

## **NETWORK AUTOCORRELATION: A SIMULATION STUDY OF A FOUNDATIONAL PROBLEM IN REGRESSION AND SURVEY RESEARCH\***

Malcolm M. DOW

*Northwestern University\*\**

Michael L. BURTON and Douglas R. WHITE

*University of California, Irvine\*\*\**

It is axiomatic to the social sciences, and an essential part of the network perspective, that human performances are intricately linked with their social and environmental contexts. Researchers in each of the disciplines have rediscovered this in the past decade with respect to a whole host of specific problem areas, under such labels as “context effects”, “index utility”, and “systems analysis”. The earliest mention of the problem with respect to quantitative research occurred, to our knowledge, in the debate between the nineteenth century cultural diffusionists and the evolutionists. The latter regarded individual societies as independent instances of uniform causation, and hoped to learn about causation from correlational studies. The former regarded their observations as embedded in an interactive network of historical relationships such as diffusion, migration, conquest, and competition, where the historical, evolutionary and ecological context of each society and the network of interconnectedness between societies plays a major role in multiple causation. In this view, events cannot be regarded as

\* This research was supported by a grant from the National Science Foundation to Michael Burton and Douglas White. The two Principal Investigators made major and equal contributions to this paper. We are grateful to Linton Freeman, Patrick Doreian, and Karl Reitz for their critical comments on this paper.

\*\* Northwestern University, Evanston, IL 60201, U.S.A.

\*\*\* University of California, Irvine, CA 92717, U.S.A.

isolated or independent as if each were a context-free “independent invention” of a single society.<sup>1</sup>

The same arguments, of course, apply to the interpretation of data collected in social or opinion surveys. Political science offers a recent example of the discovery of “context effects” in voting behavior (*e.g.* Jackson 1975). How much of voting behavior is affected by attributes of the voting unit (whether individuals or aggregates), and how much is the result of interactions between them: of the communication process, bandwagon effects, reference group behavior, or other forms of “symbolic interactionism”?

Our purpose in this paper, however, is not to attempt a review of the vast literature on context effects. Rather, we focus on the costs and benefits of either neglecting context or else incorporating it in the research design. Statistical methods such as multiple regression analysis necessarily contain mathematical axioms which either assert or deny the existence of context effects. We will explore here through simulation studies the following related questions:

- (1) What are the consequences of ignoring context effects, should they be present, or ordinary least squares regression estimates, and
- (2) what are some of the properties of a recently developed maximum likelihood procedure which permits context effects to be included in a regression model as network autocorrelated disturbance terms?

Autocorrelation is the technical term which means, within the regression framework, that some variable, or the error term, is correlated with itself, either directly or indirectly over time, through space or across a network. Temporal autocorrelation is a fairly common occurrence in time series analysis, and methods to deal with it have been developed for some time (Hibbs 1974). Spatial autocorrelation, on the other hand,

<sup>1</sup> In his classic paper Tylor (1889) marshalled data on over 350 preindustrial societies to illustrate the kinds of cultural “adhesions” (*i.e.* functional relationship) that he found between, for example, postmarital residence and descent reckoning. Commenting on Tylor’s paper, Francis Galton raised the possibility that due to various processes of cultural diffusion, some of the cases cited by Tylor might be “duplicate copies of the same original,” and that the independence of Tylor’s sample observations was thus in doubt. For more than half a century cross-cultural research was held in disrepute in anthropology because of the vulnerability of any findings to this criticism. Over the past 25 years, however, “Galton’s Problem” has become a major area of methodological concern in anthropology, and many different solutions have been suggested (see Naroll 1976; Schaefer 1974; Dow, White and Burton 1981; White, Burton and Dow 1981 for further discussion of Galton’s Problem).

has only very recently begun to receive serious attention (Cliff and Ord 1973; Ord 1975; Hepple 1976; Doreian 1980, 1981). Generalization of the one-dimensional time-series approach to a two-dimensional situation is not straightforward, and new analytical and computational problems arise with these more general models. A number of these difficulties have recently been overcome, and several models which allow different specifications of spatial autocorrelation are now within computational reach. As a result of our experience with previous research employing two of these models we focus on one of them in this paper, the disturbances model, where a network autocorrelation scheme is embedded in the error term of a multiple regression model.

Although the model we explore in this paper was developed within the context of spatially distributed and autocorrelated data, we refer to this class of models generally as “network” autocorrelation models (Dow, White and Hansen 1979; White, Burton and Dow 1981). As we discuss in a later section, the kinds of data structure which require the use of these models are, as we suggested above, common in the social sciences, particularly when “context” is a substantively relevant research concern.

### **The Foundational Problem**

If autocorrelated “context effects” are a pervasive feature of naturally occurring social phenomena, so that the cases under observation are non-independent, in what sense does this represent a foundational problem for testing hypotheses or drawing inferences from correlational or regression analysis? The ordinary least squares approach to regression generally assumes independence of cases, and this assumption helps to maintain the additional assumption that the regression residuals are also independent, *i.e.* not autocorrelated. If the cases under study are in fact interdependent and the residuals autocorrelated, then there are two main consequences for OLS estimation. First, the ordinary least square (OLS) estimates of the  $\beta$ s are unbiased, but they are highly dispersed relative to those obtained using the alternate maximum likelihood (ML) estimation procedures. Second, OLS estimates of the sampling variances of the  $\beta$ s will generally underestimate the true variances. Hence, with respect to estimates based on a single sample, underestimating the variances of the regression coefficients will lead to

spurious attribution of significance to particular independent variables. On the other hand, in replication studies across several interdependent samples, the investigator would tend to conclude that a valid model fails to replicate because of large differences in the magnitude of the same  $\beta$ s due to their unreliability. Thus, both single and multiple sample replication studies are biased towards finding differences where none exists (type I inferential error) if OLS estimation procedures are used with interdependent samples.

While the maximum likelihood estimator outlined below has the desirable properties of consistency and normality that the ordinary least squares estimator does not have when the disturbances are network autocorrelated, it nonetheless has the usual drawback of ML procedures: the desirable properties are asymptotic. It is well known that ML estimators need not retain these properties in finite, small samples. Since no exact analytical results are available for the small sample properties of the ML disturbances model, its small sample behavior has to be investigated *via* Monte Carlo methods. A major goal of the present study, then, is to use Monte Carlo methods to examine the small sample bias and efficiency of the ML network disturbances model relative to OLS procedures. That is, we are interested in whether or not the analytical differences in the two estimation strategies actually show up in small samples. Since the ML computational procedures are considerably more complex and expensive than OLS, it would be desirable to be sure that the additional costs really result in some significant gain in precision and confidence in ones estimates. Before going on to present the results of the simulation study, we first discuss some of the problems that could be expected to arise if data generated by an underlying network process are analyzed without consideration of these processes. Then we outline the ML computational procedures employed in this study.

## Specification of Context

Most physical and social phenomena are embedded within elaborate networks of interdependencies which taken together make up their entire "context". In the experimental sciences it is often possible to gain control over this complexity through randomization and attempts to isolate effects, but this is usually not possible with naturally occurring

social or behavioral phenomena. Since the obvious complexity of the total context surrounding any interesting social phenomena precludes any attempt to include the “whole” context in the analysis, it is necessary to direct attention to selected aspects of contextual interaction, usually through simplifying assumptions about the substantive nature of the process under study.

The principal contextual interaction which we examine in this paper is one where sample units are in some way differentially connected, and the connections between units specific differential interactions between them. When interacting units tend to become either more alike – through diffusion, contagion, imitation, assimilation, cooptation, convergent competition or a host of other processes – or more dissimilar – through repulsion, divergent competition, differentiation, *etc.* – as a result of interaction, we have a particular kind of contextual effect resulting from the network of relations between sample units rather than from their attributes.

Typically, however, purely “local” effects such as might be specified in a multiple regression framework in terms of independent and dependent variable occur together with “contextual” or “interactive” effects. For example, sample units may be affected by units to which they are connected (the “contextual effect”), while in addition there are local effects in the sense that one variable in the set is affected by changes in the other variables. In a multiple regression framework specifying regression coefficients for the independent variables in predicting the dependent variable, the contextual or interactive effect will show up in the presence of autocorrelated error terms (“disturbances”) over the network of interactive connections between the units.

This type of autocorrelated or interactive “disturbance” in addition to local regression effects is common in many different types of social survey studies, including cross-cultural studies. A number of agricultural variables, for example, may tend to diffuse together as a cultural “packet,” yet in addition certain of these variables may tend to affect others on a purely local basis. Wherever there is significant interaction between sample units which affects a packet of interdependent variables, a network disturbances model is appropriate.<sup>2</sup>

<sup>2</sup> Erbring and Young (1979) provide an excellent discussion of the theoretical and methodological issues surrounding the “contextual effects” debate over the correct model specification of academic achievement as a function of ability plus ability of academic peers. They suggest that the second type of autocorrelation model, the “endogenous feedback” or “effect” model is more

Probably the most common approach to this problem with respect to data distributed over space is to assume a first order Markov scheme for interactive effects based on the notation of spatial adjacency (Cliff and Ord 1973; Hepple 1976). Given a collection of  $N$  mutually exclusive subregions which exhaustively subdivide a large geographical region, it is straightforward to construct an  $N \times N$  binary matrix  $C$  with elements

$$c_{ij} = \begin{cases} 1 & \text{if regions } i \text{ and } j \text{ are contiguous,} \\ 0 & \text{if regions } i \text{ and } j \text{ are not contiguous or } i = j. \end{cases}$$

This zero-one adjacency matrix can then be converted to a matrix of weights,  $W$ , by dividing each  $c_{ij}$  by its row sum. That is, a weighting matrix  $W$  is formed using as weights  $w_{ij} = c_{ij} / \sum_j c_{ij}$ . Each row of the  $W$  matrix thus sums to unity, and the weights are simple proportions based on the number of adjacent regions each region has. These weights indicate the degree of probable interaction between each pair of regions.

Selection of adjacency as the relevant contextual characteristic in constructing the weighting matrix is, of course, a choice made with respect to the substantive question at hand. Doreian (1980, 1981) chose this spatial representation in his analyses of the Phillipine Huk rebellion based on prior theoretical notions concerning the ability of government or rebel forces to move troops and weaponry within an area, and the implications of this for adjacent regions. Doreian (1980) also notes that adjacency is actually a special case of "accessibility", which may in fact be the key spatial characteristic. Bodson and Peters (1975) also argue that accessibility between regions, defined by minimum transportation time between them, is the crucial process underlying their study

---

appropriate than the disturbances model for this situation, since only the dependent variable is theoretically assumed to be autocorrelated with respect to some relational network. This "effects" model can be stated as

$$Y = \rho WY + X\beta + \epsilon, \quad \epsilon \sim NID(0, \sigma^2 I).$$

Because an independent variable ( $WY$ ) is thus a function of the error variable ( $\epsilon$ ), OLS parameter estimates of this model are biased and inconsistent (Johnston 1972). Appropriate ML estimation procedures for this model similar to those examined in this paper are presented and discussed by Erbring and Young (1979) and Doreian (1981).

of the labor-demand relations among 44 Belgian arrondissements. Gattrell (1978) has similarly employed a "communications matrix" based on distance and number of telephone links to express the interrelationships among 27 Swedish towns.

Positional similarities in a social network are relevant contextual characteristics which are conceptually independent of spatial adjacencies. Social units may interact with others in their social environment in ways that are similar, and that have similar effects on their behavior, because of their positional similarities rather than their direct connections. Thus, positional similarity is a kind of mediated or "global" interaction in a system. Positional similarity of actors in a social network may be estimated from data on their social relationships, roles, or positional attributes (Burt 1980; Lorrain and White 1970; White and Reitz 1981). A weighting matrix  $W$  could easily be constructed by normalizing the positional similarity matrix.

Similarity matrices can also be derived from the overall similarity of social units or regions with respect to relevant social, demographic, political, and ecological characteristics.  $W$  matrices derived from such measures of attribute similarity can be used to test autocorrelation models (Cliff and Ord 1973). In a regression context, however, it seems that the attributes from which such similarities are derived are themselves suitable for use as variables in the regression model. Thus, it would seem that autocorrelation models of contextual effects are best suited to theories of specific network processes.

Clearly many choices can be made in specifying of context. Unfortunately, in any empirical situation the researcher usually will not know the true processes which generate the observed interdependence. Only substantive knowledge can give guidance to possibly appropriate specifications. This indeterminacy in specifying the weighting matrix  $W$  has led some investigators (Aurora and Brown 1975) to argue that this approach to the concept of autocorrelation should be abandoned in favor of other econometric techniques, such as joint generalized least squares, and random-coefficient regression models. However, these alternative procedures either require panels of observations or interaction-dependent variables (*e.g.* migration between areas) which are relatively rare even in econometrics. Since Aurora and Brown give no concrete examples of their applying procedures to real data, their suggestions remain speculative.

If the specification of a network matrix  $W$  is theoretically well

founded, its inclusion in the disturbance term of a regression model may give the researcher some insight into the nature of the underlying processes, and perhaps offer some guidance about how the models might be respecified to explicitly include relevant autocorrelated variables. As Doreian (1980) argues, some specifications of context will be more compelling and soundling based than others. To the extent that the investigator can specify relevant contexts and corresponding  $W$  matrices the procedure outlined below become applicable.

It is important to note, however, that autocorrelated errors may be present where there is not true *interactive* context effect, and the autocorrelated errors result from model misspecification. Three instances come to mind:

- (1) An autocorrelated variable has been mistakenly left out of the regression equation;
- (2) The dependent variable is nonlinearly related to the independent variable(s);
- (3) Various subgroups have been aggregated in estimating the model where in fact the regression model is different in each group.

Each of the above misspecifications is rather straightforwardly handled, and simple modifications of the regression model will in general remove the autocorrelation from the error terms. It is entirely possible, however, that the model is correctly specified even though autocorrelation is present in the errors, and there is still no underlying interactive effect. For example, if sample units have originated by dispersal from a common origin (families, schools, social classes, language communities, *etc.*) thus being positively correlated with genetically linked units over many variables (the “contextual effect” of common history), autocorrelated errors could easily arise from the pervasive autocorrelation of other variables which are not properly part of the model under study. In this case the autocorrelation is treated as a nuisance factor to be dealt with technically, but having itself no substantive import.

### Effects of Network Disturbances on OLS Regression

The ordinary least squares population regression model is

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

$$E[\epsilon] = 0, \quad E[\epsilon\epsilon'] = \sigma_\epsilon^2 I \quad (2)$$

where

$Y$  is an  $N \times 1$  column vector of dependent variable observations,  
 $X$  is an  $N \times K$  matrix of independent variable observations,  
 $\beta$  is a  $K \times 1$  column vector of regression coefficients,  
 $\epsilon$  is an  $N \times 1$  column vector of multivariate normal errors,  
 $I$  is an  $N \times N$  identity matrix.

Assume that we are given an  $N \times N$  network matrix of interdependencies,  $W$ , and that the error terms are autocorrelated with respect to this  $W$ . The OLS model can be respecified to include this matrix as follows:

$$Y = X\beta + \epsilon \quad (3)$$

$$\epsilon = \rho W\epsilon + \nu \quad (4)$$

$$E[\nu] = 0, \quad E[\nu\nu'] = \sigma_\nu^2 I \quad (5)$$

Here  $\nu$  is a vector of multivariate normal errors with mean and covariances given in equation (5).  $\rho$  is the network autocorrelation parameter which specifies the overall degree of autocorrelation in the system. It is discussed further below.

The major computational tasks associated with the disturbances model are to obtain estimates of  $\beta$ ,  $\epsilon$ ,  $\rho$  and the variances of these estimates. ML estimation procedures are outlined below; however, taking expectations of these population parameters clearly pinpoints the problems that arise if the model corresponding to equations (1) and (2) is equations estimated when the data are generated by processes corresponding to equations (3), (4) and (5).

From equation (4) we get

$$\epsilon = (I - \rho W)^{-1} \nu, \quad (6)$$

assuming that the matrix  $(I - \rho W)$  is invertible. Hence,

$$E[\epsilon] = E[(I - \rho W)^{-1} \nu] = (I - \rho W)^{-1} E[\nu] = 0$$

and the residuals from OLS have a mean of zero. The expected variance

of the residuals is obtained from

$$\begin{aligned}
 E[\epsilon\epsilon'] &= E[(I - \rho W)^{-1} \nu \nu' (I - \rho W')^{-1}] \\
 &= \sigma_v^2 (I - \rho W)^{-1} (I - \rho W')^{-1} \\
 &= \sigma_v^2 [(I - \rho W')(I - \rho W)]^{-1} \\
 &= \sigma_v^2 V.
 \end{aligned} \tag{7}$$

Since  $V$  is not an identity matrix, the OLS regression model has a non-scalar variance-covariance matrix, thus violating the usual OLS assumptions. The effects of this violation on the OLS estimation of other parameters is easily seen. The OLS estimate of  $\beta$  is obtained from

$$\hat{\beta} = (X'X)^{-1} X'Y,$$

hence,

$$\begin{aligned}
 E[\hat{\beta}] &= E[(X'X)^{-1} X'(X\beta + \epsilon)] \\
 &= E[(X'X)^{-1} X'X\beta + (X'X)^{-1} X'\epsilon] \\
 &= E[\beta + (X'X)^{-1} X'\epsilon] \\
 &= \beta + E[(X'X)^{-1} X'\epsilon] \\
 &= \beta + (X'X)^{-1} X'E[\epsilon] \\
 &= \beta.
 \end{aligned} \tag{8}$$

The OLS estimate  $\hat{\beta}$  is thus an unbiased estimate of the population  $\beta$ .

The variance of  $\hat{\beta}$  is obtained by first noting that

$$\begin{aligned}
 \hat{\beta} - \beta &= (X'X)^{-1} X'Y - \beta \\
 &= (X'X)^{-1} X'(X\beta + \epsilon) - \beta \\
 &= \beta + (X'X)^{-1} X'\epsilon - \beta \\
 &= (X'X)^{-1} X'\epsilon,
 \end{aligned} \tag{9}$$

and so

$$\begin{aligned}
 E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] &= E[(X'X)^{-1}X'\epsilon\epsilon'X(X'X)^{-1}] \\
 &= (X'X)^{-1}X'E[\epsilon\epsilon']X(X'X)^{-1} \\
 &= \sigma_v^2(X'X)^{-1}[(X'VX(X'X)^{-1})]. \tag{10}
 \end{aligned}$$

In the absence of any autocorrelation in the system, the term in the square brackets will reduce to the identity matrix and the remaining term,  $\sigma_v^2(XX)^{-1}$ , will be identical to the OLS formula for the variance of  $\beta$ . This occurs only when  $\rho = 0$ , or when there is no autocorrelation with respect to the independent  $X$  variables, since expanding the term inside the square brackets result in terms corresponding to the autocorrelation coefficients of the  $X$  variables at successive lags (Johnston 1972; Martin 1974). The extent and direction of this estimation error depends on the autocorrelation in the  $X$  variables and on  $\rho$ . Thus the bias occurring in this part of the estimation procedure could be positive or negative. If, for example, any of the  $X$  variables are positively autocorrelated with respect to  $W$ , and also  $\rho > 0$ , the term in square brackets almost certainly contains terms greater than unity and so the OLS formula will underestimate, perhaps very seriously, the true sampling variance of  $\beta$ .

A second source of error in using equation (10) above stems from using  $\sigma_v^2$  to estimate  $\sigma_\epsilon^2$  when the errors are in fact autocorrelated. The usual OLS estimator of the error variance is (Johnston 1972: 128)

$$S^2 = \frac{\hat{e}'\hat{e}}{N - K},$$

where

$$\begin{aligned}
 \hat{\epsilon} &= Y - X\hat{\beta} \\
 &= X\beta + \epsilon - X(X'X)^{-1}X'(X\beta + \epsilon) \\
 &= \epsilon - X(X'X)^{-1}X'\epsilon \\
 &= I - X(X'X)^{-1}X'\epsilon
 \end{aligned}$$

$$= M\epsilon \text{ (Where } M \text{ is symmetric idempotent),} \quad (11)$$

hence,

$$\begin{aligned} E[s^2] &= \frac{1}{N-K} E[\hat{\epsilon}'\hat{\epsilon}] \\ &= \frac{1}{N-K} E[\epsilon' M' M \epsilon] \\ &= \frac{1}{N-K} E[\epsilon' M \epsilon] \\ &= \frac{1}{N-K} E\{v' [(I - \rho W')^{-1} M (I - \rho W)^{-1}] v\} \\ &= \frac{\sigma_v^2}{N-K} E \text{tr} \{ [(I - \rho W')^{-1} M (I - \rho W)^{-1}] \} \\ &= \frac{\sigma_v^2}{N-K} E \text{tr} \{ (I - \rho W')^{-1} [I - X(X'X)^{-1} X'] (I - \rho W)^{-1} \} \\ &= \frac{\sigma_v^2}{N-K} E \text{tr} \{ [(I - \rho W')^{-1} (I - \rho W)^{-1}] \\ &\quad - (I - \rho W')^{-1} X(X'X)^{-1} X' (I - \rho W)^{-1} \} \\ &= \frac{\sigma_v^2}{N-K} E \{ \text{tr} V - \text{tr} [(X'X)^{-1} X' V X] \}. \quad (12) \end{aligned}$$

Unless the term inside the curly brackets is equal to  $N - K$ , bias will exist in estimating the error variance using OLS procedures.

The above results suggest that when autocorrelation is present, but is ignored, and the usual OLS procedure are applied, the regression coefficients will be unbiased but estimates of their sampling variances will be incorrect. In general, the estimates will be underestimates of the true variances, thus tending to result in inflated  $t$ - and  $F$ -ratios, and, hence, misleading inferences.

## Maximum Likelihood Estimation Procedures

Ordinary least squares procedures are thus problematic when the disturbances are autocorrelated according to some network scheme. Ord (1975) has examined the more general context of maximum likelihood estimation for this situation. More detailed treatments of the maximum likelihood approach are given by Hepple (1976) and Doreian (1980). The advantage of maximum likelihood estimates in this context are their properties of consistency, that is, asymptotic unbiasedness and efficiency, and asymptotic multivariate normality. Coefficients which maximize the likelihood function can be estimated and the variance-covariance matrix of these estimates can be obtained, so that valid asymptotic significance tests can be performed.

The computational algorithms employed in the simulation study reported below were developed using results stated in Ord (1975) and derived in detail by Doreian (1980).

For the network disturbances model the appropriate likelihood function is

$$L(\nu) = \left( \frac{1}{2\pi\sigma^2} \right)^{N/2} \exp \left[ - \left( \frac{1}{2\sigma^2} \right) \nu' \nu \right]. \quad (13)$$

When no network process is operative the estimates which maximize this likelihood function are identical to OLS estimates. However, there is no such equivalence in the case of network disturbances, as the following discussion shows.

Changing variables in the likelihood function from  $\nu$  to  $\epsilon$ , where  $\nu = (I - \rho W)\epsilon = A\epsilon$ , gives

$$L(\epsilon) = \left( \frac{1}{2\pi\sigma^2} \right)^{N/2} \exp \left[ - \left( \frac{1}{2\sigma^2} \right) (A\epsilon)'(A\epsilon) \right] |A|, \quad (14)$$

where  $|A|$  is the jacobian of the change of variable from  $\nu$  to  $\epsilon$ . Since the  $\epsilon$  are unobserved we make another change of variable from the  $\epsilon$  to the observed  $Y$ . After changing variables, and switching to the log-likelihood function, the function to be maximized is

$$\begin{aligned} \ell(Y) = & \text{const} - (N/2) \cdot \ln \sigma^2 \\ & - \frac{1}{2\sigma^2} [Y'A'AY - 2\beta'X'A'AY + \beta'X'A'AX\beta] + \ln |A|. \end{aligned} \quad (15)$$

The necessary conditions for maximizing the log-likelihood function are that the partial derivatives with respect to each of the unknown parameters equal zero. Thus,

$$\frac{\partial \ell(Y)}{\partial \beta} = -\frac{1}{2\sigma^2} [-2X'A'AY + 2X'A'AX\beta] = 0, \quad (16)$$

which gives

$$\hat{\beta} = [X'A'AX]^{-1} X'A'AY. \quad (17)$$

For a known  $\rho$ , this amounts to a straightforward generalized least squares procedure where  $AY$  is regressed on  $AX$ .

To obtain an estimating equation for  $\hat{\sigma}^2$ , we again differentiate the log-likelihood function:

$$\frac{\partial \ell(Y)}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4} [Y'A'AY - 2\beta'X'A'AY + \beta'X'A'AX\beta] = 0, \quad (18)$$

which gives

$$\hat{\sigma}^2 = 1/N [Y'A'AY - 2\beta'X'A'AY + \beta'X'A'AX\beta]. \quad (19)$$

Estimating equations are thus available for both  $\hat{\beta}$  and  $\hat{\sigma}^2$ , but each depends on the unknown parameter  $\rho$ . This parameter is estimated by a direct search procedure.<sup>3</sup> Ord (1975) notes that the major difficulty in implementing a direct search of the likelihood function lies with the evaluation of the determinant of the real-valued but non-symmetric matrix  $A$ , *i.e.*  $|A| = |I - \rho W|$ . This determinant would have to be evaluated at every search point over a given search interval, which would be extremely burdensome computationally. Ord suggests an alternate procedure which considerably reduces the computational ef-

<sup>3</sup> Hepple (1976) suggests various numerical optimization procedures to estimate  $\rho$ , and he employs a conjugate gradient algorithm method in this empirical example. Ord (1975) also states that some formal iterative procedure is probably desirable here. Doreian (1981), on the other hand, finds that numerical optimization is not really superior to the direct search procedure for this problem. Because of its computational simplicity, we employed a direct search procedure.

fort required. Given the eigenvalues of matrix  $W$ ,  $\{\lambda_1, \lambda_2, \dots, \lambda_N\}$  the eigenvalues of  $(I - \rho W)$  are  $\{1 - \rho\lambda_1, 1 - \rho\lambda_2, \dots, 1 - \rho\lambda_N\}$ . (Ord 1975: 121). Since the determinant of a matrix is the product of its eigenvalues

$$|A| = |I - \rho W| = \prod_{i=1}^N (1 - \rho\lambda_i). \quad (20)$$

The eigenvalues of  $W$  thus need by computed only once, and this simple product function evaluated at each iteration.

Substituting the expression for  $\hat{\sigma}^2$  into the log-likelihood function gives

$$\ln \hat{\sigma}^2 - \left( \frac{2}{N} \right) \sum_{i=1}^N \ln(1 - \rho\lambda_i), \quad (21)$$

as the function which  $\hat{\rho}$  must minimize. From (19) we see that

$$N\hat{\sigma}^2 = Y'A'AY - Y'A'AX\hat{\beta} - \hat{\beta}'X'A'AY + \hat{\beta}'X'A'AX\hat{\beta}, \quad (22)$$

and, after substitution for  $\hat{\beta}$  from equation (18) this gives, after some simplification

$$N\hat{\sigma}^2 = Y'A'PAY, \quad (23)$$

where  $P = I - (AX)[(AX)'(AY)]^{-1}(AX)'$ .

Substituting (23) into (20) yields the final expression that  $\hat{\rho}$  must minimize as:

$$\ln(Y'A'PAY) - \frac{2}{N} \sum_{i=1}^N \ln(1 - \rho\lambda_i) \quad (24)$$

$\hat{\rho}$  is located by a direct search over the interval  $|\rho| < (1/|\lambda_{max}|)$ , where  $|\lambda_{max}|$  is the greatest absolute value of the eigenvalues of  $W$  (Ord (1975). When  $W$  is row normalized to unity,  $|\lambda_{max}| < 1$ . Thus the search need only be carried out over the interval  $(-1, 1)$ . With the appropriate  $\hat{\rho}$  found from equation (24),  $\hat{\beta}$  and  $\hat{\sigma}^2$  are obtained from their estimating equations.

The next computational task is to estimate the corresponding variance-covariance matrix of these three parameter estimates. A consistent estimate of the asymptotic variance-covariance matrix is ob-

tained from the negative inverse of the matrix of second-order partial derivatives of the log-likelihood function:

$$I(\theta) = -E \left[ \frac{\partial^2 \ell}{\partial \theta \partial \theta'} \right]^{-1}, \quad \hat{\theta} = (\hat{\sigma}^2, \hat{\rho}, \hat{\beta}). \tag{25}$$

Detailed derivations of this variance–covariance matrix are available in Doreian (1980).

Briefly, given

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

and

$$B = WA^{-1},$$

the information matrix was estimated from

$$I(\hat{\sigma}^2, \hat{\rho}, \hat{\beta}) = \sigma^4 \begin{bmatrix} N/2 & \sigma^2 \text{tr}(B) & 0' \\ \sigma^2 \text{tr}(B) & \sigma^4(\text{tr}(B'B) - \alpha) & 0' \\ 0 & 0 & \sigma^2 X'A'AX \end{bmatrix}^{-1}. \tag{26}$$

### Design of the Simulation

The following model was used to generate the observations for the simulation:

$$Y = 0 + 1 * X + \epsilon \tag{27}$$

$$X = \lambda WX + u \tag{28}$$

$$\epsilon = \rho W\epsilon + v \tag{29}$$

where

$$E(u) = 50, \quad E(v) = 0, \quad E(uu') = 64I, \quad E(vv') = 81I. \tag{30}$$

Since our main interest here is only with the relative bias and efficiency of the OLS and ML procedures with respect to the regression coefficient, the intercept was set equal to zero. For each combination of  $\rho$  and  $\lambda$ , we first generated a random vector  $u$  with mean and variance as given in equation (30). The independent  $X$  variable was then calculated using

$$X = (I - \lambda W)^{-1} u. \quad (31)$$

As we mentioned previously, the extent and direction of the estimation error is a function of the degree of autocorrelation in the independent variable ( $\lambda$ ) and  $\rho$ .

For each independent  $X$  vector we generated 50 error vectors by first drawing random vectors  $v$  and then transforming them into autocorrelated error vectors as before:

$$\epsilon = (I - \rho W)^{-1} v \quad (32)$$

Then, given the  $X$  vector and the 50  $\epsilon$  vectors, we constructed the dependent  $Y$  vectors using equation (27) above. These steps were repeated for values of  $\lambda = 0, 0.4, 0.8$  and  $\rho = 0.2, 0.4, 0.6, 0.8$  and for sample sizes 20, 30, and 40. Only positive (or zero) values of  $\lambda$  and  $\rho$  were employed, since negative autocorrelation is less interesting theoretically and less likely empirically, and the costs of running the ML procedure precluded examination of all possible autocorrelation values.

At each sample size,  $W$  matrices conformable with the  $u$  and  $v$  vectors were required to generate the simulation data *via* equations (28) and (32). We constructed two connectivity matrices for each sample size, for a total of six matrices. Three of the matrices, one at each sample size, we constructed to have density equal to 0.1. Here, density refers to the simple ratio of actual links to the total number of possible links. Links were added randomly to the rows of a null matrix until the specified density was reached, after which the matrix was row normalized to unity. Although not all of the effects of employing different  $W_R$  matrices are thus removed, effects due to varying sample size are nevertheless quite clear in the results reported below.

Another three connectivity matrices provide considerable contrast to the low density random matrices. The conceptual image used in constructing the second set of matrices was that of a tree-like structure, such

as might be generated by a set of languages of varying degrees of historical relatedness. Each  $W_L$  matrix was formed as a block diagonal, each block being an identical 10 by 10 submatrix with each row containing five 1s, four 2s, and 0s on the main diagonal. This corresponds to a tree-structure where the entire sample is divided into blocks of 5 units, within which each unit is given a relatedness score of 2 for each member of its own block, 1 for each member of one other block of 5 units, and zero relatedness with all other units.

The  $W_R$  and  $W_L$  matrices are thus clearly distinct in terms of both structure and density. Thus in comparing results using different connectivity matrices any differences noted will be due in part to the joint effects of structure and density. It was not possible, given the expense of the ML procedure at this time, to generate sets of matrices varying only on density, and then compare the performance of each procedure over all autocorrelation parameters and sample sizes. Rather, the aim at this point is simply to look at the nature of the difference produced by employing alternate connectivity schemes. Results of a study of the effects of varying network density over a range of autocorrelation parameters for a fixed sample size will be presented in a later paper.

## Results

### *Comparison of OLS and ML in Estimating $\beta$ and $\text{var}(\beta)$*

(1) Average bias in estimating  $\beta$  ( $\hat{\beta} - \beta$ ). Table 1 gives the results of both procedures. Since there were no notable differences across levels of  $\rho$ , the results were averaged over this parameter. Each figure then represents the outcome of 600 replications. The results indicate that the ML procedure is generally superior to OLS on this criteria. However, the amount of bias is very small using either procedure, from 1–5 percent of the magnitude of the coefficient, as we might expect from our previous discussion. There is a slight tendency for the bias to move from positive to negative with increasing sample size, and to be slightly less for the  $W_R$  matrices than for the  $W_L$  matrices.

(2) Relative efficiency. The overall performance of an estimator depends not only on its average bias but also on its variance. It is possible for a relatively unbiased estimator to have high variance, negating the

Table 1  
Average bias in estimating  $\beta$ .

	$W_R$		$W_L$		
	$\lambda$	OLS	ML	OLS	ML
$N = 20$	0	0.03600	0.02258	0.03953	0.02690
	0.4	0.03836	0.02573	0.06035	0.03275
	0.8	0.02546	0.02165	0.06679	0.03860
$N = 30$	0	0.00683	0.00558	0.01400	0.01673
	0.4	0.01196	0.01261	0.01107	0.01419
	0.8	0.01550	0.01648	0.01842	0.00858
$N = 40$	0	-0.05983	-0.05610	-0.05849	-0.04935
	0.4	-0.05771	-0.05076	-0.06434	-0.05098
	0.8	-0.05222	-0.05406	-0.04537	-0.05263

gains made from low bias. Since the mean square error (MSE), defined as  $E(\hat{\beta} - \beta)^2 = \text{bias}^2 + \text{variance}^2$ , combines both bias and variance, it is generally used to compare the performance of various estimators.

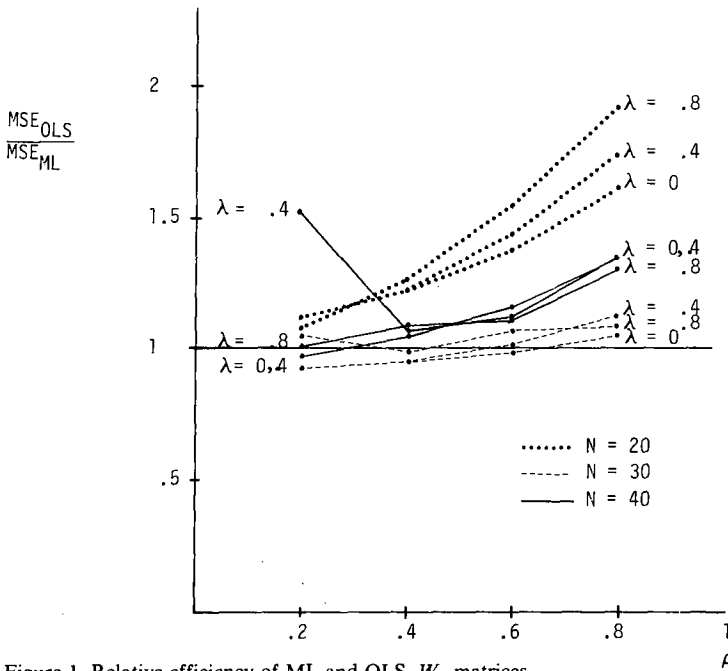


Figure 1. Relative efficiency of ML and OLS.  $W_R$  matrices.

Figure 1 compares the performance of OLS and ML using the random  $W_R$  matrices. Since it is of interest to know how the results are affected by varying the autocorrelation in the independent variable, the

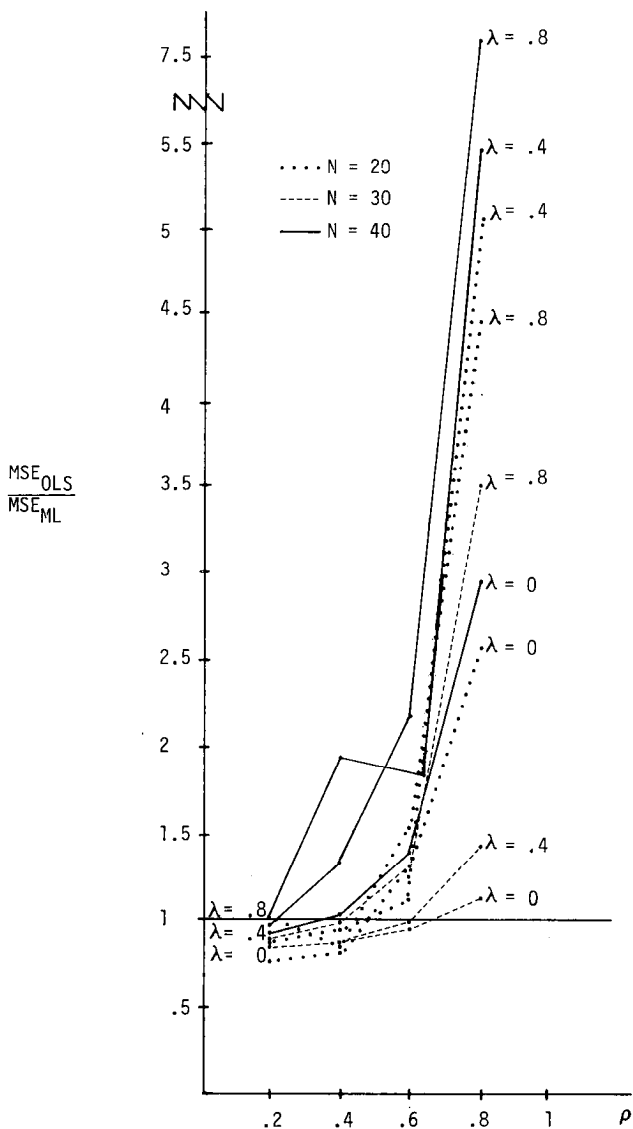


Figure 2. Relative efficiency of ML and OLS.  $W_L$  matrices.

results are presented for each level of  $\lambda$ . For  $\rho \geq 0.4$ , Fig. 1 suggests that using ML offers potential reductions in MSE of 5–50 percent with no serious potential costs. There is also a tendency for the relative efficiency of ML to improve with increasing  $\lambda$ . Figure 2 shows similar results for the  $W_L$  matrix. The advantages of ML are quite clearly demonstrated when  $\rho > 0.6$ : the potential gains are anywhere from about 10–700 percent, especially if the independent variable is also autocorrelated. Where  $0.5 < \rho < 0.6$ , there would also seem to be little potential loss in efficiency using ML, particularly if  $\lambda > 0$  given the strong tendency for efficiency to improve with increasing autocorrelation in the independent variable.

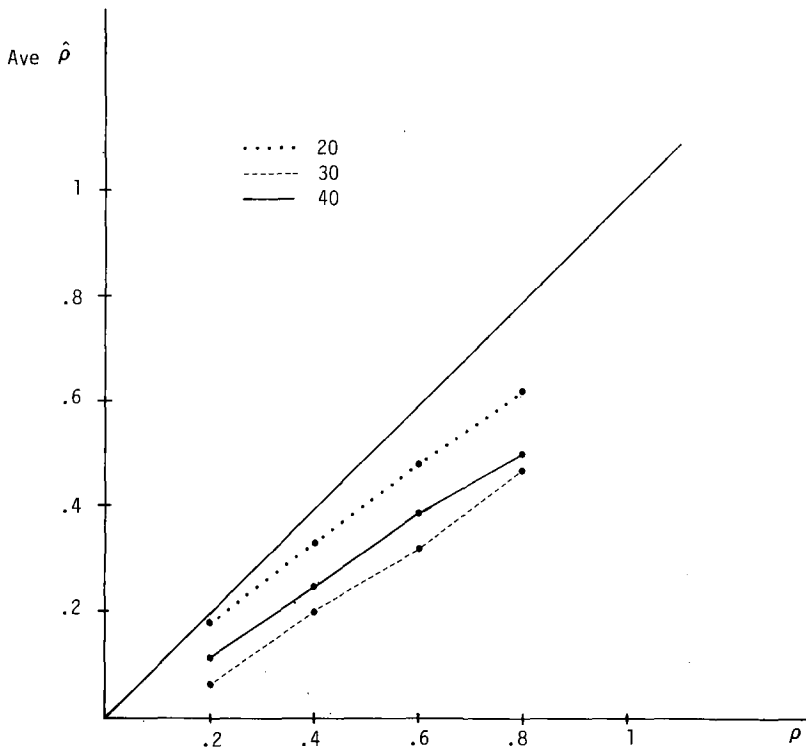


Figure 3. Average bias in ML estimation of  $\rho$ .  $W_R$  matrix.

### ML Estimation of $\rho$ and $\text{var}(\rho)$

(1) Average bias in estimating  $\rho$ . Figure 3 shows that using the  $W_R$  matrix, the ML estimate  $\hat{\rho}$  consistently underestimates the true autocorrelation parameter. Hepple (1976) reports a similar tendency towards negative bias in some preliminary unpublished results. Doreian (1981) has also noted this result in an empirical context. The amount of negative bias appears to have no relationship to the sample sizes reported here, and appears to increase with increasing  $\rho$ . Since the results do not seem to be sensitive to variation in  $\lambda$ , we averaged the replications over this parameter.

Figure 4 shows the same results using the  $W_L$  matrix. Here, there is a tendency to have a negative bias at lower values of  $\rho$ , although the bias now tends to decrease in absolute value as  $\rho$  increases, and becomes

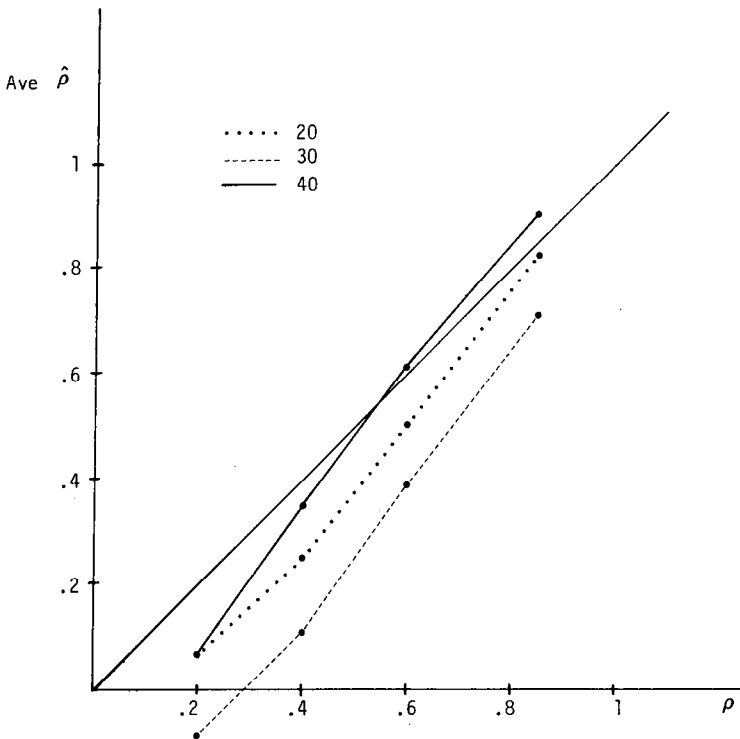


Figure 4. Average bias in ML estimation of  $\rho$ .  $W_L$  matrix.

Table 2  
Average MSE ( $\rho$ ) and average var( $\rho$ ).

Sample size	$W_R$									
	MSE( $\rho$ )					Var( $\hat{\rho}$ )				
	$\rho$	0.2	0.4	0.6	0.8	$\rho$	0.2	0.4	0.6	0.8
20	0.01164	0.18264	0.29557	0.25682	0.13147	0.11673	0.06850	0.02116		
	0.23514	0.24017	0.32867	0.60921	0.35979	0.26765	0.16374	0.06043		
	0.08799	0.20741	0.49612	0.77327	0.15163	0.11550	0.05590	0.01275		
30										
40										
Sample size	$W_L$									
	$\rho$	0.2	0.4	0.6	0.8	$\rho$	0.2	0.4	0.6	0.8
	0.09841	0.11352	0.15900	0.21113	0.07182	0.06260	0.05176	0.03808		
0.07442	0.11416	0.18317	0.30249	0.06369	0.05874	0.05204	0.04186			
0.05994	0.11138	0.19722	0.30502	0.06984	0.06568	0.05717	0.04704			

slightly positive for sample size 40. Overall, there appears to be less bias associated with a sample size of 40, than with the smaller samples.

(2) Average bias in estimating  $MSE(\rho)$ . The negative bias found in estimating  $\rho$  suggests that there will be problems in estimating  $MSE(\rho)$ , since this latter figure combines both bias and variance of  $\rho$ . With a substantial bias, the MSE will appear large, even though the variance may be relatively small.

Table 2 gives the results of averaging the variance of  $\rho$  [ $MSE(\rho)$ ] and the estimated  $var(\hat{\rho})$  over  $\lambda$ . For the  $W_R$  matrix, there is a clear increase in MSE with increasing  $\rho$  for each sample size. However, there is an opposite trend with respect to  $var(\hat{\rho})$ , which decreases with increasing  $\rho$ . These results appear to be a reflection of the tendency noted above for the negative bias in estimating  $\rho$  to increase with  $\rho$  for the  $W_R$  matrix. Hence, the MSE could also be expected to increase as  $\rho$  increase, and the  $var(\hat{\rho})$  should thus increasingly underestimate  $MSE(\rho)$ . Figure 5 illustrates that the bias in ML estimates of the true variance of  $\rho$  is an approximately positive linear function of  $\rho$ .

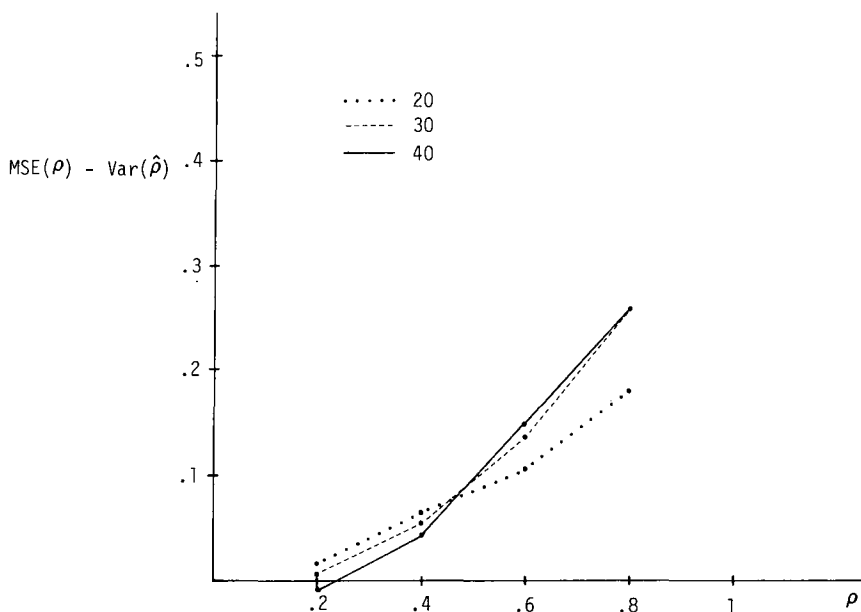


Figure 5. Average bias in ML estimation of  $var(\rho)$ .  $W_R$  matrix.

The results are generally similar for the  $W_L$  matrix, although the average MSEs are rather larger and the average  $\text{var}(\hat{\rho})$  rather smaller, than previously. Since there was no observed tendency towards increasing negative bias in estimate of  $\rho$  as  $\rho$  increases (for the WL matrix), in

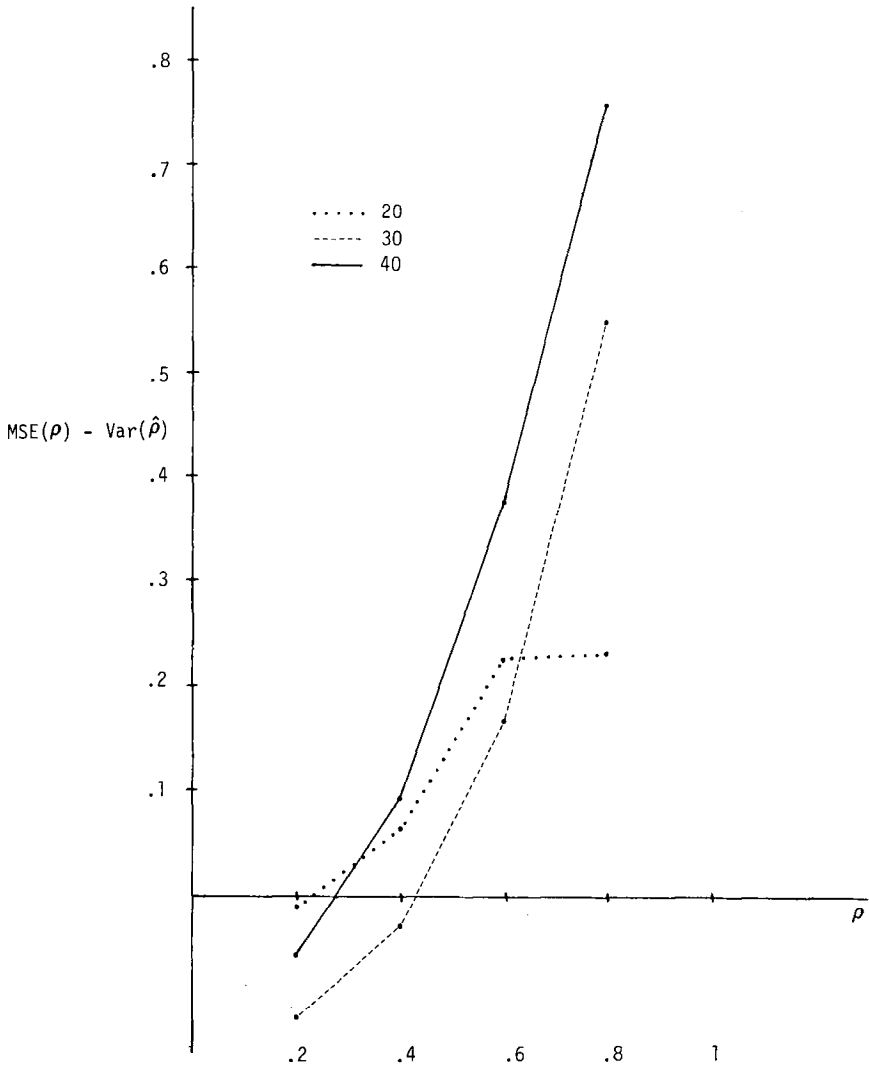


Figure 6. Average bias in ML estimation of  $\text{var}(\rho)$ .  $W_L$  matrix.



0.8	0.91 N = 40	1.02 N = 30	2.98 N = 20	7.73 N = 40	2.92 N = 30	5.51 N = 20	8.50 N = 40	2.86 N = 30	1.84 N = 20
0.2	0.81	0.74	0.75	0.41	0.22	0.63			
0.4	1.23	1.21	1.51	0.94	0.80	1.30			
0.6	1.37	1.43	2.13	1.57	1.40	2.08			
0.8	1.47	1.48	2.69	2.27	2.28	3.16			
4	0.2	0.86	0.61	0.45	0.23	0.73			
0.4	1.20	1.18	1.49	0.98	0.84	1.60			
0.6	1.35	1.40	2.17	1.62	1.48	2.20			
0.8	1.45	1.45	2.76	2.35	2.35	3.24			
0	0.2	0.79	0.58	0.47	0.30	0.82			
0.4	1.17	1.16	0.90	1.03	0.87	1.45			
0.6	1.33	1.38	1.06	1.68	1.52	2.26			
0.8	1.42	1.43	1.18	2.44	2.41	3.30			
Average									
All	0.82	0.73	0.68	0.44	0.25	0.73	0.54	0.43	1.07
$\lambda$	0.4	1.18	1.50	0.98	0.84	1.45	0.82	0.71	0.97
	0.6	1.40	2.15	1.62	1.46	2.18	1.20	1.04	1.01
	0.8	1.45	2.72	2.35	2.35	3.24	1.62	1.62	1.19
	N = 40	N = 30	N = 20	N = 40	N = 30	N = 20	N = 40	N = 30	N = 20
Ratios $\hat{Z}/Z$									

fact just the opposite occurred, it appears that the ML estimator of  $\text{var}(\rho)$  is problematical here. Figure 6 illustrates that the bias in estimating  $\text{var}(\rho)$  is a more rapidly increasing function of  $\rho$ . Significance tests of  $\hat{\rho}$  using the ML  $\text{var}(\hat{\rho})$  would thus appear to be unreliable

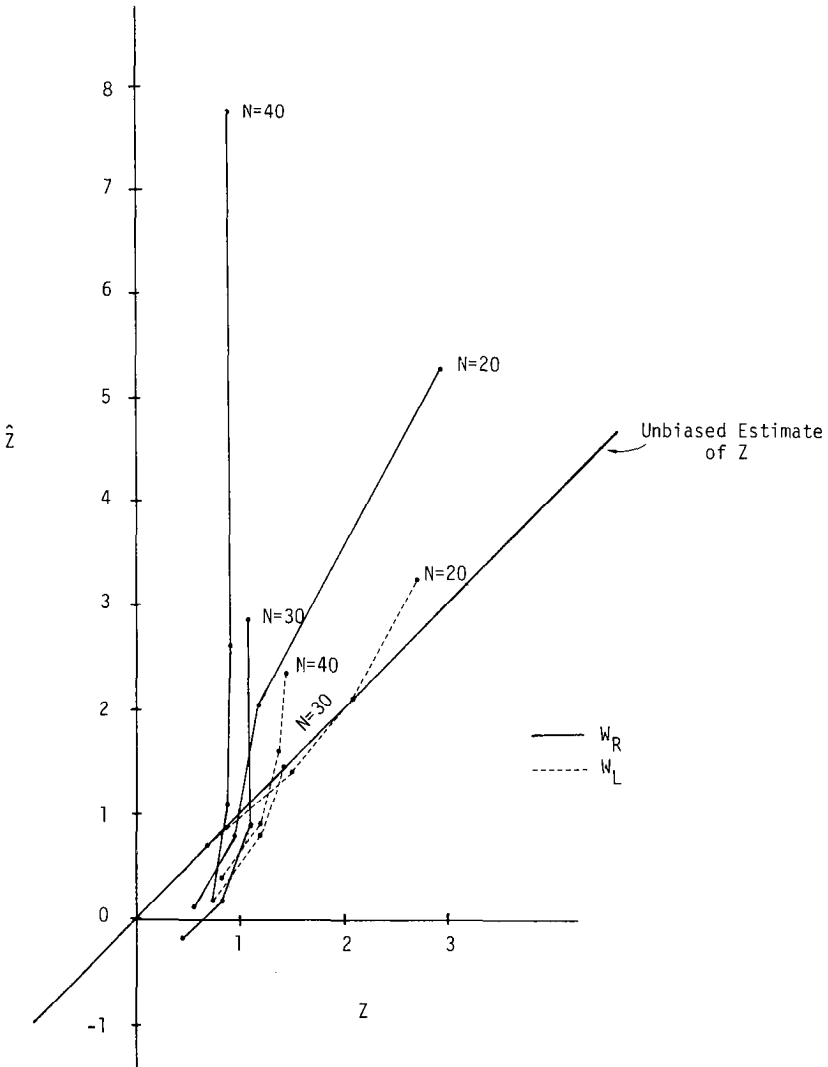


Figure 7.  $\hat{Z}$  versus  $Z$ .

for high  $\hat{\rho}$  and complexly structured connectivity matrices.

To assess the degree of reliability of such significance tests, we have computed true  $Z$  scores and estimated  $Z$  scores for the various combinations of the parameters. These appear in Table 3. Here we see that  $Z$  and  $\hat{Z}$  are not affected by  $\lambda$ , but that they both increase with  $\rho$ . The ratio of  $\hat{Z}/Z$  also increases with  $\rho$ . It reaches a value of 1 at a  $\rho$  of about 0.50 for all combinations of  $X$  and  $N$ . There is no simple relationship to  $N$ . The highest values of  $\hat{Z}/Z$  are for  $N=40$ , but the values of  $N=20$  are not consistently smaller than the values for  $N=30$ . This is graphed in Fig. 7. The conclusion from these calculations is that the maximum likelihood technique will overestimate significance for large values of  $\rho$ , and underestimate significance for low values of  $\rho$ . Hence the significance tests associated with the maximum likelihood procedure must be used conservatively.

## Conclusions

The ML solution to the autocorrelated disturbances model of context effects in a multiple regression framework is proposed as superior to OLS for investigation of the substantive nature of the underlying network autocorrelation processes. Specifying such models involves the need to make clear theoretical decisions as to whether autocorrelated residuals in OLS should be conceived of as due to (1) interaction effects, (2) unspecified systemic independent variables which are autocorrelated, or (3) nuisance factors having no substantive import (note that only 1 and 3 are properly specified by the model).

Simulation tests of the proposed superiority of ML over OLS estimation procedures in the face of autocorrelated disturbances showed clear dominance of the ML model (Figs. 1 and 2) with moderate or high levels of autocorrelation ( $\lambda > 0, \rho \geq 0.5$ ) in estimating the variance of the regression coefficients. Both procedures are relatively unbiased in estimating these coefficients (Table 1), and differ insignificantly in this regard. Since it is usually of major importance to obtain the best possible estimates of the regression coefficients and their variances, the relative efficiency results indicate that the ML procedure will generally be preferred. In particular, if highly significant autocorrelation of the residuals were initially detected using Cliff and Ord's (1973)  $I$ -statistic, then ML should be used over OLS.

The simulation results for estimation of the autocorrelation parameter  $\rho$  (Fig. 3 and 4) and bias in variance (Fig. 5 and 6) indicate the difficulties in using ML estimation procedures to investigate network autocorrelation processes. For random autocorrelation matrices (Fig. 3), ML consistently underestimates the  $\rho$  parameter, roughly in proportion to its magnitude, and independently of sample size. On the positive side, however, the ML estimates for structured autocorrelation matrices gain in accuracy with increases in the magnitude of  $\rho$ . For the language  $W_L$  matrix, there is very little bias in the estimate  $\rho$  at a true  $\rho = 0.8$ . Bias in estimating the variance of  $\rho$  increase, however, at higher levels of autocorrelation, for more structured autocorrelation matrices, and for larger samples sizes. Thus, estimates of the significance of  $\hat{\rho}$  are unreliable from ML procedures.

Problems in estimating the significance of  $\hat{\rho}$ , however, are much more severe for random  $W$  matrices than for structured autocorrelation, which is the more usual case in naturally occurring phenomena. With low sample size ( $N = 20$ ), such estimates are fairly reliable (overestimating Z-scores of  $\hat{\rho}$  by about 25 percent at the highest levels of autocorrelation, and unbiased at lower levels where  $\rho \geq 0.6$ ). With larger samples ( $N = 40$ ), the estimates are over by about 50 percent at  $\rho = 0.8$ , and reliable or underestimated elsewhere.

There are two other routes to circumvent the problems in estimating the variance and significance of the autocorrelation parameter. One is to use the magnitude of the estimate  $\hat{\rho}$  (e.g.  $\rho \geq 0.40$ ) as indicating preference for ML over OLS solutions, or to use the  $I$ -statistics (Cliff and Ord 1973) to test significance of autocorrelation in the independent or dependent variables. The other is to explore other mathematical models for estimating the autocorrelation parameter and its variance.

One other result of the simulation is gratifying. In generating the simulated data, we had to specify two autocorrelation parameters, only one of which can be solved for in the ML procedures. Rho ( $\rho$ ), which can be estimated, is the amount of autocorrelation in the error terms. Lambda ( $\lambda$ ), which cannot be estimated, is the amount of independent autocorrelation in the independent variables. It turns out, however, that varying lambda makes no difference to the ML estimation results except in Fig. 2, where for structured  $W$  matrices, higher levels of  $\lambda$  simply increase the efficiency of ML over OLS procedures.

Overall, the problems in significance testing of the autocorrelation coefficient are dwarfed by the gains in using ML over OLS procedures

for estimating regression coefficients and their variances, when moderate or high levels of autocorrelation are present. Our simulation results on random *versus* structured autocorrelation matrices, however, clearly illustrate the need for further study of the effects of network structure on the autocorrelation models and the results of estimation procedures.

## References

- Aurora, S.S and M. Brown  
 1977 "Alternative approaches to spatial autocorrelation: an improvement over current practice." *International Regional Science Review* 2: 67–78.
- Bodson, P. and D. Peters  
 1975 "Estimation of the coefficients of a linear regression in the presence of spatial autocorrelation: an application to a Belgian labor-demand function." *Environment and Planning* 7: 455–472.
- Burt, R.  
 1980 "Models of network structure." *Annual Review of Sociology* 6: 79–141.
- Cliff, A and K. Ord  
 1973 *Spatial Autocorrelation*. London: Pion.
- Doreian, P.  
 1980 "Linear models with spatially distributed data: spatial disturbances on spatial effects?" *Sociological Methods and Research* 9: 29–60.  
 1981 "Estimating Linear Models with Spatially Distributed Data." In S. Leinhardt (ed.), *Sociological Methodology*. San Francisco: Jossey-Bass.
- Dow, M., D. White and D. Hansen  
 1979 "Regression analysis and network autocorrelation." Ms. University of California, Irvine.
- Dow, M., D. White and M. Burton  
 1982 "Multivariate modeling with interdependent network data." *Behavior Science Research* (in press).
- Erbring, L. and A. Young  
 1979 "Individuals and social structure: contextual effects as endogenous feedback." *Sociological Methods and Research* 7: 396–430.
- Gatrell, A.C.  
 1979 "Autocorrelation in spaces." *Environment and Planning* 11: 507–516.
- Hepple, L.  
 1976 "A maximum-likelihood model for econometric estimation with spatial series." In I. Masser (ed.), *Theory and Practice in Regional Analysis*. London: Pion.
- Hibbs, D.  
 1974 "Problem of statistical estimation and causal influence in dynamic time series models." In H. Costner (ed.), *Sociological Methodology*. San Francisco: Jossey-Bass.
- Jackson, J.  
 1975 "Issues, party choices, and presidential votes." *American Journal of Political Science* 19: 161–185.
- Johnston, J.  
 1972 *Econometric Methods*. New York: McGraw-Hill.

Lorrain, F.P and H.C. White

- 1971 "Structural Equivalence of Individuals in social networks." *Journal of Mathematical Sociology* 1: 49–80.

Martin, R.

- 1974 "On spatial dependence, bias, and the use of first spatial differences in regression analysis." *Area* 6: 185–194.

Naroll, R.

- 1976 "Galton's Problem and HRABLIB." *Behavior Science Research* 11: 123–146.

Ord, K.

- 1975 Estimation methods for models of spatial autocorrelation. *Journal of the American Statistical Association* 70: 120–126.

Schaefer, J. (ed.)

- 1974 *Studies in cultural diffusion: Galton's Problem*. New Haven: Hraflex Books.

Taylor, E.

- 1889 "On a method of investigating the development of institutions applied to the laws of marriage and descent." *Journal of the Royal Anthropological Institute* 18: 272ff.

White, D., M. Burton and M. Dow

- 1981 "Sexual division of labor in African agriculture: a network autocorrelation analysis." *American anthropologist* 83: 824–849.

White, D. and K. Reitz

- 1981 "Rethinking the role concept: Social networks, homomorphisms and equivalence of positions." Ms. University of California. Irvine.