Generalized Estimating Equation Models for Correlated Data: A Review with Applications
Author(s): Christopher J. W. Zorn
Source: *American Journal of Political Science*, Vol. 45, No. 2 (Apr., 2001), pp. 470-490
Published by: Midwest Political Science Association
Stable URL: http://www.jstor.org/stable/2669353
Accessed: 13/05/2009 17:01

Generalized Estimating Equation Models for Correlated Data: A Review With Applications

Christopher J. W. Zorn Emory University

Political scientists are often called upon to estimate models in which the standard assumption that the data are conditionally independent can be called into question. I review the method of generalized estimating equations (GEE) for dealing with such correlated data. The GEE approach offers a number of advantages to researchers interested in modeling correlated data, including applicability to data in which the outcome variable takes on a wide range of forms. In addition, GEE models allow for substantial flexibility in specifying the correlation structure within cases and offer the potential for valuable substantive insights into the nature of that correlation. Moreover, GEE models are estimable with many currently available software packages, and the interpretation of model estimates is identical to that for commonly used models for uncorrelated data (e.g., logit and probit). I discuss practical issues relating to the use of GEE models and illustrate their usefulness for analyzing correlated data through three applications in political science.

Empirical political science is most often interested in estimating the effect of some set of explanatory covariates on an outcome variable of interest. At the same time, in many cases, the data on which we observe the phenomena of interest are likely to be correlated. The most common instances of correlated data are those involving repeated observations over time, either in the form of panel studies or time-series of cross-sections.[1] Correlated data can also arise in other ways, including dyadic studies (e.g., Oneal and Russett 1997; Hojnacki and Kimball 1998; Huckfeldt, Sprague, and Levine 2000) and examinations of individual decisions in a collegial context, for example, voting decisions in a legislature (e.g., Levitt 1996; Snyder and Groseclose 2000) or a court (e.g., Traut and Emmert 1998). While issues relating to temporal correlation have recently received a good deal of attention among political scientists (Beck and Katz 1995; Beck, Katz, and Tucker 1998; Box-Steffensmeier and Jones 1997; see also Stimson 1985), the latter form of correlation often goes unaddressed. Moreover, few of the methods for dealing with correlation over time are appropriate for use with data where the interdependence is not of a temporal nature, and fewer still are capable of dealing with noncontinuous dependent variables (for example, dichotomous variables or event counts).

In this article, I review a general method for dealing with correlated data: the technique of generalized estimating equations (GEEs).[2] GEEs of-

Christopher J. W. Zorn is Assistant Professor of Political Science, Emory University, Atlanta, GA 30322 (czorn@emory.edu).

Thanks to David Davis for providing some of the data used here, to Huiman Barnhart, Neal Beck, Mike Giles, Dan Reiter, Elliot Slotnick, John Williamson, the editor, and several anonymous reviewers for helpful comments and discussions, and to the John M. Olin Foundation for their support of this research. Any remaining errors are my own. An earlier version of this article was presented at the 1998 Annual Meeting of the Midwest Political Science Association.

[1]Studies of this kind are widespread; a quick review of articles in recent journals finds such data in analyses of American political institutions (e.g., Flemming and Wood 1997; Krehbiel 1997; Meier, Polinard, and Wrinkle 2000; Scholz and Wood 1998), political behavior and elections (e.g., Bartels 1999; James and Lawson 1999; Lowry, Alt, and Feree 1998; Sears and Funk 1999), comparative politics (e.g., Brown and Hunter 1999; Finkel and Muller 1998; Clark and Hallerberg 2000; Huber 1998; Radcliffe and Davis 2000) and international relations (e.g., Mansfield, Milner, and Rosendorff 2000; Morrow, Siverson, and Tabares 1998; Reed 2000).

[2]Recent applications of GEEs in political science include Bennett and Stam (2000), Bliss and Russett (1998), Caldeira, Wright, and Zorn (1999), Leeds and Davis (1997) and Oneal and Russett (1999a, 1999b).

American Journal of Political Science, Vol. 45, No. 2, April 2001, Pp. 470–490

fer a number of advantages for researchers interested in modeling correlated data, including applicability to continuous, dichotomous, polychotomous, ordinal and event-count response variables. In addition, GEEs allow for a range of substantively-motivated correlation patterns within clusters and offer the potential for valuable substantive insights into the dynamics of that correlation. Finally, GEEs are estimable with a host of currently available software packages, and the interpretation of those estimates is identical to that for commonly used models for uncorrelated data (e.g., logit and probit).

In the next section, I introduce the technique of generalized estimating equations and discuss its statistical properties. The following section addresses the strengths and weaknesses of these models, as well as a number of practical matters regarding their use. I then apply these models to three applications in political science. The first presents a comparison of GEE and random-effects models of House voting on the impeachment of President Clinton. Next, I demonstrate the utility of these models for analyzing temporally correlated data in which the dependent variable takes the form of an event count through an analysis of data on joint memberships in international intergovernmental organizations. Finally, I illustrate the usefulness of GEEs for modeling non-temporally-correlated data by examining Supreme Court decision making in civil rights and liberties cases during three recent Court terms. I conclude with a discussion of the strengths and weaknesses of GEE models in general and an appendix on software issues.

Generalized Estimating Equation Models: An Overview

The GEE approach has its roots in the quasi-likelihood methods introduced by Wedderburn (1974) and Nelder and Wedderburn (1972) and developed and extended by McCullagh and Nelder (1983, 1989; see also Heyde 1997) and others.[3] While standard maximum-likelihood analysis specification of the full conditional distribution of the dependent variable, quasi-likelihood requires only that we postulate the relationship between the expected value of the outcome variable and the covariates and between the conditional mean and variance of the response variable. This generalized linear model (GLM) approach has received widespread use in cross-sectional analyses (see, e.g., Gill 2000).

The GLM approach was first extended to correlated data by Liang and Zeger (1986; also Zeger and Liang 1986) in the context of repeated observations over time.[4] Using notation similar to that of Zeger and Liang (1986), consider a model of observations on a dependent variable Y_{it} and k covariates $\mathbf{X}_{it}$, where i indexes the N units of analysis ("cases" or "clusters") $i = \{1, 2, \ldots N\}$ and t indexes the T time points (or repeated measurements) $t = \{1, 2, \ldots T\}$. Let $\mathbf{Y}_i = [Y_{i1}, Y_{i2}, \ldots Y_{iT}]$ denote the corresponding column vector of observations on the outcome variable for observation i, and $\mathbf{X}_i$ indicate the $T \times k$ matrix of covariates for observation i.[5] Writing $E(\mathbf{Y}_i) = \mu_i$, we define a function h which specifies the relationship between $\mathbf{Y}_i$ and $\mathbf{X}_i$:

$$\mu_i = h(\mathbf{X}_i \boldsymbol{\beta}) \quad (1)$$

where $\boldsymbol{\beta}$ is a $k \times 1$ vector of parameters; the inverse of h is known as the "link" function. Likewise, the variance $\mathbf{V}_i$ of $\mathbf{Y}_i$ is specified as a function g of the mean. In the cross-sectional case (i.e., $T = 1$), we write:

$$V_i = \frac{g(\mu_i)}{\phi} \quad (2)$$

where ϕ is a scale parameter which may or may not be of substantive interest. The quasi-likelihood estimate of $\boldsymbol{\beta}$ is then the solution to a set of k "quasi-score" differential equations:

$$U_k(\boldsymbol{\beta}) = \sum_{i=1}^{N} \mathbf{D}_i' \mathbf{V}_i^{-1} \left(\mathbf{Y}_i - \mu_i \right) = 0 \quad (3)$$

where $\mathbf{D}_i = \ \boldsymbol{\mu}_i / \ \boldsymbol{\beta}$. If the model is properly specified, then, asymptotically, $E[U_k(\boldsymbol{\beta})] = 0$ and $\mathrm{Cov}[U_k(\boldsymbol{\beta})] = \mathbf{D}_i' V^{-1} \mathbf{D}_i$. The function $U(\boldsymbol{\beta})$ thus behaves like the derivative of a log-likelihood (i.e., a score function); estimation may be accomplished either via generalized weighted least-squares or through an iterative process.[6]

[3] This section draws extensively on the presentation and notation of Zeger and Liang (1986) and Fitzmaurice, Laird, and Rotnitzky (1993). The literature on GEEs, particularly in biostatistics, is vast; good reviews of these models can be found in Liang, Zeger, and Qaqish (1992), Zeger and Liang (1992), Diggle, Liang, and Zeger (1994) and Ziegler, Kastner, and Blettner (1998).

[4] While this exposition is the standard one, it should be noted that the correlation within observations over time need not be temporal in nature; I address this further, and provide examples, below.

[5] For notational simplicity, I assume here that $T_i = T_{i'} \ \forall \ i \ \ i'$, i.e., that the "panels" are balanced. This need not be the case for the models presented here; balanced panels are, however, necessary for likelihood-based "mixed parameter" models (Fitzmaurice, Laird, and Rotnitzky 1993).

[6] The standard reference for such models is McCullagh and Nelder (1989).

For cases where $T > 1$, some provision must be made for dependence across t. Liang and Zeger's (1986) solution was to specify a $T \times T$ matrix $\mathbf{R}_i(\boldsymbol{\alpha})$ of the "working" correlations across t for a given $\mathbf{Y}_i$.[7] While $\mathbf{R}_i(\boldsymbol{\alpha})$ can thus vary across cases, it is assumed to be fully specified by the vector of unknown parameters $\boldsymbol{\alpha}$, which has a structure determined by the investigator and which is constant across clusters. This correlation matrix then enters the variance term of equation (3):

$$V_i = \frac{(A_i)^{1/2} R_i(\alpha)(A_i)^{1/2}}{\phi} \tag{4}$$

where the $\mathbf{A}_i$ are $T \times T$ diagonal matrices with $g(\mu_{it})$ as the tth diagonal element (Zeger and Liang 1986, 124). Substitution of (4) into (3) yields the GEE estimator. In this light, it is clear that the GEE is an extension of the GLM approach, and that the former reduces to the latter when $T = 1$.[8]

Since the $\mathbf{V}_i$'s are functions of both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, estimation is typically accomplished by an iterative procedure, e.g., Fisher scoring (McCullagh and Nelder 1989, 42–43). Liang and Zeger (1986) note that we can reexpress (3) as a function of $\boldsymbol{\beta}$ alone by substituting $N^{1/2}$-consistent estimates of $\boldsymbol{\alpha}$ and ϕ into (3). One can then solve for $\boldsymbol{\beta}$ using Fisher scoring and calculate standardized residuals, which are in turn used to consistently estimate $\boldsymbol{\alpha}$ and ϕ. These two steps are iterated until the estimates reach convergence. The intuition of estimating $\boldsymbol{\beta}$ via GEE is thus to "choose $\boldsymbol{\beta}$ so that $\mu(\boldsymbol{\beta})$ is close to Y_i on average and to optimally weight each residual $\mathrm{Y}_i - \mu_i$ by the inverse of $\mathrm{Cov}(\mathrm{Y}_i)$" (Zeger and Liang 1988, 1833).

The GEE model outlined here has a number of attractive properties for applied researchers. Because the first two terms of (3) do not depend on $\mathbf{Y}_i$, the score equations converge to zero (and thus have consistent roots) so long as $\mathrm{E}[\mathbf{Y}_i - \mu_i] = 0$. Assuming that the model for μ is correctly specified, GEE estimates of $\boldsymbol{\beta}$ (here, $\hat{\beta}_{\mathrm{GEE}}$) will be consistent in N (Liang and Zeger 1986). Moreover, $N^{1/2}(\hat{\beta}_{\mathrm{GEE}} - \beta)$ is asymptotically multivariate normal, and the covariance matrix of the estimates can be consistently estimated by the inverse of the derivative of the quasi-score function, evaluated at $\hat{\alpha}$ and $\hat{\beta}$.[9] Most important, the asymptotic consistency of $\hat{\beta}_{\mathrm{GEE}}$ holds even in the presence of misspecification of the "working" correlation structure $\boldsymbol{\alpha}$; thus, GEEs offer the potential of providing asymptotically unbiased estimates of the parameters of primary interest even in cases where the exact nature of the intracluster dependence is unknown. Note, however, that the consistency of the variance estimate for $\hat{\beta}$ *does* depend on proper specification of the working correlation structure; misspecification of the working correlation structure yields estimates of $\hat{\beta}$ which are still asymptotically normal, but for which $\mathrm{Var}(\hat{\beta})$ does not equal $(\mathbf{D'V^{-1}D})^{-1}$. In such cases, Liang and Zeger (1986) propose a "robust" estimate of the variance-covariance matrix of $\hat{\beta}$:

$$Var\left(\hat{\beta}_{\mathrm{GEE}}\right) = N\left(\sum_{i=1}^{N} \hat{D}_i'\hat{V}_i^{-1}\hat{D}_i\right)^{-1}\left[\sum_{i=1}^{N} \hat{D}_i'\hat{V}_i^{-1}S_i\hat{V}_i^{-1}\hat{D}_i\right]\left(\sum_{i=1}^{N} \hat{D}_i'\hat{V}_i^{-1}\hat{D}_i\right)^{-1} \tag{5}$$

where $S_i = (Y_i - \hat{\mu}_i)(Y_i - \hat{\mu}_i)'$ is a simple empirical covariance estimate. This "robust" (or "empirically-corrected") estimate is analogous to that derived by White (1980) and is consistent even under misspecification of the correlation matrix. In practice, this robust estimator is nearly always used, since a key reason for using GEEs in the first place is the belief that the observations in the data are nonindependent.[10]

An additional advantage of the GEE approach is the broad range of options available for specifying the within-cluster correlation structure. Fitzmaurice, Laird, and Rotnitzky (1993) discuss four common specifications of the "working" correlation matrix $\mathbf{R}_i(\boldsymbol{\alpha})$:

1. $\mathbf{R}_i(\alpha) = \mathbf{I}$, a $T \times T$ identity matrix. This "working independence" assumption is equivalent to assuming no intracluster correlation and yields estimates equivalent to those from simple "pooled" models. No estimate of $\boldsymbol{\alpha}$ is obtained, since the intra-cluster correlation is assumed to be zero.

[7] Alternatively, for binary outcomes the within-cluster association can be modeled as an odds ratio (cf. Lipsitz, Laird, and Harrington 1991).

[8] Even more generally, both the GLM and GEE estimators are special cases of the generalized method of moments estimator (Hansen 1982; Davidson and MacKinnon 1993, chapter 17.5) in which no overidentifying moment restrictions are present. This equivalence is not widely recognized, though see Cameron and Trivedi (1998, 39–42) and Zeigler (1995) for exceptions. I thank an anonymous referee for pointing out this connection.

[9] Zeger and Liang (1986) note that in most cases it is possible to estimate $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ without estimating ϕ directly, provided that the elements of the correlation matrix $\mathbf{R}$ are multiples of the parameters $\boldsymbol{\alpha}$.

[10] Another alternative is the jackknife (e.g., Lipsitz, Laird, and Harrington 1990), which is asymptotically equivalent to White's robust estimator (Lipsitz, Dear, and Zhao 1994). Note, however, that the usefulness of the robust estimator decreases as the ratio of T to N increases (Prentice 1988, 1040). In the limit, robust standard errors "are not meaningful when there is only one subject" (Mugglestone, Kenward, and Clark 1999, 8). Bootstrap or jackknife techniques might also be used to provide standard error estimates for the estimated as in the GEE1 context; at this time, however, the properties of such an estimator remain unknown.

2. $\mathbf{R}_i(\alpha) = \rho$, $t \quad s$. This is the "exchangeable" correlation structure; values of Y_i are assumed to covary equally across all observations within a cluster. In this specification, α is a scalar, which is estimated by the model. In the Gaussian case (i.e., with continuous normal data), this is similar to a "random-effects" model.
3. $\mathbf{R}_i(\alpha) = \rho^{|t-s|}$, $t \quad s$, an autoregressive (here, AR(1)) specification. Here, the within-observation correlation over time is an exponential function of the lag length; as is typically the case in AR models, we assume $|\rho| < 1.0$. In an autoregressive specification, is a vector $[1, \rho, \rho^2, \ldots]$. which is the same across observations. Higher-order autoregressive specifications are also available.
4. $\mathbf{R}_i(\alpha) = \alpha_{st}, t \quad s$, an unstructured (or "pairwise") correlation structure. Here, no constraints are placed on the correlations across observations within a cluster; instead, they are estimated from the data without restriction. In this context, $\boldsymbol{\alpha}$ is a $T \times T$ matrix containing the $\frac{T(T-1)}{2}$ unique pairwise correlations for all possible combinations of time points.

In addition to these, a number of other specifications of the working correlation matrix are possible, including stationary ("banded") and nonstationary models of varying orders. Alternatively, the researcher may specify $\mathbf{R}_i(\boldsymbol{\alpha})$ explicitly; this option is valuable for testing the robustness of estimates to the correlation specification.

More recently, a several scholars have considered other means by which the $m = \frac{T(T-1)}{2}$ intracluster correlations may be estimated. Prentice (1988) suggests a second, orthogonal set of estimating equations for the correlations:

$$U_m(\alpha) = \sum_{i=1}^{N} \mathbf{E}_i{}' \mathbf{W}_i^{-1}(\mathbf{Z}_i - \eta_i) = 0 \tag{6}$$

Here, $\mathbf{Z}_i' = (Z_{i12}, Z_{i13}, \ldots Z_{i1T}, Z_{i23}, \ldots Z_{iT,(T-1)})$ are the $\frac{T(T-1)}{2}$ observed "sample" pairwise correlations, $\boldsymbol{\eta}_i$ is the column vector of expected values of the pairwise intracluster correlation for observation i, $\mathbf{E}_i = \quad \boldsymbol{\eta}_i / \quad \boldsymbol{\alpha}$, and $\mathbf{W}_i$ is a square diagonal matrix of rank $\frac{T(T-1)}{2}$ containing the variances and covariances of the Z_is (see Prentice 1988 for more detail). The equations in (3) and (6) can then be considered part of a system of equations; if the estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are treated as orthogonal, this yields the equations:

$$U(\alpha,\beta) = \sum_{i=1}^{N} \begin{pmatrix} \mathbf{D}_i' & 0 \\ 0 & \mathbf{E}_i' \end{pmatrix} \begin{pmatrix} \mathbf{V}_i' & 0 \\ 0 & \mathbf{W}_i' \end{pmatrix} \begin{pmatrix} \mathbf{Y}_i - \mu_i \\ \mathbf{Z}_i - \eta_i \end{pmatrix} = 0 \tag{7}$$

(e.g., Zhao and Prentice 1990). These equations, as well as the original moment-based approach of Liang and Zeger, are often collectively referred to as "GEE1" (Liang, Zeger, and Qaqish 1992), since both assume the conditional independence of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. In contrast, Zhao and Prentice (1990) broadened this model to allow for joint estimation of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$:

$$U(\alpha,\beta) = \sum_{i=1}^{N} \begin{pmatrix} \mathbf{D}_i' & 0 \\ \mathbf{F}_i' & \mathbf{E}_i' \end{pmatrix} \begin{pmatrix} \mathbf{V}_i^{-1} & \mathrm{Cov}(\mathbf{Y}_i, \mathbf{W}_i) \\ \mathrm{Cov}(\mathbf{W}_i, \mathbf{Y}_i) & \mathbf{W}_i^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{Y}_i - \mu_i \\ \mathbf{Z}_i - \eta_i \end{pmatrix} = 0 \tag{8}$$

where $\mathbf{F}_i = \quad \boldsymbol{\alpha}_i / \quad \boldsymbol{\beta}$ (see also Prentice and Zhao 1991). This "GEE2" specification (Liang, Zeger, and Qaqish 1992) uses information about the intracluster correlations to improve the efficiency with which we can estimate $\boldsymbol{\beta}$, as well as providing direct estimates of $\boldsymbol{\alpha}$ and their standard errors. Direct estimation of GEE2 models requires specification of the third and fourth moments of the distribution; in practice, however, a natural extension of Liang and Zeger's notion of "working" covariance matrices (e.g., Zhao and Prentice 1990; Prentice and Zhao 1991) yields estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ which are nearly fully efficient relative to MLE (Liang, Zeger, and Qaqish 1992).

Provided that the specification of both the mean and the correlation structure are correct, the GEE2 approach permits more efficient estimation of the parameters of interest, as well as allowing model-building on the intracluster correlations. Its primary drawback is that, because of the simultaneous estimation process, consistent estimates of $\boldsymbol{\beta}$ depend on proper specification of the correlation matrix; thus, the GEE2 estimator lacks the robustness of GEE1 in estimating the regression parameters when the form of the intracluster dependence is unknown. That is, GEE2 estimates depend critically on the correct specification of the intracluster correlations; in the event that those dependencies are misspecified, *neither* $\boldsymbol{\alpha}$ nor $\boldsymbol{\beta}$ will be estimated consistently. Moreover, work by Liang, Zeger, and Qaqish (1992) demonstrates that using GEE2 results in minimal efficiency gains in estimating $\boldsymbol{\beta}$ as compared to GEE1. This fact, coupled with the inconsistency of GEE2 in the face of covariance misspecification, leads Prentice and Zhao to warn against using GEE2 simply to achieve more efficient estimates of $\boldsymbol{\beta}$ (1991, 831–832). Conversely, "GEE1 can be extremely inefficient for estimation of $\boldsymbol{\alpha}$" (Liang, Zeger, and Qaqish 1992, 16), suggesting that, when the intracluster correlations are of substantive interest, GEE2 may provide the better approach.

The potential advantages of the GEE approach for estimating models with correlated data are several. First, these models allow for explicit incorporation of knowledge regarding within-unit interdependencies through specification of the working correlation matrix. At the same time, the parameter estimates obtained through application of these models are robust to misspecification of those correlations, an important trait, since our understanding of those relationships is often imperfect at best. Additionally, the models in (7) and (8) also provide an opportunity to gain substantive insight into the determinants of within-cluster correlation; in particular, they allow analysts to assess the effect of covariates on the conditional correlation among observations within a cluster.

GEE Models in Practice

In this section I address several practical issues in the use of GEE models. These include general matters relating to the choice of GEE versus other kinds of models for correlated data, specification of the correlation structure, interpretation of coefficients, measures for goodness-of-fit, and issues concerning the use of GEEs in the presence of missing data. I conclude with an overview of a few recent developments in GEE modeling.

Conditional and Marginal Models for Correlated Data: A Comparison and Discussion

The method of GEE is a *marginal* (or *population-averaged*) approach to estimation with correlated data. Marginal models differ from more commonly-used *cluster-specific* (also called *conditional*, or *subject-specific*) approaches to correlated data analysis. Cluster-specific approaches model the probability distribution of the dependent variable as a function of the covariates and a parameter specific to each cluster. This latter term may either be estimated concurrently with the model (as in the fixed-effects approach) or be assumed to follow some stochastic distribution (as in a random-effects specification). Marginal models, by contrast, model the marginal (or population-averaged) expectation of the dependent variable as a function of the covariates.[11] These differences have ramifications for the usefulness of each type of model in certain conditions and on the interpretation of the estimated parameters.

Population-averaged and cluster-specific models represent two fundamentally different ways of thinking about covariate effects on the phenomenon of interest and about the nature of the correlation among observations within a cluster. Conditional models take the general form:

$$\Pr(Y_{it}) = f(X_{it}\beta_C + v_i) \tag{9}$$

where the v_is represent unit-specific effects. These individual effects can be thought of as latent, subject-specific propensities towards the outcome variable which are independent of the model's covariates. Fixed-effects models estimate these parameters directly, while random-effects models assume that the v_is follow some distribution (often a Normal), the variance of which is then estimated along with the parameters of the model.[12] In either case, in subject-specific models these individual-effects constitute the primary mechanism by which intracluster correlation is handled.

The analogous population-averaged model is simply:

$$\Pr(\mathrm{Y}_{\mathrm{it}}) = g(\mathrm{X}_{\mathrm{it}}\beta_{\mathrm{M}}) \tag{10}$$

Marginal models, then, "model ... the average response over the sub-population that shares a common value of X" (Diggle, Liang, and Zeger 1994, 131). Note that no individual (i.e., cluster-specific) effects are included in the model. Instead, intracluster correlation is accounted for by adjusting the covariance matrix of the estimated parameters to account for nonindependence across observations or time points; this is the approach adopted by the GEE method, among others.

The distinction between conditional and population-averaged models is critical, because the conditional model's parameter β_C is, in fact, a completely different quantity from the parameter β_M in (10). The former represents the effect of a change in X_{it} on $\Pr(Y_{it})$ *for the same individual i*, that is, for an observation with the same unit-specific effect ν_i. By contrast, the population-averaged coefficient β_M represents the *average* effect,

[11]For good discussions of this important distinction, including differences in interpretations of odds ratios and other characteristics of the two kinds of models, see Hu et. al. (1998), Neuhaus (1992); Neuhas, Kalbfleish, and Hauck (1991), Pendergast et al. (1996) and Zeger and Liang (1992).

[12]Typically σ_v^2 is not directly estimated; instead, most standard software packages estimate $\rho = \sigma_v^2 / (\sigma_v^2 + 1)$. In the probit case, since the variance of the error term is typically normalized to 1.0, this has an easy interpretation as the proportion of the total variance which is due to variability in the νs.

across the entire population, of a one-unit shift in X_{it} on $\Pr(Y_{it})$. Intuitively, the reason for the difference lies in the fact that, because of the nonlinear nature of the models in question, the mean of the differences in the log-odds of a particular outcome is not equal to the differences of the means of the same log-odds (see Pendergast et al. 1996, 112–113).[13] The extent of differences between the two are illustrated clearly in Neuhas, Kalbfleish, and Hauck (1991, 27–31). In general, $\beta_M < \beta_C$; that is, the coefficient for the marginal model will be smaller than that for the conditional model. This difference increases as σ_ν^2 increases, and decreases as β_M and $\beta_C \to 0$.[14] In the limit, $\beta_M = \beta_C$ when $\sigma_\nu^2 = 0$, i.e., when no individual-specific effects are present, or under independence of observations within each cluster.

The choice between conditional and marginal models should thus be seen as primarily a substantive, rather than a statistical, one. While some properties of each type of model make them more or less desirable in certain situations,[15] as a substantive matter the former are more useful when the primary question of interest is the effect of changes in covariates within a particular observation, while the latter are more valuable for making comparisons across groups or subpopulations. Hu et al.'s (1998) examination of the effects of a particular smoking prevention plan on the incidence of smoking among secondary school students provides a nice illustration:

> In our example, if the differences in the increasing trends of smoking between treatment and control groups are of interest, then random-effects models estimating the changes in individuals' smoking behavior across time are more appropriate. . . . On the other hand, the GEE approach is more desirable when the objective is to make inferences about group differences . . . if we want to estimate the averaged treatment effect, regardless of individual change over time, then the population-averaged parameter is of more interest. (Hu et al. 1998, 701)

In a political science context, we might consider the example of the "democratic peace" (e.g., Oneal and Russett 1997). If one were interested in, say, the effect of democratization on the propensity for a particular nation or pair of nations to go to war, then the conditional approach would be more appropriate. If, instead, we wished to assess the general propensity of autocracies and democracies to engage in interstate conflict, a marginal approach (such as the GEE) would be called for. The central point is that the model chosen should be driven first and foremost by the nature of the question posed by the researcher, rather than by software availability, computational convenience, or data issues.

Specification of the Working Correlation Matrix

As noted previously, a key difference between the GEE and other approaches to correlated data is the necessity of specifying, a priori, the structure of the "working" correlation matrix. The decision regarding specification of the correlation matrix has been called "more art than science" (Williamson 1999) and has received comparatively scant treatment in the literature on GEEs. As a result, little can be said, in general terms, about the sensitivity of GEE estimates to variation in the way in which the intracluster correlations are specified, and there is some debate as to the importance of that specification.

On one hand, accurately modeling the correlation structure of the data improves the efficiency of the GEE estimates (e.g., Liang, Zeger, and Qaqish 1992). The extent of that improvement may vary substantially depending on the nature of the data at hand; efficiency losses will be greatest, for example, when the extent of within-cluster correlation is high (Zhao, Prentice, and Self 1992) when the model includes covariates which vary within a cluster (Fitzmaurice 1995), when sample sizes are small, and when there are large variations in cluster size (see, generally, Mancl and Leroux 1996). Most authors agree, however, that "careful modeling of the correlation structure leads to improved estimation of the regression and the correlation parameters" (Shults and Chaganty 1998, 1623; but see Lumley 1996a). In many if not most cases, the researcher has some a priori expectation about the nature of the within-cluster variation in Y. Moreover, in instances where the within-cluster correlation is of some substantive interest, specifying and estimating a correlation matrix which accurately reflects the data-generating process offers the possibility of gaining additional insights into the substantive question at hand. In addition,

[13] A different way of thinking about the difference is as an issue of measurement error (e.g., Judge et al. 1985, §17.2), in which the ν_is are treated as individual-specific errors in the independent variables. In this light, the difference between the marginal and conditional estimators can be thought of as attenuation bias due to the marginal model's failure to account for the errors in estimation (Diggle, Liang, and Zeger 1994, 141).

[14] The precise nature of the relationship varies. In the logit case, for example, Zeger et al. (1988) show that $\beta_M \approx \beta_C / (1 + 0.35\,\sigma_\nu^2)^{1/2}$ (see also Zeger and Liang 1992). Note, however, that numerous studies (e.g., Zeger et al. 1988; Neuhas, Kalbfleish, and Hauck 1991; Neuhaus 1993) indicate that significance levels and inferences from both kinds of models will generally be similar.

[15] For example, the random-effects model's requirement that the νs be uncorrelated with the X_{it}s is often unrealistic in practice; see Wawro (2000) for an example in which this assumption is relaxed.

recent research has indicated that, for binary dependent variables, considerable care needs to be taken in those instances where the intercluster correlations may be negative (Hanley, Negassa, and Edwardes 2000). As a general rule, then, choosing a correlation structure on the basis of theoretical considerations is recommended.

On the other hand, there are a number of reasons not to take this prescription too seriously. First, as noted above, GEE1 estimates of the main parameters of interest $\boldsymbol{\beta}$ are consistent even if the correlation structure chosen is incorrect. This robustness is, in fact, one of the primary advantages of the GEE, since it means that (asymptotically) good estimates of the regression parameters can be obtained even when the researcher is unsure of the exact nature of the correlation among the Y_{it}s. In addition, numerous simulation studies have shown that the efficiency advantage gained through correct specification of the correlations is modest (Liang and Zeger 1986; Liang, Zeger, and Qaqish 1992). Moreover, as Lumley (1996a, 354) notes, when sample sizes are small, the loss of efficiency and the potential for singularities in the correlation matrix makes large, complicated covariance structures inadvisable.[16]

Two additional points are worth remembering as well. First, GEE1 models treat intercluster correlation as a problem regarding accurate inferences about the regression estimates, not as a phenomenon of intrinsic interest. This is captured in the fact that the Liang and Zeger "estimates" of the intracluster correlation(s) are in fact nuisance parameters; they lack standard errors and are thus not available for hypothesis testing or statistical inference. As a result, if the correlation parameters are of substantive interest, other techniques (such as GEE2) may be better suited to the problem at hand. Second, because the responsiveness of GEE estimates to different imposed correlation structures varies widely from one application to the next, sensitivity analyses are highly recommended. One means by which this can be done is to estimate several GEE models, each time changing the correlation specification, and look for differences in the results. In this vein, Diggle, Liang, and Zeger recommend that:

> When the regression coefficients are the scientific focus . . . one should invest the lion's share of time in modelling the mean structure, while using a reasonable approximation to the covariance. The robustness of the inferences about $\boldsymbol{\beta}$ can be checked by fitting a final model using different covariance assumptions and comparing the two sets of estimates and their robust standard errors. If they differ substantially, a more careful treatment of the covariance model may be necessary. (1994, 145)

Model Estimation, Interpretation, and Goodness-of-Fit

Because of their wide use in biostatistics and epidemiology, routines for estimating GEEs are available in a broad range of standard statistical packages.[17] In general, the Fisher scoring algorithm behaves well; as with all such iterative procedures, however, under certain circumstances it may fail to converge. As a rule, convergence of GEEs becomes more difficult as sample size decreases, as the number of correlation parameters being estimated increases, and as the size of intracluster correlations increases. In their simulation study, Lipsitz et al. (1994) found that, for $N = 15$ and $r = 0.60$, 65 percent of their convergence problems with the fully iterated GEE model could be traced to singularities in the estimated variance-covariance matrix, while the remaining 35 percent were due to exploding estimates of $\boldsymbol{\beta}$. As a practical matter, then, one may be forced to trade off complexity in specifying the working correlation matrix in favor of computational tractability, particularly when the data are sparse or intracluster correlations are high.

As noted in above, GEEs are simply one form of marginal (or population-averaged) model, similar to those widely used in cross-sectional analysis. As a result, most standard approaches for interpreting these models are also useful for discussing GEE estimates. These include examining odds ratios, first derivatives, and changes in predicted values, along with their associated measures of uncertainty (e.g., Herron 1999; King, Tomz, and Wittenberg 2000). Thus, familiarity with (for example) logit or probit is all that is required to interpret the results of a binary-dependent-variable GEE model. One significant difference between GEEs and cross- sectional models lies in the fact that GEEs are not based on full-information maximum likelihood. This means that the widely-used likelihood-ratio tests for model fit and block significance testing are not available for GEEs. Fortunately, such models are fully amenable to other, asymptotically identical tests (e.g., score and Wald tests), and most statistical packages which implement GEEs also provide procedures for conducting such tests.

[16]An alternative, proposed by Lipsitz et al. (1994) is to use the one-step estimator; that is, the estimates obtained from one iteration of the GEE. They show that this estimator is often similar to the fully iterated GEE results and is particularly useful in small samples and when the intracluster correlation is high, i.e., when the fully-iterated GEE may not converge.

[17]A review of software for estimating GEEs is presented in the appendix.

The issue of summary goodness-of-fit statistics for GEE models is a problematic one, in part because the residuals from these models are correlated and thus do not lend themselves readily to such statistics (Chang 2000). Recently, however, a few measures for model goodness-of-fit have been developed for GEE models, though thus far they have been limited to the case of binary dependent variables. Barnhart and Williamson (1998) propose a test based on partitioning the covariate space into distinct "regions" defined by combinations of covariate values, and then forming score statistics which are asymptotically chi-squared. The major drawback of this method lies in the necessity of partitioning the covariate space, particularly when sample sizes are small and/or many or continuous covariates are included.

An alternative approach is that of Horton et al. (1999), which extends the method of Hosmer and Lemeshow (1980) to the GEE context. The statistic is based on estimating a model and generating predicted probabilities $\hat{p}_{it}$. The researcher then divides the data into G (typically ten) groups based on deciles of the predicted probabilities and defines $G - 1$ dichotomous indicator variables corresponding to these deciles. These additional variables can then be included in a supplementary model and score or Wald statistics derived; a significant statistic indicates a lack of model fit. Other alternatives include measures based on reductions in entropy and deviance and on indices of concordance between ordinal rankings of predicted versus actual values (Zheng 2000). Comparisons of these different approaches in practical situations awaits future research.

GEEs and Missing Data

The issue of missing data deserves special treatment in the context of GEE models. Little and Rubin (1987; see also Sherman 2000) outline what has become the standard classification for missing data, one which depends on the mechanism by which the data are missing. For data which are missing completely at random (MCAR), the probability of missingness is completely independent of Y, whether missing (Y_m) or observed (Y_o). Data which are missing at random (MAR) have a probability of missingness which is independent of the vector of missing outcomes Y_m, though it may depend on the observed outcomes Y_o; collectively, MCAR and MAR are often referred to as *ignorable* nonresponse mechanisms. Finally, nonignorable (or informative) missing data are those for which the probability of missingness depends on other missing values Y_m. For inference based on complete likelihoods, the important distinction is between ignorable and informative missingness: if the data are MCAR or MAR, valid inferences are possible without explicitly modeling the missing data mechanism.[18]

For GEE models, however, this is not the case. Because GEEs do not specify the full conditional likelihood, data which are missing at random (i.e., in which the probability of missingness may depend on past values of Y) will not necessarily yield consistent estimates of $\boldsymbol{\beta}$.[19] In particular, the GEE's robustness to misspecification of the working covariance matrix is compromised by data which are MAR; under these circumstances, it is necessary that the working correlation matrix $\mathbf{R}_i(\boldsymbol{\alpha})$ be the true correlation matrix for consistency to obtain (Zeger and Liang 1986, 129). While simulations have shown that GEEs perform "remarkably well" (Fitzmaurice, Laird, and Rotnitzky 1993, 299) in the presence of data which are MAR, the extent to which GEE estimates of $\boldsymbol{\beta}$ are biased when the data are MAR depends on a number of factors, including the extent of missing data, the accuracy of the model's specification, the presence of time-varying covariates, and the specification of the working correlation matrix (Fitzmaurice, Laird, and Rotnitzky 1993).

Given both the pervasiveness of missing data in longitudinal studies and its potential importance in the GEE context, a great deal of effort has gone into addressing this issue. Because data which are MCAR pose no problem for the GEE estimator, a number of tests have been developed to assess whether the missing data mechanism is independent of Y (e.g., Diggle 1989; see also Diggle, Liang, and Zeger 1994, chapter 11). One recent test is that of Chen and Little (1999), which decomposes the data into groups based on patterns of missing data and constructs a Wald-type statistic based on differences in estimation results across the subgroups. Another test constructs indicator variables for each pattern of missingness and includes those covariates in the estimated model; a joint test of the significance of the missingness indicators serves as a test for data missing completely at random (Park and Lee 1997).

If data are not found to be MCAR, there are a number of options available to researchers. One approach uses corrections based on imputation (e.g., Paik 1997; Xie and Paik 1997), where missing data are imputed from available data and analyses conducted on the "complete" data. Alternatively, model-based approaches correct for

[18]Laird (1988), however, notes that in longitudinal data, the distinction between MAR and nonignorable response mechanisms is problematic; because covariates typically affect both observed and unobserved responses, and because repeated observations are correlated, covariates may have secondary effects on missing responses as well.

[19]This inconsistency is due to the fact that, conditional on the observed response, $E(Y_i - \mu_i) \quad 0$ when data are MAR.

missing data by explicitly modeling the missingness, using an approach akin to a selection model (see, e.g., Hogan and Laird 1997 for a review). Carlin et al. (1999) illustrate one simple model-based approach, wherein an initial logit model of the presence (0) or absence (1) of missing data for each subject i at each time point t is used to generate weights corresponding to the inverse probability of nonmissingness. These weights are then used in the GEE model to correct for potential biases due to missing data.

Extensions of GEE

While the bulk of work on GEE models has been done in the context of binary response variables, a number of recent developments have taken GEEs beyond the binary case. GEEs for event counts, suggested by Zeger and Liang (1986), were initially outlined by Thall and Vail (1990). GEEs for unordered polychotomous responses are presented in Lipsitz, Kim, and Zhao (1994) while GEEs for ordinal dependent variables have been discussed by Stram, Wei, and Ware (1988), Liang, Zeger, and Qaqish (1992), Lumley (1996a), and others. More recently, Wild and Lee (1996) have integrated GEEs and generalized additive models (GAMs) (Hastie and Tibshirani 1990), thus combining the flexibility of the latter with the desirable properties of the former when the data are correlated. GEE2 methods have also been used to investigate questions where spatial correlation is a concern (e.g., Albert and McShane 1995; Mugglestone, Kenward, and Clark 1999). And GEE models continue to be an active area of development in biostatistics and epidemiology, with new generalizations and applications appearing rapidly.

GEE Models for Political Science: Three Applications

In this section, I illustrate the usefulness of GEEs for modeling correlated data in three diverse circumstances. The first compares GEE and random-effects models for binary outcomes, using data on the House of Representatives' votes on the four articles of impeachment against President Clinton. The second illustrates the applicability of GEE models to data containing temporal correlation, and to event count data, by reexamining time-series cross-sectional data on joint membership in international organizations. The final example shows how GEEs can be useful in situations where the correlation is of a nontemporal or spatial nature, here in the context of decision making in a collegial body (in this case, the U.S. Supreme Court).

A GEE Model of Impeachment Voting

On December 11th, 1998, the House Judiciary Committee voted out four articles of impeachment against President Clinton—perjury before the grand jury (Article I), perjury in Paula Jones' lawsuit (Article II), obstruction of justice (Article III), and abuse of power (Article IV). On December 19th, the House approved Articles I and III by votes of 228–206 and 221–212, respectively, while defeating Articles II and IV by margins of 205–229 and 148–285. Because the articles dealt with similar issues, and were voted on in a single day, it is unlikely that member's votes on any one article were unrelated to those on the other three. Here, I compare GEE models to the more familiar "random effects" specification for panel data.

As noted, the outcome of interest is each House member's vote on each of the four articles of impeachment (N = 1734), coded 1 for yeas and 0 for nays. For illustrative purposes, I examine a simple model of the determinants of members' votes, comprised of constituency, partisan, and ideological factors. The first is operationalized as support for Clinton among each member's constituency, measured as the percentage of the vote received by Clinton in that member's district in the 1996 election ($\bar{x}$ = 50.25, σ = 12.73). The influence of partisanship is captured as a dummy variable for political party, 0 for Democrats and 1 for Republicans ($\bar{x}$ = 0.53, σ = 0.50). Each representative's ideology is captured by his or her lifetime D-NOMINATE score (Poole and Rosenthal 1997), with higher values indicating greater conservatism ($\bar{x}$ = 0.16, σ = 0.56). For comparison purposes, I estimate four logit models:[20] a *random-effects* logit, where the individual-specific effects are assumed to be distributed as N(0, σ^2_ν), a GEE model assuming that members' votes are *independent*, an *exchangeable* GEE model (where all cross-vote correlations are assumed to be equal), and a GEE model in which the *pairwise* correlations among votes are allowed to vary without restriction. The primary interest is in assessing the effects of the covariates on members' votes, as well as on making comparisons across models; accordingly, I use Liang and Zeger's GEE1 method, which treats the correlations as nuisance parameters.[21] Estimates from each model are presented in Table 1.

Comparing the results from the four models, several things stand out. First, the inferences one would make regarding variable effects do not change substantially across

[20]All analyses in this section were done using the -xt- battery of commands in Stata 6.0. Data and commands for replicating these analyses are available at http://www.emory.edu/POLS/zorn/.

[21]Model results using Prentice's (1988) independent estimating equations for the correlation parameters yielded nearly identical results and are available upon request from the author.

TABLE 1 Models of Impeachment Voting

Variables	Random-Effects Logit	GEE Models Independent	Exchangeable	Pairwise	Estimate Ratio†
(Constant)	−0.540 (1.194)	−0.032 (1.193)	−0.032 (1.193)	−0.369 (1.123)	1.47
Clinton's 1996 Vote Percentage	−0.059 (0.025)	−0.056 (0.023)	−0.056 (0.023)	−0.054 (0.021)	1.09
Republican Member	1.006 (0.736)	0.619 (0.635)	0.619 (0.635)	0.673 (0.645)	1.49
D–NOMINATE	7.136 (1.043)	6.185 (0.829)	6.185 (0.829)	5.939 (0.828)	1.20
Estimated ρ	0.549 (0.097)	—	0.179 (n/a)	‡	—
Wald test ($\chi^2(3)$)	158.38	216.24	216.31	222.65	—

Note: N = 1734. Cell entries are logit estimates; numbers in parentheses are robust standard errors. Wald tests are for H_0: $\beta = 0$.

† This column reports the ratio of the random-effects estimate to the GEE estimate from the pairwise model.

‡ See below.

Estimated Correlation Matrix: Pairwise GEE Model

	Article I	Article II	Article III
Article II	0.418		
Article III	0.438	0.362	
Article IV	−0.133	0.359	−0.109

the four models: members in districts where Clinton received high levels of electoral support were less likely to vote in favor of impeachment, while more conservative members were more likely to favor it.[22] In the random-effects model, the estimated ρ indicates that, beyond the variables included in the model, substantial individual-level effects are present; this might lead us to believe that the votes of each member themselves were not independent, but rather related to one another. As noted by numerous authors, however, interpretation of the random-effects coefficients is problematic in this context; those estimates indicate the effect of a change in the independent variables on the dependent variable *for a particular member*. In this example, however, it is unclear what it means to talk about the effect of, e.g., a one-unit change in Clinton's 1996 electoral success in a specific member's district. Under these circumstances, marginal models (such as the GEE) provide a more natural means for assessing variable effects.

The largest differences in the results occur between the random-effects and GEE estimates; as we would expect, the random-effects estimates are uniformly larger in magnitude. Recall that Zeger and Liang (1988) show that, in the logit case, the ratio of random-effects to marginal model estimates should be approximately $(1 + 0.35\,\sigma^2_\nu)^{1/2}$. Here, the random-effects estimate of σ^2_ν equals 1.215, suggesting a ratio of random-effects coefficients to marginal estimates of 1.19. Examining the last column of Table 1, we find that the differences between the coefficients vary somewhat around this figure, with those for political party exhibiting the greatest divergence and that for ideology the smallest difference.

Examining GEE estimates from the different correlation structures reveals that those from the independence and exchangeable models are identical to three decimal places. This is to be expected, for two reasons. First, the estimated value of the correlation parameter in the exchangeable model is relatively small (0.18), indicating only a low-to-moderate level of correlation among the votes. Second, as noted in several studies (e.g., McDonald 1993; Fitzmaurice 1995), the independent model is close to fully efficient, particularly in cases where (as here) none of the covariates vary within clusters. The results differ slightly between the pairwise and exchangeable models, though again the relative magnitude of these differences is small, and in no case are statistical inferences affected. For

[22] The effects of party are undoubtedly swamped by the other two variables; the party variable correlates at −0.61 and 0.93 with Clinton's vote percentage and D-NOMINATE scores, respectively.

the pairwise model, it is also interesting to examine the matrix of correlations among the different articles generated by the GEE estimation. The highest, positive correlations are found among articles I through III, while article IV exhibits the lowest correlations with the other three. Also, two of the correlations (between article IV and articles I and III) are negative. These results are evidence toward the widely held view that article IV was the weakest of the four articles, and that, in the minds of members of the House, it was qualitatively different from the other three (e.g., McGrory 1998). Note, however, that if these correlations were of primary substantive interest, a GEE2 approach would be recommended for estimation.

In this example, application of the GEE model presents us with an easily implemented means of accounting for correlation across members' votes on impeachment. Where, as is the case here, the primary quantity of interest is the effects of the independent variables on the outcome, GEEs provide an attractive means of accounting for that correlation, particularly when covariates remain fixed within clusters and cross-unit variable effects are of primary concern.

GEE Models for Temporally Correlated Event Counts: Joint Memberships in International Organizations

My second example draws on the study of international relations and, in particular, the subject of international organizations. In their recent article, Russett, Oneal, and Davis (1998) consider the factors influencing joint membership in international intergovernmental organizations (IGOs). In their view, such organizations comprise the "third leg of the Kantian tripod for peace" (Russett, Oneal, and Davis 1998, 441), and understanding the factors which increase their membership is an important task. Moreover, in addition to its intrinsic importance, their data provide a means of illustrating the usefulness of GEE models to temporally correlated data and data in which the outcome variable consists of event counts.

The data examined are those of Oneal and Russett (1997) and comprise all "politically relevant" international dyads during the years 1950–1985. Following Russett, Oneal, and Davis (1998), I consider a model of the number of joint IGO memberships shared by two nations in a dyad; this variable ranges from a low of zero to a high of 130, with a mean of 34.5 and a standard deviation of 18.1. Russett, Oneal, and Davis model these joint memberships as a function of six variables common in quantitative work on international disputes: the presence of an intradyadic *dispute* (defined as a militarized interstate dispute and based on the Correlates of War data), POLITY III *democracy* scores, economic *interdependence* (measured as dyadic trade as a proportion of GDP), the presence of an *alliance* in the dyad, the (log of the) *distance* between capitals or major ports, and per capita *GDP*.[23] Both the dispute and dependence scores are lagged one period.

For comparative purposes, I again estimate several models, beginning with a simple Poisson model with robust standard errors, grouped by dyad. The results, presented in column one of Table 2, correspond closely to those reported for Russett, Oneal, and Davis's OLS regression: democracy, interdependence, alliances, and growth increase joint IGO membership, while previous disputes and greater geographical distance decrease them. Of these, previous disputes and dependence are, by traditional standards, estimated somewhat imprecisely ($p = .06$ and $.09$, respectively, two-tailed), but the substantive implications of the results remain unchanged.[24]

An analysis which takes account of intradyadic autocorrelation presents a somewhat different picture. I present two alternative models, both estimated via GEE1. Because the dependence within each dyad is temporal in nature, I specify the working within-observation correlation matrix to follow an AR(1) format: for a particular dyad, the correlation between Y_{it} and Y_{is} ($t > s$) is specified to be $\rho^{|t-s|}$. Results from reestimation of Russett, Oneal, and Davis's model are presented in column two of Table 2, and several differences from the independent estimates are clear. The influence of previous disputes on memberships decreases markedly once intradyadic correlation is accounted for. Similarly, the effect of democracy, while remaining statistically significant at conventional levels, decreases in size by a factor of nearly twenty, while that for interdependence both declines and reverses sign; after considering intracluster correlation, such dependence is actually negatively related to joint IGO memberships. The effects of the other three variables remain consistent with the basic Poisson model, though the size of the alliance variable's effect decreases

[23] See Russett, Oneal, and Davis (1998) for the details of variable codings.

[24] Russett, Oneal, and Davis are sensitive to the fact that their findings are of a preliminary nature, particularly due to the fact that IGO membership, militarized disputes, and trade are likely to be endogenous to one another. They note that "(F)urther support for (their findings) . . . will require additional investigation as more sophisticated techniques become available" (1998, 461). This same caveat applies to these results: the GEE estimates given here deal only with intradyadic correlation, leaving unaddressed the matter of reciprocal causality among the variables in the model. As a result, these findings should be seen as primarily illustrative, and certainly not as the last word on the topic of international organizations and conflict.

TABLE 2 Models of Joint IGO Memberships

Variables	Poisson Model	GEE Models Poisson	GEE Models Poisson w/lag
(Constant)	3.563	3.592	2.993
	(0.024)	(0.023)	(0.026)
Joint IGO Membership$_{t-1}$	—	—	0.015
			(0.001)
Dispute Involvement$_{t-1}$	–0.066	0.001	0.003
	(0.035)	(0.002)	(0.003)
Democracy Score	0.018	0.001	0.002
	(0.002)	(0.0005)	(0.0004)
Dependence Score$_{t-1}$	2.239	–0.500	–0.838
	(1.310)	(0.246)	(0.442)
Allies	0.221	0.033	0.047
	(0.021)	(0.008)	(0.007)
Distance /10000	–0.640	–0.781	–0.278
	(0.043)	(0.059)	(0.030)
Per Capita GDP/10000	0.677	0.590	0.121
	(0.033)	(0.032)	(0.016)
Estimated ρ	—	0.989	0.958
N	18657	18656	18656

Note: Numbers in parentheses are robust standard errors. GEE models assume an AR(1) correlation structure.

considerably. Also interesting to note is that, particularly in the case of variables which vary both within and across dyads, the standard error estimates for the GEE model are generally lower, and in some cases much lower, than those obtained from the standard Poisson model, indicating that accounting for intradyadic correlation significantly improves the precision with which we are able to estimate covariate effects on joint IGO memberships (Fitzmaurice 1995).

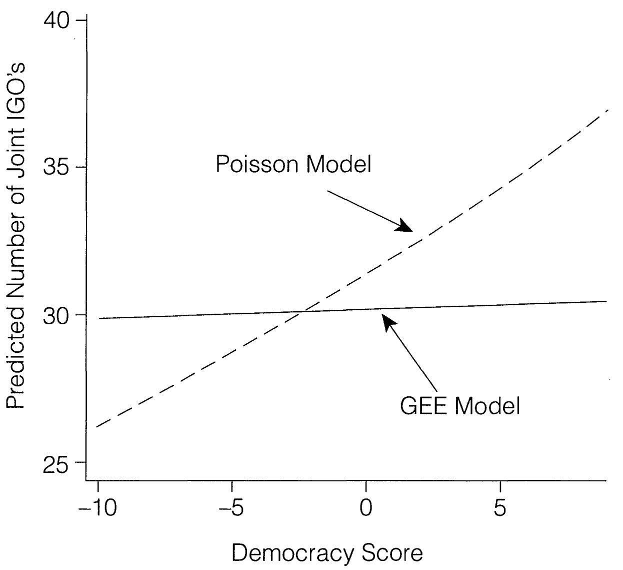
The differences between the independent Poisson and GEE estimates are illustrated graphically in Figure 1, which plots the predicted number of joint IGO memberships as a function of the democracy variable. The estimated effect of democracy in the standard Poisson model is quite large, increasing the number of joint IGOs from twenty-six to thirty-seven across the range of its variation. In contrast, after accounting for within-dyad autocorrelation via the GEE, the effect of higher levels of democracy is to increase joint IGO memberships by only one across the whole range of the democracy variable. Substantively, this result is consistent with the fact that membership in international organizations is "sticky," typically persisting from one year to the next.[25] To the extent that there was a general trend toward higher levels of democracy in the period studied, an analysis which fails to account for the natural temporal dependence in IGO memberships will be likely to overstate the effect of increases in levels of democracy on those memberships.

As a check on the robustness of the results, I estimate a third model, this time including a one-period lagged value of the dependent variable among the regressors.[26] If the findings are sensitive to the high level of correlation in the data, inclusion of a lagged dependent variable may reduce this dependence. Results from this analysis, however, are similar to that obtained in the initial GEE estimation, albeit with a few minor differences. Lagged memberships exert a substantial effect on current memberships, but their inclusion in the model

[25] Not surprisingly, the estimate of the intracluster AR(1) coefficient suggests the presence of a substantial amount of temporal correlation within dyads; again, if this autocorrelation were of primary substantive interest, an approach based on the GEE2 method would be recommended.

[26] This is a less-than-optimal solution to the problem at hand and is done primarily as a sensitivity check on the results in column 2 of Table 2. More detailed analyses of the problems associated with count data in a temporal setting can be found in Cameron and Trivedi (1998, chapter 9), Diggle, Liang, and Zeger (1994, chapter 10), Brandt et al. (2000), and Zeger and Qaqish (1988).

FIGURE 1 **Predicted Joint IGO Memberships, by Level of Democracy, Poisson and GEE Models**



Note: Predicted counts are based on estimates presented in Table 2, with all other variables held constant. See text for details.

does little to reduce the temporal correlation present in the data. The influence of previous disputes increases slightly, but remains statistically undifferentiable from zero, and the effect of democracy remains small. With the exception of the variable for interdependence, inclusion of lagged disputes increases the precision with which we are able to estimate the other variable effects, in some cases dramatically.

This second example illustrates two important characteristics of GEE models. First, unlike many other models for dealing with temporally correlated data, GEEs provide an easy means for handling data in which the outcome variable is noncontinuous—in this instance, a nonnegative event count. Second, this illustration shows how the use of models which are able to account for conditional correlation in the data can yield results which differ considerably from those obtained when such correlation is ignored. While this is not a clear argument in favor of GEEs as against other kinds of models, it does suggest that, when we have reason to expect that our data are nonindependent, we would do well to consider statistical approaches which take that correlation into consideration.

Interjustice Correlations in Supreme Court Voting

For a third illustration of the usefulness of GEEs, I turn to an analysis of the U.S. Supreme Court. Largely neglected in quantitative studies of Supreme Court voting has been the influence of other justices on each member's votes: the impact of the informal, and largely unobserved, bargaining, persuasion, and even coercion among the Court's members. On a small, collegial decision-making body such as the Court, we would expect such influences to be substantial, and previous biographical and anecdotal accounts in this area have found that justices do influence one another to a large degree (e.g., Mason 1956, 1964; Murphy 1964). More empirical and quantitative studies have also documented the impact of interjustice influence on Supreme Court decision making (e.g., Spaeth and Altfield 1985; Maltzman and Wahlbeck 1996a,b; Wahlbeck, Spriggs, and Maltzman 1998), yet most examinations of Supreme Court voting ignore this potentially critical influence.

I illustrate the usefulness of GEEs to this question by examining Supreme Court voting during three recent terms of the Rehnquist Court. This example differs from the previous sections in that, as suggested above, the nature of the correlation here is more "spatial" than temporal; this example thus shows how GEE models can address correlated data outside the temporal context.[27] In particular, modeling Supreme Court voting points out one of the most useful characteristics of the GEE2 approach: its ability to simultaneously model the mean, variance, and marginal intracluster correlation of the response variable. The data used are all civil rights and liberties cases decided by the Court during the 1994–1996 terms and are drawn from Harold Spaeth's Supreme Court Judicial Database (Spaeth 1998), treating the docket number as the unit of analysis. Excluding missing data, the number of such cases is 180; with an average of 8.9 votes per case, this yields a total of 1600 individual votes for analysis. The dependent variable is the justice's votes in civil rights and liberties decisions, coded 0 for a conservative outcome and 1 for a liberal one.

I examine an extended attitudinal model of Supreme Court decision making (e.g., Segal and Spaeth 1993). The model posits that justices' votes are determined primarily by their political ideologies, but also that ideology both *conditions* and is *conditioned by* other factors specific to the case in question (e.g., Traut and Emmert 1998). I therefore include normed Segal-Cover (1989) measures of judicial ideology (Epstein and Mershon 1996), which range from 0.0 for the most conservative justices to 1.0 for the most liberal, with the expectation that this measure will have a positive impact on the justices' votes. This effect, however, is tempered by a number of case-specific variables; accordingly, I include indicator variables for the ideological direction of the lower court deci-

[27]A good general reference for spatial statistics is Cressie (1993).

sion (conservative = 0, liberal = 1) as well as for which position, if any, the United States took as a party to the case. These three variables are then interacted with justice's liberalism, to assess the extent to which each moderates the influence of ideology on voting behavior. Note that the ideology variable varies only across justices, not across cases, while the reverse is true for the case-specific factors, while the interactions vary across both.

A more sophisticated model might also consider the influence of potentially conflicting beliefs on the votes of the justices. Recent work in political behavior suggests that the presence of conflicting values vis-a-vis a particular issue leads to increases in the variance of survey respondents' answers to questions on that issue (e.g., Feldman and Zaller 1992; Alvarez and Brehm 1995). In the context of the Court, we might expect that the variance of the underlying latent variable will be greater for moderate justices (who, presumably, will more often be conflicted over the issues presented in a case than their more extreme counterparts). Thus, I posit a curvilinear effect for ideology on the variance of the responses. The GEE approach allows for a number of variance "links"; following Alvarez and Brehm (1995), I adopt an exponential approach, such that the variance σ_{ij} for the jth justice in the ith case is:

$$\sigma_{ij} = \exp\left(\gamma_0 + \gamma_1 \text{Liberalism}_{ij} + \gamma_2 \text{Liberalism}_{ij}^2\right)$$

In this formulation, a significant, positive sign on γ_1, and a significant negative sign on γ_2, are consistent with the hypothesis of greater variance for moderate justices.

Finally, it is of some interest to model the marginal correlation among justices' votes, as a function of their ideologies. If interpersonal influence does impact the justices' voting, it would not be unreasonable to speculate that justices who are more ideologically similar would, on average, have higher marginal correlations among their votes (that is, even after controlling for the direct effects of ideology) than those who are ideologically dissimilar. Consider that, in any given case i, there are J voting justices, each of whom may be indexed by j and k; as noted above, each such observation has some variance σ_{ij} and σ_{ik}. The quantity of interest is the pairwise marginal correlation r_{ijk} between the jth and kth justices (j, k = 1,2,..m, where $m = \frac{J(J-1)}{2}$ and j k) in the ith case. Here, I operationalize this correlation in terms of the absolute value of the difference between the ideology scores of the two justices who make up the pair in question, again using an exponential link:

$$r_{ijk} = \exp\left(\delta_0 + \delta_1 \left|\text{Liberalism}_{ij} - \text{Liberalism}_{ik}\right|\right)\left(\sigma_{ij}\sigma_{ik}\right)^{1/2}$$

A significant, negative estimate of δ_1 is evidence that ideological divergence is negatively related to the degree of correlation among justices' votes on the Court.[28]

As in previous examples, I begin with a basic GEE model for comparison purposes.[29] Column 1 of Table 3 illustrates that, as expected, ideology has the expected positive effect on the probability of a justice's vote; a change in judicial liberalism from zero (its lowest value) to one (its highest) corresponds to a $exp(1.901) = 6.7$-fold increase in the probability of voting in a liberal direction. Conversely, the tendency for the Court to accept and reverse cases is indicated by the justices' substantially lower probability of voting liberally in a case decided liberally in the lower court. Estimates of the direct effects for the two U.S. variables indicate that, for the most conservative justices, the presence of the U.S. as a litigant has little effect on the votes of those justices. In both cases, however, those variables' effects interact with the ideology of the justice in question: the effect of ideology is exacerbated in cases where the U.S. argues in favor of a liberal outcome, but all but disappears if the government advocates for a conservative decision. The same is true, albeit to a lesser extent, when the lower court decided the case liberally; in such cases, the influence of liberalism on justices' votes is increased, suggesting that justices' ideology plays a larger role in judicial votes to reverse than to affirm.

The inclusion of a model for the variance does little to change the effects of the covariates on the probability of a liberal vote, but reveals an interesting aspect of Supreme Court decision making: moderate justices do, in fact, exhibit higher marginal variances in their voting than do more extreme jurists. Estimates for both the linear and quadratic terms are in the expected direction, indicating that the variance first increases, then decreases as a function of judicial ideology. This effect is illustrated in Figure 2, which plots the estimated variance from column 2 of Table 3 as a function of judicial liberalism. The estimated variance in Figure 2 reaches its maximum value at a Segal-Cover score of 0.42, very near the center of the distribution for this variable.

[28] An alternative to the exponential link is the hypergeometric function:

$$r_{ijk} = \frac{1 - \exp\left(-\delta_0 - \delta_1 \left|\text{Liberalism}_{ij} - \text{Liberalism}_{ik}\right|\right)}{1 + \exp\left(-\delta_0 - \delta_1 \left|\text{Liberalism}_{ij} - \text{Liberalism}_{ik}\right|\right)}\left(\sigma_{ij}\sigma_{ik}\right)^{1/2}$$

which has the advantage of restricting the estimates of the correlations to the range (–1,1). Results estimated using a hypergeometric link for the correlation are nearly identical to those using an exponential link and are available upon request from the author.

[29] Models in this section were estimated using the *EE* software (QGE 1994).

TABLE 3 Model Comparisons for Civil Rights and Liberties Voting on the Burger Court, OT1994-96

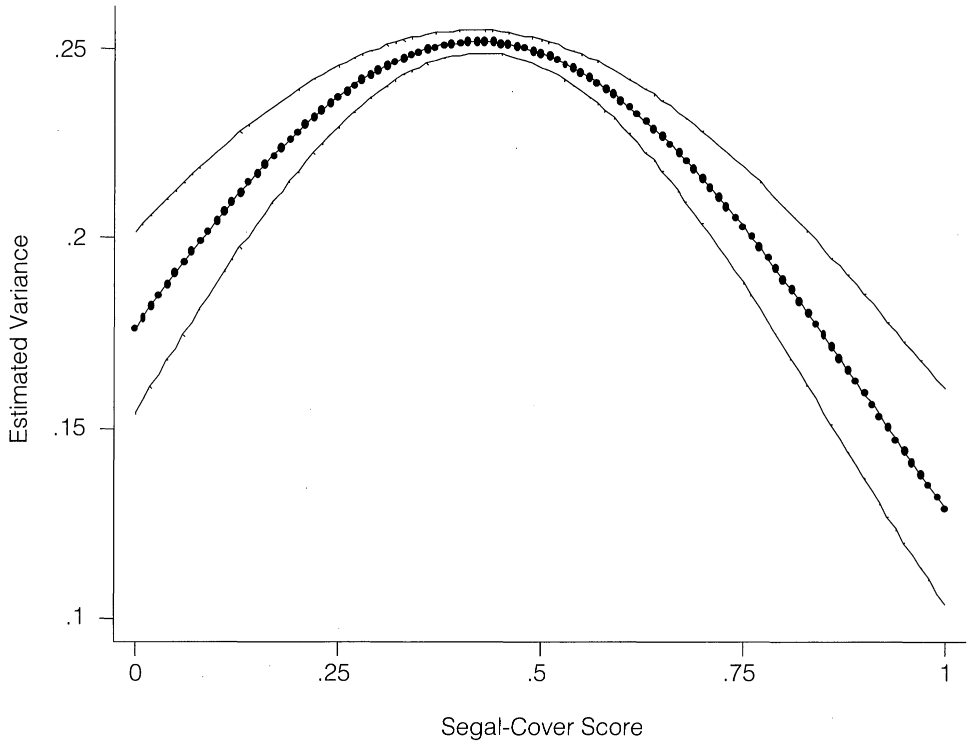
Variables	GEE Model, Mean Effects Only	GEE Model, Variance Effects	GEE2 Model, Variance and Covariance Effects
Mean Effects			
Constant	–0.592	–0.613	–0.684
	(0.193)	(0.197)	(0.211)
Justice Liberalism	1.901	1.959	1.898
	(0.347)	(0.352)	(0.347)
Liberal Lower Court Decision	–0.851	–0.828	–0.843
	(0.272)	(0.273)	(0.301)
U.S. Supports Liberal Outcome	–0.654	–0.712	–0.383
	(0.455)	(0.433)	(0.415)
U.S. Supports Conservative Outcome	0.205	0.208	0.107
	(0.322)	(0.336)	(0.377)
Liberalism × Liberal Lower Court Decision	0.776	0.679	0.682
	(0.496)	(0.496)	(0.508)
Liberalism × U.S. Liberal	5.071	5.599	6.094
	(0.461)	(0.790)	(0.798)
Liberalism × U.S. Conservative	–1.704	–1.702	–1.620
	(0.509)	(0.528)	(0.540)
Variance Effects			
Constant	—	–1.737	–1.834
		(0.069)	(0.079)
Justice Liberalism	—	1.696	2.918
		(0.289)	(0.454)
Justice Liberalism Squared	—	–2.008	–3.788
		(0.331)	(0.667)
Covariance Effects			
Constant	—	—	–0.602
			(0.067)
Justice Liberalism	—	—	–0.488
			(0.167)

Note: N = 1600 (180 cases averaging 8.9 votes per case). Cell entries are logit coefficients; numbers in parentheses are robust standard errors, clustered by case. See text for details.

The final column of Table 3 presents GEE2 estimates of the mean, variance, and intracase correlation. Once again, estimates of the effects of the covariates on the average probability of a liberal vote are changed little from the initial model: the direct effects of the U.S. variables both decline in size and remain small relative to their standard errors. In this respect, the results are consistent with the observation of Zhao and Prentice, who find little difference between estimates of covariate effects on the mean response between GEE1 and GEE2 specifications (1990, 646). By contrast, the quadratic effect of ideology on the variance of voting increases; the variance reaches its maximum value at a liberalism score of 0.38 and declines at more extreme values.

Most interesting, however, is the finding that differences in judicial ideology correspond to markedly lower levels of correlation among justices' votes, even after the direct effects of ideology on voting are accounted for. The large, negative estimate for δ_1 indicates that, as the ideological distance between two justices increases, the extent to which their votes are correlated decreases. This effect is presented graphically in Figure 3, which plots the estimated marginal intracase correlation among justices $\hat{r}_{ijk}$ as a function of the absolute value of the difference between their Segal-Cover scores. The results indicate that, independent of the effects of ideology on the mean probability of a liberal vote, the votes of two justices with the same ideology will correlate at 0.55; this value de-

FIGURE 2 Judicial Ideology's Influence on the Variance in Supreme Court Voting



Note: Figure plots estimated variances ($\hat{\sigma}_{ij}$) by normed Segal-Cover scores. Dotted line is the estimated value; smooth lines are 95% confidence intervals. Estimates are based on column 2 of Table 3. See text for details.

creases to 0.34 for two justices who are diametric opposites vis-a-vis ideology.

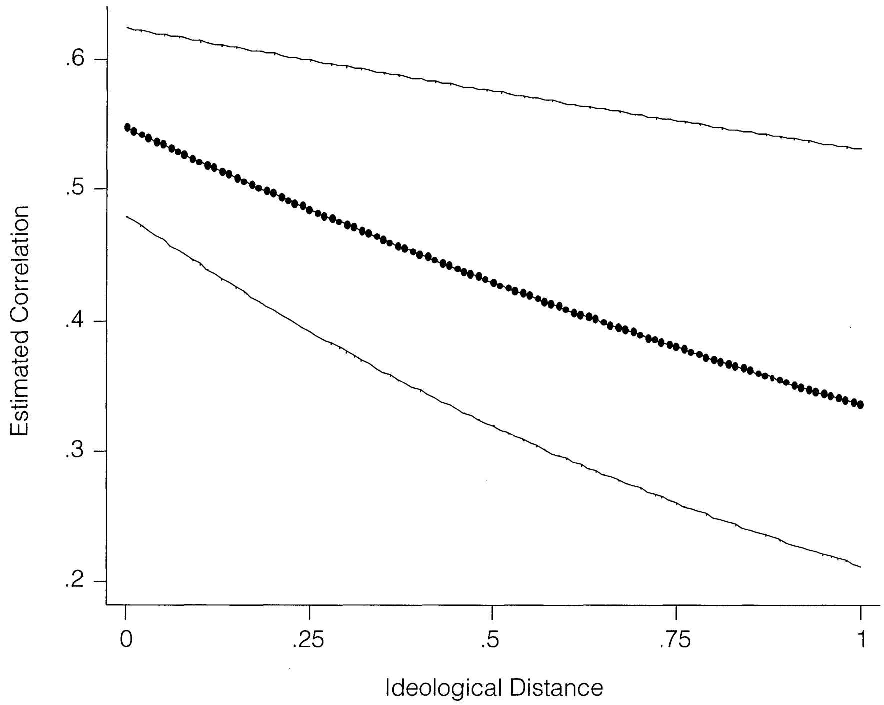
As with the examples in the preceding sections, these results suggest that GEE approaches offer the potential for greater understanding of political data and in particular for extracting additional information out of individual-level data on judicial decision making. Moreover, in instances where, as is the case here, researchers have a substantive interest in the intracluster correlations (including the possible effects of covariates on those correlations), the GEE2 method allows for flexible, consistent estimation of those effects.

Conclusions

The method of generalized estimating equations provides a unified way of modeling correlated data. The approach is particularly attractive because of its ability to estimate models in which the outcome variable is continuous, binary, ordered or unordered polychotomous, or an event count. In addition, the generality of the approach for accounting for intracluster correlation allows its use in a number of situations: repeated measures data (including panel and time-series cross-section designs), data involving sequences of related decisions by political actors, and a range of other circumstances where conditional independence across observations is unlikely. Moreover, in many cases models which account for this correlation can both yield significantly different results from those which do not and reveal substantively interesting and important insights into the process under scrutiny.

In concluding, it is important to emphasize that the usefulness of GEE models depends critically on the nature of the data being examined. In particular, in circumstances in which the intracluster correlation is of central substantive importance, the inability of the GEE1 model to obtain estimates of those correlations for inferential purposes must be weighed against GEE2's lack of robustness in the face of misspecification of that correlation. In some instances, models other than GEEs may be better suited for the problems at hand; this is true, for example, in traditional time-series data, where variability is limited to temporal changes in a single series or group of series, as well as in temporally-dominated time-series cross-sectional data. In many other cases, however, GEE

Figure 3 **Ideological Distance and the Marginal Interjustice Vote Correlation**



Note: Figure plots the estimated marginal correlation among justices' votes within a particular case, as a function of the absolute value of the difference between the justices' Segal-Cover scores. Dotted line is the estimated value; smooth lines are 95% confidence intervals. Estimates are based on column 3 of Table 3. See text for details.

models offer a flexible, versatile, and easily-implemented means of estimating models with correlated data, as well as the possibility of better understanding of the empirical properties of such dependencies.

Manuscript submitted June 1, 1998.
Final manuscript received September 22, 2000.

Appendix
Software for Estimating Generalized Estimating Equations

There is a wide range of software options available for estimating generalized estimating equations. Here I present an overview of several programs for estimating GEE models; other recent reviews include Zeigler and Grömping (1998) and Horton and Lipsitz (1999). Also, because these programs are being constantly revised and updated, other options may be available by the time this goes to press.

Stata 7.0 (http://www.stata.com) will estimate GEE models for continuous (Gaussian and inverse Gaussian), binary (binomial), and count (Poisson and negative binomial) outcomes, as well as gamma models. Allowable links include identity, log, complimentary log-log, logit, probit, power, negative binomial, and reciprocal. Allowed correlation structures include independent, exchangeable, AR(p), nonstationary(p), stationary(p), and pairwise **R** matrices, as well as user-determined fixed correlations. Robust standard errors are available and may be clustered by observation, time point, or other user-specified variable.

SAS 8.0 (http://www.sas.com) will estimate GEEs via the GENMOD procedure (after Release 6.12). Families include normal, binomial, Poisson, and gamma; SAS also provides routines for estimating GEEs with an ordinal-level dependent variable. Available correlation structures are independent, exchangeable, AR(1), m-dependent, and pairwise, as well as user-specified; SAS also implements the odds-ratio (alternating logistic regression) approach to estimating correlations of Lipsitz, Laird, and Harrington (1991). User-written SAS procedures for estimating GEEs are also available, including programs for estimating Prentice and Zhao's (1991) GEE2 model (Williamson, Lipsitz, and Kim 1999).

S-Plus 6.0 (http://www.mathsoft.com). The user-written GEE module (formerly OSWALD) will estimate Gaussian, binomial, Poisson, and gamma GEEs. Available links are identity, log, logit, probit, reciprocal, and complimentary log-log. Correlation structures include independent, exchangeable, fixed (user-defined), stationary (m-dependent), nonstationary-m, AR(p), and pairwise.

GAUSS (http://www.aptech.com). A module for estimating GEEs is available through the Riemann Library (Källén 1999). It allows for binomial, Poisson, and multinomial families, including models for ordered and unordered polychotomous variables. Available correlation options include independent, exchangeable, AR(1) and pairwise, as well as parameterization and estimation of the correlation matrices via the odds-ratio approach (Lipsitz, Laird, and Harrington 1991) using alternating logistic regression.

SUDAAN Release 7.5 (http://www.rti.org) is produced by the Research Triangle Institute. Its MULTILOG procedure implements GEE model estimation of continuous, binary logit and probit, Poisson, and ordered, and unordered polychotomous probit and logit GEEs. Available correlations are limited to independent and exchangeable. The software also provides robust variance estimates via Huber/White, jacknife, and balanced repeated replication methods.

XLISP-Stat Release 2.0 is an object-oriented statistics program based on the Lisp language and developed by Luke Tierney of the University of Minnesota (http://stat.umn.edu/~luke/xls/tutorial/techreport/techreport.html). The user-written routine GEE will estimate these models with the full complement of available families, links, and correlation structures. In addition, the program allows use of any family-link combination and provides a number of potentially useful diagnostics (e.g., plots of Pearson residuals). See Lumley (1996b) and Tierney (1989) for details.

EE Version 1.0, created by the Quantitative Genetic Epidemiology Group of the Fred Hutchinson Cancer Research Center, is available as both a free-standing program and as a set of SAS macros and S- Plus functions. EE will estimate GLM, GEE1, and GEE2 models and allows for covariate effects on the means, variances, and correlations of the response variable. Available link functions for the mean include linear, binomial, and exponential. Variance links include constant, fixed, binomial, and Poisson, or linear, exponential or logistic functions of covariates. Available correlation links are the linear, exponential, and hypergeometric.

In addition to these, a number of other specialized programs exist for estimating GEEs. Many of these are capable of estimating generalized linear models and survival models as well. Examples include SPIDA 6.0 (http://www.efs.mq.edu.au/statlab/spida.html) and QUATOR (ftp:odin.mdacc.tmc.edu).

References

Albert, Paul S., and Lisa M. McShane. 1995. "A Generalized Estimating Equations Approach for Spatially Correlated Binary Data: Applications to the Analysis of Neuroimaging Data." *Biometrics* 51:627–638.

Alvarez, R. Michael, and John Brehm. 1995. "American Ambivalence Towards Abortion Policy: Development of a Heteroskedastic Probit Model of Competing Values." *American Journal of Political Science* 39:1055–1082.

Barnhart, Huiman X., and John M. Williamson. 1998. "Goodness-of-Fit Tests for GEE Modeling with Binary Responses." *Biometrics* 54:720–729.

Bartels, Larry M. 1999. "Panel Effects in the American National Election Studies." *Political Analysis* 8:1–20.

Beck, Nathaniel, and Jonathan N. Katz. 1995. "What To Do (And Not To Do) With Time-Series Cross-Section Data." *American Political Science Review* 89:634–647.

Beck, Nathaniel, Jonathan N. Katz, and Richard Tucker. 1998. "Taking Time Seriously: Time- Series-Cross-Section Analysis with a Binary Dependent Variable." *American Journal of Political Science* 42:1260–1288.

Bennett, D. Scott, and Allan Stam. 2000. "Research Design and Estimator Choices in the Analysis of Interstate Dyads: When Decisions Matter." *Journal of Conflict Resolution* 44:653–685.

Bliss, Harry, and Bruce Russett. 1998. "Democratic Trading Partners: The Liberal Connection, 1962–1989." *Journal of Politics* 60:1126–1147.

Box-Steffensmeier, Janet M., and Bradford Jones. 1997. "Time is of the Essence: Event History Models in Political Science." *American Journal of Political Science* 41:1414–1461.

Brandt, Patrick T., John T. Williams, Richard Fordham, and Brian Pollins. 2000. "Dynamic Modelling For Persistent Event Count Time Series." *American Journal of Political Science* 44:823–843.

Brown, David S., and Wendy Hunter. 1999. "Democracy and Social Spending in Latin America, 1980–92." *American Political Science Review* 93:779–790.

Cameron, A. Colin, and Pravin K. Trivedi. 1998. *Regression Analysis of Count Data*. New York: Cambridge University Press.

Caldeira, Gregory A., John R. Wright, and Christopher J. W. Zorn. 1999. "Strategic Voting and Gatekeeping in the Supreme Court." *Journal of Law, Economics and Organization* 15:549–572.

Carlin, John B., Rory Wolfe, Carolyn Coffey, and George C. Patton. 1999. "Analysis of Binary Outcomes in Longitudinal Studies Using Weighted Estimating Equations and Discrete-Time Survival Methods: Prevalence and Incidence of Smoking in an Adolescent Cohort." *Statistics in Medicine* 18:2655–2679.

Chang, Yue-Cune. 2000. "Residual Analysis of the Generalized Linear Models for Longitudinal Data." *Statistics in Medicine* 19:1277–1293.

Chen, Hua Yun, and Roderick Little. 1999. "A Test of Missing Completely At Random for Generalised Estimating Equations with Missing Data." *Biometrika* 86:1–13.

Clark, William Roberts, and Mark Hallerberg. 2000. "Mobile Capital, Domestic Institutions, and Electorally Induced Monetary and Fiscal Policy." *American Political Science Review* 94:323–346.

Cressie, Noel A. 1993. *Statistics for Spatial Data*, Revised Edition New York: Wiley.

Davidson, Russell, and James G. MacKinnon. 1993. *Estimation and Inference in Econometrics.* Oxford: Oxford University Press.

Diggle, Peter J. 1989. "Testing for Random Dropouts in Repeated Measurements Data." *Biometrics* 45:1255–1258.

Diggle, Peter J., Kung-Yee Liang, and Scott L. Zeger. 1994. *Analysis of Longitudinal Data.* New York: Oxford University Press.

Epstein, Lee, and Carol Mershon. 1996. "Measuring Political Preferences." *American Journal of Political Science* 40:261–294.

Feldman, Stanley, and John Zaller. 1992. "A Simple Theory of the Survey Response: Answering Questions Versus Revealing Preferences." *American Journal of Political Science* 36:579–616.

Finkel, Steven E., and Edward N. Muller. 1998. "Rational Choice and the Dynamics of Political Action: Evaluating Alternative Models with Panel Data." *American Political Science Review* 92:37–50.

Fitzmaurice, Garrett M. 1995. "A Caveat Concerning Independence Estimating Equations with Multivariate Binary Data." *Biometrics* 51:309-317.

Fitzmaurice, Garrett M., Nan M. Laird, and Andrea G. Rotnitzky. 1993. "Regression Models for Discrete Longitudinal Responses." *Statistical Science* 8:284–309.

Flemming, Roy B., and B. Dan Wood. 1997. "The Public and the Supreme Court: Individual Justice Responsiveness to American Policy Moods." *American Journal of Political Science* 41:468–498.

Gill, Jeff. 2000. *Generalized Linear Models: A Unified Approach.* Thousand Hills, Calif.: Sage Publications.

Hanley, James A., Abdissa Negassa, and Michael D. deB. Edwardes. 2000. "GEE Analysis of Negatively Correlated Binary Responses: A Caution." *Statistics in Medicine* 19:715–722.

Hansen, Lars Peter. 1982. "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica* 50:1029–1054.

Hastie, Trevor J., and R. J. Tibshirani. 1990. *Generalized Additive Models.* London: Chapman and Hall.

Herron, Michael C. 1999. "Postestimation Uncertainty in Limited Dependent Variable Models." *Political Analysis* 8:83–98.

Heyde, Christopher C. 1997. *Quasi-Likelihood And Its Application: A General Approach to Optimal Parameter Estimation.* New York: Springer-Verlag.

Hogan J. W., and Nan M. Laird. 1997. "Model-Based Approaches to Analysing Incomplete Longitudinal and Failure Time Data." *Statistics in Medicine* 16:259–272.

Hojnacki, Marie, and David C. Kimball. 1998. "Organized Interests and the Decision of Whom to Lobby in Congress." *American Political Science Review* 92:775–790.

Horton, Nicholas J., Judith D. Bebchuk, Cheryl L. Jones, Stuart R. Lipsitz, Paul J. Catalano, Gwendolyn E. Zahner, and Garrett M. Fitzmaurice. 1999. "Goodness-of-Fit for GEE: An Example with Mental Health Service Utilization." *Statistics in Medicine* 18:213–222.

Horton, Nicholas J., and Stuart R. Lipsitz. 1999. "Review of Software to Fit Generalized Estimating Equation Regression Models." *The American Statistician* 53:160–169.

Hosmer, D. W., and S. Lemeshow. 1980. "Goodness of Fit Tests for the Multiple Logistic Regression Model." *Communications in Statistics Part A: Theory and Methods* 9:1043–1069.

Hu, Frank B., Jack Goldberg, Donald Hedeker, Brian R. Flay, and Mary Ann Pentz. 1998. "Comparison of Population-Averaged and Subject-Specific Approaches for Analyzing Repeated Binary Outcomes." *American Journal of Epidemiology* 147:694–703.

Huber, John D. 1998. "How Does Cabinet Instability Affect Political Performance? Portfolio Volatility and Health Care Cost Containment in Parliamentary Democracies." *American Political Science Review* 92:577–591.

Huckfeldt, Robert, John Sprague, and Jeffrey Levine. 2000. "The Dynamics of Collective Deliberation in the 1996 Election: Campaign Effects on Accessibility, Certainty, and Accuracy." *American Political Science Review* 94:641–651.

James, Scott C., and Brian L. Lawson. 1999. "The Political Economy of Voting Rights Enforcement in America's Gilded Age: Electoral College Competition, Partisan Commitment, and the Federal Election Law." *American Political Science Review* 93:115–131.

Judge, George G., W. E. Griffiths, R. Carter Hill, Helmut Lütkepol, and Tsoung-Chao Lee. 1985. *The Theory and Practice of Econometrics*, 2nd ed. New York: Wiley.

Källén, Anders. 1999. *Generalized Estimating Equations: Marginal Models for the Analysis of Clustered Data.* The Riemann Library.

King, Gary, Michael Tomz, and Jason Wittenberg. 2000. "Making the Most of Statistical Analyses: Improving Interpretation and Presentation." *American Journal of Political Science* 44:362–374.

Krehbiel, Keith. 1997. "Restrictive Rules Reconsidered." *American Journal of Political Science* 41:919–944.

Leeds, Brett Ashley, and David R. Davis. 1997. "Domestic Political Vulnerability and International Disputes." *Journal of Conflict Resolution* 41:814–834.

Levitt, Steven D. 1996. "How Do Senators Vote? Disentangling the Role of Voter Preferences, Party Affiliation, and Senator Ideology." *American Economic Review* 86:425–442.

Laird, Nan M. 1988. "Missing Data in Longitudinal Studies." *Statistics in Medicine* 7:305–315.

Liang, Kung-Yee, and Scott L. Zeger. 1986. "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika* 73:13–22.

Liang, Kung-Yee, Scott L. Zeger, and B. Qaqish. 1992. "Multivariate Regression Analyses for Categorical Data (with Discussion)." *Journal of the Royal Statistical Society Series B* 54:3–40.

Lipsitz, Stuart R., Keith B. G. Dear, and Lueping Zhao. 1994. "Jackknife Estimator of Variance for Parameter Estimates from Estimating Equations with Applications to Clustered Survival Data." *Biometrics* 50:842–846.

Lipsitz, Stuart R., Garrett M. Fitzmaurice, E. J. Orav, and Nan P. Laird. 1994. "Performance of Generalized Estimating Equations in Practical Situations." *Biometrics* 50:270–278.

Lipsitz, Stuart R., Kyungmann Kim, and Lueping Zhao. 1994. "Analysis of Repeated Categorical Data Using Generalized Estimating Equations." *Statistics in Medicine* 13:1149–1163.

Lipsitz, Stuart R., Nan M. Laird, and David P. Harrington. 1990. "Using the Jackknife to Estimate the Variance of Regression Estimators from Repeated Measures Studies." *Communications in Statistics A: Theory and Methods* 19:821–845.

Lipsitz, Stuart R., Nan M. Laird, and David P. Harrington. 1991. "Generalized Estimating Equations for Correlated Binary Data: Using the Odds Ratio as a Measure of Association." *Biometrika* 78:153–160.

Little, Roderick. J. A., and Donald B. Rubin. 1987. *Statistical Analysis with Missing Data.* New York: Wiley.

Lowry, Robert C., James E. Alt, and Karen E. Feree. 1998. "Fiscal Policy Outcomes and Electoral Accountability in American States." *American Political Science Review* 92:759–774.

Lumley, Thomas. 1996a. "Generalized Estimating Equations for Ordinal Data: A Note on Working Correlation Structures." *Biometrics* 52:354–361.

Lumley, Thomas. 1996b. "XLISP-Stat Tools for Building Generalized Estimating Equations Models." Manuscript: Department of Biostatistics, University of Washington.

Maltzman, Forrest, and Paul J. Wahlbeck. 1996a. "May it Please the Chief? Opinion Assignments in the Rehnquist Court." *American Journal of Political Science* 40:421–443.

Maltzman, Forrest, and Paul J. Wahlbeck. 1996b. "Strategic Policy Considerations and Voting Fluidity on the Burger Court." *American Political Science Review* 90:581–592.

Mancl, Lloyd A., and Brian G. Leroux. 1996. "Efficiency of Regression Estimates for Clustered Data." *Biometrics* 52:500–511.

Mansfield, Edward D., Helen V. Milner, and B. Peter Rosendorff. 2000. "Free to Trade: Democracies, Autocracies and International Trade." *American Political Science Review* 94:305–322.

Mason, Alpheus T. 1956. *Harlan Fiske Stone, Pillar of the Law.* New York: Viking.

Mason, Alpheus T. 1964. *William Howard Taft: Chief Justice.* New York: Simon and Schuster.

McCullagh, P., and J. A. Nelder. 1983. "Quasi-Likelihood Functions." *Annals of Statistics* 11:59–67.

McCullagh, P., and J. A. Nelder. 1989. *Generalized Linear Models.* 2nd ed. London: Chapman and Hall.

McDonald, B. W. 1993. "Estimating Logistic Regression Parameters for Bivariate Binary Data." *Journal of the Royal Statistical Society, Series B* 55:391–397.

McGrory, Mary. 1998. "And All Through The House." *Washington Post* 12/24/98, p. A4.

Meier, Kenneth J., J. L. Polinard, and Robert D. Wrinkle. 2000. "Bureaucracy and Organizational Performance: Causality Arguments About Public Schools." *American Journal of Political Science* 44:590–602.

Morrow, James D., Randolph M. Siverson, and Tressa E. Tabares. 1998. "The Political Determinants of International Trade, 1907–90." *American Political Science Review* 92:649–661.

Mugglestone, M.A., M.G. Kenward, and S.J. Clark. 1999. "Generalized Estimating Equations for Spatially Referenced Binary Data." Presented at the Conference on Correlated Data Modeling, University of Trieste, October 22–23, 1999.

Murphy, Walter F. 1964. *Elements of Judicial Strategy.* Chicago: University of Chicago Press.

Nelder, J. A., and R. W. M. Wedderburn. 1972. "Generalized Linear Models." *Journal of the Royal Statistical Society Series B* 54:3–40.

Neuhaus, J. M. 1992. "Statistical Methods for Longitudinal and Clustered Designs with Binary Responses." *Statistical Methods in Medical Research* 1:249–273.

Neuhaus, J. M. 1993. "Estimation Efficiency and Tests of Covariate Effects with Clustered Binary Data." *Biometrics* 49:989–996.

Neuhaus, J. M., J. D. Kalbfleisch, and W. W. Hauck. 1991. "A Comparison of Cluster-Specific and Population-Averaged Approaches for Analyzing Correlated Binary Data." *International Statistical Review* 59:25–35.

Oneal, John R., and Bruce Russett. 1997. "The Classical Liberals Were Right: Democracy, Interdependence, and Conflict, 1950–1985." *International Studies Quarterly* 41:267–294.

Oneal, John R., and Bruce Russett. 1999a. "Assessing the Liberal Peace with Alternative Specifications: Trade Still Reduces Conflict." *Journal of Peace Research* 36:423–442.

Oneal, John R., and Bruce Russett. 1999b. "The Kantian Peace: The Pacific Benefits of Democracy, Interdependence, and International Organizations, 1885–1992." *World Politics* 52:1–37.

Paik, Myunghee Cho. 1997. "The Generalized Estimating Equation Approach When Data Are Not Missing Completely at Random." *Journal of the American Statistical Association* 92:1320–1329.

Park, Taesung, and Seung-Yeoun Lee. 1997. "A Test of Missing Completely at Random for Longitudinal Data with Missing Observations." *Statistics in Medicine* 16:1859–1871.

Pendergast, Jane F., Stephen J. Gange, Michael A. Newton, Mary P. Lindstrom, Mari Palta, and Marian R. Fisher. 1996. "A Survey of Methods for Analyzing Clustered Binary Response Data." *International Statistical Review* 64:89–118.

Poole, Keith, and Howard Rosenthal. 1997. *Congress: A Political Economic History of Roll-Call Voting.* New York: Oxford University Press.

Prentice, Ross L. 1988. "Correlated Binary Regression with Covariates Specific to Each Binary Observation." *Biometrics* 44:1033–1048.

Prentice, Ross L., and L. P. Zhao. 1991. "Estimating Equations for Parameters in Mean and Covariances of Multivariate Discrete and Continuous Responses." *Biometrics* 47:825–839.

Q.G.E. 1994. "EE: Estimating Equations." Technical Report 126, Fred Hutchinson Cancer Research Center, Quantitative Genetic Epidemiology.

Radcliff, Benjamin, and Patricia Davis. 2000. "Labor Organization and Electoral Participation in Industrial Democracies." *American Journal of Political Science* 44:132–141.

Reed, William. 2000. "A Unified Statistical Model of Conflict Onset and Escalation." *American Journal of Political Science* 44:84–91.

Russett, Bruce, John Oneal, and David R. Davis. 1998. "The Third Leg of the Kantian Tripod for Peace: International Organizations and Militarized Disputes, 1950–1985." *International Organization* 52:441–468.

Scholz, John T., and B. Dan Wood. 1998. "Controlling the IRS: Principals, Principles, and Public Administration." *American Journal of Political Science* 42:141–162.

Sears, David O., and Carolyn L. Funk. 1999. "Evidence of the Long-Term Persistence of Adults' Political Predispositions." *Journal of Politics* 61:1–28.

Segal, Jeffrey A., and Albert D. Cover. 1989. "Ideological Values and the Votes of U.S. Supreme Court Justices." *American Political Science Review* 83:557–565.

Segal, Jeffrey A., and Harold J. Spaeth. 1993. *The Supreme Court and the Attitudinal Model.* New York: Cambridge University Press.

Sherman, Robert P. 2000. "Tests of Certain Types of Ignorable Nonresponse in Surveys Subject to Item Nonresponse or Attrition." *American Journal of Political Science* 44:356–368.

Shults, Justine, and N. Rao Chaganty. 1998. "Analysis of Serially-Correlated Data Using Quasi- Least Squares." *Biometrics* 54:1622–1630.

Snyder, James M., and Tim Groseclose. 2000. "Estimating Party Influence in Congressional Roll-Call Voting." *American Journal of Political Science* 44:187–205.

Spaeth, Harold J. 1998. *The United States Supreme Court Judicial Database, 1953–1996.* Ann Arbor: Inter-University Consortium for Political and Social Research.

Spaeth, Harold J., and Michael F. Altfield. 1985. "Influence Relationships Within the Supreme Court: A Comparison of the Warren and Burger Courts." *Western Political Quarterly* 37:70–83.

Stimson, James A. 1985. "Regression in Time and Space: A Statistical Essay." *American Journal of Political Science* 29:914–947.

Stram, D. O., L. J. Wei, and J. H. Ware. 1988. "Analysis of Repeated Ordered Categorical Outcomes with Possibly Missing Observations and Time-Dependent Covariates." *Journal of the American Statistical Association* 79:118–124.

Thall, Peter F., and Stephen C. Vail. 1990. "Some Covariance Models for Longitudinal Count Data with Overdispersion." *Biometrics* 46:657–671.

Tierney, Luke. 1989. "XLISP-STAT: A Statistical Environment Based on the XLISP Language (Version 2.0)," Technical Report No. 528, School of Statistics, University of Minnesota.

Traut, Carol Ann, and Craig F. Emmert. 1998. "Expanding the Integrated Model of Judicial Decision Making: The California Justices and Capital Punishment." *Journal of Politics* 60:1166–1180.

Wahlbeck, Paul J., James F. Spriggs II, and Forrest Maltzman. 1998. "Marshaling the Court: Bargaining and Accommodation on the U.S. Supreme Court." *American Journal of Political Science* 42:294–315.

Wawro, Gregory. 2000. "A Panel Probit Analysis of Campaign Contributions and Roll Call Votes." Unpublished manuscript. Columbia University.

Wedderburn, R. W. M. 1974. "Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method." *Biometrika* 61:439–447.

White, Halbert. 1980. "A Heteroscedasticity-Consistent Covariance Matrix and a Direct Test for Heteroscedasticity." *Econometrica* 48:817–838.

Wild, C. J., and T. W. Lee. 1996. "Additive Extensions to Generalized Estimating Equation Methods." *Journal of the Royal Statistical Society, Series B* 58:711–725.

Williamson, John. 1999. Personal communication with the author, Centers for Disease Control and Prevention, Atlanta, GA.

Williamson, J. M., S. R. Lipsitz, and K. R. Kim. 1999. "GEECAT and GEEGOR: Computer Programs for the Analysis of Correlated Categorical Response Data." *Computer Methods & Programs in Biomedicine* 58:25–34.

Xie, Fang, and Myunghee Cho Paik. 1997. "Generalized Estimating Equation Model for Binary Outcomes with Missing Covariates." *Biometrics* 53:1458–1466.

Zeger, Scott L., and Kung-Yee Liang. 1986. "Longitudinal Data Analysis for Discrete and Continuous Outcomes." *Biometrics* 42:121–130.

Zeger, Scott L., and Kung-Yee Liang. 1992. "An Overview of Methods for the Analysis of Longitudinal Data." *Statistics in Medicine* 11:1825–1839.

Zeger, Scott L., and B. Qaqish. 1988. "Markov Regression Models for Time Series: A Quasi- Likelihood Approach." *Biometrics* 44:1019–1031.

Zeger, Scott L., Kung-Yee Liang, and P. S. Albert. 1988. "Models for Longitudinal Data: A Generalized Estimating Equation Approach." *Biometrics* 44:1049–1060.

Zeigler, Andreas. 1995. "The Different Parameterizations of the GEE1 and GEE2." In *Statistical Modelling*, Vol. 104 of *Lecture Notes in Statistics*, ed. G. U. H. Seeber, B. Francis, R. Hatzinger, and G. Steckel-Berger. New York: Springer.

Zeigler, Andreas, and Ulrike Grömping. 1998. "The Generalised Estimating Equations: A Comparison of Procedures Available in Commercial Statistical Software Packages." *Biometrical Journal* 40:245–260.

Ziegler, Andreas, Christian Kastner, and Maria Blettner. 1998. "The Generalised Estimating Equations: An Annotated Bibliography." *Biometrical Journal* 40:115–139.

Zhao, Lue Ping, and Ross L. Prentice. 1990. "Correlated Binary Regression Using a Quadratic Exponential Model." *Biometrika* 77:642–648.

Zhao, Lue Ping, Ross L. Prentice, and S. G. Self. 1992. "Multivariate Mean Parameter Estimation by Using a Partially Exponential Model." *Journal of the Royal Statistical Society, Series B* 54:805–811.

Zheng, Beiyao. 2000. "Summarizing the Goodness of Fit of Generalized Linear Models for Longitudinal Data." *Statistics in Medicine* 19:1265–1275.