Developments in Data Analysis
A. Ferligoj and A. Kramberger (Editors)
Metodološki zvezki, 12, Ljubljana: FDV, 1996

When the Data Points are not Independent

Patrick Doreian*

Abstract

Social actors are interdependent and jointly embedded within larger social structures. This insight is lost freqently when data concerning these actors are analyzed. Regression models seem particularly vulnerable, a comment that extends to virtually all of the family of linear models. It is known that OLS can perform very badly if there are interdependencies in data. If the interdependencies can be represented in the form of a network autocorrelation model, it is then possible to incorporate them into regression type analyses. There are alternative methods for estimating these models, including MLE - which is an attractive framework for establishing estimation equations and standard errors. However, as the desirable properties of MLE are asymptotic, we do not know how well these models fare in small samples. One approach to generating this knowledge is by simulation. A design for such a simulation is presented that focuses on models with two regimes of network autocorrelation.

I have two goals for this presentation. First, I will try to persuade you of the importance of a class of inter-related problems. These problems are very general, have a definate structure and can occur in a wide variety of data analytic situations. Second, I hope to enlist your help, and support, in an effort to solve these problems.

The term 'network autocorrelation models' is a convenient label for this broad class of specified models and estimation methods. It is meant to capture the idea of inter-dependencies among data points. Tackling network autocorrelation problems amounts to importing some extant methods from other fields, developing new estimation methods where none exist and assessing the utility of all such methods.[1]

The technical point of departure is the specification of a linear model and the use of ordinary least squares (OLS) as a method for estimating the parameters of the specified model. The model is written as:

$$y = X\beta + \varepsilon \tag{1}$$

where y is the vector of observations for the predicted variable, X is the matrix of observations for the predictor variables (including the intercept) and β is the vector

*Department of Sociology, University of Pittsburgh, Pittsburgh, PA 15260, USA

[1] Many methods for dealing with inter-dependent objects have their foundations in spatial autocorrelation, a term familiar to geographers, and in efforts to deal with Galton's problem which are found in the anthropological literature.

of parameters. The disturbance terms are in ε are assumed to be independentally and identically distributed normal random variables with mean 0 and constant variance ω. More precisely, $\varepsilon \sim IN(0, \omega I)$. The OLS estimating equation for β is

$$\hat{\beta} = (X'X)^{-1}X'y \tag{2}$$

With β estimated, the estimation equation for $\hat{\omega}$ is given by:

$$\hat{\omega} = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{(n-k-1)} \tag{3}$$

where $\hat{\varepsilon}$ is the OLS residual and the variance-covariance matrix for the estimators is given by:

$$V(\hat{\omega}, \hat{\beta}) = \omega^2 \begin{bmatrix} n/2 & 0' \\ 0 & \omega X'X \end{bmatrix}^{-1} \tag{4}$$

When the assumptions of the model are satisfied, these estimating equations are the best possible - they yield unbiased estimates of the parameters and are the most efficient of the linear estimators. Alas, such theoretical results may be of little value if these assumptions are not met. In general, however, methods can be proposed to handle situations where one or more of the assumptions are violated. My focus here is on the assumption that the data points are independent of each other.

1 Some substantive domains

I will use five examples of substantive concerns where, empirically, it seems prudent to anticipate that the assumption of independent units of analysis will be problematic.

1.1 Networks of scientists

Imagine a study of scientists in a scientific field or scientific specialty. Suppose that the dependent (predicted) variables of interest include the definition of the significant problems of the field, the most fruitful way of pursuing solutions to these problems and which journals will contain useful information for tackling the key problems of the field.[2] Not all scientists will agree on the identity of problems, solutions and journals (and within the social sciences, such disagreements are even more frequent and, perhaps, more contentious.) Suppose we wanted to account for the differences among scientists in their scientific beliefs. We could try and account for them (the values of y) in terms of, for example, where the scientists were trained, how long they have been in the field, their publication record and the prestige of their places of employment. All of these variables would be examples of the X variables, and we could imagine that the linear model of equation (1) is specified.

[2] The last of these can be stated in terms of learning which journals, in general, are important enough to merit attention. See, for example Burt and Doreian (1982).

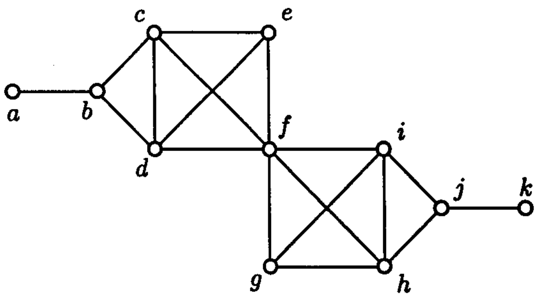

Figure 1: Hypothetical Network

The obvious problem here is that scientists do not work in complete isolation. They are members of, so called, invisible colleges. As members, they communicate with each other, they share data and methods and, to some extent, evaluate each other's work. In short, there are networks of ties linking them in their invisible colleges. Additionally, we can note that some scientists are trained by some other scientists and scientists can work together at a common location.[3] These forms of interaction are network ties. So, for example, if you wanted to predict my ideas about the major problems of my field, my views of the most fruitful avenues of inquiry and which journals I regard as important, you would need to know also the corresponding items from *some* other members of the invisible college(s) to which I belong. Once it is recognized that for predicting such items for any scientist, it is clear that the data points (scientists) are not independent. Members of an invisible college mutually socialize each other over networks of ties and it seems reasonable to include these interependencies in some fashion.

This can be made more precise with an example. Consider the network shown in Figure 1 and imagine it represents a social tie linking some scientists. For a given scientist, say d, the only other scientists that need to be considered for predicting some attribute of d are those directly linked to this actor. So to predict for d, it is necessary to consider only b, c, e and f. These would be the only sources of influence over d in the network. Models of this sort have been termed *cohesion* models. See, for example, Burt (1983). They are local in the sense that, for a given scientist, it is enough to consider those other scientists linked directly to that individual.[4] A different network model could take the form of stating that actors in identical structural locations of a network share certain characteristics. Two actors are *structurally equivalent* if they are connected to the rest of the network in the same manner. More formally, for a set, N, of nodes and a relation, R, two actors, x

[3]In an era of electronic communication collaborations can involve researchers who are geographically dispersed.

[4]Of course, a similar statement can be constructed for each actor in the network, so, in some sense, the whole network is considered.

and y are structurally equivalent ($x \equiv y$) if and only if :

$$\begin{aligned}
xRy &\Leftrightarrow yRx \\
xRx &\Leftrightarrow yRy \\
\forall z &\in N\backslash\{x,y\} : xRz \Leftrightarrow yRz \\
\forall z &\in N\backslash\{x,y\} : zRx \Leftrightarrow zRy
\end{aligned}$$

where z is any other actor in the network. In Figure 1, c and d are structurally equivalent as they are both connected to b, e and f. Similarly, i and h are structurally equivalent. It is possible to measure the extent to which actors are structurally equivalent and use this measure of the interdependence of actors as part of the prediction apparatus - although some care is needed here. (See Batagelj, Ferligoj and Doreian, 1992.) Another conceptualization of position in a network, and hence equivalence, takes the form of defining two actors as equivalent if they are equivalently connected to equivalent others. More precisely see White and Reitz (1983), $\approx$ is a *regular equivalence* on $N \Longleftrightarrow \forall x,y,z,w \in N, x \approx y$ implies both

$$\begin{aligned}
xRz &\Rightarrow \exists\, w \in N : yRw \text{ and } w \approx z \\
zRx &\Rightarrow \exists\, w \in N : wRy \text{ and } w \approx z
\end{aligned}$$

In Figure 1, a and k are regularly equivalent. The nodes b and j are regularly equivalent. The set of nodes, c, d, i and h are all regularly equivalent (to each other). The pair of nodes, e and g are equivalent with f being structurally unique. If there is a 'regular equivalence mechanism' then those actors that are regularly equivalent will be most like each other.

1.2 World system theory and dependency theory

Since World War II, considerable attention has been given to the 'development' of Third and Fourth World nations. Within the modernization literature, the impact of capital investment and economic aid (among other variables) on rates and levels of economic development was studied. In general, the relationship of the predictors on the predicted variables was positive (within regression models estimated by OLS). World sytem theorists - for example, Wallestein (1974) - and dependency theorists - for example Chase-Dunn (1975) and Rubinson (1976) - called these findings into question. Rather than having a world of disconnected nations (implicit in the regression models), they argued that there is a single (world) system made up of interconnected nations. Among the relational ties among nations are 'trades with', 'is a former colony of', 'diplomatically recognizes', 'is an ally of' etc.. Within this approach, the theory takes the form claiming that these links between nations worked to the advantage of industrialized societies (and former colonial powers) and to the disadvantage of the Third World nations (and former colonies). Third World nations, in this view, fare badly from their structural location in the networks linking countries. In the main, the research findings of this literature showed that capital investment and aid had (and continue to have) negative impacts on levels and rates of economic development of the Third World nations and positive impacts on inequality.

The theoretical arguments of modernization theorists and dependency theorists flatly contradict each other. The former group used regression methods extensively. Most strangely, the statistical method of choice among dependency theorists is also OLS! Even though their substantive propositions explicitly include relations between nations, the estimation methods for their linear models implicitly excludes those relations. This holds for most of the more recent studies - for example, Crenshaw (1991, 1992) and Simpson (1990). There are some notable exceptions. Snyder and Kick (1979) may have been the first to include relations among nations when estimating dependency theory models. They used blockmodeling techniques based on structural equivalence. Using four relations - trade, diplomatic recognition, treaties and military incursions - they empirically identified positions in the World System.[5] These positions mapped fairly cleanly into a theoretical partition of nations into core nations, semi-peripheral nations and peripheral nations. Their partition was more fine grained having eight positions (which could be aggregated into these three categories). They used a set of dummy variables[6] to represent the clusters and used OLS. On (other) methodological grounds, Jackman (1980) called their results into question but Nolan (1983), using dummy variables to represent just the three positions showed that the substantive results of Synder and Kick held up and were robust.

My concern here is not which brand of theory best explains the dynamics of change in the world system of nations. It is with the estimation methods used to estimate linear models specified within each of the perspectives. For reasons made clear below, all of the estimated results obtained from the use of OLS are questionable. There are two *possible* exceptions: the Snyder and Kick (1979) and the Nolan (1983) studies mentioned already. Another tack was taken by Smith and White (1992) who also sought partitions of the nation states in terms, not of structural equivalence, but regular equivalence. It is clear, for example, that most of the former colonies of Great Britain are structurally equivalent. As are the former colonies of France. But the former colonies of France are not structurally equivalent with the former colonies of Britain. If each of the colonial powers had unique ways of treating their colonies, then the use of structural equivalence makes sense. But if the real mechanism is colonialization, then regular equivalence is a more powerful conceptualization if the former colonies of Britain and of France are regularly equivalent. The regular equivalence based partitions of Smith and White could be incorporated via dummy varaibles into regression models and may be another possible exception. We need better methods for incorporating the interdependencies among nations into regression type models. Only when this is done, can we assess the utility of using network based partitions of the nation states into positions that, in turn, are represented by dummy variables. It may even be possible that, when

[5]This alone was a major accomplishment.

[6]Always using one less dummy variable than the number of clusters of nations - or one less dummy variable than identified positions.

the interdependencies are incorporated into the analyses, the substantive results of the earlier work hold - but their methods are sufficiently vulnerable that we cannot take those results on faith alone.

1.3 Inter-organizational networks

Organizations can be studied in a variety of ways. Although attention can be focussed on organizations as units of analysis, it is impossible to consider them as a set of independent entities. At face value, they can be conceptualized as if they are independent: "Organizations are goal-directed, boundary maintaining, activity systems" (Aldrich, 1979). Yet, as Aldrich makes very clear, the pursuit of goals, the efforts to maintain boundaries and the organizational activities are actions contingent on the environment within which organizations are located. In particular, the environment includes other organizations. Imagine a study of social service organizations providing mental health services to children and youth.[7] At face value, this is clearly delimiited. Indeed, there are organizations providing "just" mental health services. Yet, many children in need of such services do not receive them and there are children who receive them in settings that have some other primary purpose. Moreover, children receiving mental health services, disproportionally, receive them as teenagers at a time when treatments are much harder and more expensive - at least in the United States. I will digress slightly at this point.

Children are both dependent and immature. A child's immaturity means that services have to be geared to the age of the child and a child's dependence means that any problem has to be recognized by an adult. Children may be seen as having health problems and would, via the actions of an adult, enter the social service system through the health sector. The child could be born into poverty and, albeit indirectly, receive services in the poverty/social welfare sector. If there are mental health problems, they can remain un-noticed or be seen as secondary. When a child is in school, a teacher may notice unruly behavior, or attention deficits or bizarre behavior. These problems are likely to be seen as, for example, learning problems with attempts to deal with them at that level. At some point, school counsellers or psychologists could be involved and a child could be referred to a mental health agency. As either victims or perpetrators, children can enter the social service sector through agencies in the judicial sector. The police, court officials and the Juvenile Probation Office can all be involved and can, in principle, call upon mental health organizations for particular services if there appears to be mental health problems. One implication is that there is a wider network, beyond the mental health agencies, that requires attention - even if the focus is one mental health services for children. See Woodard and Doreian (1994).

At a minimum, to understand the provision of mental health services, it is necessary to consider the many ways a child can enter the 'system'. This has several implications. Certainly, it means that more organizations have to be included. More importantly, these organizations are linked to each other via explicit network ties. If

[7]Henceforth, I will use the term 'children' to cover both children and youth - people aged up to 18 years.

we focus on the inter-organizational actions taken by agencies with regard to their clients, most fall within three categories: referrals of children, provision of services to children and the creation of coordinating mechanisms (to promote more effective referrals and for better service provision). Doreian and Woodard (1994) suggest the use of k-cores (Seidman, 1983) as a snowball selection procedure for including agencies that are linked to those agencies already included. It is a method designed to reach a population of organizations in a given geographic area. Having a *network* of organizations means that the survival chances of any one organization is contingent on its location in the network. Moreover, its effectiveness is also dependent on the organizations in the wider network: it needs other organizations to achieve its goals and it competes with other organizations in the network for resources. These arguments all point to the conclusion that organizations are greatly interdependent and a study of these organizations that ignores their interdependencies will be extremely limited.

1.4 Political insurgencies in geographic space

There are many parts of the globe where insurgents (rebels) battle their governments (and each other). These battles are fought for the control of land and the people on that land. Mitchell (1969) presents an analysis of the HUK rebellion in the Philippines. The Luzon region was divided into a set of barrios that were under control of either the rebels or the government. The dependent variable, y, was the extent to which barrios are under control of the rebels. Among the predictors, X, are variables that represent land tenure arrangements, types of economic production and ethnic composition. There were also dummy variables representing the presence of mountains and swamps. Efforts to predict the control of, say, area i, via equations of the form $y_i = f(X_{ij})$, where there are a set of predictor variables indexed by j, seem incomplete. While a regression model of the sort represented by equation (1) could be specified as a form of f , Mitchell argued that the level of control in area i (by either the government or rebels), is contingent on the level of control in adjacent areas. A more useful specification is $y_i = f(X_{ij}, \{y_k\})$ where y_k is the value of the dependent variable in an adjacent area k. Mitchell's model reflected this. Whether in this general form ,or as a regression-like form, it is clear that the model is one where the data points in geographic space are interdependent. In a context like this, the inter-dependency is known as spatial autocorrelation. See, for example, Anselin (1988).

1.5 Galton's problem

Within anthropology, there have been efforts to establish (causal) models of sociocultural phenomena. Regardless of the sophistication of the statistical tools used to establish these models, it seems that the interdependence phenomenon is relevant here also. Edward Tyler (1889) presented a paper at the Royal Anthropological Institute reporting a cross-cultural study. Sir Francis Galton was in the audience and pointed out that the societies in Tyler's data were not independent. Hence the label Galton's problem. As expressed by Dow et al. (1984): "the non-independence of sample societies stems from the fact that humans have evolved from a common evolutionary stock and from the diffusion of cultural traits among societies. Societies in neighboring or historically related regions tend to be duplicates of one another in terms of a wide variety of traits that are spread by historical fission, diffusion or migration of peoples." As Dow et al. point out, the implications of this are serious: "The result is that neither the actual number of 'independent' cases nor the effect of the interdependencies on trait correlations is generally known for any cross-cultural sample".

Some of the ideas in the spatial autocorrelation literature have been used to formulate possible solutions to Galton's problem. See, for example, Loftin (1972) and Narroll (1976). Loftin and Ward (1981) used the linear regression model to frame the problem and used spatial autocorrelation ideas to modify the regression results - along with a comparison with some other proposals for solving this far-reaching problem. Dow et al. (1984) adapted spatial aurtocorrelation models for an extensive discussion of, and offered a potential solution to, Galton's problem.

2 Common features of examples

Obviously, the common feature linking all of these examples is that the units of analysis are interdependent. The major development in the spatial autocorrelation literature (section 1.4) was the idea that the interdependencies can be represented in a matrix, W, usually called the weight matrix. W is a square matrix with as many rows and columns as the number of areas. As a point of departure for the geographic example, we could use S where:

$$S = [s_{ij}]$$

with

$$s_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are contiguous} \\ 0 & \text{otherwise} \end{cases}$$

W can be consructed from S in a variety of ways. The most freqently used method is to make W row stochastic: $w_{ij} = \frac{s_{ij}}{s_{i+}}$ where s_{i+} is the row sum of the i^{th} row. An alternative is to specify W in terms of the length of the common border for two adjacent areas: $w_{ij} = \frac{b_{ij}}{b_i}$ where b_{ij} is the length of the common border between areas i and j with b_i the total border length of i. Another example from the spatial autocorrelation autocorrelation literature has W specified in terms of distance decay functions: $w_{ij} = e^{\frac{-(d_{ij}^{\alpha})}{k}}$ where α is a decay parameter to be estimated, k is a constant and d_{ij} is the distance between the centroids of two areas.[8]

Once the recognition is made that the interdependencies can be represented in terms of a square matrix, it is clear how the network examples can be represented. Indeed, the cohesion example in section 1.1 can utilize directly the row stochastic version of W obtained from a sociomatrix, S. At face value, using the distance

[8] Alternatively, the geographic locations of the capitals of the areas, for example county capitals for counties in the US.

decay function is less plausible in the network context. However, for a graph that is strongly connected, the graph distance between i and j may be useful for the cohesion model with actors further away be weighted less than close actors. The model for a structural equivalence process (sections 1.2 and 1.3) can use a weight matrix where w_{ij} is a measure of the extent to which actors i and j are structurally equivalent. White et al. (1981), for a sample of tribal societies in Africa, used a linguistic tree connecting 43 African languages as the basis for calculating graph distances between local societies whose languages were in the tree. This seems a very appealing way of incorporating the form of the interdependencies discussed in section 1.5.

The general strategy pursued here for network autocorrelation models is one where the interdependencies among the units is conceptualized in terms of weight matricies that can be used within a regression approach.

3 Single regime models

Incorporating W can be done in two ways - by changing the specification of the disturbance term or by re-specifying equation 1. For the former option,

$$\varepsilon = \rho W \varepsilon + \nu \qquad \text{with} \qquad \nu \sim IN(0, \omega I) \tag{5}$$

The specification for the linear equation is not changed and, together, equations (1) and (5) form a *disturbances model.* For the second option, $\varepsilon \sim IN(0, \omega I)$ and the linear equation is

$$y = \rho W y + X\beta + \varepsilon \tag{6}$$

This specification is described as an *effects model.* In both models, ρ is the autocorrelation parameter and has to be estimated along with β and ω. Estimating ρ complicates the whole process.

3.1 Disturbances models

The likelihood function for ν is straightforward to state but the transformation from ν to ε creates a Jacobian term that must be included. [9] The log-likelihood function for y is

$$l(y) = const - \left(\frac{n}{2}\right)\ln(\omega) - \left(\frac{1}{2\omega}\right)[y'A'Ay - 2\beta'X'A'AXy + \beta'X'A'AX\beta] + \ln|A| \tag{7}$$

where $A = I - \rho W$ and $|A|$ is the Jacobian of the transform from ν to ε. The estimation equations for β and for ω are straightforward to establish:

$$\hat{\beta} = (X'A'AX)^{-1}X'A'Ay \tag{8}$$

and

$$\hat{\omega} = \left(\frac{1}{n}\right)\hat{\varepsilon}'A'A\hat{\varepsilon} \tag{9}$$

[9] For the transformation of the transformation of the ε to the y the Jacobian is I, the identity matrix.

However, these equations cannot be mobilized yet as they involve ρ. Substituing for $\hat{\beta}$ and $\hat{\omega}$ into the log-likelihood function leads to a concentrated log-likelihood function which, in turn, leads to $\hat{\rho}$ as the value of ρ that minimizes $\ln(y'A'PAy) - \left(\frac{2}{n}\right)\sum_i \ln(1-\rho\lambda_i)$ where $\{\lambda_i\}$ are the eigenvalues of A. In this expression, $P = I - AX\left[(AX)'(AX)\right]^{-1}(AX)'$. The numerical task is simplified by using $\ln|A| = \sum_i \ln(1-\rho\lambda_i)$ (See Ord (1975) or Doreian (1980) for details.) The variance-covariance matrix for the estimated parameters is obtained from the second order partial derivatives of $l(y)$ with respect to pairs of parameters and is written as

$$V(\hat{\omega},\hat{\rho},\hat{\beta}) = \omega^2 \begin{bmatrix} \frac{n}{2} & \omega tr(B) & 0' \\ \omega tr(B) & \omega^2\{tr(B'B) - \alpha\} & 0' \\ 0 & 0 & \omega X'A'AX \end{bmatrix}^{-1} \qquad (10)$$

where $\alpha = -\sum_i \lambda_i^2/(1-\rho\lambda_i)^2$.

3.2 Effects models

The effects model, as specified in equation (6), also entails the use of a Jacobian. In this case, the salient transformation is from the ε terms to the y terms. The log-likelihood function is

$$l(y) = const - \left(\frac{n}{2}\right)\ln\omega - \left(\frac{1}{2\omega}\right)[y'A'Ay - 2\beta'X'Ay + \beta'X'X\beta] + \ln|A| \qquad (11)$$

where $A = I - \rho W$. Solving the equations obtained by setting the partial derivatives of $l(y)$ to zero gives the estimation equations for β and ω:

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1}X'Ay \\ \hat{\omega} &= \left(\frac{1}{n}\right)y'A'MAy \end{aligned} \qquad (12)$$

where $M = I - X(X'X)^{-1}X'$. Again, see Ord (1975) or Doreian (1981) for details. The estimation equations in (12) cannot be used directly as they both invlove ρ which is not known. However, when these equations are substituted into (11) , maximizing the log-likelihood function can be shown to be equivalent to minimizing:

$$-\left(\frac{2}{n}\right)\sum_i \ln(1-\rho\lambda_i) + \ln(y'My - 2\rho y'MWy + \rho^2 y'W'Wy)$$

with regard to ρ. With ρ estimated, the equations in (12) can be used to get the MLE estimates of β and ω. The variance-covariance matrix for the estimators is established in the same way as for the disturbances model. It is:

$$V(\hat{\omega},\hat{\rho},\hat{\beta}) = \omega^2 \begin{bmatrix} \frac{n}{2} & tr(B) & 0' \\ tr(B) & \omega^2\{tr(B"B) - \alpha\} + \omega\beta'X'B'BX\beta & \omega\,(X'BX\beta)' \\ 0 & \omega X'BX\beta & \omega X'X \end{bmatrix}^{-1} \qquad (13)$$

4 Two regime models

Conceptually, it is straightforward to specify multiple network autocorrelation regimes in a single model. Here, we consider only a model with an effects regime and a disturbances regime and a model with two effects regimes. The case of two disturbance regimes is contained in Brandsma and Ketellaper (1979).

4.1 One effects regime and one disturbances regime

This type of model combines equations (5) and (6):

$$\begin{aligned} y &= \rho_1 W_1 y + X\beta + \varepsilon \\ \varepsilon &= \rho_2 W_2 \varepsilon + \nu \end{aligned} \tag{14}$$

where $\nu \sim IN(0, \omega I)$. This model involves two variable transformations with two Jacobians. The log-likelihood function for the observed y is:

$$l(y) = const - \left(\frac{n}{2}\right) \ln \omega - \left(\frac{1}{2\omega}\right) [y' A_1' A_2' A_2 A_1 y - 2\beta' X' A_2' A_2 A_1 y + \beta' X' A_2' A_2 X \beta] + \\ + \ln |A_1| + \ln |A_2| \tag{15}$$

which is maximized with respect to ω, ρ_1, ρ_2, and β. The estimation equations are:

$$\begin{aligned} \hat{\beta} &= (X' A_2' A_2 X)^{-1} X' A_2' A_2 A_1 y \\ \hat{\omega} &= \left(\frac{1}{n}\right) [y' A_1' A_2' A_2 A_1 y - 2\beta' X' A_2' A_2 A_1 y + \beta' X' A_2' A_2 X \beta] \end{aligned} \tag{16}$$

where $A_1 = I - \rho_1 W_1$ and $A_2 = I - \rho_2 W_2$. The variance-covariance matrix for the estimated parameters is more complex:

$$V(\hat{\omega}, \rho_1, \rho_2, \beta) = \omega^2 \times$$

$$\times \begin{bmatrix} \frac{n}{2} & \omega tr(B_1) & \omega tr(B_2) & 0' \\ \omega tr(B_1) & \omega^2 \left\{ tr(B_1' B_1) - \alpha_1 \right\} + \omega \beta' X' C X \beta & \omega^2 tr(D) & \omega (X' A_2' A_2 W_1 A_1^{-1} X \beta)' \\ \omega tr(B_2) & \omega^2 tr(D) & \omega^2 tr \left\{ tr(B_2' B_2) - \alpha_2 \right\} & 0' \\ 0 & \omega X' A_2' A_2 W_1 A_1^{-1} X \beta & 0 & \omega X' A_2' A_2 X \end{bmatrix}^{-1} \tag{17}$$

where $B_1 = A_2 W_1 A_1^{-1} A_2^{-1}$, $B_2 = W_2 A_2^{-1}$ and $C = (A_1')^{-1} W_1' A_2' A_2 A_1^{-1}$. Further, with $V = W_2' + W_2 - 2\rho_2 W_2' W_2$ we have $D = (A_2')^{-1} (A_1')^{-1} W_1 V A_2^{-1}$. Finally, if $\{\lambda_i\}$ are the eigenvalues of W_1 and $\{\mu_j\}$ are the eigenvalues of W_2, then $\alpha_1 = \sum_i \lambda_i^2 / (1 - \rho_1 \lambda_i)^2$ and $\alpha_2 = \sum_j \mu^2 / (1 - \rho_2 \, \mu_j)^2$. See Doreian (1982) for details. As for the equations discussed earlier, equaticn (17) is used to obtain the standard errors of the parameter estimates.

4.2 Two effects regimes

The model with two regimes of network effects is

$$y = \rho_1 W_1 y + \rho_2 W_2 y + \varepsilon \tag{18}$$

with $\varepsilon \sim IN(0, \omega I)$. Defining $z = Ay$ where $A = I - \rho_1 W_1 - \rho_2 W_2$, the log-likelihood for the y is

$$l(y) = const - \left(\frac{n}{2}\right) \ln \omega - \left(\frac{1}{2\omega}\right) [z'z - 2\beta' X' z + \beta' X' X \beta] + \ln |A| \tag{19}$$

The estimation equations for β and ω are:

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1} X'z = (X'X)^{-1} X'Ay \\ \hat{\omega} &= \left(\frac{1}{n}\right) [z'z - 2\beta' X'z + \beta' X' X \beta] \end{aligned} \tag{20}$$

and from these a concentrated log-likelihood function can be constructed which is then minimized with respect to ρ_1 and ρ_2. See Doreian (1989). The variance-covariance matrix for the estimators is

$$V(\hat{\omega}, \rho_1, \rho_2, \beta) = \omega^2 \begin{bmatrix} \frac{n}{2} & \omega tr(B_1) & \omega tr(B_2) & 0' \\ \omega tr(B_1) & v_{22} & v_{23} & \omega X' B_1 X \beta \\ \omega tr(B_2) & v_{32} & v_{33} & \omega X' B_2 X \beta \\ 0 & \omega \beta' X' B_1' X & \omega \beta' X' B_2' X & \omega X' X \end{bmatrix}^{-1} \tag{21}$$

In (21) $B_1 = W_1 A^{-1}$ and $B_2 = W_2 A^{-1}$. The v_{ij} terms are

$$v_{22} = \omega^2 \left[tr(B_1' B_1) + tr(B_1^2) \right] + \omega \beta' X' B_1' B_1 X \beta$$

$$\begin{aligned} v_{23} &= \omega^2 [tr(B_1' B_2) + tr(B_1 B_2)] + \omega \beta' X' B_2' B_1 X \beta \\ v_{32} &= \omega^2 [tr(B_2' B_1) + tr(B_2 B_1)] + \omega \beta' X' B_1' B_2 X \beta \\ v_{33} &= \omega^2 \left[tr(B_2' B_2) + tr(B_2^2) \right] + \omega \beta' X' B_2' B_2 X \beta \end{aligned}$$

In the same fashion as for the previous model, (21) is used to obtain the standard errors of the parameter estimates. At face value, the derived formulae for all four types of models can be used empirically to estimate network autocorrelation models. Unfortunately, there remain many unresolved problems that require attention before these types of models can be used with confidence. One avenue for generating this confidence is via simulation where the properties of these estimation methods can be explored systematically. Before outlining a general simulation framework, it is necessary to consider some numerical issues.

5 Some numerical issues

For the single regime models it is clear that there is a clear admissible region for the parameter ρ. Consider a model where the weight matrix, W, is row stochastic. The maximum eigenvalue is 1 which has serious implications for the computation of $\ln |A|$. In the form $\sum_i \ln(1 - \rho\lambda_i)$ it is clear that if ρ is 1 and for the maximum value of λ_i, this term is $-\infty$. There is a discontinuity at $\rho = 1$. In the interval for ρ between -1 and +1, the log-likelihood function is continuous with a unique maximum. At these boundaries there are dicontinuities. Further, as ρ increases beyond 1 (and decreases beyond -1) there are multiple such discontinuities (Mellott, 1994). Additionally, the value of the log-liklihood increases indefininately as ρ increases beyond 1 and decreases beyond -1, still with the many discontinuities. See Mellott (1994). The region defined by $-1 < \rho < +1$ is called the admissible region. Numerical estimates that stray outside the admissible region are likely to fail to produce estimates or produce estimates that are worthless. These problems seem even more acute for the two regime models. Clearly, for ρ_1=1 and for ρ_2=1 there will be the same discontinuity problems. However, the admissible region is not defined by $-1 < \rho_1, \rho_2 < +1$. In preliminary simulation studies it seems that the admissible region is defined by $|\rho_1| + |\rho_2| < 1$. At a minimum, it is necessary to determine the admission region completely and to constrain numerical estimates of the ρ_i within that region. Similar arguments hold for alternative normalizations of S that can be used to create W.

At face value, the use of equations (10), (13), (17) and (21) solves the problem of obtaining the standard errors of the parameter estimates. This may be too optimistic for two reasons. First, there is some evidence that, numerically, these equations may be inferior to the use of alternative estimators of the variance-covariance matrix for the estimated parameters. These alternatives are based on either the cross product form or the heteroskedastic consistent form of the covariance matrix. See White (1980, 1981, 1982) and MacKinnon and White (1985). It seems necessary to examine these alternatives in a systematic fashion. The second reason for concern over the analytically derived covariance matrices is found in the statement of the Cramer-Rao inequality concerning the information matrix[10]. It provides *lower* bounds for the standard errors as its desirable properties are asymptotic. Clearly, for the network autocorrelation models considered here, the number of network nodes is small and n is far from the asymptote. Together, these concerns suggest the use of simulation methods to reach a better understanding of when these models can be estimated successfully when n is small.

6 A simulation design for network autocorrelation models

The approach I am taking, in collaborative work John Mellott, is to define a set of parameters that can be varied systematically across simulation runs. In broad terms, the idea is to generate data that are consistent with a wide range of parameters and to estimate models with a variety of alternative estimation methods. The objective is to reach an understanding of the relative merits of alternative numerical methods and the extent to which MLE methods are useful in estimating these types

[10]The asymptotic covariance matrix for the estimated parameters, used to generate estimates of the corresponding standard errors is

$$V = -E\left[\frac{\partial^2 l}{\partial\theta_i \partial\theta_j}\right]^{-1}$$

where θ_i and θ_j are any two parameters.

of autocorrelation models.

6.1 Alternative models

Five alternative models will be used to generate simulated data and these will all be estimated in each generated body of data. The 'rival' models are: (a) the OLS model of equation (1); (b) the single disturbances regime of equation (5); (c) the single effects regime model of equation (6); (d) the model with both an effects regime and a disturbances regime of equation (14) and (e) the two effects regime of equation (18). Clearly, when any one of these models is used to generate data, all of the remaining models will be mis-specified. One issue to explore is the costs of these mis-specifications.

6.2 Weight matricies

Perhaps the most difficult specification issue is the form of the weight matrices for the autocorrelation models. Indeed, this may be the Achilles' heel of the whole approach. I will confine my attention to *network* autocorrelation models. If the focus is on cohesion models then some direct transformation of the sociomatrix (for example, making it row stochastic) seems appropriate. As an alternative, dividing the sociomatrix by its largest eigenvalue seems reasonable. Neither transformation, however, has a strong substantive rationale.

If attention is on structural equivalence, the elements w_{ij} can be constructed to capture the extent to which two actors, i and j, are structurally equivalent. One suggestion is provided by Burt and Doreian (1982) but others are possible. Batagelj et al. (1992) point out that, for partitioning purposes, it is essential to have measures of structural equivalence that are compatible with this kind of equivalence. Whether network autocorrelation models are robust enough to permit the use of other (dis)similarities that are not compatible is not known. Establishing measures of the extent to which two actors are regular equivalent is much more difficult than for regular equivalence. Perhaps the best progress will come from explicit conceptualizations of the processes involved rather than the arbitrary choice of weight matrices.

Another way in which weight matrices can differ is in their complexity. At an intuitive level, this can be viewed in terms of network density. If W is little different from the identity matrix, I, there is little point in estimating autocorrelation models. At some (unknown) density, there is a threshold which, when crossed, means that autocorrelation models ought to be considered. At the other extreme, if W is close to the universal matrix, U, with every every actor equivalent, then the model will be intractable. There will be another threshold which, when crossed, means there is no point in even attempting to use network autocorrelation models. It would be useful to know where these thresholds are located.

6.3 Autocorrelation parameters

These are the core parameters that capture the extent to which autocorrelation is present. At noted above, for the single regime models $-1 < \rho_i < 1$. There may be thresholds here also. For values of ρ that are close to 0, it seems reasonable to ignore autocorrelation issues. Of course, the unknown is what constitutes 'close to 0'. This is complicated somewhat by the estimated standard errors. Previous simulations (Doreian et al. 1984) suggest that only at values of ρ at or above 0.3 is it possible to detect the presence of autocorrelation. It seems that as ρ increases, the standard errors decrease making inference about ρ a sharp instrument at high values of ρ (where it is obvious that there is autocorrelation) and a blunt instrument where we really need a sharp instrument. Perhaps different numerical methods will make a difference here.

The case with multiple regimes is even more complicated. In part, the admissible region is bounded by $\frac{1}{\min(\lambda)} < \rho_1 < \frac{1}{\max(\lambda)}$ and $\frac{1}{\min(\mu)} < \rho_2 < \frac{1}{\max(\mu)}$. In preliminary simulation efforts we have obverved that every time $\rho_1 + \rho_2$ is 1, the estimation method breaks down (at least once for a given set of parameters where the sum of the two autocorrelation parameters is 1). Note, these are stated for the values used to simulate the data. Not all batches of simulated data, with such a set of parameters, leads to a breakdown in the estimation. In a specific data set, the nature of the admissible region may be defined by $|\hat{\rho}_1| + |\hat{\rho}_2| = 1$ but we do not know this. Nor do we have a good understanding of the performance of these estimators as the boundary of the admissible region is approached.

6.4 Error variance

For regression models we know that as the error variance, ω, increases the standard errors for the estimates rise and confidence intervals are broader for a chosen significance level. We expect that the same will happen with network autocorrelation models. We do not know how this behavior plays out with different configurations of other parameters. It seems prudent, then, to include ω as a parameter that varies across simulations.

6.5 Size of the network

Throughout the derivation of equations to yield estimates of the standard errors for the estimated parameters, use has been made of the inverse of the information matrix. Yet, the Cramer-Rao inequality tells us that this use of the information matrix gives us lower bounds for the standard errors. We do not know how far our estimates of the standard errors are above the lower bounds and it seems that simulation is the only way of establishing this. The complication here is that changing n may also change the structure of S and hence W. As a first step, we have gone back to the spatial autocorrelation literature and the use of lattices. More specifically, we can lay out a set of 'areas' in a chess board configuration. The weight matrices can be constructed in terms of rook moves and, say, white bishop moves. If the values of n^2 are taken as, say, 16, 25, 36, 64, 81, 100,..., 256 etc., it is possible to keep the structure of the weight matrices fixed while changing n. A second approach would be to fix density parameters and generate W randomly with the fixed densities. The problem here is that this is only an indirect control of the structure of W. And even if this is successful, we know that sociomatrices are far from random.

6.6 Regressors

As a first step it seems reasonable to use Xs that are uncorrelated or, more stringently, independent. With that done, the next step will be to examine non-zero covariance structures of the regressors to see how this operates in the context of the autocorrelation models. Intuitively, it seems that collinearity will be a problem here also.

On the basis of the knowledge generated from these simulations, it seems appropriate to examine the impact of having some or all of the regressors autocorrelated also. Indeed, if we are ready to contemplate both y and ε being autocorrelated, it seems unreasonable to consider the regressors, X, as not being autocorrelated. As we consider this set of issues, it is reasonable to have only some of the regressors autocorrelated. Additionally, the network autocorrelation regime for one (or more) Xs could be defined in terms of either W_1 or W_2. As a step beyond this, the regressors could have some *other* autocorrelation regime.

6.7 Boundary problems

Locating the boundaries of networks is a very difficult problem. See Laumann et al. (1983) for a statement of the problem and Doreian and Woodard (1994) for a suggested partial solution. It is always possible, indeed likely, that a network is studied that is really a subnetwork of a larger empirical entity. This is a direct parallel with spatial autocorrelation models. In that contect, a region is located within a wider region. In both cases the boundary problem is serious. The following is taken from Anselin (1988). Let G be a subnetwork (or subregion) that is located in a wider network (region) where H is the complement of G in the wider network (region). Confining attention to the single effects model, the model can be written in the following partitioned form:

$$\begin{bmatrix} y_g \\ y_h \end{bmatrix} = \rho \begin{bmatrix} W_{gg} & W_{gh} \\ W_{hg} & W_{hh} \end{bmatrix} \begin{bmatrix} y_g \\ y_h \end{bmatrix} + \begin{bmatrix} X_g \\ X_h \end{bmatrix} \beta + \begin{bmatrix} \varepsilon_g \\ \varepsilon_h \end{bmatrix} \tag{22}$$

in obvious notation. Suppose the wider network is ignored. Then the model that would be estimated, given the model in (22), is:

$$y_g = \rho W_{gg} y_g + X_g \beta + \varepsilon_g$$

while, on the basis of (22), the model that should be estimated is:

$$y_g = \rho W_{gg} y_g + \rho W_{gh} y_h + X_g \beta + \varepsilon_g$$

which clearly differs. This last equation can be re-written as:

$$y_g = \rho W_{gg} y_g + X_g \beta + (\rho W_{gh} y_h + \varepsilon_g)$$

where the term in parentheses is unlikely to be $\sim IN(0, \omega I)$. Simulation would seem a reasonable way in which the magnitude of the boundary problem could be gauged. The prior work in the spatial autocorrelation literature provides a good point of departure. See Anselin (1986), Anselin and Griffith (1988) and the references therein.

7 Discussion

I have couched my treatment of network autocorrelation in terms of the simple OLS models. It seems reasonable that similar problems will occur for other members of the family of linear models. For example, Griffith (1992)extends the treatment of spatial autocorrelation models to N-way ANOVA models. McMillen (1992) treats probit models by including spatial autocorrelation tems. I doubt that log-linear models are immune from these problems. For spatial categorical data, Fingleton (1983) suggests adjustments in the computation of chi-square. Reitz and Dow (1989)continue this line of thought for network autocorrelation models. "Analytical, empirical and simulation evidence suggest that chi-square and likelihood ratio tests reject at rates substantially lower than the nominal rate whenever the data have been generated using a sampling scheme that does not ensure independence of sample units". For some of their examples, chi-square values are hugely inflated for some regimes of network autocorrelation and they suggest heuristic-based ways of deflating these figures to compensate for autocorrelation problems. I suspect also that structural equation models may need to be reexamined whenever network autocorrelation is present.

There is clearly much to be done: when all of the options are laid out, I am talking of thousands of simulation runs. Thus far, the software for network autocorrelation estimation and preliminary simulations (Mellott, 1994) are written in GAUSS (Aptech, 1992). Even with 486 DX machines running at 66 mhertz these simulations are very time consuming for the two regime models. Two people clearly cannot do this alone without serious support. If you agree that autocorrelation is a serious problem, I invite you to join the effort and put your shoulder to the wheel.

References

[1] Aldrich H. (1979): *Organizations and Environments.* Prentice-Hall, Englewood Cliffs.

[2] Anselin L. (1986): Non-nested tests on the weight structure in spatial autoregressive models: some monte carlo results. *Journal of Regional Science*, **26**, 267–284.

[3] Anselin L. (1988): *Spatial Econometrics: Methods and Models.* Kluwer, Dordrecht.

[4] Anselin L. and Griffith D.A. (1988): Do spatial effects really matter in regression analyses? *Papers of the Regional Science Association*, **65**, 11–34.

[5] Aptech (1992): *The Gauss Programming Language.* Aptech.

[6] Batagelj V., Ferligoj A., and Doreian P. (1992): Direct and indirect methods for structural equivalence. *Social Networks*, **14**, 63–90.

[7] Brandsma A. and Ketellaper R. (1979): A biparametric approach to spatial autocorrelation. *Environment and Planning A*, **11**, 51–58.

[8] Burt R.S. (1983): Cohesion versus structural equivalence as a basis for network groups. In R. S. Burt and M. J. Minor (Eds.), *Applied Network Analysis*, Sage, Beverly Hills, 262–282.

[9] Burt R.S. and Doreian P. (1982): Testing a structural model of perception: conformity and deviance with respect to journal norms in elite sociological methodology. *Quality and Quantity*, **16**, 109–150.

[10] Chase-Dunn C. (1975): The effects of international economic dependence on development and inequality: A cross national study. *American Sociological Review*, **40**, 720–738.

[11] Crenshaw E. (1991): Foreign investment as a dependent variable: determinants of foreign investment and capital penetration in developing nations, 1967-1978. *Social Forces*, **69**, 1169–1182.

[12] Crenshaw E. (1992): Cross-national determinants of income inequality: A replication and extension using ecological-evolutionary theory. *Social Forces*, **71**, 339–363.

[13] Doreian P. (1980): Linear models with spatially distributed data: spatial disturbances or spatial effects? *Sociological Methods and Research*, **9**, 29–60.

[14] Doreian P. (1981): Estimating linear models with spatially distributed data. *Sociological Methodology, 1981*, 359–388.

[15] Doreian P. (1982): Maximum likelihood methods for linear models: spatial effects and spatial disturbance terms. *Sociological Methods and Research*, **10**, 243–269.

[16] Doreian P. (1989): Two regimes of network autocorrelation. In M. Kochen (Ed.), *The Small World*, Ablex, Norwood, 280–29.

[17] Doreian P., Teuter K., and Wang C.-H. (1984): Network autocorrelation models: some monte carlo results. *Sociological Methods and Research*, **13**, 155–200.

[18] Doreian P. and Woodard K.L. (1994): Defining and locating cores and boundaries of social networks. *Social Networks*, **16**, 267–293.

[19] Dow M., Burton M., White D.R., and Reitz K.P. (1984): Galton's problem as network autocorrelation. *American Ethnologist*, **11**, 754–770.

[20] Fingelton B. (1983): Independence, stationarity, categorical data and the chi-squared test. *Environment and Planning A*, **15**, 483–499.

[21] Griffith D.A. (1992): A spatially adjusted n-way anova model. *Regional Science and Urban Economics*, **22**, 347–369.

[22] Jackman R. (1980): A note on the measurement of growth rates in cross-national research. *American Journal of Sociology*, **86**, 604–617.

[23] Laumann E.O., Marsden P.V., and Prensky D. (1983): The boundary specification problem in network analysis. In R. S. Burt and M. J. Minor (Eds.), *Applied Network Analysis: A Methodological Introduction*, Sage, 18–34.

[24] Loftin C. (1972): Galton's problem as spatial autocorrelation: comments on ember's empirical test. *Ethnology*, **11**, 425–435.

[25] Loftin C. and Ward S.K. (1981): Spatial autocorrelation models for galton's problem. *Behavior Science Research*, **16**, 105–128.

[26] MacKinnon J.G. and White H. (1985): Some heteroskedastic-consistent covariance matrix with improved finite sample properties. *Journal of Econometric*, **29**, 305–325.

[27] McMillen D.P. (1992): Probit with spatial autocorrelation. *Journal of Regional Science*, **32**, 335–348.

[28] Mellott J.L. (1994): The numeric estimation of maximun likelihood spatial models with more than one spatial parameter. 1994.

[29] Mitchell E.J. (1969): Some econometrics of the huk rebellion. *American Political Science Review*, **64**, 1159–1171.

[30] Narroll R. (1976): Galton's problem and hraflib. *Behavior Science Research*, **2**, 123–148.

[31] Nolan P.L. (1983): Status in the world system: income inequality and economic growth. *American Journal of Sociology*, **89**, 410–419.

[32] Ord K. (1975): Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, **70**, 120–126.

[33] Reitz K.P. and Dow M.M. (1989): Network interdependence of sample units in contingency tables. *Journal of Mathematical Sociology*, **14**, 85–96.

[34] Rubinson R. (1976): The world economy on the distribution of wealth within states: a cross-national study. *American Sociological Review*, **41**, 638–659.

[35] Seidman S.B. (1983): Network structure and minimum degree. *Social Networks*, **5**, 269–287.

[36] Simpson M. (1990): Political rights and income inequality: A cross-national test. *American Sociological Review*, **55**, 682–693.

[37] Smith D.A. and White D.R. (1992): Structure and dynamics of the global economy: network analysis of international trade 1965-1980. *Social Forces*, **71**, 857–893.

[38] Snyder D. and Kick E. (1979): Structural position in the world system and economic growth 1955-1970: A multiple network analysis of transnational interactions. *American Journal of Sociology*, **84**, 1096–1126.

[39] Tyler E.B. (1889): On a methods of investigating the development of institutions applied to the laws of marriage and descent. *Journal of the Royal Anthropological Institute*, **18**, 272.

[40] Wallestein I. (1974): *The Modern World System.* Academic.

[41] White D.R., Burton M.L., and Dow M.M. (1981): Sexual division of labor in african agriculture: A network autocorrelation analysis. *American Anthropologist*, **83**, 824–849.

[42] White D.R. and Reitz K.P. (1983): Graph and semigroup homomorphisms on networks of relations. *Social Networks*, **5**, 193–234.

[43] White H. (1980): A heteroskedastic-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, **48**, 817–838.

[44] White H. (1981): Consequences and detection of misspecified nonlinear regression models. *Journal of the American Statistical Association*, **76**, 419–433.

[45] White H. (1982): Maximum likelihood estimation of misspecified models. *Econometrica*, **50**, 1–25.

[46] Woodard K.L. and Doreian P. (1994): Child service systems: Life cycle based models for differential identification and treatment of mental health problems.