Cross-Cultural Research

The Journal of Comparative Social Science

Volume 41 Number 4 November 2007

SAGE Publications
Los Angeles • London • New Delhi • Singapore

Cross-Cultural Research
Volume 41 Number 4
November 2007 336-363
© 2007 Sage Publications
10.1177/1069397107305452
http://ccr.sagepub.com
hosted at
http://online.sagepub.com

Galton's Problem as Multiple Network Autocorrelation Effects

Cultural Trait Transmission and Ecological Constraint

Malcolm M. Dow
Northwestern University

Empirical evidence that cultural traits are often nonrandomly distributed because of the individual or combined effects of common history, diffusion, borrowing, and/or other types of cultural transmission processes has been accumulating for decades. Because many cultural traits have recently been shown to be influenced by more than one transmission process, it has become a methodological priority in comparative research to develop statistical methods that can simultaneously incorporate multiple transmission processes. This article proposes a multiple network autocorrelation effects model and associated two-stage least squares (2SLS) estimation procedures. The network autocorrelation effects model offers an alternative interpretation of how cultural trait transmission processes operate than does the network autocorrelation disturbances model. Conceptual differences between the two classes of models suggest that the network effects specification will be more generally applicable in comparative studies. An empirical example demonstrates the substantive value of the multiple network autocorrelation effects model and the widely available 2SLS estimation procedures.

Keywords: *Galton's Problem; network autocorrelation effects model; two-stage least squares; trait transmission; ecological constraint*

Author's Note: The present article is the first in a new series of papers that deal with issues related to Galton's Problem and network autocorrelation modeling in cross-cultural research. A second article (under review) assesses the nature and extent of network autocorrelation in cross-cultural data, and a third article, currently underway, extends the network autocorrelation effects model two-stage least squares approach to deal with categorical dependent variables. Please address correspondence to Malcolm M. Dow, Department of Anthropology, Northwestern University, 4247 N. Paulina St, Chicago, IL 60613; e-mail: mmd383@northwestern.edu.

Empirical evidence has been steadily accumulating throughout many decades to show that a large percentage of cross-cultural traits are nonrandomly distributed because in part of the individual or combined effects of common history, diffusion, borrowing, and/or other types of cultural transmission. In a recent article that clearly demonstrates such effects, Eff (2004) reports the percentages of 1,700 Standard Cross-Cultural Sample (SCCS) variables that display statistically significant levels of autocorrelation, that is, similarities of same-variable scores on spatially or otherwise related societies. Using the well-known Moran's *I* statistic as the measure of autocorrelation, Eff reports that about 44% of the variables exhibit spatial autocorrelation and about 43% exhibit linguistic autocorrelation.[1] Although Eff does not report the percentage of variables that show at least one of these types of autocorrelation, more recent work suggests that approximately 53% of the SCCS variables exhibit either spatial or language autocorrelation globally (Dow & Eff, 2007).

Other studies have reported similar findings using similar methods. For example, White (1993) also used Moran's *I* statistic to assess levels of spatial autocorrelation in 1,178 variables for a half sample of the SCCS societies and reports that "35% of the variables showed moderate to strong autocorrelation and 26% showed severe or extreme autocorrelation" (p. 460).[2] Using a regional subset of data from the *Ethnographic Atlas*, Guglielmino, Viganotti, Hewlett, and Cavalli-Sforza (1995) assessed the fit of 47 cultural traits across 277 African societies to language, spatial, and environmental similarity matrices. Nineteen variables (40%) grouped into economic, social stratification, and house-building categories were reported to be associated with all three autocorrelation processes, whereas 28 (60%) of the variables grouped into family and kinship, division of labor by sex, and the "various others" categories were associated with a single process. In a similar type of regional study, Hewlett, de Silvestri, and Guglielmino (2002) examined the fit of 109 "semes" (cultural traits) to the following four network autocorrelation matrices for 36 African societies: demic diffusion, cultural, linguistic, and spatial distances. Forty-five (41%) of the semes were found to be associated with a single autocorrelation process and 29 (27%) potentially with two processes. Although all of the above studies clearly demonstrate extensive lack of independence within variables associated with the most commonly used comparative data sets, they are of course simply the latest generation of studies to report such empirical findings. One only has to look at the maps in Driver and Coffin (1975), Jorgensen (1980), and Burton, Moore, Whiting, and Romney (1996) to appreciate the extent of clustering (i.e., autocorrelation) of cross-cultural variables in both regional and world-wide samples.

Although the statistical inference problems associated with the analysis of such autocorrelated variables are by now well understood—biased and inconsistent parameter estimates, inflated Type I inferential errors—there is nonetheless a surprising resistance among cross-cultural researchers to using either sample selection methods or more recent autocorrelation methods for dealing with this data-analytic problem. Korotayev and de Munck (2003) comment that there are eight solutions to nonindependence based on sampling restrictions "which virtually none of the practicing cross-cultural researchers actually ever use" (p. 354). These authors go on to note that "the one technique used by almost all cross-cultural researchers . . . is to select one tribe from each culture area" (p. 354) to control for nonindependence of cultural traits. That is, despite clear demonstrations that such sampling restrictions do not begin to control for nonindependence in cross-cultural data (Dow, 1993; Loftin & Hill, 1974; Murdock & White, 1969), this inadequate approach is still commonly accepted by cross-cultural researchers (e.g., Barry & Yoder, 2002) as providing a sufficient safeguard against the kinds of inferential errors commonly encountered with the analysis of autocorrelated data. And not only are the older sampling approaches seldom employed, the network autocorrelation disturbances regression model introduced into the cross-cultural literature by Dow (1979), Dow White, and Burton (1983), and Dow, Burton, White, and Reitz (1984) has been applied only twice (Dow, 1984; White, Burton, and Dow, 1981).

The present article extends the discussion of network autocorrelation beyond network disturbances models to consideration of network effects models. Importantly, the suggested two-stage least squares (2SLS) estimation procedure for the network effects models is available in most of the commonly used statistical software packages. No specialized maximum likelihood (ML) computational software is required, unlike previously proposed network disturbances autocorrelation models. In addition, it is a straightforward matter to include more than one network autocorrelation process simultaneously with ecological and functional associations in the effects model and estimate it using 2SLS regression procedures. The ability to simultaneously test hypotheses about language, distance and other transmission processes, such as trade networks (if data are available), in addition to ecological constraints and functional associations, represents an important advance over other comparative methods that are restricted to single phylogenetic language trees when controlling for nonindependence (Borgerhoff Mulder, George-Cramer, Eshleman, & Ortolani, 2001; Mace & Pagel, 1994).

Expanding the choice of appropriate autocorrelation methods and simplifying the estimation procedures should help to overcome the apparent reluctance among cross-cultural researchers to embrace appropriate statistical models and perhaps help change the apparent perception that the lack of independence within traits (as Galton's Problem is known in anthropology) is a mere statistical nuisance with little or no bearing on the validity of statistical results. Network dependencies potentially have a dual impact on the analysis of cross-cultural data. On one hand, they can be viewed as a source of nuisance and error when traditional statistical methods requiring independence are employed. On the other hand, as suggested by Witkowski (1974) and Korotayev and de Munck (2003), these interdependencies also represent valuable information that may be exploited for further understanding of the relative importance of the various transmission processes that are thought to be responsible for much of the cultural heterogeneity so clearly evident in cross-cultural data sets. The role of various transmission processes in shaping the cultural trait diversity that is apparent across societies is still far from clear, as is their interrelationships with ecological constraints. By helping to disentangle the relative effects of vertical transmission (parental to offspring populations) and horizontal transmission (borrowing, trade, copying, etc.) while controlling for possible ecological constraints and at the same time separating out their effects from functional associations, the multiple network autocorrelation effects model presented here represents an important advance over previously proposed methods for handling network data dependencies.

The organization of the article is as follows. The first section describes how network autocorrelation effects for language, spatial distance, or any other hypothesized cultural transmission process can be formally expressed as connectivity matrices between all pairs of societies in the sample. Next, the network disturbances and network effects models are described, and underlying differences in interpretation between the two models are outlined. Then, the 2SLS estimation procedure for the network effects model is described, and the construction of the required instrumental variables is discussed. The multiple transmission network effects with ecological constraints and functional associations are then given substantive meaning and illustration through a detailed reanalysis of Eff's (2004) regression results.

Formal Expression of Network Autocorrelation

Simply stated, spatial autocorrelation implies that what happens at one location in space is in some way related to what happens at nearby locations. More generally, network autocorrelation implies that the attributes of each node in a network can be predicted in part from knowledge of the attributes at related nodes. Hence, the first step prior to specifying any network autocorrelation regression model in comparative studies is to construct the network matrices that correspond to the sociocultural transmission processes of interest. In the network autocorrelation regression models described later, autocorrelation depends on an explicit specification of the strength of relatedness between each pair of societies in the sample. Early approaches to this problem using one-dimensional "diffusional arcs" were introduced by Naroll (1961) and Loftin and Hill (1974), and more realistic two-dimensional matrix specifications were proposed by Wirsing (1974) and Pryor (1976). The weakness in the one-dimensional arcs approach is that the multiway network dependencies are clearly misspecified, and in the latter two matrix proposals, the dependency matrix is not incorporated into the estimation procedures for the regression (or any other multivariate) model.

Correct specification of the N-by-N pairwise dependency matrix is important. A regression model is correctly specified when the equations in the model adequately correspond to the processes generating the observed variation in the data. Model specification is a matter of degree: The better the theoretical understanding of the generating processes, the closer the correspondence between the postulated equations and the empirical data, and the more precise the estimates of the model parameters. Probably the most frequently hypothesized cause of similarities of cultural traits across societies is common history. Attempts to control for such interdependencies within traits typically base the measurement of common history on some form of language relatedness between sample units. Jorgensen (1969), for example, constructed a matrix of similarities for 31 Salish societies based on number of common words in an extensive word list, and White et al. (1981) and Dow (1984) measured linguistic distance using a count of the number of steps between societies on a hypothesized phylogenetic tree of languages for a regional sample of 43 SCCS African societies.

Network autocorrelation regression models assume that the N sample units are somehow differentially interrelated and that the interrelations can be operationalized as an N-by-N connectivity weights matrix C, where the individual matrix elements c_{ij} correspond to some measure of relationship societies i and j. Several points should be noted about the construction of a connectivity C matrix. First, by convention, a node, or society, is not connected to itself, so the C matrix will have 0s along the main diagonal. Second, although no restrictions need be placed on the weights c_{ij}, the interpretation of the autocorrelation parameter estimate is more interpretable if each row of C is rescaled to sum to 1. That is, each element of the C matrix is divided by its row sum to yield a new stochastic matrix W, where $w_{ij} = c_{ij} / \sum_j c_{ij}$ for all i. One result of this scaling is that in general, $w_{ij} \neq w_{ji}$ as i and j will not usually be connected to the same number of other societies in the same way. A second result of this scaling is to facilitate the interpretation of the autocorrelation estimate as a "correlation," as the original y dependent variable and the autocorrelated variable Wy will be on the same scale. Third, this matrix formulation permits rich expression to the idea of interdependence. Any society may be connected to any other society, not simply those that are actually contiguous, spatially or otherwise. Fourth, any relevant linkage between units that can be quantified can be represented by the appropriate elements of the W matrix.

Selection of the weights is of major importance, as spurious results may arise if the hypothesized weights do not correspond to any real social process (Cliff & Ord, 1981). Commonly, in geographical research, the weights (c_{ij}) are estimated according to some notion of space-friction constraints on the possibility of effects from one unit to another. The simplest function in this case is an exponentially decaying distance function such as d_{ij}^{-a}, where d_{ij} is the distance from society i to society j and a is a suitable exponent. More general notions of interaction between nodes are also possible. For example, in their study of state budgetary expenditures in the United States, Case, Rosen, and Hines (1993) construct W matrices based on per capita income and percentage of population that is Black for each state, and Conley and Topa (2002) use differences in ethnic and occupational distributions between Chicago census tracts in their study of unemployment rates.

In addition to operationalizing linkages and flows between sample units, which is one way that trait transmission processes clearly operate, it may also be useful to construct a W matrix that represents similarities in ecological constraints faced by each pair of societies. Ecological setting is often seen as a constraint on possible trait configurations, which may modify or even rule out adoption of certain cultural practices (Whiting, Sodergren, & Stigler, 1982). There is a six-category ecological type variable coded for the SCCS, and a similarity matrix can be constructed using this code, as suggested below. Row-normalizing such a matrix to unity and including a corresponding Wy predictor variable in the regression model yields an estimate of the constraint exerted on the trait of interest (i.e., y) by the environment and controls for this effect with respect to transmission processes and functional associations. There is nothing in principal that prohibits the inclusion of "constraint" types of matrices in these network autocorrelation models. Using a distance measure between societies based on ecological setting is precisely the same idea as using percentage of population that is Black as a measure to identify similarly situated states.

Network Autocorrelation: Effects and Disturbances Models

All of the network autocorrelation models of interest here are extensions of the usual linear regression model, which can be compactly stated in matrix form as

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

where y is an $N \times 1$ column vector of dependent variable observations, X is an $N \times k$ matrix of independent variables plus an initial column of 1s for the intercept term, β is a k x 1 column vector of regression coefficients, and ε is an $N \times 1$ column vector of independent normally distributed error terms with equal variances—that is, $\varepsilon \sim IN(0,\sigma^2\mathrm{I})$. Note that this model completely ignores any dependency or network structure in the data. When the assumption about independence in this model is not met but the above regression model is nonetheless estimated using the usual ordinary least squares (OLS) procedures, then the regression coefficient estimates are either unbiased but inefficient or biased and inconsistent, depending on how the underlying network autocorrelation mechanism is assumed to enter into the regression model (Ord, 1975). Two different assumptions are made in this regard, both of which are examined next.

The simplest linear model that specifies network effects, as represented in an appropriate W matrix, assumes that the values on the dependent variable for each society i, y_i, are a function of the values of the societies to which i is linked via the W matrix. Expressing this in a matrix notation gives the following linear network autocorrelation model:

$$y = \rho Wy + \varepsilon, \tag{2}$$

where as before the error terms are normally distributed with zero mean and equal variances. Without loss of generality, we assume that both variables are mean corrected and the usual intercept term is thus 0. Rho is a scalar autocorrelation parameter, analogous to a correlation coefficient. This is the pure network effects model, where the score on a cultural trait y for society i is predicted from a weighted average of the scores on the same trait for the societies that i is related to through the process embodied in the W matrix (i.e., the ith row of W). Although it might seem that because Wy is simply a vector of scores and that the rho parameter could be estimated using OLS regression procedures, the fact that the Wy variable and the ε term are correlated means that the OLS estimate of rho will be biased and inconsistent (Johnston, 1984).

In most research situations, of course, there will be additional independent predictor variables that are also hypothesized as contributing causes of the dependent variable scores. These variables are easily introduced simultaneously into the pure network effects model to give the following network autocorrelation effects model:

$$y = \rho Wy + X\beta + \varepsilon, \tag{3}$$

where again the errors are assumed to be independently normally distributed with zero means and equal variances. Note that when $\rho = 0$ and there are no network effects, this model defaults back to the OLS regression model (1), and when $\beta = 0$, the pure network effects model (2) results. Again, because the Wy variable and the error terms are correlated, OLS estimation of this model yields biased and inconsistent estimates of both rho and the regression coefficients for the X variables. Details on the derivation of ML estimation (MLE) procedures for models (2) and (3), which yield consistent parameter estimates, are given in Ord (1975) and Anselin (1988). 2SLS estimation for model (3) is discussed in Land and Deane (1992) and Kelejian and Prucha (1998).

Frequently in cross-cultural research, it is hypothesized that more than one network transmission process is operating simultaneously—that is, both language and geographic distance transmission effects (among others) are thought to be relevant to the observed distribution of trait scores. The network autocorrelation effects model is readily extended to incorporate two regimes of network transmission processes as follows:

$$y = \rho_1 W_1 y + \rho_2 W_2 y + X\beta + \varepsilon \tag{4}$$

Doreian (1989a, 1989b) provides a derivation of the MLE procedures for this model, and Lacombe (2004) provides a nicely detailed empirical example. No specialized software routines to implement the nonlinear optimization MLE approach are currently available to estimate this complicated model. However, the 2SLS procedure used to estimate the single network effects model is easily extended to handle this model. Indeed, generalization of model (4) to incorporate additional network transmission processes is also straightforward, and estimation of even more general models using 2SLS is limited only by availability of data required to construct the W matrices and by the sample size. Details on the 2SLS methods are provided below.

A different class of network autocorrelation regression models was previously proposed for cross-cultural data analysis by Dow (1979), Dow et al. (1983), and Dow et al. (1984). In these models, network autocorrelation is hypothesized to affect the disturbances term in the usual regression model (1). The network autocorrelated disturbances model is specified as

$$y = X\beta + \varepsilon, \quad \varepsilon = \rho W\varepsilon + \nu \tag{5}$$

Here, the disturbance term for each society is expressed as a weighted average of the disturbances at related societies, plus an independently normally distributed error term ν. Derivation of the MLE procedures for this model are given in Ord (1975). As before, this model can be generalized to incorporate multiple network transmission processes as

$$y = X\beta + \varepsilon, \quad \varepsilon = \rho_1 W_1 \varepsilon + \rho_2 W_2 \varepsilon + \nu \tag{6}$$

This more general model is discussed in Dow (1984) as the biparametric autocorrelated disturbances model, and an empirical example is presented. Again, there is no available estimation software for this model, and generalization of the model beyond two regimes of autocorrelation appears to be completely intractable using MLE approaches.

A natural generalization of the above models incorporates both network effects and disturbances models as follows:

$$y = \rho_1 W_1 y + X\beta + \varepsilon, \quad \varepsilon = \rho_2 W_2 \varepsilon + \nu \tag{7}$$

Note that if either of the ρ_1 or ρ_2 autocorrelation parameters is equal to 0, the effects or disturbances model becomes appropriate. Anselin (1988) provides an extended discussion of this effects/disturbances model and outlines MLE methods. Kelejian, Prucha, and Yuzefovich (2004) combine a 2SLS approach with a generalized method of moments estimation procedure to obtain estimates for this model.[3]

Interpreting Trait Transmission as Network Autocorrelation

Although the above typology of network autocorrelation models clearly provides for rich expression of the idea of sociocultural transmission processes appropriate for use with comparative data, it is worth reviewing

briefly what the two conceptually distinct models, effects and disturbances, imply about how the transmission processes are hypothesized to operate. To more clearly understand what these models imply about the underlying network processes they embody, it will be helpful to take another look at the usual OLS regression model with no network dependencies assumed. In the usual OLS model,

$$y = X\beta + \varepsilon \tag{8}$$

the $X\beta$ term can be thought of as the intrinsic position of each sample unit. Absent any sociocultural transmission process operating between sample units, the predicted score on the dependent y variable for each sample unit is a function of the X "local effects" (i.e., other attributes of that sample unit). Without any social context that might influence scores on the y cultural trait of interest, the only hypothesized influences on y are the presence and levels of other cultural traits within the same society. That is, cultural traits are functionally interrelated within societies and uninfluenced by the configuration of cultural traits in related societies.

The interpretation of the network effects model is, however, quite different. Looking again at the network autocorrelation effects model,

$$y = \rho Wy + X\beta + \varepsilon \tag{9}$$

the y variable is now hypothesized to be sensitive to, and influenced by, the "social context" of each society (i.e., the other societies to which it is related through common history, trade, conquest, demic migration, and so forth). That is, the presence and level of the y trait for each society is now hypothesized to be a function of the scores on that same trait within the set of societies to which each society is somehow related, in addition to local X factors. This network autocorrelation effects model has been described as an "endogenous feedback model of social conformity" by Erbring and Young (1979). In their conceptualization, network effects arise out of ongoing social processes, such as contagion, competition, and conformity, that derive from the reciprocal influences of social actors interacting with one another. In the analysis of cross-cultural data, the network effects model would also appear to be applicable to "contagion" brought about both through "vertical" inheritance of cultural traits and through "horizontal" effects caused by conquest, trade, borrowing, and other such mechanisms.

The network disturbances model, on the other hand, represents a quite different mechanism through which network processes can operate. Starting again with the usual OLS regression model, $y = X\beta + \varepsilon$, it is clear that the $X\beta$ term represents the expected, intrinsic value on the dependent y trait, whereas the error term, ε, represents all of the unmeasured, latent factors that tend to cause a deviation from the intrinsic position. As Leenders (2002) notes, autocorrelation in the disturbances term can be due to at least three different factors. First, incorrect specification of the $X\beta$ term as linear when one or more variables are nonlinearly related to the dependent variable can also result in autocorrelation of the error terms. Correcting the model specification removes the observed autocorrelation. Second, because the set of independent X variables seldom includes all of the important effects on the y variable, autocorrelation of the error terms may arise because of an omitted variable. Including the omitted variable(s) causes the autocorrelation to vanish. This is the approach adopted by Pryor (1976) and White et al. (1981), where dummy variables were included as predictors to control for autocorrelation in the residuals from an initial OLS regression analysis. That is, in both of these situations, autocorrelation is treated as indicating a misspecification of the regression model, which can be corrected either by the transforming one or more variables or by adding an additional independent variable. A third, network-based interpretation for the autocorrelated disturbances is a model of deviation from the intrinsic position because of insecurity, level of risk acceptance, and uncertainty. Here, individuals (or societies) adjust the amount of deviation from their intrinsic position based on the deviations observed in others to whom the focal actor is related by the mechanism represented in the W matrix. Leenders (2002) characterizes the substantive differences between the network autocorrelation effects and disturbances model specifications in terms of interaction and reaction. The network autocorrelation effects model is understood as a model of interaction between societies, where the interactions are hypothesized to lead to mutual adjustments and adoption of modifications, including inheritance, with respect to cultural configurations of traits. The network autocorrelation disturbances model is a model of reaction, or deviation from an intrinsic position in the face of perceived risk and uncertainty. Given the nature of the cultural transmission processes (e.g., inheritance of traits from ancestor populations, trade, borrowing, copying, and the like), it appears that the multiple network autocorrelation effects model will be a more generally appropriate model specification than the network disturbances model when analyzing cross-cultural data, especially if relative contributions of different cultural trait transmission processes are of theoretical interest.

Estimating the Multiple Network Effects Model: 2SLS

In this article, the general network autocorrelation effects model with multiple forms of dependency that is of primary interest is

$$y = \rho_1 W_1 y + \rho_2 W_2 y + \ldots + \rho_t W_t y + X\beta + \varepsilon \tag{10}$$

Extension of the 2SLS estimation procedures suggested by Land and Deane (1992) and Kelejian and Prucha (1998) for a single network autocorrelation effect to two or more network effects, as in model (10), is straightforward. The primary limitations to employing the model are the data required to construct the separate weight matrices and sample size.

The estimation problem with the multiple network effects model (10) arises from the fact that all of the $W_i y$ endogenous network variables are correlated with the error term ε, as all are functions of the dependent y variable. This lack of independence leads to biased and inconsistent estimates for both the autocorrelation parameters and the regression coefficients if the model is estimated using OLS regression methods (Ord, 1975). However, this problem can be overcome if, for each $W_i y$ variable, one or more variables can be found that are correlated with it but are uncorrelated with the error term ε. Such "instrumental" variables can then be used to obtain predicted scores on each of the corresponding $W_i y$ variables, and the vectors of predicted scores then substituted into the original equation in place of the original variable. In the statistical and econometrics literature, the terms instrumental variables estimation and 2SLS are treated as synonyms.

The 2SLS estimation of the autocorrelation parameters and regression coefficients can be quite easily carried out using the following two-step OLS procedure. For simplicity, we initially rewrite model (10) as

$$y = Y_1\gamma + X_1\beta + \varepsilon \tag{11}$$

where

y is the dependent variable
Y_1 is the $N \times t$ matrix of observations on the endogenous variables (all of the $W_i y$ network autocorrelated variables) in the equation,
γ is a $t \times 1$ column vector of the t autocorrelation coefficients $(\rho_1, \rho_2, \ldots \rho_t)'$, where $'$ denotes transpose,
X_1 is the $N \times k$ matrix of observations on the k predictor variables,
β is a $k \times 1$ vector of regression coefficients, and
ε is a $N \times 1$ vector of disturbances.

Stage 1. Compute the following: $\hat{Y}_1 = X(X'X)^{-1}X'Y_1$, where $X = [X_1\ X_2]$ is the $N \times (k + p + q + \cdots + s)$ matrix of observations on all k predictor variables, that is, the original set of predictors X_1, plus the complete set $(p + q + \cdots + s)$ instrumental variables for each of the t autocorrelated predictor variables X_2. Each of the W_iy may have different numbers of instruments associated with them. Note that the original set of predictor variables is included in the first stage, as each of them is treated as an instrument for themselves. Also note that this first stage consists of computing t OLS regressions, where each of the W_iy variables is regressed singly on the original set of predictor variables X_1 plus the set of instruments for the entire set of autocorrelated endogenous variables. The predicted scores from each of these OLS regressions are saved as the $N \times t$ matrix $\hat{Y}_1$.

Stage 2. The following equation is estimated using OLS: $\hat{y} = \hat{Y}_1 c + X_1 b$, where the estimated vectors of b and c coefficients are now the two-stage least squares estimates of β and γ.

Although the Stage 2 OLS regression will produce correct (consistent) estimates for the autocorrelation and regression parameters, their standard errors and the R^2 will be incorrect. This is the result of using predicted scores for the autocorrelation effects variables in the OLS estimation rather than the true W_iy scores. It is, however, relatively straightforward to generate the correct error variance–covariance matrix, and thus standard errors, by using the b and c estimates from this Stage 2 regression. First, calculate a vector of residuals from

$$e = y - Y_1 c - X_1 b, \tag{12}$$

using the W_iy scores in Y_1 rather than the predicted scores. Next, calculate the estimated error variance as usual

$$s^2_e = e'e/\,(N - k - t) \tag{13}$$

Finally, insert the estimated error variance into the usual expression for the estimated coefficient variance–covariance matrix

$$S_{bc} = s_e^2 \begin{vmatrix} X_1'X_1 & X_1'Y_1 \\ Y_1'X_1 & Y_1'X(X'X)^{-1}X'Y \end{vmatrix}^{-1} \tag{14}$$

The usual R^2 statistic is easily obtained from $1 - (e'e/\Sigma(y - y_m)$, where y_m is the mean of y. Note that because all of the known properties of the 2SLS procedures are asymptotic, OLS procedures are uncorrected for the loss of degrees of freedom that results from using instrumental variables to get estimates of the endogenous network effects variables, that is, the denominator in equation (13) uses t and not $(p + q + \ldots + s)$ degrees of freedom. Fortunately, this correction procedure is automatically carried out using the 2SLS routines available in the widely used statistical packages.

The 2SLS procedures yield consistent estimates, that is to say, as the sample size increases, the estimates will converge to the population parameters. In general, nothing is known about how fast the convergence takes place. Hence, the behavior of this procedure with finite samples has to be determined via simulation studies. Both Das, Kelejian, and Prucha (2003) and Kelejian et al. (2004) report that for small samples, the 2SLS procedure for autocorrelated effects performs virtually as well as the much more complex ML procedures, and both of these dominate the usual OLS methods. Obviously, however, a model with multiple network autocorrelated variables and several instrumental variables for each will require a reasonably large sample to yield reliable estimates.

The 2SLS approach is, of course, only as good as the "instruments" employed. There are some suggestions in the literature about how well they should be correlated with the endogenous variables to be useful in practice. The choice of appropriate instruments in comparative research is addressed in the next section.

Substantive Example: Predicting Average Female Contribution to Subsistence

In addition to reporting Moran's I autocorrelation results for 1,700 SCCS variables, Eff (2004) also reports an OLS regression equation and corresponding p values associated with Moran's I statistics for language and spatial network matrices.[4] The p value reported for the language autocorrelation I statistic is highly significant ($p = .0102$), but the p value for distance is not quite significant ($p = .0650$).[5] Eff suggests that these p values render the usual interpretation of the OLS estimated regression equation invalid, and he suggests that some form of lagged dependent variable should be included into the regression equation to remove the autocorrelation indicated by the associated I statistics. Unfortunately, Eff does not actually re-estimate his regression equation using any such lagged variable, noting that doing so generally requires estimation using ML rather than OLS methods, as the latter result in biased and inconsistent regression coefficient estimates. Because variables are almost certainly missing from the specified regression equation, however, autocorrelation could be viewed here as entering the regression equation through the error term. In that case, the network autocorrelation disturbances model would be appropriate, as the autocorrelation distances matrix is embedded in the error term of the regression equation. On the other hand, if the detected autocorrelation can be linked to an hypothesized response mechanism, such that localities with a similar dependent variable trait score will tend to cluster together, then autocorrelation should be entered into the analysis as a lagged dependent variable. The network effects model outlined in this article is in the spirit of the latter approach and thus seems more consistent with Eff's statement.

Construction of Language, Spatial, and Ecological *W* Matrices

The language similarity matrix employed by Eff (2004) in computing the Moran's *I* statistic is based on earlier SCCS language codes. Specifically, Eff constructs a language distance matrix *C* by scoring pairs of societies as 0 if they are on different language continents, 1 if they are in the same language continent, 2 if in the same language phylum, and 3 if in the same language family.

The language *C* matrix employed in the re-analysis of Eff's regression model presented below was constructed using Burton's (1999) more recent language codes for the SCCS sample as follows:

This *C* matrix was then row-normalized to 1 to yield the appropriate language weights *W* matrix as described above.[6]

$$c_{ij} = \begin{cases} 1 & \text{if societies are in the same language family} \\ 2 & \text{if societies are in the same subfamily 1} \\ 3 & \text{if societies are in the same subfamily 2} \\ 0 & \text{otherwise} \end{cases}$$

One immediately apparent difference between this language matrix and that employed by Eff deserves comment. First, Burton's language codes do not include the notion of language continents. This has important implications for how the 14 societies that are coded by Burton as being the single representative member of their language family, and the 9 societies he codes as linguistic "isolates," are treated in the two regression analyses. In Eff's language matrix, these 23 societies are each assigned a connection of 1 to each other society in their language continent: Thus, there are no rows containing all 0s in the initial connectivity *C* matrix. On the other hand, using the Burton codes, each of these 23 societies is represented by a row of 0s in the language similarity matrix. Because these rows cannot be row-normalized to 1 to yield an appropriate *W* matrix, they in fact correspond to missing data for those societies. Hence, in the current analysis, these 23 societies are treated as missing values and thus are dropped from the linguistic section of the analyses reported below.[7]

The spatial weights matrix employed by Eff is based on the great circle distances between each pair of societies using the latitude and longitude coordinates for each society given in the SCCS database. The squared inverse of these distances was taken as the final weighting scheme. In the current reanalysis, great circle distances between each pair of societies were calculated using the geographical software package Passage V. 1.0 (Rosenberg, 2001). However, this global coordinates approach assigns positive weights between all pairs of societies no matter how distant their separation and thus no matter how unlikely the possibility of interaction. That is, societies in Southern Africa are positively related to societies in the Northwest Americas, despite the improbability of any meaningful interactions having occurred between such societies, even indirectly, throughout many millennia. Therefore, in the current analysis, all distances in the spatial matrix greater than 2,500 kilometers were set to 0. The inverse of the remaining distances were taken as the connectivity weights, and the matrix was then row-normalized to 1. This latter approach has the effect of emphasizing the network effects of societies within a 2,500 *K* radius of each focal society and ignoring the effects of more distant societies. Getis and Aldstadt (2004) suggest that it is better to employ a matrix with fewer connections than one with too many, as the power of autocorrelation tests is reduced in the latter situation.

Evolutionary and ecologically oriented anthropologists have long shown concern with ecology and how it constrains vertical and horizontal trait transmission processes. The notion of ecology as "constraint" and not "cause" in cultural adaptation, however, is an old issue in anthropology (Driver, 1956; Whiting, 1964). There is a 6-point ordinal scale coded for climate type in the

SCCS data set that represents ease of access to rich ecological resources. Using this ordinal scale, a similarity score for each pair of societies i and j was constructed using the following measure: $c_{ij} = e^{-|a_i - a_j|}$, that is, e raised to the negative of the absolute value of the difference in climate type scores, where a_i and a_j are the scores for societies i and j. The ecology matrix was then row-normalized to unity.[8]

Obviously, there are additional metrics that could be used to construct each of these three autocorrelation weight matrices. Unfortunately, in any empirical study, it is seldom the case that a single metric can be unambiguously identified as somehow offering the "best" representation of the dependency process of interest. Even in the seemingly straightforward case of spatially located sample units, for example, numerous specifications of pairwise relationships are possible: contiguity (rook and queen metrics), percentage of boundary in common, distance decay with a suitably chosen exponent, nearest neighbors, and so on (see Cliff & Ord, 1981, for a review). Griffith (1996) offers some guidance and rules of thumb when considering how to construct a spatial weights matrix for inclusion into an autocorrelation model, and Leenders (2002) suggests various useful considerations in the construction of weight matrices based on social influence processes. Getis and Aldstadt (2004) propose constructing a spatial weights matrix using a local statistic to detect the cluster of observations surrounding each individual observation that shows statistical clustering. Only those societies are used to assign weights in the row of W corresponding to each society; all others are set to 0. In most empirical studies, however, the final choice about W is made on substantive grounds, based on the assumption that the metric chosen adequately represents the theory the investigator has about the nature of the process generating the data dependencies.[9]

Choice of Instrumental Variables

The superiority of the 2SLS estimation procedure over the usual OLS estimation methods for the network autocorrelation effects model depends critically on the choice of instrumental variables for the network effects variable(s). To the extent that the chosen instruments are "weak" and do not do a good job of predicting the corresponding *Wy* effects variable, the 2SLS procedures may not perform very well. This is especially true in small samples, as all of the 2SLS properties are asymptotic, and thus, it may perform relatively poorly in small samples with poor instruments.[10] One compensating factor, however, is that the 2SLS method does not require the assumption of normality, which may not be reasonable in comparative data.

There are two suggestions for instruments in empirical studies that appear to be well suited for research using cross-cultural data. In their study of the effects of religious pluralism and social conditions on church adherence rates using a sample of 300 county groups across the continental United States, Land and Deane (1992) employed eight regional dummy (0, 1) variables as instruments for the spatial network effects variable constructed from the product of an inverse distance weight matrix W and a y vector of county adherence rates. Because Burton (1999) has provided a new regional code for the SCCS data, it is a straightforward matter to generate dummy variables for the nine regions and employ these dummies as instruments for any W matrix based on geographic distance. Similarly, because language codes are available for language families for the SCCS data, it is also quite straightforward to generate a set of language family dummy variables as instruments for an endogenous network effects variable based on a W language similarity matrix.[11] In general, using an appropriate set of dummy variables appears to offer one solution to the choice of instrumental variables for cross-cultural research involving language and spatial transmission effects.

Kelejian and Robinson (1993) suggest a second way to generate instruments for network effects variables that also appears to be viable for comparative data. In this case, instruments for a *Wy* variable are obtained by multiplying the set of predictor X variables by the W matrix, with the resulting "lagged" variables (i.e., *WX*) being used as instruments for *Wy*. That is, the weighted averages of scores on the independent variables are used as instruments for the weighted average of the dependent variable at these same societies. This also appears to offer a very general way to generate a suitable set of instruments in comparative research. These latter lagged predictor instruments are used in the empirical analysis presented next.[12]

Regression Results

Table 1 reports Eff's results for an OLS regression of the "average female contribution to subsistence" variable on a set of 10 predictor variables and compares them to the results from two autocorrelation effects models that include language and spatial autocorrelation effects separately, a third that includes ecological constraints, and a fourth equation that includes all three variables simultaneously.[13] The latter regression models were all estimated using 2SLS and the weights matrices described above, with instrumental variables for each of the effects and constraint variables generated by multiplying the set of predictor variables by their respective W matrices.[14]

Table 1
OLS Regression and 2SLS Regression Using Language and Spatial Network Autocorrelation Effects and Ecological Constraint: Dependent Variable–Average Female Contribution to Subsistence; Unstandardized Coefficients B; W_LX, W_SX, and W_EX as Instruments

		Eff's OLS Regression (N = 179)		Language Autocorrelation (N = 158)		Spatial Autocorrelation (N = 179)		Ecological Constraint (N = 179)		Language and Spatial Autocorrelation Plus Ecological Constraint (N = 158)	
Variable	Label	B	p	B	p	B	p	B	p	B	p
V816	Importance of fishing	–0.199	.003	–0.200	.011	–0.184	.008	–0.210	.002	–0.212	.007
V817	Importance of hunting	–0.430	.000	–0.399	.000	–0.418	.000	–0.430	.000	–0.413	.000
V1260	Total pathogen stress	–1.241	.002	–1.242	.005	–1.317	.001	–1.194	.002	–1.183	.009
V855	Niche rainfall	–2.281	.000	–1.515	.041	–2.372	.000	–1.800	.008	–1.491	.053
V79	Polygamy	2.951	.095	2.387	.236	2.753	.121	2.835	.101	2.634	.181
V353	Sex of parent in residence	5.409	.003	2.094	.315	2.973	.007	4.545	.012	2.686	.176
V150	Fixity of residence	–1.419	.123	–1.559	.126	–1.397	.130	–1.461	.105	–1.561	.123
V154	Land transport	–2.351	.047	–2.186	.108	–2.033	.095	–2.403	.038	–2.418	.081
V157	Political integration	2.565	.056	1.903	.209	2.415	.079	3.153	.021	2.059	.184
V158	Social stratification	–2.457	.033	–2.703	.041	–0.249	.032	–3.004	.010	–2.674	.043
W_Ly	Language autocorrelation			0.755	.001					0.668	.006
W_Sy	Spatial autocorrelation					0.242	.319			–0.168	.569
W_Ey	Ecological constraint							0.665	.021	0.039	.909
	R^2	0.312		0.347		0.316		0.337		0.360	

Note: OLS = ordinary least squares; 2SLS = two-stage least squares.

Eff's OLS regression results (third and fourth columns in Table 1) suggest that "fixity of residence" is clearly a statistically nonsignificant predictor (p = .123), and "polygamy" (p = .095) and "political integration" (p = .056) are not quite significant at the usual .05 level. All of the other seven predictor variables are clearly statistically significant. Introducing the language autocorrelation effects variable W_Ly yields some major changes with respect to several variables. First, note that the coefficient for "land transportation" changes from significant to nonsignificant (p= .047 to .108). Also, the magnitude of the regression coefficient for "sex of parent in residence" is now more than halved and has also switched from being significant to nonsignificant (p = .003 to .315). In addition, "polygamy" has become much less significant (p = .095 to .236), as has "political integration" (p = .056 to .209). Note too that the language autocorrelation effects coefficient for W_Ly is highly statistically significant (p = .001), as suggested by Eff's reported language I statistic.

Entering just the spatial autocorrelation effects variable W_Sy into the regression model along with Eff's predictors gives roughly similar results as using the W_Ly variable alone, with the exception of "sex of parent in residence," which remains statistically significant (p = .007.) The spatial effects coefficient is positive though statistically nonsignificant (p = .319), as might be expected from Eff's nonsignificant spatial I statistic.

Although entering the ecological constraints variable into the regression model returns a statistically significant estimate of the rho parameter (ρ = .665, p = .021), the results for the 10 predictor variables are in general quite similar to the results generated by Eff's OLS regression. The major difference is that with the 2SLS regression model, the coefficient for "political integration" is now clearly significant (p = .056 to .021).

In the final regression model, both of the autocorrelation effects variables and the ecological constraint variable were entered simultaneously. Here, the language effects variable is still fairly large and highly significant (ρ = .668, p = .006), and not unexpectedly, the spatial effects coefficient is still nonsignificant (p = .569). However, the magnitude of the ecological constraint coefficient is greatly reduced and is now nonsignificant (ρ = .039, p = .909). This is almost certainly the result of multicollinearity between the language and ecological constraint variables and the "niche rainfall" variable. The latter variable, another type of ecological constraint, has now become statistically nonsignificant (p = .053), and the spatial effects coefficient has switched from positive to negative (.242 to –.168), which by itself is a relatively unimportant indication of multicollinearity, as spatial effects are nonsignificant in any case. However, the results from entering language, distance, and ecological effects separately, and then from entering them simultaneously, clearly suggests that, in this particular study, language autocorrelation is the most important transmission effect. That is, adding the spatial effect and ecological constraint variables to the language effects regression leads to virtually the same conclusions as language effects alone, with the minor exception of the significance levels associated with the "niche rainfall" variable (p = .041 vs. .053). Finally, dropping the "niche rainfall" variable from the two regression equations that include ecological constraint yields virtually no differences in the results. Thus, although it is obvious that multicollinearity problems are likely to be more difficult to disentangle as more network effects are added, in the present analysis, the difficulties encountered did not obscure the main findings concerning the major importance of common history effects.[15]

To summarize, with respect to the hypothesized functional associations represented by Eff's 10 predictor variables, introducing spatial and language network effects and an ecological constraint variable into the regression model results in 3 predictors becoming nonsignificant and a further 2 variables moving from marginally nonsignificant to clearly nonsignificant. This finding, that 5 of the 10 predictor variables (50%) are clearly affected by autocorrelation processes, is consistent with earlier empirical studies that show similar percentages of changes from significant to nonsignificant results after autocorrelation terms are brought into the analysis (Borgerhoff Mulder et al., 2001; Dow, 1993).

From the perspective of the relative importance of the trait transmission processes and ecological constraint, there is little doubt that common history, as represented by the language matrix, is the most important factor with respect to female contribution to subsistence. Horizontal transmission, as represented by the spatial effects matrix, had very little effect on the dependent variable, although it did have a similar effect as language on the significance levels of some of the predictor variables. And although ecological setting had a significant effect on female contribution to subsistence, that effect appears to be the result of the well-known fact that daughter populations often end up in similar ecological settings as their parent population, so the effect of ecological setting disappears when common history is controlled for.

As a final check on the above autocorrelation regression results, the Moran's I statistic was computed using the standardized 2SLS residuals from the four 2SLS regression equations. In all cases, the Moran's I statistic was close to 0 and statistically very nonsignificant. These results confirm the validity of including network effects as trait transmission variables, and the ecological constraint variable, into the model and estimating it using 2SLS regression procedures.

Conclusions

Cultural traits are in general neither uniformly nor randomly distributed across societies. As Burton et al. (1996) demonstrate, cultural variation of social structural traits at the global level largely occurs across nine contiguous though nonoverlapping regions of the world, each of which shows more trait similarities—that is, are more homogeneous—among societies within the region than with societies in other regions. This finding alone suggests that both global and/or local autocorrelation of traits should be expected when dealing with cross-cultural data. Indeed, studies showing that a large percentage of traits display one or more types of network autocorrelation at both the global (Dow & Eff, 2007; Eff, 2004) and regional (Guglielmino et al., 1995; Hewlett et al., 2002) levels confirms the expectation that network autocorrelation of cultural traits must be anticipated and incorporated into studies by cross-cultural researchers. And the fact that some traits are influenced by more than one network transmission process makes the development of methods that can simultaneously incorporate multiple networks a priority. The current article contributes to the solution of this problem.

Clearly, results reported from cross-cultural studies that do not take into account trait dependencies should be skeptically received. In addition to the reanalysis reported here, other studies have confirmed that trait associations reported as statistically significant frequently become nonsignificant when appropriate statistical methods are employed (Borgerhoff Mulder et al., 2001; Dow, 1993). Both of these latter studies report that approximately 50% or more of significant associations become clearly nonsignificant once some appropriate statistical method is applied. The current results are consistent with these previous studies. The implications of these findings for cross-cultural research are huge. For example, Levinson (Ember & Levinson, 1991, appendix) reports that he located 577 hologeistic studies that appeared in print between 1889 and 1987, but he gives no indication of how many of these studies addressed trait dependence problems in their analysis. The fact that most cross-culturalists either ignore data dependencies altogether or rely on a simple sampling procedure, which has been shown here and elsewhere to be completely inadequate, suggests the possibly that as many as half of these reported findings may be incorrect.

This article has introduced a new statistical model to comparative research that allows the estimation of conventional regression models using 2SLS procedures that are implemented in most of the widely available statistical software packages. This offers considerable advantages, as regression is a highly developed theory with a battery of well-developed diagnostic test statistics that makes interpretation of output easy and straightforward. The relatively straightforward implementation of the network autocorrelation effects model suggests that a paradigm shift in the analysis of cross-cultural data, long overdue, may perhaps now be underway.

Notes

1. Moran's I statistic is a widely employed measure of the spatial autocorrelation of a single variable where the observations are in some known spatial arrangement such as intersections on a grid or at a set of known locations. The spatial arrangement is operationalized by assigning some measure of relationship to all pairs of observations, such as geographical distance, and entering each pairwise measure into a square sample unit by sample unit matrix, W. The W matrix is a set of "weights" that represent the underlying structure presumed to influence the observed data values at each location. Moran's I statistic assesses the extent to which high values on a variable tend to occur between close neighbors (positive autocorrelation) or the extent to which high values tend to alternate with low values (negative autocorrelation). The w_{ii} elements of W are all set to 0—that is, relationships of sample units with themselves are disregarded. Moran's I auto(self)correlation statistic for a single variable X can be expressed as follows:

$$I = \frac{N\Sigma_i\Sigma_j w_{ij}(x_i - x_m)(x_j - x_m)}{S\,\Sigma_i(x_i - x_m)^2}$$

where $S = \Sigma_i\Sigma_j w_{ij}$ is simply the sum of the weights, N is the sample size, and x_m is the mean value of the X variable. Clearly, this statistic depends heavily on how the connections between pairs of sample units are operationalized (i.e., the w_{ij}). Cliff and Ord (1981) provide a detailed discussion of the derivation of the I statistic and related statistical inference issues. It should also be noted that the measure of pairwise relationships represented by the elements of the W matrix can be based on very general concepts of relatedness (e.g., linguistic similarity, religion, trading relations, similarity of ecological niche).

2. White's (1993) assessment of moderate, strong, and extreme autocorrelation is based on the relative magnitudes of the Moran's I statistic for each variable rather than formal tests of significance. In addition, the differences in overall percentages reported by White and by Eff for spatial autocorrelation appears to be due to three sources: (a) use of inverse distance versus inverse of distance squared in the computation of the I statistic, (b) differences in recoding/collapsing of categories of variables prior to estimating Moran's I statistic, and (c) use of the full SCCS sample by Eff and a half sample by White.

3. Various software packages are available that implement MLE methods for network autocorrelation models containing a single autocorrelation weight matrix. An overview of several packages is given in Rey and Anselin (2006). MLE software for model (7), network effects and network disturbances, is available in the software package SpaceStat. Currently, the only estimation procedure available for the multiple effects model of interest in this article is 2SLS.

4. This regression model was selected for re-analyses primarily to provide an illustration of the substantive value of the multiple network autocorrelation effects model and associated 2SLS methodology. Because Eff reports autocorrelation statistics for two network matrices, language and distance, and suggests that an autocorrelation regression model would have been more appropriate than his use of OLS, it seemed particularly appropriate to re-estimate this model.

5. It is not clear from Eff's discussion whether the Moran I statistics and their corresponding significance levels are estimated using the residuals from the regression equation—the most common use of the I statistic in regression analysis—or using the dependent variable, average female contribution to subsistence.

6. As part of a recent study (Dow & Eff, 2007), a new language matrix based on Burton's language codes was constructed. The language similarity measure treated the New World as one language continent and the Old World as a second continent. All pairs of societies in the same language continent were scored as 1, 2 if in the same language family, 3 if in the same Subfamily 1, and 4 if in the same Subfamily 2. For the present study, however, this language matrix was subsequently edited by subtracting 1 from each score, then deleting all rows/columns containing only 0s, and it was then row-normalized.

7. Coding the 14 singleton language families and 9 isolates as being related to other societies in the same language continent at Level 1 of the coding is equivalent to replacing each corresponding 0 value on the Wy vector with the average score of all societies within the respective language continent. Replacing missing data by average values is not an uncommon procedure and is generally fairly conservative in its effect on subsequent parameter estimation and associated inferences.

8. The metric for the ecological similarity matrix was motivated in part by the metric used in Case et al. (1993), where a weights matrix W was constructed using $w_{ij} = |q_i - q_j|^{-1}$ as the metric, where the qs were percentage of population that is Black in states i and j. However, with a 6-point ordinal scale of ecological similarity, all pairs of societies with the same rank score would have an undefined weight using this metric. Hence, the negative of the absolute difference in rank scores was exponentiated. This transformation constrains the weights to lie between 1 (for pairs of societies with the same score on the ordinal ecological setting scale) and .0067 (for pairs of societies at the opposite end of the scale). Again, the weights were row-normalized.

9. There are two related issues here. First, the Getis and Aldstadt (2004) procedure for constructing a "best" matrix based on their local statistic is extremely computationally burdensome. For each sample unit, the local statistic must be computed for all societies that fall within an increasing critical distance d, until the statistic becomes nonsignificant, indicating no further autocorrelation with respect to the focal sample unit. Societies within the radius of the final d value are then assigned positive weights in the row of W corresponding to that sample unit. This procedure must be carried out for every sample unit. It should be noted that it is not always possible to slowly increase the required d value until the local statistic becomes nonsignificant. For the language matrix based on the SCCS codes, for example, only a few increases in the d values are possible, and these usually result in vastly different numbers of positive weights being assigned within rows of the corresponding W matrix, resulting in lower autocorrelation statistics than does using the original language W matrix (Dow & Eff, 2007).

A second issue arises when the researcher has narrowed down the possibilities for a W matrix but is unable to choose between them on theoretical grounds. Statistical procedures are available to test a single W chosen as the null hypothesis against a range of alternative Ws and in the case where there are a (small) fixed number of equally plausible Ws representing the same dependency process (Anselin, 1988; Anselin & Bera, 1998; Leenders, 2002). The utility of these testing procedures is beyond the scope of the current article, but this is another important area for future research by cross-cultural researchers.

10. The effectiveness of the instrumental variable(s) may be assessed by inspecting the correlation between the endogenous *Wy* and the predicted endogenous variable, pred(*Wy*), obtained using the instrumental variable(s). In the empirical example presented below, the following highly significant Pearson correlations were obtained: $r(W_L y, \text{Pred}(W_L y)) = .605$; $r(W_D y, \text{Pred}(W_D y)) = .740$; $r(W_E y, \text{Pred}(W_E y)) = .796$, where the subscripts denote language, distance, and ecology, respectively, and Pred() indicates the predicted values obtained using the set of independent variables X premultiplied by the corresponding weight matrix. The magnitude of these correlations suggests that the sets of lagged independent variables are very acceptable as instruments.

11. To recode the language family code as a set of dummy variables requires an initial decision as to how many societies there have to be within a language family before it gets coded as a separate dummy variable. That is, Burton's language family code gives 11 families with 5 or more members. Lowering the minimum number of societies within a language family that gets recoded as a separate dummy variable obviously leads to a larger number of dummies, or instruments, and a corresponding tendency to "overfit" when regressing the autocorrelated *Wy* variable in the first stage regression. In the present study, using a set of dummy variables with five as the minimum number produced results that were reasonably close to the results using $W_L X$ instruments. One possibility for reducing the total number of instruments is to regress each of the *Wy* variables on the corresponding set of instruments using a stepwise procedure and then use only the statistically significant instruments in the full 2SLS procedure. In the present study, stepwise regression reduced the number of $W_L X$, $W_S X$, and $W_E X$ instrumental variables from 10 to 5 in all three cases. The resulting 2SLS results were fairly close to the results in Table 1, with the exception that the "niche rainfall" became marginally nonsignificant in the estimation using only the language effects variable, $W_L y$.

12. The question of which instruments are generally preferable for use in cross-cultural research is an area for further research. In the current study, both *WX* variables and sets of dummy variables based on region/language family were used: Only the results based on the lagged *WX* variables are reported here. It is also worth noting that the advantage of row-normalizing *W* weight matrices to unity is that any variable that is premultiplied by such a *W* matrix (i.e., *Wy* and *WX*) is on the same scale as the original variables. Hence, the interpretability of the resulting autocorrelation coefficients is not complicated by scale changes.

13. Several data preprocessing tasks are required prior to carrying out the 2SLS estimation procedures for the network autocorrelated effects model. For example, the *C* matrices have to be row-normalized to unity to generate the *W* matrices, and then the autocorrelated *Wy* effects variables also have to be generated. Similarly, lagged instrumental variables *WX* may have to be created. These tasks can be carried out in Excel using macros or in any statistical software package such as STATA that has matrix programming capabilities. In the current article, these matrix operations were carried out using the matrix algebra module available in UCINET V.6 (Borgatti, Everett, & Feeman, 2002).

14. All of the networks effects estimated coefficients reported in Table 1 were obtained using the 2SLS procedures available in Statistical Package for the Social Sciences Version 11.5.

15. The fact that "history and geography (and also ecology) often covary, as has been shown in practically every study that looks at history and geography together," does not necessarily imply that this covariance will invariably generate the "huge problems of multicollinearity" that Borgerhoff Mulder, Nunn, and Towner (2006, p. 59) appear to anticipate in comparative research. The Moran's *I* statistic autocorrelation results cited at the beginning of this article show clearly that global and regional autocorrelation levels of cultural traits are highly variable with respect to different types of transmission processes. In general, the interpretive problems raised by multicollinearity will depend heavily on the constellation of trait transmission, ecological, and functional variables and the dependent variable of interest in each particular study.

References

Anselin, L. (1988). *Spatial econometrics: Methods and models.* Dordrecht, the Netherlands: Kluwer Academic Publishers.

Anselin, L., & Bera, A. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. In A. Ullak & D. Giles (Eds.), *Handbook of applied economic statistics* (pp. 237-289). New York: Marcel Dekker.

Barry, H., & Yoder, B. L. (2002). Multiple predictors of contribution by women to agriculture. *Cross-cultural Research, 36*, 286-297.

Borgatti, S. P., Everett, M. G., & Feeman, L. C. (2002). *UCINET V.6. for Windows: Software for social network analysis.* Harvard, MA: Analytic Technologies.

Borgerhoff Mulder, M., George-Cramer, M., Eshleman, J., & Ortolani, A. (2001). A study of East African kinship and marriage using a phylogenetically based comparative method. *American Anthropologist, 103*, 1059-1082.

Borgerhoff Mulder, M., Nunn, C., & Towner, M. (2006). Cultural macroevolution and the transmission of traits. *Evolutionary Anthropology, 15*, 52-64.

Burton, M. L. (1999). Language and region codes for the Standard Cross Cultural Sample. *Cross-Cultural Research, 33*, 63-83.

Burton, M., Moore, C., Whiting, J. M., & Romney, A. K. (1996). Regions based on social structure. *Current Anthropology, 37*, 87-123.

Case, A., Rosen, H., & Hines, J. (1993). Budget spillovers and fiscal policy interdependence: Evidence from the states. *Journal of Public Economics, 52*, 285-307.

Cliff, A., & Ord, J. K. (1981). *Spatial processes: Models and applications.* London: Pion Press.

Conley, T., & Topa, G. (2002). Socio-economic distance and spatial patterns in unemployment. *Journal of Applied Economics, 17*, 303-327.

Das, D., Kelejian, H., & Prucha, I. (2003). Finite properties of estimators of spatial autoregressive models with autoregressive disturbances. *Papers in Regional Science, 82*, 1-26.

Doreian, P. (1989a). Models of network effects on social actors. In L. Freeman, D. White, & A. K. Romney (Eds.), *Research methods in social network analysis* (pp. 295-318). Fairfax, VA: George Mason University Press.

Doreian, P. (1989b). Two regimes of network autocorrelation effects. In M. Kochen (Ed.), *The small world problem* (pp. 280-295). Norwood, NJ: Ablex.

Dow, M. (1979). *Structure and process in cultural-ecological complexes.* Unpublished doctoral thesis, University of California at Irvine.

Dow, M. (1984). A bi-parametric approach to network autocorrelation: Galton's Problem. *Sociological Methods and Research, 13*, 201-217.

Dow, M. (1993). Saving the theory: On chi-square tests with cross-cultural survey data. *Cross-Cultural Research, 27*, 247-276.

Dow, M., Burton, M., White, D., & Reitz, K. (1984). Galton's problem as network autocorrelation. *American Ethnologist, 11*, 754-770.

Dow, M., & Eff, E. A. (2007). Global, regional, and local network autocorrelation in the Standard Cross-Cultural Sample. *Cross-Cultural Research.* Under review.

Dow, M., White, D., & Burton, M. (1983). Multivariate modeling with interdependent network data. *Behavior Science Research, 17*, 216-245.

Driver, H. (1956). An integration of functional, evolutionary, and historical theory by means of correlations. *Indiana University Publications in Anthropology and Linguistics, 12*, 1-36.

Driver, H. E., & Coffin, J. L. (1975). *Classification and development of North American Indian cultures: A statistical analysis of the Driver-Massey sample* (Transactions of the American Philosophical Society, Vol. 1.65, Part 3). Philadelphia: American Philosophical Society.

Eff, E. A. (2004). Does Mr. Galton still have a problem? Autocorrelation in the Standard Cross-Cultural Sample. *World Cultures, 15*, 153-170.

Ember, C., & Levinson, D. (1991). The substantive contributions of worldwide cross-cultural studies using secondary data. *Behavior Science Research, 25*, 79-140.

Erbring, L., & Young, A. (1979). Individuals and social structure: Contextual effects as endogenous feedback. *Sociological Methods and Research, 7*, 396-430.

Getis, A., & Aldstadt, J. (2004). Constructing the spatial weights matrix using a local statistic. *Geographical Analysis, 36*, 90-104.

Griffith, D. (1996). Some guidelines for specifying the geographic weights matrix contained in spatial statistical models. In S. L. Arlinghaus (Ed.), *Practical handbook of spatial statistics* (pp. 65-82). Boca Raton, FL: CRC.

Guglielmino, C. R., Viganotti, C., Hewlett, B., & Cavalli-Sforza, L. L. (1995). Cultural variation in Africa: Role of mechanisms of transmission and adaptation. *Proceedings of the National Academy of Sciences, 92*, 7585-7589.

Hewlett, B. S., de Silvestri, A., & Guglielmino, C. R. (2002). Semes and genes in Africa. *Current Anthropology, 43*, 313-321.

Johnston, J. (1984). *Econometric methods* (2nd ed., International Student Edition). New York: McGraw-Hill.

Jorgensen, J. (1969). *Salish language and culture: A statistical analysis of internal relationships, history, and evolution* (Indiana University Language Science Monograph 3). Bloomington: Indiana University.

Jorgensen, J. (1980). *Western Indians*. San Francisco: W. H. Freeman and Co.

Kelejian, H., & Prucha, I. (1998). A generalized two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics, 17*, 99-121.

Kelejian, H., Prucha, I., & Yuzefovich, Y. (2004). Instrumental variable estimation of a spatial autoregressive model with autoregressive disturbances: Large and small sample results. In J. P. LeSage & R. Pace (Eds.), *Spatial and spatiotemporal econometrics* (Vol. 18, Advances in Econometrics, pp. 163-198). Oxford, UK: Elsevier.

Kelejian, H., & Robinson, D. (1993). A suggested method of estimation for spatial interdependent models with autocorrelated errors, and an application to a county expenditures model. *Papers in Regional Science, 72*, 297-312.

Korotayev, A., & de Munck, V. (2003). "Galton's asset" and "Flower's problem": Cultural networks and cultural units in cross-cultural research. *American Anthropologist, 105*, 353-358.

Lacombe, D. (2004). Does econometric methodology matter? An analysis of public policy using spatial econometric techniques. *Geographical Analysis, 36*, 105-118.

Land, K., & Deane, G. (1992). On the large-sample estimation of regression models with spatial- or network-effects terms: A two-stage least squares approach. In P. Marsden (Ed.), *Sociological methodology 1992* (pp. 221-248). Oxford, UK: Blackwell.

Leenders, R. T. (2002). Modeling social influence through network autocorrelation: Constructing the weight matrix. *Social Networks, 24*, 21-47.

Loftin, C., & Hill, R. (1974). A comparison of alignment procedures for tests of Galton's Problem. In J. Schaeffer (Ed.), *Studies in cultural diffusion: Galton's Problem* (pp. 23-61). New Haven, CT: HRAF Press.

Mace, R., & Pagel, M. (1994). The comparative method in anthropology. *Current Anthropology, 35*, 549-564.

Murdock, G. P., & White, D. (1969). Standard Cross-Cultural Sample. *Ethnology, 8*, 329-369.

Naroll, R. (1961). Two solutions to Galton's Problem. *Philosophy of Science, 28*, 15-39.

Ord, J. K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association, 70*, 120-126.

Pryor, F. (1976). The diffusion possibility method: A more general and simpler solution to Galton's Problem. *American Ethnologist, 4*, 731-749.

Rey, S., & Anselin, L. (2006). Recent advances in software for spatial analysis in the social sciences. *Geographical Analysis, 38*, 1-4.

Rosenberg, M. S. (2001). *PASSAGE: Pattern analysis, spatial statistics, and geographic exegesis*. Tempe: Department of Biology, Arizona State University.

White, D. (1993). Spatial levels in cultural organization: An empirical study. In T. Schweitzer, M. Schweitzer, & W. Kokot (Eds.), *Handbuch der ethnologie* (pp. 459-488). Berlin, Germany: Reimer Verlag Press.

White, D., Burton, M., & Dow, M. (1981). Sexual division of labor in African agriculture: A network autocorrelation analysis. *American Anthropologist, 83*, 824-849.

Whiting, J. M. (1964). The effects of climate on certain cultural practices. In W. Goodenough (Ed.), *Explorations in cultural anthropology* (pp. 511-544). New York: McGraw-Hill.

Whiting, J. M., Sodergren, J., & Stigler, S. (1982). Winter temperature as a constraint to the migration of preindustrial peoples. *American Anthropologist, 84*, 279-298.

Wirsing, R. (1974). Measuring diffusion: The Geary method and the Dacey method. In J. Schaefer (Ed.), *Studies in cultural diffusion: Galton's Problem* (pp. 198-219). New Haven, CT: HRAF Press.

Witkowski, S. (1974). Galton's opportunity: The hologeistic study of historical processes. In J. Schaefer (Ed.), *Studies in cultural diffusion: Galton's Problem* (pp. 84-112). New Haven, CT: HRAF Press.

Malcolm M. Dow is professor emeritus of anthropology and mathematical methods in the social sciences at Northwestern University, where he taught from 1980 to 2002. He has published more than 50 articles, about a dozen of which deal with aspects of Galton's Problem in cross-cultural and cross-species research.