On the Large-Sample Estimation of Regression Models with Spatial- Or Network-Effects Terms: A Two-Stage Least Squares Approach
Author(s): Kenneth C. Land and Glenn Deane
Source: *Sociological Methodology*, Vol. 22 (1992), pp. 221-248
Published by: American Sociological Association
Stable URL: http://www.jstor.org/stable/270997
Accessed: 07/07/2009 14:53

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode=asa.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit organization founded in 1995 to build trusted digital archives for scholarship. We work with the scholarly community to preserve their work and the materials they rely upon, and to build a common research platform that promotes the discovery and use of these resources. For more information about JSTOR, please contact support@jstor.org.

American Sociological Association is collaborating with JSTOR to digitize, preserve and extend access to *Sociological Methodology*.

http://www.jstor.org

7

ON THE LARGE-SAMPLE ESTIMATION OF REGRESSION MODELS WITH SPATIAL- OR NETWORK-EFFECTS TERMS: A TWO-STAGE LEAST SQUARES APPROACH

*Kenneth C. Land**
Glenn Deane†

In this paper we apply the two-stage least squares (2SLS) estimation technique to produce regression estimators for linear models that incorporate spatial- or network-effects terms. Such models are often necessary to adequately represent social processes in data that have been aggregated for politically or administratively defined areas such as census tracts, cities, counties, or states. They are also necessary to represent data from individuals or organizations connected in explicit social networks. We review maximum likelihood (ML) estimators of the spatial-

This research was supported in part by National Science Foundation grant SES-8907665 and by the Yale University Program on Non-Profit Organizations. The paper was presented at the Annual Meetings of the American Sociological Association, Cincinnati, 1991. We thank Yu Xie, Noah E. Friedkin, Peter Marsden, and three anonymous reviewers for helpful comments on an earlier version. Members of the Applied Statistics Seminar organized by Kenneth A. Bollen at the Institute for Research in the Social Sciences of the University of North Carolina, Chapel Hill, also provided an insightful discussion of several issues in a presentation based on this paper.

*Duke University
†State University of New York, Albany

effects model (Ord 1975; Doreian 1981) and sample-size limitations thereto. We discuss Roncek and Montgomery's (1984) suggested use of generalized population potentials to measure spatial effects, and we demonstrate the inconsistency of ordinary least squares (OLS) estimators of spatial-effects models with potential variables. We use the 2SLS technique to derive consistent estimators in spatial-effects models with generalized potential variables. We derive the asymptotic (large-sample) variance of the 2SLS estimator and compare it with the ML estimator. We also compare the 2SLS estimator and its properties with similar estimators for spatial econometric models developed independently by Anselin (1988). We apply the 2SLS estimators to a large sample of county-level data on church adherence rates and find that our 2SLS estimator is much more computationally efficient than the ML estimator and yields numerical estimates of comparable statistical efficiency.

Sociologists and other social scientists study a diverse array of social, political, demographic, and economic phenomena that take place in geographical space. When the social processes that generate these phenomena are analyzed with data that have been aggregated for politically or administratively defined areas such as census tracts, cities, counties, or states, the geographical constraints and impacts of these arbitrary units on the social processes are implicitly retained. Yet standard applications of linear statistical models rarely take geographical dependency into account.[1] This point has been made most forcefully in the sociological methodology literature by Doreian (1980, 1981), who applied modifications of Ord's (1975) maximum likelihood (ML) estimation procedures for spatial-interaction models.

A shortcoming of Ord's estimators is that they become unwieldy or impossible to apply in larger samples, a point made by Roncek and Montgomery (1984). This—and the absence of computer software specifically written for Ord's procedures—may account for the few empirical applications in the sociological literature.

[1]Similar problems arise in the linear model analysis of data on individuals or organizations connected in explicit social networks when the network interdependencies are ignored. Because the empirical applications that motivated this paper involve spatially distributed rather than network data, however, the present discourse is limited to the spatial context. We briefly discuss applications of the methods developed herein to the related network-effects model in the "Conclusion."

To circumvent the sample-size limitations of ML estimators, Roncek and Montgomery (1984) introduced the generalized population-potential variable and recommended using this variable as a regressor in the conventional linear model. As we show below, a problem with this procedure is that the resulting regressor is correlated with the error term of the linear model, which means that the usual least squares regression estimator of its effect parameter is statistically inconsistent (i.e., biased in large samples).

In this paper we apply the two-stage least squares (2SLS) technique to produce a regression estimator for Ord's spatial-effects model that is statistically consistent and can accommodate large samples. In the following sections, we review the specification of the spatial-effects model, its ML estimators, and its sample-size limitations. We then discuss Roncek and Montgomery's (1984) suggested use of generalized population potentials in spatial-effects models and the inconsistency of ordinary least squares (OLS) estimation of linear models with potential variables. We use the 2SLS estimation technique to produce consistent estimators in spatial-effects models with potential variables. We derive the asymptotic (large-sample) variance of the 2SLS estimator and compare it with the ML estimator. We also compare similar 2SLS estimators developed by Anselin (1988) for spatial econometric models. Finally, we apply the 2SLS estimators to a large sample of county-level data on church adherence rates.

1. THE SPATIAL-EFFECTS MODEL: SPECIFICATION AND MAXIMUM LIKELIHOOD ESTIMATION

Ord (1975) developed two types of regression models for spatial interaction to accommodate situations in which spatially aggregated social data violate the standard linear model's assumption of random errors that are independently distributed among the sampled units. These violations are due either to errors that are correlated across some systematic ordering of the spatial units because of the influence of unmeasured variables or to spatial processes for the dependent variable that are not fully captured by the regressors included in the model. In the latter case, the values of the dependent variable in one spatial unit are systematically related to values of this variable in adjacent units. Doreian (1980, 1981) called a regression

model of the former situation the *spatial-disturbances model* and a regression model of the latter the *spatial-effects model.*

The spatial-effects model is the focus of this paper. It is specified by

$$\mathbf{y} = \rho \mathbf{W} y + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{1}$$

with

$$\mathrm{E}[\boldsymbol{\epsilon}] = 0 \quad \text{and} \quad \mathrm{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}'] = \sigma_{\epsilon}^2 \mathbf{I},$$

where $\mathbf{y}$ is a column vector of n sample observations on a dependent or response variable, $\mathbf{X}$ is an n-by-k matrix of observations on k fixed exogenous or explanatory variables (including a unit vector), $\boldsymbol{\beta}$ is a k-row column vector of regression (intercept and slope) parameters, $\boldsymbol{\epsilon}$ is an n-row column vector of unobserved random errors or disturbances with expected values of zero and variance σ_{ϵ}^2, $\mathbf{W}$ is an n-by-n matrix of weights with elements w_{ij} describing the "nearness" of sample units i and j and a zero principal diagonal (indicating that the distance of unit i to itself is zero), and $\mathbf{I}$ is an identity matrix of order n.

In the specification of the spatial-effects model, the random errors must be normally distributed with mean and variance as given above (Ord 1975; Doreian 1981). This assumption (or some suitable alternative parametric assumption) is necessary for the derivation of ML estimators. But only the weaker error specifications stated above are necessary for the 2SLS estimation techniques applied here. This adds substantial flexibility to the error processes permitted by model (1). All that is required is that the errors have zero expectation, have constant variance, and be independent (random). Even the latter two assumptions can be relaxed under suitable generalizations of 2SLS techniques (which are not pursued in this paper).

The parameter ρ is a *spatial-effects coefficient* that measures the relation of the value of y in any given geographical unit of analysis to the values of y in other areas. If ρ is significantly different from zero, a spatial-interaction or diffusion process is operating. Doreian (1981) described several empirical applications in which the estimation of regression models that do not contain the spatial-effects term results in coefficient estimates for the exogenous variables in the $\mathbf{X}$ matrix of model (1) that are inflated (biased upward) relative to their values when the spatial effects are taken into account. Thus, the

accuracy of inferences may depend upon the incorporation of the spatial-effects term.

The statistical inference problem for the spatial-effects model is to develop statistically efficient estimators of ρ, $\boldsymbol{\beta}$, and σ^2, and the standard errors of these parameters. Ord (1975) developed ML estimators for this purpose, which Doreian (1981) exposited for sociologists, modified, and applied empirically. The ML estimators can be written as follows:

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,(\mathbf{I} - \hat{\rho}\mathbf{W})\mathbf{y} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{z},\end{aligned} \tag{2}$$

where $\mathbf{z} = \mathbf{A}\mathbf{y} = (\mathbf{I} - \rho\mathbf{W})\mathbf{y}$,

$$\hat{\sigma}^2 = (1/n)\mathbf{z}'\mathbf{M}\mathbf{z}, \tag{3}$$

where $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, and $\hat{\rho}$ is the value of ρ that minimizes the quantity

$$-(2/n)\sum_{i=1}^{n} \ln(1 - \rho\lambda_i) + \ln\,[\mathbf{y}'\mathbf{M}\mathbf{y} - 2\rho\mathbf{y}'\mathbf{M}\mathbf{W}\mathbf{y} + \rho^2(\mathbf{W}\mathbf{y})'\mathbf{M}\mathbf{W}\mathbf{y}], \tag{4}$$

where the $(1 - \rho\lambda_i)$ are the eigenvalues of the matrix $\mathbf{A} = \mathbf{I} - \rho\mathbf{W}$.

When they are written in terms of the transformed dependent variable z, we can see that the ML estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ in equations (2) and (3) look similar to the ML expressions for these parameters in the standard linear model. However, a numerical value of $\hat{\rho}$ is required to compute each of these estimators, and $\hat{\rho}$, in turn, must generally be computed by a direct-search numerical-minimization procedure applied to equation (4). For datasets of almost any realistic complexity, this requires a computer. In small to medium-size samples (up to 200, say), while computationally demanding, this should be feasible and relatively quick with current statistical packages (SAS, SPSS^{x}) and mainframe computers.

Roncek and Montgomery (1984) noted, however, that the ML estimators of the spatial-effects regression model can be increasingly difficult to compute in larger samples. The difficulties relate partly to the storage and processing capacity of computers and to the limitations of statistical software. The source of these difficulties is the spatial weight matrix $\mathbf{W}$ of model (1), which is an n-by-n matrix. This matrix is used in matrix addition and multiplication operations in the

ML estimators (2), (3), and (4) above. To obtain the variance-covariance matrix of these estimators (see Doreian 1981, pp. 383–84), which is necessary for hypothesis testing and the construction of confidence intervals, we must compute the inverse of the n-by-n matrix $\mathbf{A} = \mathbf{I} - \rho\mathbf{W}$.

Roncek and Montgomery (1984) observed that specifying a **W** matrix (and hence the corresponding **A** matrix) to analyze census-tract data from the city of Chicago would require approximately a 900-by-900 matrix, and that analyzing Chicago city blocks would require a 15,000-by-15,000 matrix. Using the double-precision arithmetic necessary for multiplying and inverting matrices would require 128 megabytes of main memory just to store the latter matrix. Large mainframe computers typically have about 100 megabytes of main memory. Thus, the addition, multiplication, and inversion of matrices of these sizes would require the use of supercomputers.[2]

2. GENERALIZED POPULATION POTENTIALS

To circumvent the sample-size limitations of ML procedures for the spatial-effects model, Roncek and Montgomery (1984) proposed an adaptation of the population-potential variable developed in population studies of migration and related spatial phenomena by Duncan, Cuzzort, and Duncan (1963). The population-potential variable measures the cumulative proximity of the populations of all the areas surrounding a particular place. The population potential for location i, PP_i, is defined by

$$PP_i = \sum_{j \neq i}^{n} (P_j/D_{ij}), \qquad i = 1, \ldots, n, \tag{5}$$

[2]Standard computer software packages that perform regression analyses also limit the sizes of matrices on which matrix operations can be performed. For instance, generally available versions of SAS, one of the most powerful statistical packages used by social scientists, limit the number of elements that can be simultaneously processed in the main memory to 32,000, which equates to a maximum square matrix of 178 rows and columns. However, the recently released version 6 of SAS allows matrices with any number of rows and columns, provided there is enough main computer memory to store them. For an example of the space-conservation strategies necessary to apply the ML estimator—on a mainframe computer with eight megabytes of main memory—to social network data on a sample of approximately 1,000 high school students, see Duke (1991, App. A).

where P_j is the population of the jth location or place in the environment of location i, D_{ij} is the distance of location j from location i, and the summation is taken over all locations j other than i. To operationalize the denominator, Duncan et al. (1963) use the Euclidean or straight-line distance between geographical locations. Other measures such as travel distance are more difficult to measure. Compared with transformed distance measures such as the logarithm of Euclidean distance or the squared distance (as in gravity models), the simple Euclidean distance measure works well in practice (Roncek and Montgomery 1984).

In brief, to calculate the population potential for a set of n locations, each place is treated successively as the point of reference, and the sum of quotients of the population of every other place divided by its distance from the reference point is computed. In classical population models of migration among a set of locations (Duncan et al. 1963), it is sufficient to use the size of populations in other locations in the numerator, as in equation (5). When the substantive phenomenon being investigated is something other than the migratory exchange of individuals among the locations, variants of the potential concept, which we call *generalized population potentials,* can be defined by replacing the population sizes in the numerator of (5) with the values of the dependent variable being studied (Roncek and Montgomery 1984). For instance, Roncek and Robinson (1984) analyzed the effects of crime in the surroundings of census tracts in Cleveland by using a generalized population-potential variable they called the crime potential.

In the spatial-effects regression model (1), the generalized population-potential variable can be obtained by multiplying the weight matrix $\mathbf{W}$ (with elements suitably defined as inverses of Euclidean distances between the geographical units in an analysis) by the dependent variable (column vector) $\mathbf{y}$:

$$\mathbf{y}^* = \mathbf{W}\mathbf{y}. \tag{6}$$

For example, in the aforementioned census-tract crime study of Roncek and Robinson (1984), the dependent column vector $\mathbf{y}$ consists of the numbers of crimes reported to the police in the various census tracts, and the column vector $\mathbf{y}^*$ contains measures of the crime potential of each census tract relative to the numbers of crimes

in the other census tracts of the city, each weighted by its distance from the focal tract.

Algebraically, Roncek and Montgomery's (1984) proposal to use generalized population potentials is the same as treating the product of the weight matrix **W** and the dependent variable vector **y** on the right side of model (1) as a new variable, $\mathbf{y}^*$, so that model (1) can be rewritten

$$\mathbf{y} = \rho\mathbf{y}^* + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}. \tag{7}$$

This new variable can be entered as just another regressor variable in an OLS estimation of model (7), without creating problems in the estimates of the spatial-effects parameter ρ (Roncek and Montgomery 1984). In general adjacency weight matrices, **W** (i.e., in matrices with elements not necessarily restricted to inverses of physical distances), Doreian and Hummon (1976) similarly suggested that $\mathbf{y}^* = \mathbf{Wy}$ can simply be included as another exogenous variable in an OLS estimation of model (7).[3]

3. THE INCONSISTENCY OF ORDINARY LEAST SQUARES ESTIMATION

We agree with Roncek and Montgomery (1984) that the treatment of $\mathbf{y}^*$ as an ordinary regressor in the estimation of model (7) has the computational benefit in large samples of replacing the large n-by-n matrix **W** with an n-element column vector, but consistent estimates of ρ are not as easy to obtain as OLS estimates. As Ord (1975) pointed out, this requires that inferences be restricted to so-called conditional inferences in which the spatial effects are presumed to be determined prior to the dependent variable. But in spatial-effects models, the spatial diffusion or interaction processes are determined simultaneously with the dependent variable. This produces a nonzero correlation between the potential variable $\mathbf{y}^*$ and the error term $\boldsymbol{\epsilon}$, which violates the assumptions under which OLS produces unbiased (and therefore consistent) estimates of the regression coefficients.

This can be demonstrated as follows. Taking expectations of model (1), we see that for the ith case

[3]Doreian (1981, p. 373) recognized that the suggested OLS procedure of Doreian and Hummon (1976) was incorrect and compared numerical estimates of OLS with ML estimates.

$$E(y_i) = \rho \Sigma_j w_{ij} E(y_j) + \mathbf{X}_i \beta, \tag{8}$$

where $\mathbf{X}_i$ denotes the row vector of observations on the k explanatory variables corresponding to the unit i. The appearance of the expected-value operator on both sides of this equation emphasizes the joint nature of the $\mathbf{y}$ and $\mathbf{y}^*$ variables. By contrast, as Ord (1975) noted, in a conditional spatial-effects specification,

$$E\{y_i \mid y_j, j \in J(i)\} = \rho \Sigma_j w_{ij} y_j + \mathbf{X}_i \boldsymbol{\beta}, \tag{9}$$

where $J(i)$ denotes the set of locations whose values on the dependent variable interact with the value of location i. The expectation on the left side of equation (9) is a conditional expectation, where the conditioning is on the observed values of the dependent variable y_j that fall within the set $J(i)$. Furthermore, the two equations are equivalent only if

$$E\{\epsilon_i \mid y_j, j \in J(i)\} = E\{\epsilon_i y_j^*\} = 0, \tag{10}$$

that is, only if ϵ_i and $y_j^* = \Sigma_k w_{jk} y_k$ are uncorrelated.

In some time-series models, the time-dependent, one-sided nature of the relationship may make equation (10) a natural restriction. But in spatial-effects models in which the dependent variable in a given location may be influenced by the dependent variable in all of the locations in a given set, equation (10) will generally not be satisfied (Ord 1975). The consequence, of course, is a violation of the assumption in the general linear regression model that the error term $\boldsymbol{\epsilon}$ is distributed independently of the regressors. In turn, this implies that the OLS estimators will be biased in samples of any size and therefore asymptotically biased or inconsistent (Johnston 1984).

The foregoing can be intuitively expressed as follows. For any given location i in a set under consideration, assume that all $w_{ij} \geq 0$, although some of these coefficients may be quite small. Assume also that $\rho > 0$, so that spatial effects are operative among the locations. Then the value of the dependent variable for location i, y_i, will be influenced (i.e., will be a function of) the values of the dependent variable at all other locations through the generalized potential variable for location i, $y_i^* = \Sigma_j w_{ij} y_j$. But in turn, y_i and therefore its random component ϵ_i will enter into the determination of the values of y_j for each other location j through the contributions of y_i to the generalized potentials, y_j^*. This implies that the value of the general-

ized potential for location i, y_i^*, will be partly affected by its own random error, ϵ_i, and therefore that the two are correlated. Again, this violates the conditions necessary for unbiased and consistent OLS regression estimators.[4]

4. A TWO-STAGE LEAST SQUARES ESTIMATOR

In brief, direct OLS estimation of the spatial-effects model (7) will result in biased and inconsistent parameter estimates because of an unknown level of correlation of the generalized potential variable and the error term of the model. Theil (1953*a*, 1953*b*, 1961) and Basmann (1957) developed 2SLS estimation to produce consistent estimators in a similar context: simultaneous equation econometric models in which some regressors affecting the dependent or endogenous variable in a given equation (namely, those corresponding to other endogenous variables in the model) are correlated with the error term. In 2SLS estimation, the regressors in question are replaced by other regressors, which are purged of the stochastic elements that correlate with the error term of the equation, and then an OLS regression of the dependent variable of the equation on the purged and exogenous variables of the equation is performed.

The problem in this paper is to construct a variable, say $\hat{\mathbf{y}}^*$, to replace the generalized potential variable $\mathbf{y}^*$ in model (7). This is similar to the problem for which the 2SLS technique was developed. It will be useful (for reasons explained below) to redefine and partition the matrix $\mathbf{X}$ of all fully exogenous variables as $[\mathbf{X}_1\ \mathbf{X}_2]$, where $\mathbf{X}_1$ consists of a set of exogenous variables that enter into (7), and $\mathbf{X}_2$ consists of those that do not enter.

In the 2SLS procedure (Johnston 1984, pp. 472–73), the column vector $\hat{\mathbf{y}}^*$ is computed in the *first-stage:* The generalized

[4]Doreian, Teuter, and Wang (1984) provided a Monte Carlo simulation study of the consequences of direct application of OLS—which they called the QAD (quick and dirty) method—to model (7). In general, they found that as the size of the ρ parameter increases from .1 to .9, the bias of the OLS estimates of the ρ and $\boldsymbol{\beta}$ parameters increases, although not dramatically in all cases. More important for statistical inferences, as the size of ρ increases, the actual standard deviations of the OLS coefficient estimates (estimated by a simulation approximation to their true variation) are often much larger than those of the ML estimates.

population-potential column vector $\mathbf{y}^*$ is regressed on the complete set of fully exogenous variables (i.e., on the variables in the full matrix $\mathbf{X}$), and the actual observations on the $\mathbf{y}^*$ variable are replaced by the corresponding regression (predicted) values:

$$\hat{\mathbf{y}}^* = \mathbf{X}\mathbf{d} = \mathbf{X}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}^*] = \mathbf{X}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y}]. \tag{11}$$

In the *second stage,* y is regressed on $\hat{y}^*$ and the submatrix X_1, yielding the following solution to the normal equations:

$$\begin{bmatrix} r \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{y}}^{*\prime}\hat{\mathbf{y}}^* & \hat{\mathbf{y}}^{*\prime}\mathbf{X}_1 \\ \mathbf{X}_1'\hat{\mathbf{y}}^* & \mathbf{X}_1'\mathbf{X}_1 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{y}}^{*\prime}\mathbf{y} \\ \mathbf{X}_1'\mathbf{y} \end{bmatrix}, \tag{12}$$

where $[r\ \mathbf{b}]'$ now denotes the 2SLS estimator of $[\rho\ \boldsymbol{\beta}]'$.

Several comments will help to clarify this application of the 2SLS method to model (7). First, to estimate $[r\ \mathbf{b}]'$, there is no need to compute the expected regression values in $\hat{\mathbf{y}}^*$ and enter them explicitly in the matrix equation (12). An alternative form of the estimator (12), which involves only the matrices of actual observations, is available (Johnston 1984, p. 473). It is the latter form that is utilized in the 2SLS programs available in many computer packages. While not reproduced here, this form may be obtained by substituting equation (11) into equation (12).

Second, we indicated above that $\mathbf{y}^*$ should be regressed on all fully exogenous variables (contained in the matrix $\mathbf{X}$) in the first stage. This produces a set of predicted values, the $\hat{\mathbf{y}}^*$ of equation (11), for $\mathbf{y}^* = \mathbf{W}\mathbf{y}$ in the spatial-effects model (7). The logic here is to treat $\mathbf{y}^*$ the same way that endogenous variables are treated in simultaneous equation systems and compute predicted values of $\mathbf{y}^*$ via equation (11). Because the resulting $\hat{\mathbf{y}}^*$ of equation (11) is a linear combination of all of the exogenous variables contained in the $\mathbf{X}$ matrix, the matrix to be inverted in equation (12) will be singular and not invertible unless equation (11) contains some exogenous variables that do not enter into the second-stage regression of equation (12).[5] Hence, we get the partition of the matrix $\mathbf{X}$ introduced above. Note that this partition requires that the analyst be able to sort exogenous variables into those (in $\mathbf{X}_1$) that have direct effects on the

[5]We are grateful to an anonymous referee for pointing out this identifiability constraint and to Peter V. Marsden for providing a numerical demonstration of the requirement.

dependent variable $\mathbf{y}$ in model (7) and those (in $\mathbf{X}_2$, which often are called *instrumental variables* or *instruments*) suitable for the identification and estimation of the predicted variable $\hat{\mathbf{y}}^*$. (Additional suggestions for choosing instrumental variables are given in later sections of the paper.)

Third, as in a simultaneous equation system, the 2SLS estimator (12) is consistent. That is, this estimator $[r\ \mathbf{b}]'$ converges in probability (or has probability limit equal) to the population-parameter vector $[\rho\ \boldsymbol{\beta}]'$.

Consistency of the 2SLS estimator (12) requires that both the predicted variable $\hat{\mathbf{y}}^*$ and the exogenous variables $\mathbf{X}$ are uncorrelated with the error term $\boldsymbol{\epsilon}$ of model (7) in the probability limit. By their very definition as exogenous variables, the X variables are uncorrelated in the limit with the model's error term. For the predicted variable $\hat{\mathbf{y}}^*$, we have

$$\begin{aligned}\operatorname{plim}[(l/n)\hat{\mathbf{y}}^{*\prime}\boldsymbol{\epsilon}] &= \operatorname{plim}[(1/n)\mathbf{y}^{*\prime}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\epsilon}] \\ &= [\operatorname{plim}(1/n)\mathbf{y}^{*\prime}\mathbf{X}][\operatorname{plim}(1/n)(\mathbf{X}'\mathbf{X})^{-1}][\operatorname{plim}(1/n)\mathbf{X}'\boldsymbol{\epsilon}] \\ &= \mathbf{0}, \qquad (13)\end{aligned}$$

where $\mathbf{0}$ denotes a column vector of dimension n, and we have utilized standard properties of the probability limit operation (see, e.g., Johnston 1984, pp. 269–74). To see that this expression indeed equals a zero vector, first note that the last term in the second line pertains to the probability limit of the covariance of the exogenous variables and the error term of model (7), which we have just noted must be zero, otherwise the X variables would not be classified as exogenous. Second, note that for the equality to hold, the other two probability limits in (13) must converge to finite matrices and the middle term must be nonsingular.

These are fairly weak conditions on the cross-product matrices of these components that must be satisfied for any meaningful regression analysis. Essentially, they require that the cross products standardized for sample size (or, equivalently, the second moments) of the X variables with themselves and with the generalized population-potential variable $\mathbf{y}^*$ stabilize at finite levels as the sample size increases without bound. Under these conditions, and assuming that the exogenous variables are truly exogenous, the 2SLS estimator (12) is consistent.

5. THE ASYMPTOTIC VARIANCE OF THE TWO-STAGE LEAST SQUARES ESTIMATOR

Under similar conditions, it has been established that 2SLS estimators are generally asymptotically normally distributed with mean vector equal to the parameter vector and a specific form for the asymptotic variance-covariance matrix (see, e.g., Theil 1971, pp. 497–99). In the present case, the estimator of this matrix has the following form:

$$\text{asy var}\begin{bmatrix} \mathbf{r} \\ \mathbf{b} \end{bmatrix} = s_\epsilon^2 \begin{bmatrix} \hat{\mathbf{y}}^{*\prime}\hat{\mathbf{y}}^* & \hat{\mathbf{y}}^{*\prime}\mathbf{X}_1 \\ \mathbf{X}_1'\hat{\mathbf{y}}^* & \mathbf{X}_1'\mathbf{X}_1 \end{bmatrix}^{-1}, \tag{14}$$

where

$$s_\epsilon^2 = \frac{(\mathbf{y} - \hat{\mathbf{y}}^* r - \mathbf{X}_1\mathbf{b})'(\mathbf{y} - \hat{\mathbf{y}}^* r - \mathbf{X}_1\mathbf{b})}{n}$$

is a consistent estimator of σ_ϵ^2, the error variance of model (7).[6] Like the 2SLS estimator (12), the variance-matrix estimator (14) has a more computationally convenient expression that involves only the matrices of actual observations (see, e.g., Johnston 1984, p. 479), which can be obtained by substituting the expression for $\hat{\mathbf{y}}^*$ from equation (11) into equation (14). We will compare the asymptotic variance matrix (14) of the 2SLS estimator (12) with that of Ord's (1975) ML estimator.

For this purpose, it is useful to transform expression (14) into another form. Note first that the observed population-potential vector $\mathbf{y}^*$ can be written as the sum of its predicted value from the first-stage regression, $\hat{\mathbf{y}}^*$, and the OLS error vector, $\mathbf{v}$:

$$\mathbf{y}^* = \hat{\mathbf{y}}^* + \mathbf{v}. \tag{15}$$

Furthermore, premultiplying each side of this expression by its own transpose yields

$$\begin{aligned} \mathbf{y}^{*\prime}\mathbf{y}^* &= (\hat{\mathbf{y}}^* + \mathbf{v})'(\hat{\mathbf{y}}^* + \mathbf{v}) \\ &= \hat{\mathbf{y}}^{*\prime}\hat{\mathbf{y}}^* + \hat{\mathbf{y}}^{*\prime}\mathbf{v} + \mathbf{v}\hat{\mathbf{y}}^{*\prime} + \mathbf{v}'\mathbf{v} \\ &= \hat{\mathbf{y}}^{*\prime}\hat{\mathbf{y}}^* + \mathbf{v}'\mathbf{v}, \end{aligned} \tag{16}$$

[6]No correction for loss of degrees of freedom is applied to the 2SLS estimator of the error variance because its known properties refer to large samples for which such corrections have no appreciable effect.

by the usual properties of OLS residuals.

Equation (16) can be substituted for the upper-left-hand submatrix in the 2SLS variance matrix (14). Then the expectation operator can be applied to asymptotic distributions of the random components, the generalized population-potential vectors, $\mathbf{y}^*$, and their predicted values, $\hat{\mathbf{y}}^*$, and functions thereof. This yields, for instance, $\mathrm{E}(\mathbf{X}_1'\hat{\mathbf{y}}^*) = \mathrm{E}(\mathbf{X}_1'\mathbf{y}^*) = X_1'\mathrm{E}(\mathbf{y}^*)$, where the first equality follows from properties of OLS residuals and the second from factoring the fixed exogenous variables (constants) out of the expectation. This results in the following equivalences for the asymptotic variance matrix:

$$\begin{aligned}
\text{asy var}\begin{bmatrix} r \\ \mathbf{b} \end{bmatrix} &= \sigma_\epsilon^2 \begin{bmatrix} \hat{y}^{*\prime}\hat{y}^* & \mathbf{X}_1'\mathrm{E}(\hat{\mathbf{y}}^*) \\ \mathbf{X}_1'\mathrm{E}(\hat{\mathbf{y}}^*) & \mathbf{X}_1'\mathbf{X}_1 \end{bmatrix}^{-1} \\
&= \sigma_\epsilon^2 \begin{bmatrix} \mathrm{E}(\mathbf{y}^{*\prime}\mathbf{y}^* - \mathbf{v}'\mathbf{v}) & \mathbf{X}_1'\mathrm{E}(\mathbf{y}^*) \\ \mathbf{X}_1'\mathrm{E}(\mathbf{y}^*) & \mathbf{X}_1'\mathbf{X}_1 \end{bmatrix}^{-1} \\
&= \sigma_\epsilon^2 \begin{bmatrix} \mathrm{E}(\mathbf{y}^{*\prime}\mathbf{y}^*) - \mathrm{E}(\mathbf{v}'\mathbf{v}) & \mathbf{X}_1'\mathrm{E}(\mathbf{y}^*) \\ \mathbf{X}_1'\mathrm{E}(\mathbf{y}^*) & \mathbf{X}_1'\mathbf{X}_1 \end{bmatrix}^{-1} \\
&= \sigma_\epsilon^2 \begin{bmatrix} \mathrm{E}(\mathbf{y}^{*\prime}\mathbf{y}^*) - \sigma_v^2 & \mathbf{X}_1'\mathrm{E}(\mathbf{y}^*) \\ \mathbf{X}_1'\mathrm{E}(\mathbf{y}^*) & \mathbf{X}_1'\mathbf{X}_1 \end{bmatrix}^{-1}
\end{aligned} \tag{17}$$

where σ_v^2 denotes the asymptotic variance of the error term of the first-stage regression of $\mathbf{y}^*$ on $\mathbf{X}$.

To compare expression (17) with the asymptotic variance matrix of the ML estimators (2) and (4) for model (1), we cannot use the final expression derived by Doreian (1981, p. 384), because it is written in terms of the parameters of model (1) rather than model (7). Stepping back in his derivations a few paces and using Ord's (1975) asymptotic expectation notation, however, we get the following equivalent expression:

$$\text{asy var}\begin{bmatrix} \hat{\rho} \\ \hat{\boldsymbol{\beta}} \end{bmatrix} = \sigma_\epsilon^2 \begin{bmatrix} \mathrm{E}(\mathbf{y}^{*\prime}\mathbf{y}^*) - \alpha\sigma_\epsilon^2 & \mathbf{X}_1'\mathrm{E}(\mathbf{y}^*) \\ \mathbf{X}_1'\mathrm{E}(\mathbf{y}^*) & \mathbf{X}_1'\mathbf{X}_1 \end{bmatrix}^{-1}, \tag{18}$$

where $\alpha = \partial^2 \ln|\mathbf{A}|/\partial\rho^2 = \Sigma_{i=1}^n(\lambda_i^2)/(1-\hat{\rho}\lambda_i)^2$, $|\mathbf{A}| = |\mathbf{I} - \rho\mathbf{W}|$, $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of the weight matrix W, and

$$\hat{\sigma}_\epsilon^2 = \frac{(\mathbf{y} - \hat{\rho}\mathbf{W}\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \hat{\rho}\mathbf{W}\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n}$$

is an alternative expression of the ML estimator (3) of the error variance of model (1).

Comparing the submatrix components of the asymptotic variance-covariance matrices (17) and (18), we see that three of them—namely, those in the lower-diagonal and off-diagonal positions—are identical. The first of these three, the $\mathbf{X}_1'\mathbf{X}_1$ component, is the usual matrix of cross products of the fixed exogenous variables, which enters into the calculation of the variances and covariances of the estimators of the regression coefficient vector, $\boldsymbol{\beta}$, as in the conventional linear model. The off-diagonal elements in (17) and (18), $\mathbf{X}_1'\mathrm{E}(\mathbf{y}^*)$, similarly enter into the calculation of the covariances of the estimators of the regression and spatial-effects coefficients (i.e., the elements of $\boldsymbol{\beta}$ and ρ). These submatrices have computational expressions that differ, because the parameterizations of model (1) and model (7) differ. When they are written in terms of the respective asymptotic expected values of the two models, as in expressions (17) and (18), however, we can see that they are indeed identical.

It is more difficult to compare the last component submatrices of the two asymptotic variance matrices (the upper-diagonal term), which represent the asymptotic variances of the respective estimators of the spatial-effects coefficient ρ. Again, because of the different parametric representations of models (1) and (7), these elements differ, even when they are written in terms of expected values of asymptotic distributions, as in (17) and (18). Nonetheless, they do have a common component—namely, the asymptotic expected value $\mathrm{E}(\mathbf{y}^{*\prime}\mathbf{y}^*)$. But the 2SLS expression (17) subtracts from this the asymptotic variance of the error term of the first-stage regression of the 2SLS estimator, whereas the ML expression (18) subtracts a function of the asymptotic variance of the error term of the original model (1). The former is the error variance from the first-stage regression of the weighted $\mathbf{y}^* = \mathbf{W}\mathbf{y}$ on the full set of exogenous variables in $\mathbf{X}$, while the latter error variance is scaled by the second partial derivative with respect to the spatial-effects parameter ρ of the log of the determinant of the matrix $\mathbf{A} = \mathbf{I} - \rho\mathbf{W}$, which is one component of the loglikelihood function of the ML estimator of model (1). Informally, the scaling of the ML error variance by this second partial derivative is necessary to transform it to the scale of the error variance of the first-stage regression of the weighted $\mathbf{y}^* = \mathbf{W}\mathbf{y}$ on $\mathbf{X}$ in the 2SLS estimator.

Investigations comparing 2SLS and ML estimators of the parameters of single equations in classical simultaneous equation models have shown the equivalence of the variances of their asymptotic distributions (see, e.g., Theil 1971, pp. 500–507). The conditions necessary for the proof of this equivalence pertain to model accuracy (no misspecification) and stability of distributions of relevant variables. Because the spatial-effects model analyzed here is a special case of the classical model, asymptotic equivalence of the upper-diagonal components of expressions (17) and (18) follows under similar conditions.

6. COMPARISON WITH ANSELIN'S TWO-STAGE LEAST SQUARES ESTIMATORS

Anselin (1988) developed two 2SLS estimators in spatial econometrics models, one of which is identical to that developed herein.[7] Therefore, we call the estimator developed above and by Anselin the *Land-Deane/Anselin 2SLS estimator.* Anselin also described an alternative 2SLS estimator that is not identical to the estimator developed here. We call this *Anselin's alternative 2SLS estimator.* In this section we describe conceptual differences between the two estimators. In the next section, we investigate some numerical differences in a particular empirical application and discuss the relative computational efficiencies of the two estimators.

Anselin (1988, pp. 82–87) described a general instrumental variables (IV) approach to estimation that permits the specification of instruments for all variables in the regression model (7), i.e., for both the $\mathbf{X}$ and $\mathbf{y}^*$ variables. In practice, of course, the focus is on the $\mathbf{y}^*$ variable, which Anselin called a spatially lagged dependent variable. Anselin (1988, p. 84) showed that his general IV estimator can subsume an alternative 2SLS estimator in which $\mathbf{y}^*$ in the second-stage OLS regression is replaced by a transformation of the predicted values of the dependent variable $\mathbf{y}$ from a first-stage regression on a

[7]Anselin (1988), however, did not motivate the derivation of the 2SLS method in terms of the population-potential notions of spatial sociology, he did not give a detailed comparison of the variance-covariance matrix of the 2SLS estimator with that of the ML estimator, and he did not present numerical and computational efficiency comparisons of 2SLS and ML.

fixed set of exogenous variables, where the weight matrix **W** is used to perform the transformation.

Formally, Anselin's alternative 2SLS estimator proceeds by OLS estimation of the reduced form of model (7), in which the effects of the $\mathbf{y}^*$ variable are factored out and the dependent variable **y** is taken as dependent only on the fully exogenous variables **X**. The predicted values of **y** from this first-stage regression are given by $\hat{\mathbf{y}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. Anselin's alternative estimator then transforms these predicted values by the weight matrix **W** to yield the following:

$$\tilde{\mathbf{y}}^* = \mathbf{W}\hat{\mathbf{y}} = \mathbf{W}[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}], \tag{19}$$

where we use the notation $\tilde{\mathbf{y}}^*$ to distinguish these predicted values for $\mathbf{y}^*$ from those introduced in equation (11). Anselin's alternative 2SLS estimator then is completed by inserting these predicted values into a second-stage OLS regression, as in equation (12).

Comparing the right sides of equations (11) and (19), we can see, first, that the predicted variables of the two 2SLS procedures differ in the way in which the weight matrix **W** enters. In equation (11), we enter **W** into the first-stage regression by scaling the dependent variable (regressand) of the regression. In equation (19), we enter **W** after the first-stage regression by scaling the predicted values computed after the regression of **y** on **X**. In this alternative procedure, the predicted value of $\mathbf{y}^*$ is not a linear combination of **X**, because it is premultiplied by **W**, which varies across observations, so there is no perfect collinearity between **X** and the instrument.

Accordingly, if we use this alternative 2SLS estimator, we can avoid the exclusion restrictions (which can be hard to justify) required by the estimator of equation (11). On the other hand, equation (19) requires that we premultiply the predicted values of y by **W**, which can yield a very large matrix in large samples. Indeed, standard computer packages (e.g., SAS) typically require that one save the predicted values of y from the first stage, premultiply these by **W** in an auxiliary routine (e.g., in SAS/IML), and then move back into a regression program. Alternatively, one can trick SAS into performing matrix multiplication.[8] This, of course, means that the 2SLS

[8]The numerical application of Anselin's method described in the next section was programmed in both ways. The code used to perform the analyses in SAS without resort to an auxiliary routine is available from the authors on request.

estimator of equation (19) in the spatial-effects model, like the ML estimator, is not easily applicable in large samples, which is a primary motivation for the 2SLS estimator developed herein.

7. EMPIRICAL APPLICATION AND COMPARISON OF ESTIMATORS

The 2SLS estimator of equations (11), (12), and (14) was developed by Land, Deane, and Blau (1991) in a study of the effects of religious pluralism and social conditions on church membership in the U.S. in the early decades of this century. The debate on the effect of religious pluralism has recently been enlivened by Finke and Stark (1988, 1989) and Breault (1989*a*,*b*). Their findings, however, are contradictory: Finke and Stark reported that religious pluralism has a positive effect on the rate of church membership (church adherence), but Breault found a negative effect.

Land et al. (1991) recently entered this debate with an analysis of church adherence rates in U.S. counties or county groups for three decennial census periods in the early twentieth century: 1910, 1920, and 1930. County-level analyses, especially those that use data from the period before World War II, are plagued by problems of boundary comparability, because counties may split, merge, or emerge in places where there had been no internal political boundaries. To overcome this difficulty, Land et al. grouped counties involved in any kind of boundary shift with other counties implicated in the boundary change by using a county longitudinal template compiled by Horan and Hargis (1989). Consequently, the unit of analysis is a county or county group, depending on whether or not boundaries changed.

The Horan and Hargis (1989) template includes approximately 1,600 counties and county groups, from which Land et al. (1991) drew a 50 percent random sample stratified by percentage urban in 1910: 0 percent, 1 percent to 22 percent, 23 percent to 40 percent, and over 40 percent. Initial statistical analyses and models were estimated on this 50 percent sample. To assess whether inferences from these analyses were unduly influenced by statistical overfitting on this initial sample, Land et al. used the remaining 50 percent sample of county groups for replication analyses.

Explanatory variables in the Land et al. (1991) study were

aggregated to county groups from the Inter-University Consortium for Political and Social Research (ICPSR, no date) collection of demographic, economic, and social data based on county-level counts from the U.S. decennial censuses of 1910, 1920, and 1930. Church membership data were obtained from ICPSR files reporting the 1906, 1916, and 1926 censuses of religious bodies. Although the religious censues were carried out four years before each of the population censuses, this is not considered problematic because of the high stability of church membership counts for geographical units over time (Bainbridge and Hatch 1982, p. 244; Christiano 1987, p. 40). For convenience, we describe the analyses in terms of the census dates.

Variables in the Land et al. (1991) study were defined as follows. The dependent or endogenous variable to be explained is the percentage of the county group's population that are church members, as recorded in the religious censuses. The explanatory or exogenous variables include the county group's religious diversity, the percentage of the county group's population that is Catholic, ethnic diversity, population change (computed as the percentage change in total population size from the previous decade), the percentage of the population that is living in an urban area, the percentage of the adult population that is illiterate, and the average county crop value in 1910, which was used as an indicator of economic well-being. The religious and ethnic diversity variables were measured by applying the Gibbs-Martin index (Christiano 1987; Lieberson 1969; Gibbs and Martin 1962) to sets of religious denominations and ethnic groups, respectively.

In addition, Land et al. (1991) included a variable for *church-adherence potential,* which is like the generalized population-potential variables described above: It is a product of a geographical distance weight matrix **W** and a vector of county adherence rates. This generalized potential variable was introduced to capture the impact on the rate of church adherence in a given county of higher or lower rates of adherence in neighboring and more-distant counties. Land et al.'s (1991) collective influence/social diffusion interpretation of the church-adherence potential is based on the models of Granovetter (1978) and Granovetter and Soong (1983). The presence of the potential variable combined with relatively large samples (on the order of 700, depending on the decade and the number

of counties with missing data) and the inability of the mainframe computer available for regression analyses to handle weight matrices larger than 300 rows by 300 columns limited the applicability of the ML estimators described above. Accordingly, Land et al. (1991) applied the 2SLS estimators (11), (12), and (14) of model (7).

To illustrate the numerical application of the 2SLS estimators derived above and compare them with Anselin's alternative 2SLS estimators (based on equation (19) in the first stage) and with the ML estimators, we draw upon the Land et al. (1991) analysis for a random subsample of 300 county groups for 1910. (Corresponding calculations for 1920 and 1930 are available from the authors on request.) The number of cases in this subsample is limited only by the aforementioned limits of the mainframe computer.

Table 1 reports the first-stage regression (based on equation (11)) of the church-adherence potential/spatial-effects variable on a set of fully exogenous variables. In addition to the explanatory variables listed above, this set of exogenous variables includes (as instrumental variables) eight regional dummy variables (the Northeast region is the baseline region against which the effects of the other regions are estimated) and the county group's total population size (logged). The coefficient of determination, 0.68, shows that the first stage is quite successful in determining the values of the potential variable for county groups. In particular, several of the regional control variables are significantly related to the dependent variable, although the total population variable does not reach statistical significance. Our experience with applications of the 2SLS method to

TABLE 1
First-Stage Least Squares Estimates for Model of Church Adherence Rates; Random Sample of County Groups, 1910 (Metric Coefficients, Standardized Coefficients in Brackets, *t* Ratios in Parentheses

Intercept	13.92
	(3.22)**
Religious diversity	2.50
	[.07]
	(1.37)
Percentage Catholic	−.03
	[−.05]
	(−1.03)

Log total population	.29
	[.05]
	(1.08)
Ethnic diversity	−4.16
	[−.08]
	(−2.17)*
Percentage urban	−.02
	[−.07]
	(−1.39)
Log percentage illiterate	.03
	[.01]
	(.08)
Log economic conditions[a]	.95
	[.09]
	(1.95)
Middle Atlantic[b]	1.00
	[.04]
	(.82)
South Atlantic[b]	1.78
	[.10]
	(1.39)
East South-Central[b]	1.98
	[.08]
	(1.39)
West South-Central[b]	−5.32
	[−.23]
	(−4.01)**
East North-Central[b]	6.02
	[.37]
	(5.55)**
West North-Central[b]	.56
	[.04]
	(.49)
West Mountain[b]	−10.81
	[−.40]
	(−7.55)**
Pacific[b]	−14.16
	[−.41]
	(−9.21)**
N	287
R^2	.68

[a]Average crop value

[b]The reference category for regional dummy variables is Northeast.

*$p < .05$

**$p < .01$

these data, however, and the fact that potential variables will often capture regional or population size effects on generalized potential variables suggest that these are useful instrumental variables for entry in the first-stage regressions.

Table 2 reports the second-stage regressions of the county-group church-adherence rates on the predicted values of the church-adherence potential/spatial-effects variable and the set of explanatory regressors identified above. For Land-Deane/Anselin's and Anselin's alternative 2SLS estimators (recall that the former are based on equation (11) in the first-stage regressions and that the latter use equation (19) in the first stage), the estimated 2SLS metric regression coefficients, standardized coefficients, and t ratios are given in the second and third columns of the table, respectively. Sample sizes after loss of cases due to incomplete data and measures of fit are given at the bottom of the table.[9] For comparison, the corresponding ML estimates are reported in the fourth column. To assess the extent to which the spatial-effects corrections of each of these reduces bias in the estimated coefficients and increases explanatory power, we also display the OLS estimates of a nonspatial model (i.e., a regression model that ignores the spatial-effects term) in the first column of the table.

Several comments concerning these results are in order. First, the OLS estimates for the nonspatial model are clearly inferior to the 2SLS and ML estimates. In particular, the magnitude of the coefficient of religious diversity and its t ratio (the key variable in the debate on the effects of religious diversity described above) are greatly underestimated by the nonspatial OLS estimators. Coefficients and t ratios of several of the other variables also are substantially affected by omission of the spatial term. Specifically, the coefficients tend to be inflated in the nonspatial OLS model, while the t ratios tend to be deflated (see Doreian 1981).

Second, the magnitudes of the metric regression coefficient estimates from both 2SLS methods are generally quite close to those of the ML estimates. The only exception is the spatial-effects coeffi-

[9]The measures of fit reported in Table 2 are the squares of the correlation between $\mathbf{y}$ and the fitted values $\hat{\mathbf{y}}$ in each model. While this appears similar to the standard coefficient of determination of regression models, it cannot strictly be interpreted as the proportion of variance explained due to the interdependence of the observations (Doreian 1981, p. 368).

TABLE 2
Alternative Estimates for Model of Church Adherence Rates; Random Sample of County Groups, 1910. Metric coefficients, standardized coefficients [in brackets], and t-ratios (in parentheses).

		2SLS		
	OLS	Land-Deane Anselin	Anselin Alternative	ML
Intercept	60.75	41.82	41.16	42.13
	(6.09)**	(4.92)**	(4.74)**	(5.04)**
Spatial effects	—	.94	1.03	.92
	—	[.49]	[.47]	[.48]
	—	(11.10)**	(10.43)**	(11.08)**
Religious diversity	−9.31	−22.17	−22.52	−21.96
	[−.14]	[−.32]	[−.33]	[−.32]
	(−2.31)*	(−6.23)**	(−6.16)**	(−6.27)**
Percentage Catholic	.36	.38	.36	.38
	[.33]	[.34]	[.33]	[.34]
	(5.59)**	(7.09)**	(6.60)**	(7.17)**
Population change	−2.10	−1.39	−1.47	−1.40
	[−.20]	[−.13]	[−.14]	[−.13]
	(−4.20)**	(−3.30)*	(−3.42)**	(−3.37)*
Ethnic diversity	−25.24	−14.85	−15.72	−15.02
	[−.27]	[−.16]	[−.17]	[−.16]
	(−4.74)**	(−3.66)**	(−3.81)**	(−3.76)**
Percentage urban	.08	.07	.08	.07
	[.15]	[.15]	[.16]	[.14]
	(2.86)*	(3.28)*	(3.54)**	(3.18)*
Log percentage illiterate	2.14	1.66	1.79	1.67
	[.19]	[.15]	[.16]	[.15]
	(3.11)*	(2.88)*	(3.06)*	(2.94)*
Log economic conditions[a]	−1.23	−1.18	−1.16	−1.18
	[−.06]	[−.05]	[−.05]	[−.05]
	(−1.02)	(−1.17)	(−1.14)	(−1.19)
N	287	287	287	287
R^2	.39	.58	.56	.58

[a]Average crop value
*$p < .01$
**$p < .001$

cient, for which Anselin's alternative 2SLS point estimate is considerably larger than the ML and Land-Deane/Anselin 2SLS estimates. Both the Land-Deane/Anselin 2SLS and the ML estimators suggest strong spatial-diffusion effects, with highly statistically significant metric regression coefficients of 0.94 and 0.92, respectively. By comparison, Anselin's alternative 2SLS estimate of this coefficient is greater than one, a value that certainly indicates strong spatial effects but that also differs substantially from the ML estimate.

Third, the Land-Deane/Anselin 2SLS metric regression coefficient estimates for the other regressor variables are uniformly closer to the ML estimates than the Anselin alternative 2SLS estimates. The former often agree with the ML estimate to two or three digits.

Fourth, the Land-Deane/Anselin 2SLS method yields t ratios that are closer to those of the ML estimates than are the Anselin alternative estimates for seven of the eight regressors. The measure of fit (see footnote 9) of the Land-Deane/Anselin estimates is also identical to that of the ML estimates, while that of the Anselin alternative estimates differs slightly.

In addition to the greater accuracy and statistical efficiency of the 2SLS and ML estimators compared with nonspatial OLS, the Land-Deane/Anselin 2SLS estimator is much more computationally efficient than both the Anselin alternative 2SLS estimator and the ML estimator. Specifically, as noted above, the Land-Deane/Anselin 2SLS estimator can handle a wider range of sample sizes in standard computer packages than the other two estimators. This, of course, affects the amount of computer processing time required to compute the estimates. In particular, for calculations executed in SAS batch mode on an IBM 3090 mainframe computer with a vector processor, the ML estimates in Table 2 required 43.08 seconds of CPU time, about 55 times more than the Land-Deane/Anselin 2SLS estimates, which required 0.78 second. The Anselin alternative 2SLS estimates required 16.36 seconds, about 21 times more than the Land-Deane/Anselin estimates.

8. CONCLUSION

Sociologists and other social scientists will undoubtedly continue to analyze data that are aggregated into arbitrary spatial configurations. In such analyses, incorporating spatial-effects terms into

regression models is always an issue. The empirical results reported here and in Doreian (1981) indicate that when spatial effects are substantial and statistically significant, inferences can be greatly affected by the exclusion of spatial regression terms. Accordingly, analysts are well advised to take spatial-effects terms explicitly into account, especially when they can be substantively interpreted in terms of collective influences, social diffusion, or other processes.

Similar recommendations apply to analyses of data collected from individuals or organizations connected in well-defined (social, exchange, authority, communication, transportation, etc.) networks—when data on the network interrelationships (specified in the **W** matrix) are available. In such cases, the spatial-effects term of the spatial regression model is a network-effects term, and failure to estimate such effects in a regression model can have similarly deleterious effects on coefficient estimates and inferences (see, e.g., Duke's [1991] study of network effects in a high school student body).

Major limiting factors are the relatively computationally burdensome extant ML procedures and their sample-size limitations. In this paper, we have derived a corresponding 2SLS estimator for the spatial-effects regression model that is applicable to large samples. We have demonstrated that this estimator, which is equivalent to a 2SLS estimator previously derived in spatial econometrics by Anselin (1988), has desireable statistical and computational properties. Unlike the ML estimator, the 2SLS estimator does not require the assumption of normal random errors and yet is consistent and computationally efficient.

We also compared this 2SLS estimator with Anselin's alternative 2SLS estimator for the spatial regression model. In an empirical application to a moderately large sample, we found that Anselin's alternative 2SLS estimator is somewhat less computationally efficient. The Land-Dean/Anselin estimator also produced coefficient estimates closer to the ML estimates than Anselin's alternative 2SLS estimator, but evidence from other empirical applications is necessary before any generalizations can be stated. Based on the empirical analyses we have done and an empirical application[10] to data on electoral support for the Democratic presidential candidate in the 64

[10]This application was graciously provided by Peter V. Marsden.

Louisiana parishes analyzed in Doreian's (1981) article, we believe that the Land-Deane/Anselin method will perform relatively well in large samples when good exogenous identifying variables are available as predictors for the first-stage regressions.[11]

As noted earlier, the Land-Deane/Anselin 2SLS method requires such identifying variables, but Anselin's does not. Thus, we believe that when such identifying variables are available and sample sizes are large (say, 200 or more cases), the method has much to recommend it. When sample sizes are large but exogenous identifiers are unavailable, the analyst must cope with the computational complications of ML or Anselin's alternative 2SLS. In the case of small samples with or without identifying exogenous variables, ML may be the preferred method, provided that the errors are reasonably well approximated by the normal distribution assumption.

Given that appropriate estimators are now available for these various contingencies, future regression analyses of spatially distributed or network-related social data may be more likely to estimate the effects of spatial- or network-effects terms. All existing methods have been developed for the estimation of a single linear (in parameters) equation. Extensions to more-complicated error processes, nonlinear equations, and systems of simultaneous linear or nonlinear equations have yet to be fully explored.

REFERENCES

Anselin, Luc. 1988. *Spatial Econometrics: Methods and Models.* Dordrecht, The Netherlands: Kluwer Academic Publishers.

Bainbridge, William Sims, and Laurie Russell Hatch. 1982. *The Churches and the American Experience.* Grand Rapids, MI: Baker Book House.

[11]The proper choice of instruments with which to construct predicted values in the first stage is always problematic in two-stage and instrumental variables estimators. Although these estimators are consistent, their efficiency is effectively only asymptotic (i.e., applicable to large samples) and depends on the exogenous variables used as regressors in the first stage. The general textbook advice is that these variables should be highly correlated with the endogenous variable for which predicted values are sought and uncorrelated with the error term of the model. In practice, we believe that the former criterion means that the coefficient of determination obtained in the first-stage regression should be reasonably large in comparison to the amount of association in the variables being analyzed. For instance, if the variables being analyzed yield a measure of fit in the second-stage regression of approximately 0.5, then the first-stage regression should produce a coefficient of determination of greater magnitude.

Basmann, R. L. 1957. "A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation." *Econometrica* 25:77–83.

Breault, Kevin D. 1989*a*. "New Evidence on Religious Pluralism, Urbanism, and Religious Participation." *American Sociological Review* 54:1048–53.

———. 1989*b*. "Reply to Finke and Stark." *American Sociological Review* 54:1056–59.

Christiano, Kevin J. 1987. *Religious Diversity and Social Change*. Cambridge: Cambridge University Press.

Doreian, Patrick. 1980. "Linear Models with Spatially Distributed Data: Spatial Disturbances or Spatial Effects?" *Sociological Methods and Research* 9:29–60.

———. 1981. "Estimating Linear Models with Spatially Distributed Data." Pp. 359–88 in *Sociological Methodology 1981*, edited by Samuel Leinhardt. San Francisco: Jossey-Bass.

Doreian, Patrick, and Norman P. Hummon. 1976. *Modeling Social Processes*. New York: Elsevier.

Doreian, Patrick, Klaus Teuter, and Chi-Hsein Wang. 1984. "Network Autocorrelation Models: Some Monte Carlo Results." *Sociological Methods and Research* 13:155–200.

Duke, James Brian. 1991. "The Peer Context and the Adolescent Society: Making Sense of the Context Effects Paradox." Ph.D. diss., Harvard University, Department of Sociology.

Duncan, Otis Dudley, Ray P. Cuzzort, and Beverly A. Duncan. 1963. *Statistical Geography*. New York: Free Press.

Finke, Roger, and Rodney Stark. 1988. "Religious Economies and Sacred Canopies." *American Sociological Review* 53:41–49.

———. 1989. "Comment on Breault." *American Sociological Review* 54:1054–55.

Gibbs, Jack P., and Walter T. Martin. 1962. "Urbanization, Technology, and the Division of Labor." *American Sociological Review* 27:667–77.

Granovetter, Mark. 1978. "Threshold Models of Collective Behavior." *American Journal of Sociology* 83:1420–43.

Granovetter, Mark, and Roland Soong. 1983. "Threshold Models of Diffusion and Collective Behavior." *The Journal of Mathematical Sociology* 9: 165–80.

Horan, Patrick M., and Peggy G. Hargis. 1989. "The County Longitudinal Template." Paper presented at the Annual Meetings of the Social Science History Association, Chicago.

Inter-university Consortium for Political and Social Research. n.d. Historical, Demographic, Economic, and Social Data: The United States, 1790–1970. Ann Arbor: ICPSR [producer and distributor].

Johnston, J. 1984. *Econometric Methods*. 3d ed. New York: McGraw-Hill.

Land, Kenneth C., Glenn Deane, and Judith R. Blau. 1991. "Religious Pluralism and Church Membership: A Spatial Diffusion Model." *American Sociological Review* 56:237–49.

Lieberson, Stanley. 1969. "Measuring Population Diversity." *American Sociological Review* 34:850–62.

Ord, Keith. 1975. "Estimation Methods for Models of Spatial Interaction." *Journal of the American Statistical Association* 70:120–26.

Roncek, Dennis W., and Andrew Montgomery. 1984. "Spatial Autocorrelation: Diagnoses and Remedies for Large Samples." Paper presented at the Annual Meetings of the Midwest Sociological Society, Des Moines.

Roncek, Dennis W., and James Robinson. 1984. "The Effects of Crime in the Surroundings of Residential Areas." Paper presented at the Annual Meetings of the American Society of Criminology, Cincinnati.

Theil, H. 1953*a*. *Repeated Least-Squares Applied to Complete Equation Systems*. The Hague: Central Planning Bureau.

———. 1953*b*. *Estimation and Simultaneous Correlation in Complete Equation Systems*. The Hague: Central Planning Bureau.

———. 1961. *Economic Forecasts and Policy*. 2d ed. Amsterdam: North-Holland.

———. 1971. *Principles of Econometrics*. New York: Wiley.