mathbf{y}$, an initial question is whether or not $\mathbf{y}$ is spatially autocorrelated. This technical issue is dealt with by defining an appropriate test statistic. Moran (1950) proposed such a statistic that was modified by Dacey (1965). The test statistic proposed by Dacey was generalized by Cliff and Ord (1973, p. 12) as

$$I = (N/T)(\mathbf{y}'\mathbf{W}\mathbf{y}/\mathbf{y}'\mathbf{y}) \tag{27}$$

for a spatially distributed variable where N is the number of areas, T is the sum of the weights of $\mathbf{W}$, and $\mathbf{W}$ is an appropriate matrix of spatial weights.[13] Cliff and Ord (1973, pp. 13–15, 29–33) establish the distribution theory for I to test for spatial autocorrelation by treating $(I - E[I])/(V[I])^{1/2}$ as a standardized normal deviate with E and V the expected value and variance operators respectively. They extend this approach (1973, pp. 87–97) to deal with residuals from a regression analysis. While the details are omitted here, some computational formulas are included in Appendix B. On the basis of their work it is possible to test for spatial autocorrelation either in a variable of interest or in a residual resulting from a regression analysis.

In the empirical examples that follow, the analysis of spatial autocorrelation is undertaken and reported. It is only of secondary interest to my objectives, however. In one respect, spatial autocorrelation can be viewed as an annoying technical problem. By the nature of

[13]Cliff and Ord discuss other measures that are not considered here.

the models discussed earlier, it is clear that the endogenous variable **y** is hypothesized as spatially autocorrelated due to a well-defined spatial process specified in (3) or in (16). As the estimation procedures detailed earlier and in Appendix A are complex and computationally burdensome, it is reasonable to test for spatial autocorrelation prior to such an analysis. If **y** is not spatially autocorrelated, it is not fruitful to proceed further. If spatial autocorrelation does exist in **y**, it is reasonable to perform the regression analysis implied by (1) and assess the residual for spatial autocorrelation in **y** to see if it is removed by the regression analysis. If it is removed, the estimation of (16) is not warranted; but if it has not been removed, it is appropriate to proceed to the estimation of (16).[14]

EMPIRICAL EXAMPLES

To illustrate the mixed endogenous-exogenous model, we consider first the Huk example of Mitchell. The exogenous variables are:

P	Proportion of the population speaking the Pampangan dialect
FMP	Farmers as a percentage of the population
OWN	Owners as a percentage of farmers
SGR	Percentage of cultivated land given over to sugarcane
MNT	Presence of mountainous terrain (dummy)
SWP	Presence of swamps (dummy)

For further details on these variables, and on the rationale for their inclusion, see Mitchell (1969) or Doreian and Hummon (1976). The actual specification of the linear model uses *P* multiplicatively with the other variables so that the exogenous variables are P*FMP, P*OWN, P*SGR, P*MNT, and P*SWP.

Table 1 gives the results of three estimations for a linear model linking insurgent (Huk) control to the cultural, demographic, economic, and physical exogenous variables. Panel 1 gives the result of OLS applied to the specification of the standard population regression

[14]If the analysis were approached from the viewpoint of spatial autocorrelation and OLS did not remove the spatial autocorrelation, then it would be necessary to ask how the spatial autocorrelation could be removed. Given the formulation of the models in the previous section, it is obvious how the (nonspatial) OLS model should be reformulated.

function (Equation 1) where geographical space is ignored. Panel 2 gives the outcome of the MLE procedure given in the preceding pages. As far as the primary objective of this chapter is concerned, the major comparison is made between these two panels. Panel 3 gives the OLS procedure suggested (incorrectly) by Doreian and Hummon, where **Wy** is simply included as another exogenous variable.[15] However, the practical question behind the comparison of panels 2 and 3 is whether the much simpler OLS procedure (with a spatial term) will suffice as a surrogate for the more complicated and computationally burdensome MLE procedure. In each case, the figures in parentheses are estimated standard errors.

TABLE 1

Alternative Estimations for Multiplicative Model of Huk Insurgent Control

Nonspatial Model

1: OLS
$$\begin{cases} Y = \underset{(2.94)}{1.15} + \underset{(0.94)}{3.79\mathrm{P^*FMP}} - \underset{(0.44)}{1.91\mathrm{P^*OWN}} + \underset{(0.16)}{0.46\mathrm{P^*SGR}} + \underset{(7.02)}{38.38\mathrm{P^*MNT}} + \underset{(7.94)}{17.17\mathrm{P^*SWP}} \\ R^2 = 0.73 \quad (0.94) \end{cases}$$

Spatial Model

2: MLE
$$\begin{cases} Y = \underset{(2.48)}{-0.88} + \underset{(0.11)}{0.47WY} + \underset{(0.81)}{2.27\mathrm{P^*FMP}} - \underset{(0.38)}{1.07\mathrm{P^*OWN}} + \underset{(0.14)}{0.178\mathrm{P^*SGR}} + \underset{(5.89)}{30.46\mathrm{P^*MNT}} + \underset{(6.52)}{12.43\mathrm{P^*SWP}} \\ (\mathrm{FIT} = 0.80)^{a} \qquad \hat{\sigma}^2 = 126.4\,(23.9) \end{cases}$$

3: OLS
$$\begin{cases} Y = \underset{(2.62)}{-1.38} + \underset{(0.14)}{0.59WY} + \underset{(0.93)}{1.89\mathrm{P^*FMP}} - \underset{(0.45)}{0.86\mathrm{P^*OWN}} + \underset{(0.16)}{0.11\mathrm{P^*SGR}} + \underset{(6.51)}{28.49\mathrm{P^*MNT}} + \underset{(7.01)}{11.26\mathrm{P^*SWP}} \\ R^2 = 0.80 \end{cases}$$

[a]The measure FIT is not strictly comparable to R^2 (see footnote 12).

For the MLE procedure, $\hat{\rho} = 0.47$ with a standard error of 0.11. This indicates that a spatial process is operative and that the mixed endogenous-exogenous specification is appropriate. There are dramatic numerical and, more important, inferential differences between the spatial model and the nonspatial model.[16] Apart from the

[15]The estimates differ from those of Doreian and Hummon (1976) and Mitchell (1969) since a different matrix W is used in this case. Here **W** is normalized to have row sums of unity rather than using a binary matrix.

[16]The term *spatial model* refers to the specification of the mixed endogenous-exogenous model.

intercept (which is not significantly different from zero in either approach), all the numerical estimates of the regression coefficients for the nonspatial model are inflated. Moreover, the estimates, for all coefficients, of the standard errors are also inflated. These are serious deficiencies in the nonspatial model and the application of OLS when geographical space is ignored. The nonspatial model would lead to the inclusion of all the exogenous variables (defined interactively with P) whereas the spatial model, together with the use of the appropriate MLE procedure, would lead to the inclusion only of P*FMP, P*OWN, P*SGR, and P*MNT. There are clear, substantive differences in the outcomes of the two specifications and their corresponding estimation procedures.

When we turn to the two estimation alternatives for the spatial model, there appear to be few differences. The numerical values of the estimated coefficients are close. The only systematic difference is that the estimates of the standard errors of the coefficient estimates are smaller under the MLE procedure than under the OLS approach. In this case, there is one inferential difference under the two approaches: P*OWN would not be included if the OLS estimates were used whereas it would be included with the maximum-likelihood estimates.

Further examples are merited, and all are taken from the context of Louisiana politics.[17] The dependent variable is support for the Democratic presidential candidate at various elections, and the exogenous variables are limited to:

B Percentage black (in a parish)
C Percentage Catholic (in a parish)
U Percentage urban (in a parish)
BPE Measure of black political equality

The data are for all 64 Louisiana parishes (counties), and BPE operationalizes the extent to which blacks are enfranchised in a parish in relation to their numbers there. Again our attention is confined to estimation problems rather than the setup of models or their detailed interpretation.

In the tables that follow, there are again two comparisons being made: between a spatial model and a nonspatial model and between

[17]The author, in collaboration with Charles Grenier, is analyzing the dynamics of Louisiana politics from 1932 to 1976. The examples used here are for illustrative purposes only, since our concern is with estimation procedures.

TABLE 2
Alternative Linear Equations Predicting Support for Democratic Presidential Candidate (Kennedy) 1960: Louisiana

Nonspatial Model

$$\text{1: OLS}\begin{cases} Y = \underset{(4.40)}{21.03} + \underset{(0.08)}{0.01\text{B}} + \underset{(0.04)}{0.30\text{C}} - \underset{(0.04)}{0.11\text{U}} + \underset{(0.06)}{0.39\text{BPE}} \\ R^2 = 0.88 \end{cases}$$

Spatial Model

$$\text{2: MLE}\begin{cases} Y = \underset{(4.67)}{13.78} + \underset{(0.09)}{0.31WY} - \underset{(0.07)}{0.004\text{B}} + \underset{(0.05)}{0.22\text{C}} - \underset{(0.04)}{0.10\text{U}} + \underset{(0.06)}{0.29\text{BPE}} \\ (\text{FIT} = 0.90)^{a} \qquad \hat{\sigma}^2 = \underset{(8.83)}{49.78} \end{cases}$$

$$\text{3: OLS}\begin{cases} Y = \underset{(4.99)}{12.34} + \underset{(0.12)}{0.37WY} - \underset{(0.08)}{0.007\text{B}} + \underset{(0.05)}{0.21\text{C}} - \underset{(0.04)}{0.10\text{U}} + \underset{(0.07)}{0.28\text{BPE}} \\ R^2 = 0.90 \end{cases}$$

[a]The measure FIT is not strictly comparable to R^2 (see footnote 12).

two alternative ways of estimating the spatial model. Table 2 presents the equations for the 1960 presidential election. As before, the coefficient estimates for the nonspatial model are inflated relative to the maximum-likelihood estimates for the spatial model. The estimates for the standard errors of the coefficient estimates are close to each other. However, no inferential differences occur save the obvious inclusion of a spatial term for the spatial model. Percentage Catholic and BPE are positively related to percentage support for Kennedy whereas percentage urban is negatively related to support for Kennedy. Percentage black is not relevant as a predictor of that support. (The coefficient sign difference for this variable is irrelevant as it is not significant.)

Comparing the two estimation approaches for the spatial model, we see that the coefficient estimates for the exogenous variables are very close. As before, the maximum-likelihood estimates of the standard errors of the coefficient estimates are smaller than the corresponding OLS estimates (where there are differences). However, these make no difference so far as inferential decisions and substantive interpretations are concerned. In this instance, the OLS procedure is a perfectly good surrogate for the more involved MLE procedure.

The next example is for the 1972 election; Table 3 gives the estimated linear equations. As before, the OLS estimates of the coefficients and the standard errors of the estimates for the nonspatial model are inflated (with one exception) relative to the maximum-likelihood estimates for the spatial model (although the differences are very small). The one exception is the intercept, which (in magnitude) is deflated in the nonspatial model. This exception leads to the one inferential difference. For the nonspatial model, the intercept is not significant whereas it is in the spatial model. And, in the equation specified, the intercept term has a perfectly reasonable interpretation. Those (hypothetical) parishes with no blacks, no Catholics, no urban dwellers, and no black political equality were predisposed against the Democratic candidate in 1972. When the two estimation procedures are compared for the spatial model, the same pattern as for 1960 can be seen. The coefficient estimates are virtually identical; the maximum-likelihood estimates of the standard errors of the estimates are smaller; and no inferential differences exist. Again, for the spatial model, the OLS estimators are satisfactory surrogates for the maximum-likelihood estimators.

TABLE 3
Alternative Linear Equations Predicting Support for Democratic Presidential Candidate (McGovern) 1972: Louisiana

Nonspatial Model

$$\text{1: OLS}\begin{cases} Y = -7.36 + 0.41\text{B} + 0.09\text{C} + 0.001\text{U} + 0.29\text{BPE} \\ \quad\;\; (4.32) \quad (0.05) \quad (0.03) \quad (0.022) \quad (0.07) \\ R^2 = 0.75 \end{cases}$$

Spatial Model

$$\text{2: MLE}\begin{cases} Y = -11.44 + 0.29WY + 0.39\text{B} + 0.07\text{C} + 0.01\text{U} + 0.24\text{BPE} \\ \quad\;\; (4.18) \quad (0.10) \quad (0.04) \quad (0.02) \quad (0.02) \quad (0.06) \\ (\text{FIT} = 0.78)^{a} \quad \hat{\sigma}^2 = 163.2 \quad (2.9) \end{cases}$$

$$\text{3: OLS}\begin{cases} Y = -12.51 + 0.37WY + 0.39\text{B} + 0.07\text{C} + 0.01\text{U} + 0.22\text{BPE} \\ \quad\;\; (4.50) \quad (0.13) \quad (0.04) \quad (0.02) \quad (0.02) \quad (0.07) \\ R^2 = 0.78 \end{cases}$$

[a]The measure FIT is not strictly comparable to R^2 (see footnote 12).

The final example concerns the 1968 presidential election; the estimated equations are shown in Table 4. The comparisons between the nonspatial model, with its OLS procedure, and the spatial model, with its MLE procedure, are similar to the previous case. However, both the coefficient estimates and the standard error estimates are much closer. The coefficient for the spatial term in the spatial model is rather small, and one would expect that as $\hat{\rho}$ tends to zero, and no spatial process operates, the two procedures will tend to give the same results. This is, of course, obvious, since the crucial difference between the two models is the specification of the spatial term. In the 1968 election, there is again an inferential difference concerning the intercept term: It is included in the spatial model but excluded from the nonspatial model (as before). Both models judge percentage urban not to be a significant predictor of Democratic presidential support in 1968.

TABLE 4
Alternative Linear Equations Predicting Support for Democratic Presidential Candidate (Humphrey) 1968: Louisiana

Nonspatial Model

1: OLS	$Y =$	-4.54		+ 0.59B	+ 0.12C	+ 0.04U	+ 0.11BPE
		(3.0)		(0.05)	(0.03)	(0.03)	(0.04)
	$R^2 =$	0.76					

Spatial Model

2: MLE	$Y =$	-6.12	+ 0.12WY	+ 0.58B	+ 0.11C	+ 0.05U	+ 0.10BPE
		(3.67)	(0.12)	(0.05)	(0.03)	(0.03)	(0.04)
	(FIT = 0.76)[a]		$\hat{\sigma} = 22.9$		(4.1)		
3: OLS	$Y =$	-6.83	+ 0.12WY	+ 0.58B	+ 0.11C	+ 0.05U	+ 0.10BPE
		(4.26)	(0.15)	(0.05)	(0.03)	(0.03)	(0.04)
	$R^2 =$	0.76					

[a]The measure FIT is not strictly comparable to R^2 (see footnote 12).

In this final example, the OLS and MLE procedures return the same estimates of ρ. For the previous examples, $\hat{\rho}$ via OLS was higher than the corresponding maximum-likelihood estimate. The pattern is maintained: The OLS estimate of the standard error of $\hat{\rho}$ is greater than the corresponding maximum-likelihood estimate. The OLS bias of tending to overestimate the magnitude of ρ is offset by the OLS bias of tending to overestimate the magnitude of the standard error of the estimate of ρ. The nature of the two biases makes it difficult to state a general conclusion on the merits of OLS for the spatial model versus the MLE procedure. If the two procedures were to lead to different inferential processes concerning ρ, then obviously the OLS procedure would not be a good surrogate for the MLE procedure.[18]

DISCUSSION

Given that social scientists do consider social phenomena that are distributed in geographical space, this chapter has addressed three

[18]In part, this issue can be assessed in terms of spatial autocorrelation. For each of the foregoing examples, the spatial autocorrelation statistic I was computed. In all cases, the dependent variable y was spatially autocorrelated.

issues. First, should geographical space be included in modeling efforts when linear equations are used? Second, if geographical space should be included, how can we include it parsimoniously? And, third, given the inclusion of geographical space in linear models, what estimation procedures should be used?

The first issue arose in the context of the use of aggregated data for areal units. As such, geography is implicitly included in the data structure and whether or not geographical space should be included depends on whether or not some spatial process is operative. Such a decision is a theoretical one, and several examples were given in which a spatial process was at least plausible.

With regard to the second issue, we have explored a straightforward way of including geographical space by means of a matrix **W** of weights representing a spatial property. The exemplar spatial property was that of adjacency. Representing space in this fashion is, however, only one alternative. Coelen (1976) has argued that use of a single ρ means that the autocorrelation is either positive or negative across all observations and has suggested that this use is too restrictive. There may be different spatial effects in different local subregions. Specification of a differential ρ complicates the estimation procedure considerably, as the simplification provided by Equation (12) is not available; this research problem merits further attention. Another line of inquiry is to specify multiple spatial effects with multiple ρ's and **W**'s. However, the use of multiple ρs also means that the simplification found in (12) would no longer be available.

As there are difficulties in specifying **W**, Arora and Brown (1977) have proposed abandoning such an approach and have suggested, instead, using econometric methods. They outline the procedures of joint generalized least squares, equicorrelated error terms, random error component models, and random coefficient regression models. As they make no attempt to apply these approaches empirically, their suggestions remain speculative. The reader is referred to their article for details, but some preliminary remarks are in order here. The method of joint generalized least squares (see Theil, 1971) requires panels of observations. Further, distinct β's are specified for each area, which greatly enlarges the number of parameters to be estimated. For the situations discussed in this chapter, the method is not applicable. If there are a sufficiently large number of panels of observations, and if the β_i are fixed through time, the method has some appeal. Both Swamy's (1970) and Hsiao's (1975) approaches to the random coefficients model also require panels of observations. The equicorrelated error term models (without the spatial term) for a single cross-section amount to OLS if an intercept is specified (Theil, 1971, p. 243). If there are good grounds for anticipating a spatial process, however, the methods described here are preferable to OLS. Arora and Brown propose the random error components model for interaction (between areas) variables, but since the dependent variables discussed here are not such interaction variables, the approach does not appear relevant in this context. On the other hand, for interaction variables this approach can be explored further.

With regard to the third issue, given the specification of a (linear) spatial process, the mixed endogenous-exogenous model, the chapter has detailed a maximum-likelihood procedure for estimating such a model. We also discussed a test for spatial autocorrelation that can be used to guide the researcher in assessing the need to incorporate spatial characteristics into the formulation of a model.

In the examples we have used, it is clear that important differences do exist between the spatial model and the nonspatial model. This was most dramatically the case in the Huk insurgency example. While there was an instance in the set of Louisiana examples where the two models led to substantively the same conclusion, the nonspatial model is unreliable as a means of coupling endogenous variables to exogenous variables when a spatial process operates. In the nonspatial model, both coefficient estimates and estimates of the standard errors of those estimates are inflated. The examples make clear that when a spatial social process is operative, a spatial model should be specified and the MLE procedures detailed in the preceding pages should be used.

The MLE procedure is not straightforward, however, and it is computationally burdensome. If a simpler procedure will work, it seems preferable to use it. To this end, a second line of inquiry was to see if OLS applied to the spatial model would be an adequate substitute for the MLE procedure. The results of this inquiry were mixed for the Louisiana data. In most instances, it appears that OLS is satisfactory as a surrogate for MLE when a spatial process is *clearly* operating. The OLS coefficient estimates are always close to, if not identical to, the maximum-likelihood estimates. However, the OLS estimates of the standard errors of the coefficient estimates are inflated. In general, this inflation is problematic for OLS unless ρ is likely to be relatively low. In such cases, the greater precision of the maximum-likelihood estimators makes them preferable. For the Huk data, OLS is not an adequate surrogate for the MLE approach detailed here. The **W** matrix for the Huk data is more densely connected in the sense that many areas are connected to a large number of other areas relative to the Louisiana examples. It seems that the structure of **W** and the true value of ρ are critical in deciding whether OLS can be used as a surrogate for maximum-likelihood estimators. This is something that can best be investigated via Monte Carlo simulations, and such studies are under way. The clear, and perhaps conservative, advice is that for the spatial processes considered here, the MLE approach should be used. For regions with "less connected" areas and high values of ρ, OLS should probably suffice.

One further option outlined by Ord, but not discussed here, is to retain the usual population regression function $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ but to incorporate the spatial effects into the disturbance term via $\boldsymbol{\epsilon} = \rho\mathbf{W} + \boldsymbol{\nu}$, where $\boldsymbol{\nu}$ is a white noise term. This is a way of dealing with spatial autocorrelation and represents another method of incorporating geographical space into the analysis of social phenomena.

This chapter advocates that geographical space be at least explicitly considered for some social phenomena. For many social, political, and economic phenomena, geographical space may not be relevant. But for the cases where it is relevant, the procedures outlined here should lead to richer and more substantive analyses of those phenomena.

APPENDIX A: DERIVATION OF VARIANCE-COVARIANCE MATRICES FOR ESTIMATORS

Purely Endogenous Model

The log-likelihood function is given by

$$l(\mathbf{y}) = \text{const} - (N/2)\ln\sigma^2 - (1/2\sigma^2)\mathbf{y}'\mathbf{A}'\mathbf{A}\mathbf{y} + \ln|\mathbf{A}| \quad \text{(A-1)}$$

From Equation (9) we have, writing l for $l(\mathbf{y})$,

$$\partial l/\partial\omega = -N/2\omega + (1/2\omega^2)\mathbf{y}'\mathbf{A}'\mathbf{A}\mathbf{y} \tag{A-2}$$

Differentiating (A-1) with respect to ρ gives

$$\partial l/\partial\rho = (\partial/\partial\rho)\ln|\mathbf{A}| - (1/2\omega)(\partial/\partial\rho)\mathbf{y}'\mathbf{A}'\mathbf{A}\mathbf{y}$$

Using the form of $|\mathbf{A}|$ given in (12),

$$(\partial/\partial\rho)(\ln|\mathbf{A}|) = (\partial/\partial\rho)\sum_{i=1}^{N}\ln(1-\rho\lambda_i) = -\sum_{i=1}^{N}\lambda_i/(1-\rho\lambda_i)$$

and

$$\begin{aligned}(\partial/\partial\rho)(\mathbf{y}'\mathbf{A}'\mathbf{A}\mathbf{y}) &= (\partial/\partial\rho)[\mathbf{y}'\mathbf{y} - 2\rho\mathbf{y}'\mathbf{W}\mathbf{y} + \rho^2(\mathbf{W}\mathbf{y})'(\mathbf{W}\mathbf{y})]\\ &= -2\mathbf{y}'\mathbf{W}\mathbf{y} + 2\rho(\mathbf{W}\mathbf{y})'\mathbf{W}\mathbf{y}\end{aligned}$$

Hence

$$\partial l/\partial\rho = -\sum_{i=1}^{N}\lambda_i/(1-\rho\lambda_i) + (\mathbf{y}'\mathbf{W}\mathbf{y}/\omega) - (\rho\mathbf{y}'\mathbf{W}'\mathbf{W}\mathbf{y}/\omega) \tag{A-3}$$

We turn now to the second derivatives of $l(\mathbf{y})$. From (A-2):

$$\begin{aligned}\partial^2 l/\partial\omega^2 &= (N/2\omega^2) - (2/2\omega^3)\mathbf{y}'\mathbf{A}'\mathbf{A}\mathbf{y}\\ &= (1/2\omega^2)[N - (2/\omega)\mathbf{y}'\mathbf{A}'\mathbf{A}\mathbf{y}]\\ &= (1/2\omega^2)[N - 2N] \qquad \text{(at the minimum)}\\ &= -N/2\omega^2 \end{aligned} \tag{A-4}$$

From (A-3):

$$\begin{aligned}\partial^2 l/\partial\rho^2 &= (\partial/\partial\rho)\left(-\sum_{i=1}^{N}\lambda_i/(1-\rho\lambda_i)\right)\\ &\quad + (1/\omega)(\partial/\partial\rho)(\mathbf{y}'\mathbf{W}\mathbf{y} - \rho\mathbf{y}'\mathbf{W}'\mathbf{W}\mathbf{y})\\ &= -\sum_{i=1}^{N}\lambda_i^2/(1-\rho\lambda_i)^2 - (1/\omega)\mathbf{y}'\mathbf{W}'\mathbf{W}\mathbf{y}\\ &= \alpha - (1/\omega)\mathbf{y}'\mathbf{W}'\mathbf{W}\mathbf{y}\end{aligned} \tag{A-5}$$

where

$$\alpha = -\sum_{i=1}^{N}\lambda_i^2/(1-\rho\lambda_i)^2$$

From (A-2):

$$\begin{aligned}\partial^2 l/\partial\omega\partial\rho &= (1/2\omega^2)(\partial/\partial\rho)(\mathbf{y}'\mathbf{A}'\mathbf{A}\mathbf{y})\\ &= (1/2\omega^2)[-2\mathbf{y}'\mathbf{W}\mathbf{y} + 2\rho(\mathbf{W}\mathbf{y})'(\mathbf{W}\mathbf{y})]\\ &= -\ (1/\omega^2)[\mathbf{y}'(\mathbf{I} - \rho\mathbf{W}')\mathbf{W}\mathbf{y}]\\ &= -\epsilon'\mathbf{W}\mathbf{y}/\omega^2 \qquad \text{(by 4)}\end{aligned} \tag{A-6}$$

To obtain the information matrix, we need to take the expected values of these second derivatives. For (A-4):

$$E[-\ N/2\omega^2] = -\ N/2\omega^2 \tag{A-7}$$

For (A-5):

$$E[(1/\omega)\mathbf{y}'\mathbf{W}'\mathbf{W}\mathbf{y}] = (1/\omega)E[\epsilon'\mathbf{A}^{-1\prime}\mathbf{W}'\mathbf{W}\mathbf{A}^{-1}\epsilon]$$

Define $\mathbf{B} = \mathbf{W}\mathbf{A}^{-1}$; then

$$\begin{aligned}(1/\omega)E[\epsilon'\mathbf{A}^{-1\prime}\mathbf{W}'\mathbf{W}\mathbf{A}^{-1}\epsilon] &= (1/\omega)\mathrm{E}[\epsilon'\mathbf{B}'\mathbf{B}\epsilon]\\ &= (1/\omega)E[\mathrm{tr}\ \epsilon'\mathbf{B}'\mathbf{B}\epsilon]\\ &= (1/\omega)E[\mathrm{tr}\ \mathbf{B}'\mathbf{B}\epsilon\epsilon']\\ &= (1/\omega)[\mathrm{tr}\ (\mathbf{B}'\mathbf{B})E\epsilon\epsilon']\\ &= (1/\omega)[\mathrm{tr}\ (\mathbf{B}'\mathbf{B})\sigma^2\mathbf{I}]\\ &= \mathrm{tr}\ (\mathbf{B}'\mathbf{B}) \qquad \text{as } \omega = \sigma^2\end{aligned}$$

Therefore

$$E[\partial^2 l/\partial\rho^2] = \alpha - \mathrm{tr}(\mathbf{B}'\mathbf{B}) \tag{A-8}$$

For (A-6):

$$\begin{aligned}E[-(1/\omega^2)\epsilon'\mathbf{W}\mathbf{y}] &= -(1/\omega^2)E[\epsilon'\mathbf{W}\mathbf{A}^{-1}\epsilon]\\ &= -(1/\omega^2)E[\mathrm{tr}\ \epsilon'\mathbf{B}\epsilon]\\ &= -(1/\omega^2)[\mathrm{tr}\ (\mathbf{B})E\epsilon\epsilon']\\ &= -(1/\omega^2)[\mathrm{tr}\ (\mathbf{B})\sigma^2\mathbf{I}]\\ &= -\mathrm{tr}\ (\mathbf{B})/\omega\end{aligned} \tag{A-9}$$

The expressions in (A-7) to (A-9), when substituted into (14), yield

$$\mathbf{V}(\omega, \rho) = \begin{bmatrix} N/2\omega^2 & \operatorname{tr} \mathbf{B}/\omega \\ \operatorname{tr} \mathbf{B}/\omega & \operatorname{tr}(\mathbf{B}'\mathbf{B}) - \alpha \end{bmatrix}^{-1}$$
$$= \omega^2 \begin{bmatrix} N/2 & \omega \operatorname{tr} (\mathbf{B}) \\ \omega \operatorname{tr} (\mathbf{B}) & \omega^2 (\operatorname{tr} \mathbf{B}'\mathbf{B} - \alpha) \end{bmatrix}^{-1} \tag{A-10}$$

The expression given in (A-10) is the asymptotic variance-covariance matrix for the parameters estimated for the pure endogenous model.

Mixed Endogenous-Exogenous Model

The expressions for $\partial^2 l/\partial\omega^2$, $\partial^2 l/\partial\rho^2$, and $\partial^2 l/\partial\omega\partial\rho$ remain exactly as for the pure endogenous model. It is now necessary to obtain $\partial^2 l/\partial\beta^2$, $\partial^2 l/\partial\beta\partial\rho$, and $\partial^2 l/\partial\beta\partial\omega$. From (22), which gives $\partial l/\partial\beta$, we have

$$\partial^2 l/\partial\beta^2 = -(1/\omega)(\mathbf{X}'\mathbf{X}) \tag{A-11}$$

From (22) we also have

$$\begin{aligned} \partial^2 l/\partial\beta\partial\omega &= (1/\omega^2)[(\mathbf{X}'\mathbf{X})\boldsymbol{\beta} - \mathbf{X}'\mathbf{z}] \\ &= (1/\omega^2)[\mathbf{X}'\mathbf{z} - \mathbf{X}'\mathbf{z}] \\ &\qquad \text{as } \boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{z} \text{ at the minimum} \\ &= 0 \end{aligned} \tag{A-12}$$

From (22):

$$\begin{aligned} \partial^2 l/\partial\beta\partial\rho &= (1/\omega)(\partial/\partial\rho)(\mathbf{X}'\mathbf{z}) \\ &= (1/\omega)(\partial/\partial\rho)[\mathbf{X}'(\mathbf{I} - \rho\mathbf{W})\mathbf{y}] \\ &= -(1/\omega)\mathbf{X}'\mathbf{W}\mathbf{y} \end{aligned} \tag{A-13}$$

We now consider the expected values of the second derivatives in order to obtain the information matrix. As before,

$$E[\partial^2 l/\partial\omega^2] = -N/2\omega^2 \qquad \text{and} \qquad E[\partial^2 l/\partial\omega\partial\rho] = \operatorname{tr}(\mathbf{B})/\omega$$

From (A-12), $E[\partial^2 l/\partial\omega\partial\beta] = 0'$ (as a row vector). The result for $\partial^2 l/\partial\rho^2$ is not identical to the result for the purely endogenous model.

From (A-5):

$$\begin{aligned}\partial^2 l/\partial\rho^2 &= \alpha - (1/\omega)\mathbf{y}'\mathbf{W}'\mathbf{W}\mathbf{y}\\ &= \alpha - (1/\omega)[(\epsilon'\mathbf{A}^{-1\prime} + \beta'\mathbf{X}'\mathbf{A}^{-1\prime})\mathbf{W}'\mathbf{W}(\mathbf{A}^{-1}\mathbf{X}\beta + \mathbf{A}^{-1}\epsilon)]\\ &= \alpha - (1/\omega)[\epsilon'\mathbf{A}^{-1\prime}\mathbf{W}'\mathbf{W}\mathbf{A}^{-1}\mathbf{X}\beta + \beta'\mathbf{X}'\mathbf{A}^{-1\prime}\mathbf{W}'\mathbf{A}^{-1}\epsilon\\ &\quad + \epsilon'\mathbf{A}^{-1\prime}\mathbf{W}'\mathbf{W}\mathbf{A}^{-1}\epsilon + \beta'\mathbf{X}'\mathbf{A}^{-1\prime}\mathbf{W}'\mathbf{W}\mathbf{A}^{-1}\mathbf{X}\beta]\\ &= \alpha - (1/\omega)[\epsilon'\mathbf{B}'\mathbf{B}\epsilon + \beta'\mathbf{X}'\mathbf{B}'\mathbf{B}\mathbf{X}\beta + 2\epsilon'\mathbf{B}'\mathbf{B}X\beta]\end{aligned}$$

as $\epsilon'\mathbf{A}^{-1\prime}\mathbf{W}'\mathbf{W}\mathbf{A}^{-1}\mathbf{X}\beta = (\beta'\mathbf{X}'\mathbf{A}^{-1\prime}\mathbf{W}'\mathbf{A}^{-1}\epsilon)'$, a scalar. Thus

$$E[\partial^2 l/\partial\rho^2] = -\alpha + \mathrm{tr}(\mathbf{B}'\mathbf{B}) + (1/\omega)\beta'\mathbf{X}'\mathbf{B}'\mathbf{B}\mathbf{X}\beta \qquad \text{(A-14)}$$

From (A-13):

$$\begin{aligned}E[\partial^2 l/\partial\beta\partial\rho] &= -(1/\omega)\mathbf{X}'E[\mathbf{W}\mathbf{y}]\\ &= (1/\omega)\mathbf{X}'\mathrm{E}[\mathbf{W}\mathbf{A}^{-1}\mathbf{X}\beta + \mathbf{W}\mathrm{A}^{-1}\epsilon] \qquad \text{(A-15)}\\ &= (1/\omega)\mathbf{X}'\mathbf{B}\mathbf{X}\beta\end{aligned}$$

Finally, from (A-11),

$$E[\partial^2 l/\partial\beta^2] = -(1/\omega)(\mathbf{X}'\mathbf{X}) \qquad \text{(A-16)}$$

Substituting all the expressions for the expected values and noting the negative signs gives

$\mathbf{V}(\omega, \rho, \beta)$

$$= \omega^2\begin{bmatrix} N/2 & \omega\,\mathrm{tr}\,(\mathbf{B}) & \mathbf{0}' \\ \omega\,\mathrm{tr}\,\mathbf{B} & \omega^2(\mathrm{tr}\,\mathbf{B}'\mathbf{B} - \alpha) + \omega\beta'\mathbf{X}'\mathbf{B}'\mathbf{B}\mathbf{X}\beta & \omega\mathbf{X}'\mathbf{B}\mathbf{X}\beta \\ \mathbf{0} & \omega\mathbf{X}'\mathbf{B}\mathbf{X}\beta & \omega\mathbf{X}'\mathbf{X}\end{bmatrix}^{-1} \qquad \text{(A-17)}$$

APPENDIX B: TESTING FOR SPATIAL AUTOCORRELATION

Suppose a residual, $\hat{\epsilon}$, has been returned from a (nonspatial) regression analysis. To test for spatial autocorrelation, the following test statistic is defined:

$$I = (N/T)(\hat{\epsilon}'\mathbf{W}\hat{\epsilon}/\hat{\epsilon}'\hat{\epsilon}) \qquad \text{(B-1)}$$

where

$$T = \sum_{\substack{i,j\\ i\neq j}} w_{ij}$$

Cliff and Ord (1973) derive expressions for $E[I]$ and $V[I]$ in order to construct a standardized normal deviate. Define $\mathbf{D} = [d_{ij}] = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Then

$$E[I] = [1 + (N/T)\sum_{\substack{i=1\\ i\neq j}}^{N}\sum_{j=1}^{N} w_{ij}\, d_{ij}]/(N - K) \tag{B-2}$$

where there are K exogenous variables (including the column of 1's for the intercept). The expression for $V[I]$ is considerably more complex. The following preliminary definitions are required:

$$w_{i.} = \sum_{j=1}^{N} w_{ij}$$

$$w_{j.} = \sum_{i=1}^{N} w_{ij}$$

$$S_1 = \frac{1}{2}\sum_{\substack{i=1\\ i\neq j}}^{N}\sum_{j=1}^{N} (w_{ij} + w_{ji})^2$$

$$S_2 = \sum_{i=1}^{N} (w_{i.} + w_{.i})^2$$

With these definitions,

$$\begin{aligned} V[I] = [N/(N - K)T^2]\Bigg\{&(N^2S_1 - NS_2 + 3T^2)/N^2 \\ &+(1/N)\sum_{i=1}^{N}\sum_{j=1}^{N}(w_{i.} + w_{.i})(W_{j.} + w_{.j})\, d_{ij} \\ &+ 2\Bigg(\sum_{\substack{i=1\\ i\neq j}}^{N}\sum_{j=1}^{N} w_{ij}d_{ij}\Bigg)^2 \\ &-\Bigg[\sum_{i=1}^{N}\sum_{\substack{j=1\\ i\neq j\neq k}}^{N}\sum_{k=1}^{N}(w_{ik} + w_{ki})(w_{jk} + w_{kj})d_{ij} \\ &+ \sum_{\substack{i=1\\ i\neq j}}^{N}\sum_{j=1}^{N}(w_{ij} + w_{ji})^2\, d_{ii}\Bigg] \\ &+ (1/N)\sum_{i=1}^{N}\sum_{\substack{j=1\\ i\neq j\neq k}}^{N}\sum_{k=1}^{N}(w_{ij} + w_{ji})(w_{ik} + w_{ki})(d_{ii}d_{jk} - d_{ij}d_{ik})\Bigg\} \\ &- [1/(N - K)^2] \end{aligned} \tag{B-3}$$

Note that when there are no exogenous variables $E[I] = -(1/N)$ and

$$V[I] = N/(N - K)T^2 [(N^2 S_1 - NS_2 + 3T^2)/N^2]$$

which can be used to assess the spatial autocorrelation of $\mathbf{y}$.

REFERENCES

AIGNER, D. J., AND HEINS, A. J.
1967 "On the determinants of income equality." *American Economic Review* 57:175-184.

ARORA, S. S., AND BROWN, M.
1977 "Alternative approaches to spatial autocorrelation: An improvement over current practice." *International Regional Science Review* 2:67-78.

CAPECCHI, V., AND GALLI, G.
1969 "Determinants of voting behavior in Italy: A linear causal model of analysis." In M. Dogan and S. Rokkan (Eds.), *Social Ecology*. Cambridge, Mass.: M.I.T. Press.

CHIROT, D., AND RAGIN, C.
1975 "The market, tradition and peasant rebellion: The case of Romania in 1907." *American Sociological Review* 40:428-444.

CLIFF, A. D., AND ORD, K.
1970 "Spatial autocorrelation: A review of existing and new measures with applications." *Economic Geography* June(Supplement):269-292.
1973 *Spatial Autocorrelation*. London: Pion.

COELEN, S. P.
1976 "Book review: Spatial autocorrelation by Cliff and Ord." *Journal of Economic Literature* 14:924-925.

COX, K. R.
1969 "Voting in the London suburbs: A factor analysis and a causal model." In M. Dogan and S. Rokkan (Eds.), *Social Ecology*. Cambridge, Mass.: M.I.T. Press.

DACEY, M. F.
1965 "A review on measures of contiguity for two and *k*-color maps." Technical Report 2, Spatial Diffusion Study. Department of Geography, Northwestern University.

DOREIAN, P., AND HUMMON, N. P.
1976 *Modeling Social Processes*. New York: Elsevier.

DUNLAP, R., AND CATTON, W.
1979 "Environmental sociology." *Annual Review of Sociology* 5:243-273.

FRISBIE, W. P., AND POSTON, D. L.
1975 "Components of sustenance organization and non-metropolitan change: A human ecological investigation." *American Sociological Review* 40:773–784.

HIBBS, D.
1974 "Problems of statistical estimation and causal inference in time series regression models." In H. L. Costner (Ed.), *Sociological Methodology 1973–74*. San Francisco: Jossey-Bass.

HSIAO, C.
1975 "Some estimation methods for a random coefficient model." *Econometrica* 43:305–325.

INVERARITY, J. M.
1976 "Populism and lynching in Louisiana 1889–1896: A test of Erikson's theory of the relationship between boundary crises and repressive justice." *American Sociological Review* 41:262–280.

JOHNSTON, J.
1963 *Econometric Methods*. New York: McGraw-Hill.

KENDALL, M. G., AND STUART, A.
1967 *The Advanced Theory of Statistics*. Vol. 2. London: Hafner.

KERNALL, S.
1973 "Comment: A re-evaluation of black voting in Mississippi." *American Political Science Review* 67:1307–1318.

LANCASTER, K.
1971 *Mathematical Economics*. New York: Macmillan.

MATTHEWS, D. R., AND PROTHRO, J. W.
1963a "Social and economic factors and Negro voter registration in the south." *American Political Science Review* 57:24–44.
1963b "Political factors and Negro voter registration in the south." *American Political Science Review* 57:355–367.

MEAD, R.
1967 "A mathematical model for the estimation of interplant competition." *Biometrics* 23:189–205.

MITCHELL, E. J.
1969 "Some econometrics of the Huk rebellion." *American Political Science Review* 63:1159–1171.

MORAN, P. A. P.
1950 "A test for the serial independence of residuals." *Biometrika* 37:178–181.

ORD, K.
1975 "Estimation methods for models of spatial interaction." *Journal of American Statistical Association* 70:120–126.

PIELOU, E. C.
1969 *An Introduction to Mathematical Ecology*. New York: Wiley.

RAGIN, C.
1977 "Class, status, and reactive ethnic cleavages: The social basis of political regionalism." *American Sociological Review* 42:438–450.
SALAMON, L. M., AND VAN EVERA, S.
1973 "Fear, apathy, and discrimination: A test of three explanations of political participation." *American Political Science Review* 67:1288–1306.
SCHOENBERGER, R. A., AND SEGAL, D. R.
1971 "Ecology of dissent: The southern Wallace vote in 1968." *Midwest Journal of Politics* 15:583–586.
SWAMY, P. A. V. B.
1970 "Efficient-inference in a random coefficient regression model." *Econometrica* 38:311–323.
THEIL, H.
1971 *Principles of Econometrics*. New York: Wiley.
WASSERMAN, I. M., AND SEGAL, D. R.
1973 "Aggregate effects in the ecological study of presidential voting." *American Journal of Political Science* 17:177–181.