# Comparing GEE and Robust Standard Errors for Conditionally Dependent Data

CHRISTOPHER ZORN, UNIVERSITY OF SOUTH CAROLINA

In recent years political scientists have become increasingly sensitive to questions of conditional dependence in their data. I outline and compare two general, widely-used approaches for addressing such dependence—robust variance estimators and generalized estimating equations (GEEs)—using data on votes in Supreme Court search and seizure decisions between 1963 and 1981. The results make clear that choices about the unit on which data are grouped, i.e., clustered, are typically of far greater significance than are decisions about which type estimator is used.

Regression and regression-like models form the basis of most quantitative work in the social sciences. In the cross-sectional realm, advances in such models have largely occurred through the development of models for response variables which do not conform to standard linear-normal assumptions, including binary, nominal, ordinal, and count data. In recent years, however, scholars have increasingly begun to move beyond the form taken by the dependent variable to consider other issues in their data analyses. One of the most prominent has been violation of the *exchangeability* assumption, that is, data where, conditional on the influence of the model's covariates, values of the response variable are not independently and identically distributed (King 2001; McCullagh 2004). Such a situation may be especially likely to arise when data are grouped or clustered; examples include dyadic data (e.g., Hojnacki and Kimball 1998; Oneal and Russett 1999) or panel/time-series cross-sectional data with repeated observations on units (Stimson 1985).

While a number of approaches to dealing with such possible dependence exist, two merit special attention because of their generality and predominance in the field. The first, commonly referred to as robust standard errors,[1] is a general means of empirically correcting variance-covariance estimates in the presence of heteroscedasticity, clustering, and other forms of conditional dependence. The second, the method of generalized estimating equations (GEE) (Liang and Zeger 1986), offers researchers the benefits of asymptotically-consistent variance-covariance estimates when data are nonexchangeable, even when the precise nature of that dependence is unknown. Both methods have seen increasing use by applied researchers in recent years, yet most know little about their respective properties and even less about how to make an informed choice on their use.

The purpose of this article is to begin to address this lacuna. I begin with an outline and comparison of the development and general characteristics of these two approaches. I then illustrate, using a realistic example from the literature on judicial politics, the potential benefits and tradeoffs associated with each of the methods, with particular focus on how decisions over model specification and choice of clustering unit can influence one's results. I conclude with some general guidelines for applied researchers about choosing among these methods.

## ROBUST STANDARD ERRORS

Consider at the outset a basic regression-like model, in which a $N \times 1$ vector $\mathbf{Y}$ of responses is modeled as some stochastic function of $k$ covariates $\mathbf{X}$ (typically including a constant term) and a vector of disturbances $\mathbf{u}$:

$$\mathbf{Y} = f(\mathbf{X}\beta) + \mathbf{u} \tag{1}$$

This general model subsumes a number of commonly used alternatives, including exponential-family GLMs (McCullagh and Nelder 1989) as well as panel, time-series cross-sectional, and other grouped data models. In a likelihood-based estimation framework (e.g., King 1989), we can derive the log-likelihood for (1) by selecting $f(\cdot)$ and making a distributional assumption about $\mathbf{u}$. Under the usual regularity conditions (that is, given a properly specified model and conditionally independent, identically distributed observations), one can then obtain a consistent estimate of the variance of the estimated parameter vector $\hat{\beta}$ by considering the negative of the inverse matrix of second derivatives (the "information matrix"):

$$\hat{\mathbf{V}} = -\left(\frac{\partial^2 InL}{\partial\hat{\beta}\partial\hat{\beta}}\right)^{-1} \tag{2}$$

The square roots of the diagonal elements of this matrix are the parameter standard errors typically reported by statistical

[1] This usage is the predominant one in political science; other names include heteroscedasticity-consistent covariance matrix estimator, sandwich estimator, and empirical covariance matrix estimator. For simplicity, and in the interest of contributing to terminological path dependence, I adopt the robust usage herein.

NOTE: Thanks to Jeff Segal for graciously making his data available, to Andrew Martin, Chuck Smith, Steve Van Winkle, Andreas Ziegler, three anonymous reviewers for helpful comments and discussions, and to Jennifer Barnes for research assistance. Data and instructions for replication of the analyses herein are available from the author upon request.

*Political Research Quarterly,* Vol. 59, No. 3 (September 2006): pp. 329-341

software; they provide information about the precision of those estimates and form the basis for inference about the covariates' marginal effects. Increasingly, however, researchers are aware that the assumptions necessary for (2) to provide a valid basis for inference are difficult if not impossible to achieve in practice. Whether due to temporal dependence, spatial contagion, or the nature of the data (e.g., the use of dyads in research in international relations), it is increasingly the case that the standard assumption of conditional exchangeability is unlikely to be met. The desire to relax the otherwise strict conditions necessary for valid inferences in this circumstance led to the development of "robust" estimators of the variance-covariance matrix $\mathbf{V}$:

$$\hat{\mathbf{V}}_R = \hat{\mathbf{V}} \left( C \sum_{i=1}^{N} (\hat{u}_i \hat{u}'_i) \right) \hat{\mathbf{V}} \tag{3}$$

where $\hat{u}_i$ is the contribution of $i$ to the scores $\partial \ln L / \partial \beta$, i.e., $(\partial \ln L_i / \partial \beta)_{\beta = \hat{\beta}}$ and $C$ is a correction factor.[2] The origins of this estimator can be traced at least as far back as Eicker (1963), though it was more fully developed by (and is more commonly attributed to) Huber (1967); White (1980) independently derived the same estimator in an econometric context. The intuition of $\hat{\mathbf{V}}_R$ is to weight each observation's contribution to the overall variance-covariance estimate by an empirical measure of that observation's residual variability; for this reason, this estimate is sometimes referred to as the empirical or empirically-corrected variance estimate. It is also known as the sandwich estimator, since the empirical correction factor is sandwiched between two standard estimators for $\hat{\mathbf{V}}$.

The variance-covariance estimate in (3) is a very general one, and assumes that the analyst has no a priori expectations about the nature of the conditional dependence in $Y_{it}$. In many cases, however, it can be argued that the dependence in question has some structure. In panel-structured survey data, for example, we often observe $Y_{it}$, where $i$ indexes the (randomly selected) survey respondent and $t$ the interview period. In such an instance, it is likely justified to assume that the $i$s are independent, but that the responses over $t$ for a particular respondent are not. In cases like this, the estimator in (3) can be extended to consider data which are grouped or clustered in a straightforward fashion (e.g., Williams 2000). Consider data on $N$ observational units or clusters indexed by $i$, each of which has $n_i$ observations.[3] For such data, the "clustered" robust variance-covariance estimator is:

$$\hat{\mathbf{V}}_C = \hat{\mathbf{V}} \sum_{i=1}^{N} \left[ \left( \sum_{j=1}^{n_i} \hat{u}_{ij} \right) \left( \sum_{j=1}^{n_i} \hat{u}'_{ij} \right) \right] \hat{\mathbf{V}} \tag{4}$$

where $\hat{u}_{ij}$ is defined as in (3). This estimator thus treats each cluster as a "super-observation," considering first variability within that cluster and then summing across clusters for the final adjustment.

As we will see below, the differences between standard error estimates obtained in (3) and (4) can, under certain circumstances, be substantial. Also, it is important to note that, while the simple robust estimates given in (3) will generally be larger than the naïve estimates in (2), those calculated based on (4) may be either smaller or larger than either (2) or (4). If, as is most often the case, there is positive contagion within clusters (i.e., $Cov(u_{ij}, u_{i\ell}) > 0$ for $j \neq \ell$), then this increased intracluster variability will lead to larger components of the diagonal elements of $\hat{\mathbf{V}}_C$ and correspondingly larger standard error estimates. Conversely, in cases of negative contagion within clusters (i.e., $Cov(u_{ij}, u_{i\ell}) < 0$ for $j \neq \ell$), the within-cluster estimates of $\hat{u}_{ij}$ will tend to cancel each other out, such that the overall estimate $\hat{\mathbf{V}}_C$ will be smaller than $\hat{\mathbf{V}}$ alone.

Informally, the issue of conditional dependence is closely tied to the effective amount of information in the data. Intuitively, one can think of the naïve variance estimates as giving equal weight to all observations in the data. If, in contrast, observations are conditionally correlated, then the actual variability in the data may over- or under-represent the actual amount of information the data contain; observations whose conditional correlations are strongly positive can be thought of as in some sense containing less information than those which are independent, while the reverse is true for those that exhibit high negative correlations. Likewise, if within a particular unit of analysis observations are not conditionally independent, they may also contain effectively more or less information than if they were. In the limit, observations within a cluster which are exactly identical contain little or no more information than if each cluster contained only a single observation. In either case, methods that fail to account for this conditional covariation run the risk of over- or underestimating the precision of the parameter estimates.

A very simple simulation serves to illustrate this intuition. In this example, I begin with data generated such that $X_i \sim N(0,1)$ and $Y_i = \beta_0 + \beta_1 X_i + u_i$, where $\beta_0 = \beta_1 = 1.0$, $u_i \sim N(0, 1)$, and $N = 25$; consider each of these "original" observations as a cluster. I then create 100 exact copies of each observation in the data; these correspond to repeated observations on each cluster, and has the effect of increasing the number of observations while adding little if any

[2] This correction is analogous to the $N - 1$ correction commonly used when calculating $\sigma^2$. Three values predominate for $C$: White's (1980) consistency proof has $C = 1$, that is, no correction, while others have used

$$C = \frac{N}{N-1} \text{ or } C = \frac{N}{N-k};$$

see MacKinnon and White (1985) for details. The first and second are seen most commonly, and are the default option in most statistical packages (e.g., the `sandwich` package in R and the `robust` option in Stata, respectively). Other special cases of $C$ exist for the linear regression model; see Long and Ervin (2000) for a thorough Monte Carlo study of the properties of these variants. For simplicity, in the equations below I assume that $C = 1$; of course, all are asymptotically negligible as $N \to \infty$.

[3] In practice, $n_i$ may be constant across $i$ (if the data are balanced) or may vary from one cluster to another.

TABLE 1
RESULTS FOR SIMULATED DATA

| Variable | $\hat{\beta}$ | Standard Error Estimates: One Observation Per Cluster | Standard Error Estimates: All Data, Not Clustered | Standard Error Estimates: All Data, Clustered |
|---|---|---|---|---|
| (Constant) | 1.074 | 0.184 | 0.018 | 0.181 |
| X | 1.237 | 0.188 | 0.018 | 0.177 |
| *F* | | 43.26 | 4698.5 | 49.12 |
| *N* | | 25 | 2500 | 2500 |

Note: For all regressions, $R^2$ and RMSE = 0.88. All *F*-statistics are $p < 0.001$. See text for details.

additional information about the relationship between *X* and *Y*. Results of estimating three regression models on these data are presented in Table 1.[4] The first column reports the estimates of $\hat{\beta}$, which do not change across the three sets of estimates. Column two gives standard error estimates for the model using only the original 25 observations in the data, while column three presents the same standard errors for the regression using the expanded ($N$ = 2500) data. Finally, column four presents the robust standard errors, grouped by cluster, i.e., corresponding to Equation (4).[5]

Several things are immediately apparent from this simple simulation. First, and not surprisingly, the conventional standard error estimates for the $N$ = 2500 data are significantly smaller than those when $N$ = 25.[6] This is true despite the fact that, because of how the data were generated, we know relatively little more about the relationship between *X* and *Y* from the additional 2475 observations created. The results in column four illustrate how using (clustered) robust standard errors accounts for this fact; those standard errors are effectively identical to those when $N$ = 25.[7] Similar differences are observed in the *F*-tests for the joint significance of $\hat{\beta}_0$ and $\hat{\beta}_1$; once again, the results from using the formula in Equation (4) are effectively identical to those in the model where $N$ = 25.

Robust variance-covariance estimators, then, are an easily implemented, relatively general means for dealing with conditionally nonexchangeable data. This ease of use has caused them to be widely adopted in political science, as is illustrated in Figure 1, which plots the incidence of the phrase "robust standard errors" in three major political science journals between 1992 and 2002. An important (and attractive) characteristic of robust variance estimates is that they are agnostic about the nature of the interdependence in the data. That is, the estimates obtained by (3) or (4) do not require the analyst to specify whether the conditional correlation among observations is positive or negative. In contrast, the GEE models discussed below provide a means for making inferences about covariate effects in which the nature of the interdependence, if known, can be used by the researcher to obtain better estimates of the parameters of interest.

## GEE MODELS

Generalized estimating equation models were introduced into biostatistics by Liang and Zeger (1986), and Zeger and Liang (1986) and were quickly adopted by researchers in a range of fields.[8] They are a generalization of the widely used generalized linear model (GLM) approach for exchangeable data (McCullagh and Nelder 1989). In both GLMs and GEEs, only the first two moments of the outcome variable are specified; specifically, the analyst specifies the mean of $Y_i$ to be some function of the $k$ covariates $\mathbf{X_i}$:

$$g(\mu_i) = \mathbf{X_i}\,\beta_{GEE} \tag{5}$$

and the variance is assumed to be some function of the mean and, if necessary, a scale parameter $\phi$. Once the researcher has selected the distribution of *Y* and the form of the link function $g(\cdot)$, estimates of $\hat{\beta}_{GEE}$ are then obtained from the solution to the set of $k$ "quasi-score" equations:

$$\mathbf{Q}_k(\beta_{GEE}) = \sum_{i=1}^{N} \frac{\partial \mu_i}{\partial \beta_k} (\mathbf{V}_i)^{-1} (Y_i - \mu_i) = 0 \tag{6}$$

[4] Commands for exactly replicating this simulation using the Stata statistical software are presented in the Appendix.

[5] I omit the robust, unclustered standard error estimates corresponding to Equation (3), because, given the way in which the data are generated, they are essentially identical to the estimates given in column three.

[6] Recall that, in the simple bivariate case, the formula for the standard error of $\hat{\beta}_1$ is simply

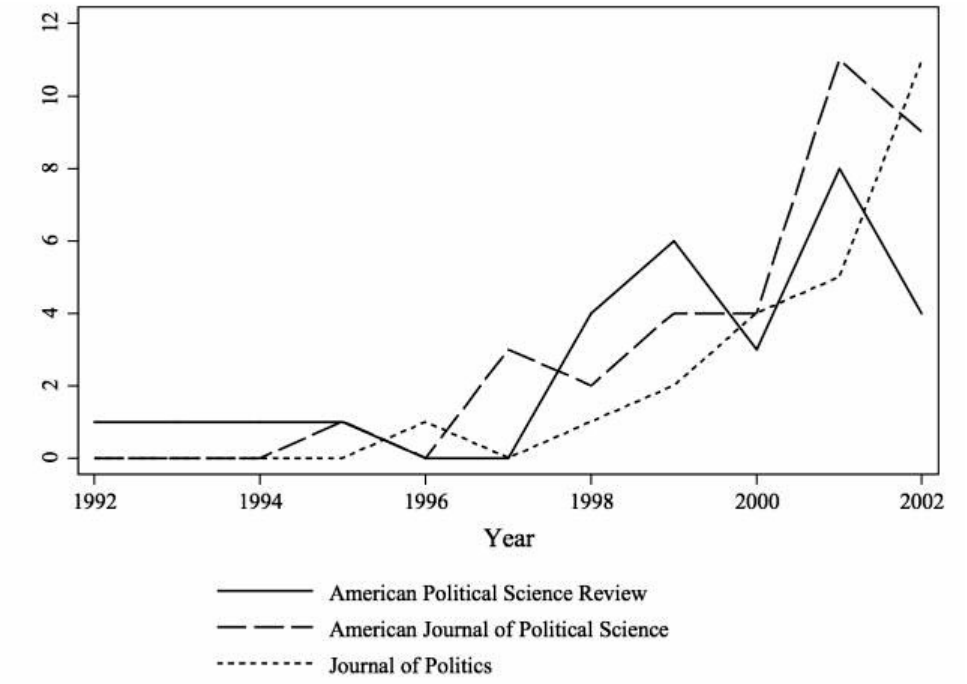
$$\sqrt{\frac{\frac{\sum \hat{u}_i^2}{N-2}}{\sum (X_i - \bar{X})^2}};$$

this suggests that increasing $N$ by a factor of 100 should result in approximately a 10-fold decrease in the standard error estimate for $\hat{\beta}_1$.

[7] Intuitively, the slight decrease in the standard errors between columns two and four is due to the fact that increasing the number of observations—even identical ones—in the data does, in fact, add some information to what we know about *Y* and *X*.

[8] Good reviews of GEE models include Diggle, Liang, and Zeger (1994) and Zorn (2001b); a now slightly-dated bibliography of these methods can be found in Ziegler, Kastner, and Blettner (1998).

FIGURE 1
INCIDENCE OF THE PHRASE "ROBUST STANDARD ERRORS" IN THREE MAJOR POLITICAL SCIENCE JOURNALS, 1992-2002




where $\mathbf{V}_i$ corresponds to the within-unit variance-covariance matrix. In cases where the data are correlated within the $N$ clusters, some provision must be made to account for that dependence. Zeger and Liang's (1986) solution was to specify a $n_i \times n_i$ matrix $\mathbf{R}_i(\alpha)$ of the working correlations across $j$ for a given unit $i$. While $\mathbf{R}_i(\alpha)$ can thus vary across observations, it is assumed to be fully specified by the vector of unknown parameters $\alpha$, which has a structure determined by the investigator and which is constant across units. This correlation matrix then enters the variance term of equation (6):

$$\mathbf{V}_i = \frac{(\mathbf{A}_i)^{1/2}\,\mathbf{R}_i(\alpha)(\mathbf{A}_i)^{1/2}}{\phi} \tag{7}$$

where the $\mathbf{A}_i$ are $n_i \times n_i$ diagonal variance matrices of $Y_i$ with $g(\mu_{ij})$ as the $j$th diagonal element. From this discussion, it is clear that the GEE is an extension of the GLM approach, and that the former reduces to the latter when $n_i = 1$. A range of possible correlation structures are possible, including independence (i.e., no within-unit correlation), exchangeable (where all observations in a cluster are equicorrelated), and autoregressive specifications of various orders; alternatively, a researcher may leave the matrix unspecified and simply estimate all

$$\frac{n_i(n_i - 1)}{2}$$

unique elements of $\mathbf{R}_i$.

If the model is properly specified, it can be shown that $Cov[\mathbf{Q}_i(\beta)] = \mathbf{D}'_i\mathbf{V}_i^{-1}\mathbf{D}_i$ (where $\mathbf{D}_i = \partial\mu_i/\partial\beta_k$), from which one can obtain a simple, model-based estimate of the parameter variances and covariances. This result depends, however, on proper specification of the correlation matrix $\mathbf{R}_i$; in the presence of misspecification of the correlation structure, the estimates $\hat{\beta}_{GEE}$ are still consistent, but $Cov[\mathbf{Q}_i(\beta)] \neq \mathbf{D}'_i\mathbf{V}_i^{-1}\mathbf{D}_i$. Under these circumstances, Liang and Zeger (1986) suggest the use of a robust estimate of the variance-covariance matrix:

$$\hat{\mathbf{V}}_{GEE} = \underset{i=1}{\overset{N}{N}} \left(\sum \hat{\mathbf{D}}'_i\hat{\mathbf{V}}_i^{-1}\mathbf{D}_i\right)^{-1} \left(\sum_{i=1}^{N} \hat{\mathbf{D}}'_i\hat{\mathbf{V}}_i^{-1}\,\hat{\mathbf{S}}_i\hat{\mathbf{V}}_i^{-1}\,\mathbf{D}_i\right) \left(\sum_{i=1}^{N} \hat{\mathbf{D}}'_i\hat{\mathbf{V}}_i^{-1}\mathbf{D}_i\right)^{-1} \tag{8}$$

where $\hat{\mathbf{S}}_i = (Y_i - \hat{\mu}_i)(Y_i - \hat{\mu}_i)'$ is a simple empirical covariance estimate. This estimator is analogous to that discussed above, in that it is consistent even if $\mathbf{R}_i$ is misspecified (Liang and Zeger 1986).

GEEs can be estimated with a range of standard software, and their parameter estimates can be interpreted in the same way as those from standard GLMs. While they were developed primarily for data involving multiple observations over time, GEEs have come to be used to address a range of other causes of correlated data as well, including spatial correlation (e.g., Albert and McShane 1995; Mugglestone, Kenward, and Clark 2002) and correlation due to the use of dyadic data (e.g., Oneal and Russett 1999). At one level, GEEs are similar to standard models with robust standard errors,[9] in that they account for dependence by simply

[9] Formally, GEEs are identical to such models when the correlation structure is specified to be independence; that is, when $\mathbf{R}_i = \mathbf{I}$.

correcting the variance-covariance matrix after the fact.[10] On the other hand, a potential advantage of GEEs over simple robust variance estimates is their ability to use information about the nature of the intracluster dependence to recover more precise estimates of the standard errors of $\hat{\beta}$. For example, if, as is often the case, we know not simply that observations in our data may be conditionally related, but how they are so (perhaps due to autocorrelation over time, or spatial effects), it would in principal be valuable to incorporate that knowledge into our analysis.

One question, then, is whether and to what extent—under practical conditions—the added complexity of GEEs is warranted, over simply using clustered or unclustered robust variance estimates. In addition, both of the two approaches leave it to the researcher to select the appropriate unit on which to cluster. In some instances (e.g., the survey example discussed above), question of what defines a cluster is effectively determined by the data structure. In most applied situations, however, the choice is not so obvious. For example, in the widely-discussed case of time-series cross-sectional data (e.g., Beck and Katz 1995, 1996), the analyst often expects conditional dependence both within cross-sectional units over time (due to persistence in unmeasured phenomena) and across units at a given point in time (e.g., due to unmeasured system-level phenomena). What, if any, effects the selection of the unit of clustering might have on a model's results remains an unanswered and largely uninvestigated question.

## AN EXAMPLE: REEVALUATING SEARCH AND SEIZURE, 1963-81

In an influential study, Segal (1986) examined individual-level voting in search and seizure cases decided by the U.S. Supreme Court between 1963 and 1981. Here, I reexamine Segal's data, considering the effects of case factors as well as measures of judicial ideology on the votes of justices in search and seizure cases. The purpose of this reanalysis is not to replicate Segal's original study.[11] Rather, it is to illustrate, using a realistic example, how choices about the variance estimators described above can have significant implications for one's findings, and to provide guidance about how applied researchers should best go about making those choices.

The data consist of a total of 1037 votes by 14 different justices in 123 search and seizure cases. The outcome of interest is each justice's vote on whether the search is found to be reasonable (coded 1) or not (coded 0). Segal (1986) examined the influence of variables relating to the nature of the search, the decision of the court below, and the participation of the United States on that outcome.[12] Six of his variables (including indicators for searches involving a *House*, a *Business*, an *Automobile*, and the defendant's *Person*, as well as those for the *Extent* of the search and whether or not the search was incident to an *Unlawful Arrest*) are expected to decrease the probability that a search would be found reasonable, while the remaining six variables (including indicators for the presence of a *Warrant*, *Probable Cause*, the *U.S. as a Party* to the case, whether the case was *Incident to* or *Following a Lawful Arrest*, and an index of *Exceptions* to the standard Fourth Amendment guarantee) should increase the probability of reasonableness. Here, I also include a variable for *Justice Liberalism*, coded as each justice's rescaled Segal-Cover (1989) score (Epstein and Mershon 1996), with the expectation that it will be negatively related to the propensity to find a search reasonable. Summary statistics are presented for the variables in Table 2. These data are especially well-suited to an examination of the differences across the various estimators outlined above: they vary both across cross-sectional units and over time, and are likely to exhibit correlation on both dimensions; they are of a size similar to many datasets used in political science research; and it is widely agreed that the factors present constitute a well-specified model of Supreme Court decision making in search and seizure cases, thus mitigating questions about model specification to the extent possible.

I begin by estimating a series of probit models, along with the standard error estimates from Equations (2)–(4) and their corresponding *z*-scores; these results are presented in Table 3.[13] In addition to the non-robust and unclustered estimates, I present two sets of clustered estimates. The first treats each Supreme Court case as the cluster, and so sums scores across votes within cases before summing across cases. This yields an effective *N* of 123, with from six to nine votes per case. The second treats each justice as a cluster, yielding an "*N*" of fourteen and between 24 and 121 votes per justice ($\mu = 74$). In practice, either of these approaches is reasonable, depending on one's beliefs about the likely source of conditional interdependence. On one hand, factors specific to each case, as well as potential interjudge influence in the form of bargaining, persuasion, and the like (e.g., Spaeth and Altfield 1985; Wawro and Farhang 2004), might lead one to the conclusion that justice's votes within a particular case are likely to be related. On the other hand, to the extent that justices attempt to maintain consistency in their voting records, and possibly because of the impact of precedent and other temporal factors, it is also reasonable to believe that a given justice's votes may be correlated across cases, but that there are few reasons to believe that different justices' votes within a case will be strongly related.

Substantively, the results square with the expectations in Segal's (1986) article: judicial ideology, the location and

[10] This is in contrast to subject-specific approaches, such as fixed- and random-effects models, which account for intrasubject correlation through explicit parameterization; see generally Hsaio (2003), or Wawro (2001) for an example.

[11] Segal's original analysis was limited to four justices who served long terms on the Court; here, I include data on all of the justices that participated in the cases in question.

[12] These variables are coded as in Segal (1986); see that study for coding details.

[13] Note that because adjustments to the standard errors take place after estimation, they have no effect on the point estimates of $\hat{\beta}$.

TABLE 2
SUMMARY STATISTICS

| Variable | Mean | Standard Deviation | Minimum | Maximum |
|---|---|---|---|---|
| Vote to Uphold Search | 0.53 | 0.50 | 0 | 1 |
| Justice Liberalism | 0.59 | 0.35 | 0.045 | 1 |
| House Search | 0.23 | 0.42 | 0 | 1 |
| Business Search | 0.15 | 0.36 | 0 | 1 |
| Auto Search | 0.20 | 0.40 | 0 | 1 |
| Person Search | 0.31 | 0.46 | 0 | 1 |
| Extent of Search | 0.86 | 0.35 | 0 | 1 |
| Warrant | 0.15 | 0.35 | 0 | 1 |
| Probable Cause | 0.32 | 0.47 | 0 | 1 |
| Incident to Lawful Arrest | 0.06 | 0.23 | 0 | 1 |
| After Lawful Arrest | 0.13 | 0.33 | 0 | 1 |
| Unlawful Arrest | 0.07 | 0.26 | 0 | 1 |
| Exception Index | 0.35 | 0.60 | 0 | 3 |
| U.S. Party | 0.45 | 0.50 | 0 | 1 |

Note: $NT$ = 1037 (123 cases and 14 justices); see text for details.

extent of the search, the occurrence of one or more exceptions, and the presence of the U.S. as a litigant all have the expected effects on the finding that a search is reasonable. In addition, several things about the various types of estimates are immediately apparent. First, there are only small differences between the non-robust and robust standard error estimates. Moreover, these differences are not systematic: for some covariates the robust estimates are larger, while for others the reverse is true. In this case, then, the choice of standard or robust but non-clustered variance estimates effectively makes no difference at all in the inferences one would make about the variables' effects.

The same cannot be said, however, about the clustered estimates. There, we see large differences in the sizes of the standard error estimates, depending on whether observations are clustered by case or by justice. In general, both sets of clustered estimates are larger than either the naive or the unclustered robust estimates; this is unsurprising, since we would expect that, in either case, any within-cluster correlation across votes would be positive. In addition, the estimated standard errors clustered by justice are generally smaller than those clustered by case, suggesting that the extent of within-case correlation across votes is greater than the cross-case correlation within each justice's set of votes.

The differences in the standard error estimates are illustrated graphically in Figure 2, which plots the estimated standard errors for each type of variance estimator against one another. Clearly, the strongest correspondence is between the naive and unclustered robust models; we see only a slight increase in the size of the standard errors from the naive to the unclustered model.[14] In contrast, the largest differences are between the case-specific and justice-specific robust models, where there is little or no correspondence between the standard error estimates.[15] Also, Figure 2 illustrates the fact that models which cluster on a particular unit have a disproportionate effect on the standard errors of covariates which vary only within those units. This is seen most graphically in the estimates for *Justice Liberalism*, a variable which is constant for any particular justice; relative to the naive estimate, the justice-specific estimate is more than twice as large. Conversely, for case-specific variables, all of which for a given case are constant across justices, we see the largest differences between the naive model and the case-specific estimates: in many instances (e.g., the estimates for *House* and *Business Search*, and the *Exception Index*) the case-clustered estimates are substantially larger than those in any of the other three models Once again, this is unsurprising: for variables that do not vary within clusters, equations (3) and (4) will yield similar results.

For comparison, we next turn to estimates using GEE models. Specifically, I estimate a series of four GEE models, all of which assume an exchangeable correlation structure within each cluster.[16] As in Table 3, I include both naive and robust estimates, and as before estimate models with both the justice and the case as the unit of clustering; these results are presented in Table 4.

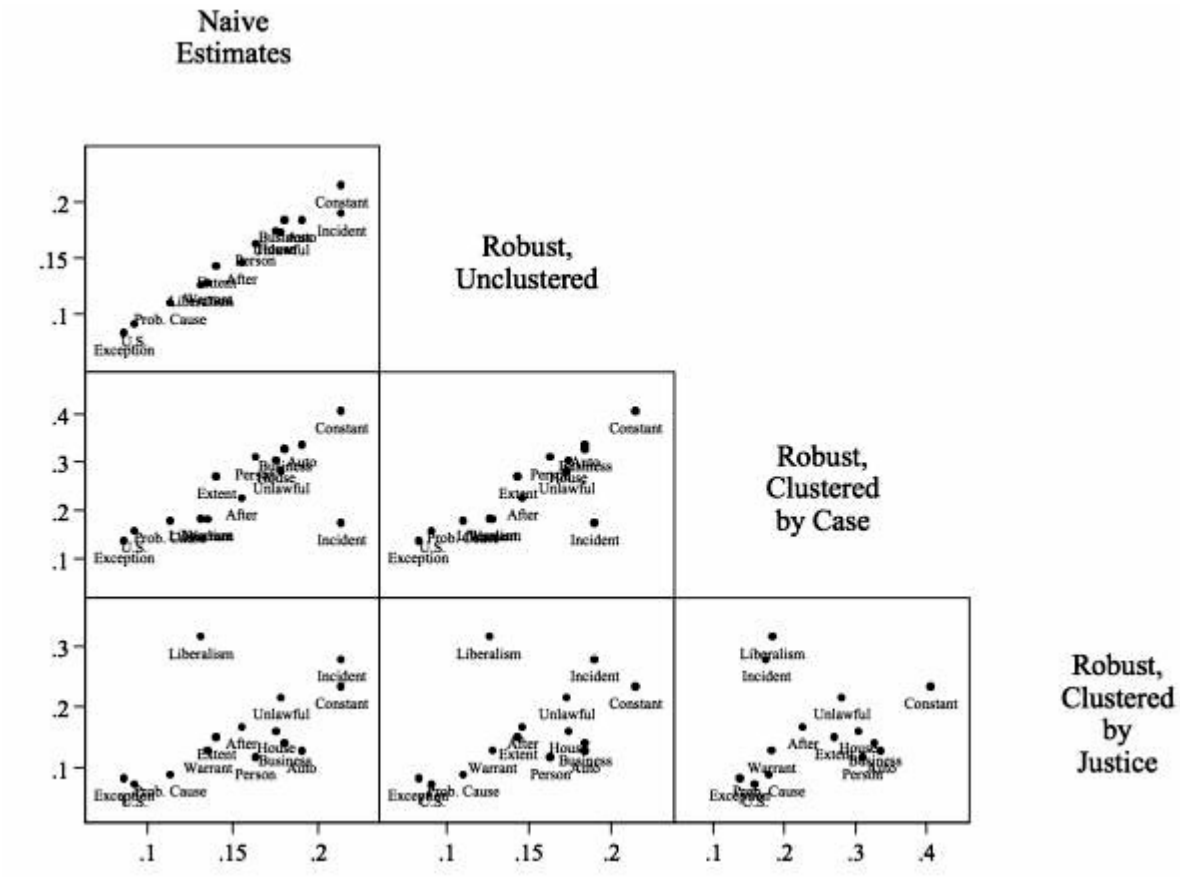
An important difference from the models presented in Table 3 is that, for GEE models, the choice of clustering unit

[14] The estimated relationship is *Naive* = 0.001 + 1.018 (*Robust*) + $\varepsilon$ ($R^2$ = 0.97, $N$ = 14). Also, a Wald test fails to reject the null that the two are the same (i.e., that $\beta = 1$) ($\chi^2_1 = 0.13$, $p = 0.72$).

[15] Formally, *Case-Specific* = 0.223 + 0.151 (*Justice-Specific*) + $\varepsilon$ ($R^2$ = 0.02, $N$ = 14); the corresponding Wald test rejects $H_0$: $\beta = 1$ ($\chi^2_1 = 7.17$, $p$ = 0.02).

[16] While other correlation structures may alter the results slightly, limiting the estimates to a single choice for $\mathbf{R}_i$ assists in the presentation. Moreover, GEE estimates are typically only slightly responsive to the choice of correlation structure; see Liang, Zeger, and Qaqish (1992) for a discussion.

FIGURE 2
ESTIMATED STANDARD ERRORS, PROBIT MODELS




Note: Figure plots standard errors for the 14 estimates (13 covariates plus the constant term), by type of estimate. See text for details.

has implications for both the point estimates $\hat{\beta}$ as well as for their standard errors. This is because the elements of $\hat{\beta}$ and $\mathbf{R}_i$ are estimated iteratively, so that the within-cluster correlations influence the main parameter estimates. With one or two exceptions, the estimates of $\hat{\beta}$ in Table 4 map closely both to one another and to those in Table 3, indicating that the choice of GEE and the selection of units has little effect on our assessment about the size of the estimated relationships. In contrast, however, we see much larger differences in the estimated standard errors, both across clustering units and between naive and robust variance estimators. Interestingly, the GEE results reveal what the standard probit results could only hint at: that the extent of intracluster correlation is higher within cases than within justices. While the lack of standard error estimates for $\rho$ make inferences impossible,[17] the estimate for intracase dependence is over twice that for dependence within justice's votes; to the extent that the divergence between naive and clustered estimates depends on the extent of intracluster correlation, this finding is consistent with the results in Table 3.

The differences in standard errors across the different GEE models are illustrated in Figure 3, which again plots the various estimates against one another. While, strictly speaking, the four panels in the lower left are not comparable (since they are based on different estimates of $\hat{\beta}$), they are presented for illustration. Within each choice of unit, the differences between naive and robust standard errors are generally slight.[18] And once again we see that, because it varies only across cases, the variable for *Justice Liberalism* is an outlier in the comparisons between the justice-specific and case-specific models. This finding again stresses the difference that the choice of unit makes when using clustered robust variance estimates, particularly for variables which vary only within clusters.[19]

[17] If the intragroup correlation were of greater substantive interest, GEE2 models could be used to estimate $\rho$ along with an associated measure of uncertainty; see Zorn (2001b) for an illustration.

[18] Formally: $Naive_{Case} = 0.40 + 0.880\ (Robust_{Case}) + \varepsilon$ ($R^2 = 0.67$; $H_0$: $\beta = 1 \rightarrow \chi^2_1 = 0.45$, $p = 0.51$) and $Naive_{Justice} = 0.038 + 0.821\ (Robust_{Justice}) + \varepsilon$ ($R^2 = 0.88$; $H_0$: $\beta = 1 \rightarrow \chi^2_1 = 4.16$, $p = 0.06$).

[19] As an additional point of comparison, I also estimated a fixed-effects model with separate intercepts for each of the 14 justices in the data. Briefly, both the $\hat{\beta}$s and the estimated standard errors from that model map closely to those using in Tables 3 and 4 in which the justice was the unit of clustering, though the fixed effects results were uniformly larger in magnitude. This latter effect is expected, because fixed-effects models are a form of conditional (or cluster-specific) model, while both ordinary

TABLE 3
PROBIT MODELS OF SUPREME COURT VOTING

| Variable | $\hat{\beta}$ | S.E. (*z*-score) | Robust S.E. (*z*-score) | Robust S.E., by case (*z*-score) | Robust S.E., by justice (*z*-score) |
|---|---|---|---|---|---|
| (Constant) | 1.531 | 0.213<br>(7.20) | 0.215<br>(7.11) | 0.406<br>(3.77) | 0.234<br>(6.55) |
| Justice Liberalism | –1.498 | 0.131<br>(–11.47) | 0.126<br>(–11.88) | 0.183<br>(–8.18) | 0.317<br>(–4.72) |
| House Search | –0.816 | 0.175<br>(–4.66) | 0.174<br>(–4.70) | 0.304<br>(–2.68) | 0.160<br>(–5.09) |
| Business Search | –0.957 | 0.180<br>(–5.32) | 0.184<br>(–5.21) | 0.327<br>(–2.93) | 0.140<br>(–6.85) |
| Auto Search | –0.863 | 0.190<br>(–4.55) | 0.184<br>(–4.70) | 0.336<br>(–2.57) | 0.127<br>(–6.82) |
| Person Search | –0.705 | 0.163<br>(–4.31) | 0.163<br>(–4.33) | 0.310<br>(–2.27) | 0.117<br>(–6.02) |
| Extent of Search | –0.390 | 0.140<br>(–2.78) | 0.143<br>(–2.73) | 0.270<br>(–1.44) | 0.150<br>(–2.60) |
| Warrant | 0.425 | 0.135<br>(3.16) | 0.128<br>(3.33) | 0.182<br>(2.34) | 0.128<br>(3.33) |
| Probable Cause | 0.028 | 0.113<br>(0.25) | 0.110<br>(0.26) | 0.178<br>(0.16) | 0.088<br>(0.32) |
| Incident to Lawful Arrest | 0.971 | 0.213<br>(4.55) | 0.190<br>(5.11) | 0.174<br>(5.57) | 0.279<br>(3.48) |
| After Lawful Arrest | 0.303 | 0.155<br>(1.95) | 0.146<br>(2.07) | 0.226<br>(1.34) | 0.167<br>(1.81) |
| Unlawful Arrest | –0.112 | 0.178<br>(–0.63) | 0.173<br>(–0.65) | 0.281<br>(–0.40) | 0.216<br>(–0.52) |
| Exception Index | 0.552 | 0.086<br>(6.45) | 0.083<br>(6.64) | 0.137<br>(4.04) | 0.082<br>(6.77) |
| U.S. Party | 0.357 | 0.092<br>(3.89) | 0.091<br>(3.92) | 0.158<br>(2.25) | 0.072<br>(4.93) |

Note: *ln*L = –582.62; *NT* = 1037. See text for details.

To grasp the above findings better, Table 5 presents the results of a principal-components factor analysis of the standard error estimates in Tables 3 and 4. Here, each of the fourteen model covariates (13 covariates plus the constant) is treated as an observation, while each of the eight types of standard error estimates constitute a separate variable. Those results clearly demonstrate what Figures 2 and 3 suggest: that the key determinant of similarity in the standard error estimators is not the technique used, but the choice of unit on which to cluster. The eight sets of standard error estimates group strongly into two factors,[20] which divide clearly according to the unit on which clustering was specified. As suggested above, the two sets of unclustered estimates load most heavily with those clustered by case, while the results clustered by justice constitute a distinct factor. For all eight, the uniqueness (defined as one minus the variance in *X* explained by the latent factor) is low, indicating that the two factors quite adequately represent the variation in those estimates.

probit and GEE are marginal (or population-averaged) models (see e.g., Neuhaus, Kalbfleisch, and Hauck 1991). Once again, the largest differences were observed for the *Justice Liberalism* variable. Interested parties can obtain those results from the author upon request.

[20] The eigenvalues for these first two factors are 5.09 and 2.29 respectively; together, they explain more than 92 percent of the variance. The same results hold if we include standard error estimates from the fixed- and random-effects models in the analysis.

TABLE 4
GEE MODELS OF SUPREME COURT VOTING

| | GEE, Grouped by Case | | | GEE, Grouped by Justice | | |
|---|---|---|---|---|---|---|
| Variable | $\hat{\beta}$ | Naïve S.E. (*z*-score) | Robust S.E. (*z*-score) | $\hat{\beta}$ | Naïve S.E. (*z*-score) | Robust S.E. (*z*-score) |
| (Constant) | 1.738 | 0.365<br>(4.76) | 0.412<br>(4.22) | 1.360 | 0.312<br>(4.36) | 0.281<br>(4.84) |
| Justice Liberalism | –1.800 | 0.123<br>(–14.66) | 0.169<br>(–10.65) | –1.232 | 0.367<br>(–3.36) | 0.393<br>(–3.13) |
| House Search | –0.816 | 0.311<br>(–2.62) | 0.285<br>(–2.86) | –0.904 | 0.164<br>(–5.52) | 0.127<br>(–7.11) |
| Business Search | –0.984 | 0.322<br>(–3.06) | 0.302<br>(–3.26) | –0.998 | 0.168<br>(–5.93) | 0.121<br>(–8.25) |
| Auto Search | –0.888 | 0.337<br>(–2.63) | 0.322<br>(–2.76) | –0.849 | 0.175<br>(–4.86) | 0.121<br>(–7.02) |
| Person Search | –0.830 | 0.293<br>(–2.83) | 0.295<br>(2.81) | –0.715 | 0.151<br>(–4.73) | 0.116<br>(–6.19) |
| Extent of Search | –0.367 | 0.254<br>(–1.45) | 0.296<br>(–1.24) | –0.415 | 0.131<br>(–3.17) | 0.144<br>(–2.88) |
| Warrant | 0.330 | 0.237<br>(1.39) | 0.205<br>(1.61) | 0.392 | 0.124<br>(3.17) | 0.117<br>(3.34) |
| Probable Cause | 0.063 | 0.200<br>(0.31) | 0.196<br>(0.32) | 0.082 | 0.104<br>(0.78) | 0.082<br>(1.00) |
| Incident to Lawful Arrest | 0.882 | 0.362<br>(2.44) | 0.220<br>(4.02) | 0.987 | 0.194<br>(5.10) | 0.254<br>(3.89) |
| After Lawful Arrest | 0.263 | 0.274<br>(0.96) | 0.248<br>(1.06) | 0.227 | 0.143<br>(1.59) | 0.151<br>(1.50) |
| Unlawful Arrest | –0.096 | 0.316<br>(–0.31) | 0.302<br>(–0.32) | –0.065 | 0.164<br>(–0.40) | 0.191<br>(–0.34) |
| Exception Index | 0.527 | 0.148<br>(3.57) | 0.146<br>(3.60) | 0.577 | 0.080<br>(7.17) | 0.065<br>(8.83) |
| U.S. Party | 0.345 | 0.165<br>(2.09) | 0.171<br>(2.02) | 0.345 | 0.085<br>(4.06) | 0.073<br>(4.76) |
| Estimated $\hat{\rho}$ | 0.303<br>(n/a) | | | 0.122<br>(n/a) | | |

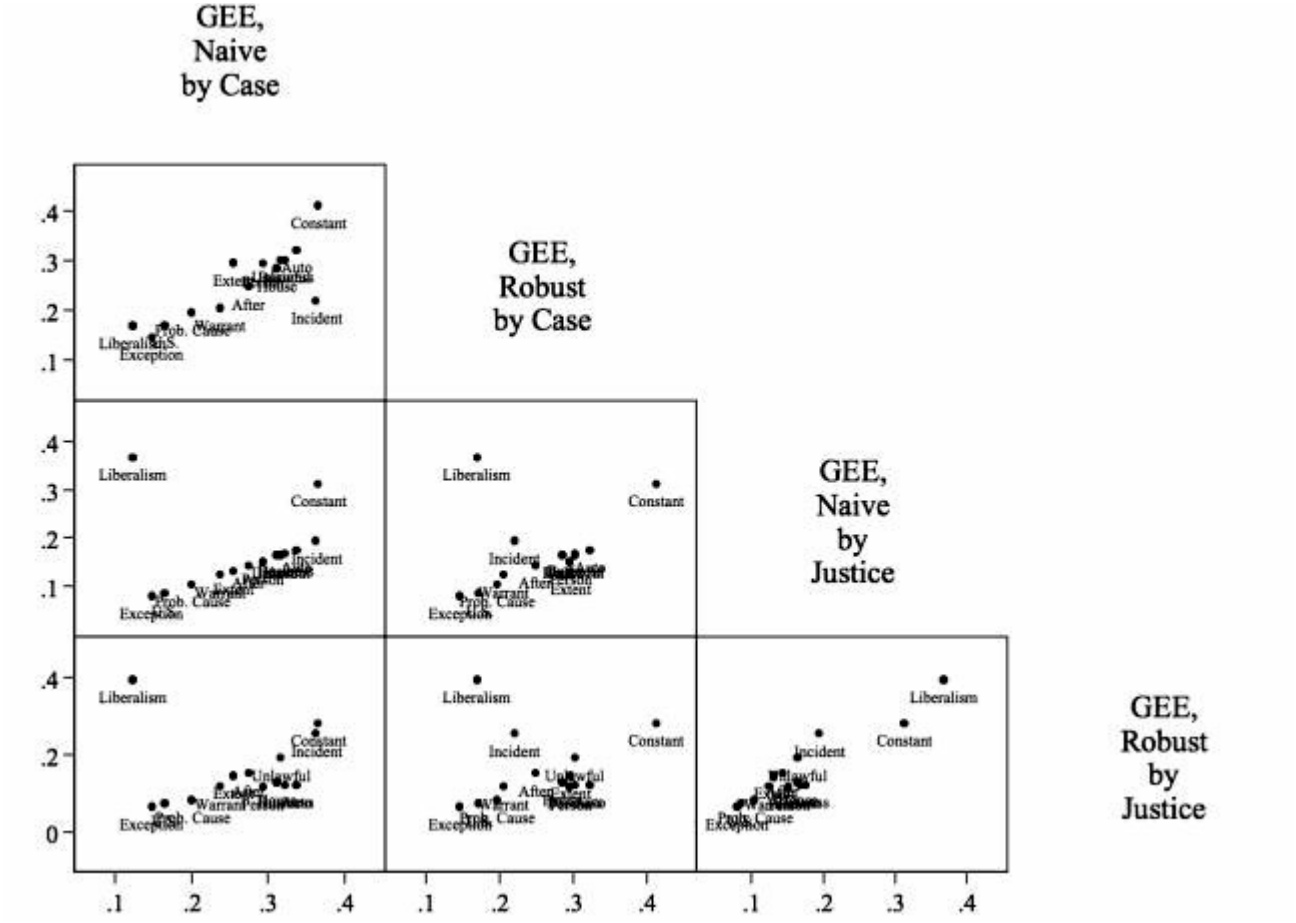
Note: All models assume an exchangeable correlation structure. *NT* = 1037. See text for details.

A final question goes to the practical effects that model choice may have on the inferences made from these data. To summarize these effects, I have grouped the findings into one of five categories, based on their significance levels: $p < .001$, $.001 \leq p < .01$, $.01 \leq p < .05$, $.05 \leq p < .10$, and $p \geq .10$, all one-tailed.[21] I then examine whether the inference one would make about the significance of each variable is consistent with that from the naive probit model (i.e., Table 3, column 1). Variables in which the estimate's significance level category is the same as for the naive probit model are indicated with a filled circle; those in which the significance level is in an adjacent category are indicated with a hollow circle. These results are illustrated in Table 6.

As in Table 5, the results in Table 6 suggest that it is the choice of clustering unit, rather than the statistical method, which has the largest impact on inferences. That is, while both ordinary probit and GEE models clustered by justice correspond closely to the naive results, both sets of case-clustered results show marked differences from the naive

[21] While this Fisherian method is, admittedly, an imperfect way to conduct statistical inference (e.g., Gill 1999), it nonetheless corresponds to the approach used by most political scientists.

FIGURE 3
GEE STANDARD ERRORS




Note: Figure plots standard errors for the 14 estimates (13 covariates plus the constant term), by type of estimate. See text for details.

model. This is particularly true for the *Extent of Search* and *U.S. Party* covariates: for those variables, all three case-clustered models yield the same inferences as the naive model, while none of the justice-clustered models do so. The exception is, predictably, the judicial ideology variable, where the results of the justice-clustered GEE models are the only ones which differ substantially from the naive model. Thus, from this example, it would appear that the unit on which clustering occurs, rather than the choice of estimator itself, is the larger factor affecting inferences.

## CONCLUSIONS

The growing awareness of issues relating to interdependence, and the increasing sophistication with which political scientists are addressing those issues, are laudable developments in the discipline. At the same time, to this point applied researchers have had little guidance in choosing among the various techniques available for modeling data in which conditional dependence is present.

From this examination of various approaches to correlated data, we may draw several conclusions. First, and most significant, the differences between GEE and more traditional GLM models with robust variance estimates appear to be less important, at least for inference, than are choices about the unit on which observations are grouped. In nearly every instance, at least for this single example, the choice of estimator made little or no difference on the substantive inferences the researcher would draw from these data. By contrast, the results showed consistently large differences in both actual estimates and substantive conclusions regarding statistical significance as a function of the unit on which the data were grouped. Relatedly, these differences varied systematically with the nature of the variation in the covariates themselves: the choice to cluster the data on a particular dimension of the data led to especially large differences in the standard error estimates for variables which exhibited most or all of their variation on that dimension.

The central practical implication of these results is that researchers should think hard about the appropriate unit for clustering variance estimates. Current practices often encourage researchers to engage in extensive hand wringing over the type of statistical model used, while giving little serious thought to defining the fundamental units of analysis in

TABLE 5
FACTOR LOADINGS FOR STANDARD ERROR ESTIMATES

| Estimator | Factor I | Factor II | Uniqueness |
|---|---|---|---|
| Naïve | 0.88 | 0.38 | 0.08 |
| Robust, Unclustered | 0.93 | 0.34 | 0.02 |
| Robust, Clustered by Case | 0.91 | 0.06 | 0.17 |
| Robust, Clustered by Justice | 0.16 | 0.96 | 0.05 |
| GEE, Naïve by Case | 0.95 | 0.03 | 0.10 |
| GEE, Robust by Case | 0.95 | 0.07 | 0.10 |
| GEE, Naïve by Justice | 0.22 | 0.92 | 0.10 |
| GEE, Robust by Justice | 0.07 | 0.99 | 0.01 |

Note: $N = 14$. Results are for principal-components factor analysis with orthogona(varimax) rotation. Results are for principal-components factor analysis with orthogonal (varimax) rotation.

their data. These findings suggest that, while estimator choice remains an important issue, it ought not necessarily to consume the level of intellectual energy it often does, while the choice of clustering unit should be given correspondingly more thought. As mentioned above, in some situations, this choice is a natural one: panel surveys of randomly selected respondents, for example, will have little or no nonrandom cross-unit variation, and so grouping the data by subject (to address potential temporal dependence) is the obvious choice. In most cases, however, the choice of a clustering unit is less clear, and may involve balancing expectations about the locus of conditional dependence.

In making that choice, several factors should come into play. As in all such decisions, the key motivating influence should be substance: to the extent that the researcher has strong, theoretically-grounded, a priori reasons to expect conditional variation within a particular unit, those beliefs should be at the center of the decision. Also, researchers should pay particularly close attention to the nature of the variation in their independent variables. In most instances, analysts are faced with data that vary both between cross-sectional units and over time or observations for example. The nature of that variation can, as seen above, have important implications for one's results (Zorn 2001a). In general, the choice to cluster on a particular unit will have the greatest impact on the standard error estimates which vary primarily over those units.

A related consideration is model specification. It is useful to recall that White's original (1980) discussion of robust variance-covariance estimators was framed, at least in part, as an approach to specification testing. One implication of this is that variability in one's variance estimates, either as a function of estimator or due to the choice of clustering unit, can serve as an indication of potential specification issues.[22] In particular, sensitivity in those estimates to such choices is a telltale sign of systematic, unit-level conditional variation, and thus can be seen as providing evidence in favor of the superiority of robust variance estimates over their naïve counterparts.

In light of these factors, several concrete prescriptions follow. For example, it is reasonable to expect that in cases

[22] I thank an anonymous reviewer for pointing this out.

TABLE 6
VARIABLE-SPECIFIC INFERENCES ACROSS MODELS

| | Probit Models | | | GEE Models | | | |
|---|---|---|---|---|---|---|---|
| Variable | Robust | Robust, by Case | Robust, by Justice | Naïve, by Case | Robust, by Case | Naïve, by Justice | Robust, by Justice |
| (Constant) | ● | ● | ● | ● | ● | ● | ● |
| Justice Liberalism | ● | ● | ● | ● | ● | ❍ | ❍ |
| House Search | ● | ❍ | ● | ❍ | ❍ | ● | ● |
| Business Search | ● | ❍ | ● | ❍ | ● | ● | ● |
| Auto Search | ● | ❍ | ● | ❍ | ❍ | ● | ● |
| Person Search | ● | — | ● | ❍ | ❍ | ● | ● |
| Extent of Search | ● | — | ● | — | — | ● | ● |
| Warrant | ❍ | ● | ❍ | — | ❍ | ● | ❍ |
| Probable Cause | ● | ● | ● | ● | ● | ● | ● |
| Incident to Lawful Arrest | ● | ● | ● | ❍ | ● | ● | ● |
| After Lawful Arrest | ● | ❍ | ● | — | — | ❍ | ❍ |
| Unlawful Arrest | ● | ● | ● | ● | ● | ● | ● |
| Exception Index | ● | ● | ● | ● | ● | ● | ● |
| U.S. Party | ● | — | ● | — | — | ● | ● |

Note: Table indicates whether the same inferences would be drawn about the variable effects as in the naïve model. ● indicates categorical agreement about variable significance; ❍ indicates agreement within one category. Categories are $p < .001$, $.001 < p < .01$, $.01 < p < .05$, $.05 < p < .10$, and $p > .10$ (all one-tailed). See text for details.

where the choice is nonobvious, authors should provide some substantive justification for their choice of clustering/grouping unit. Relatedly, in those instances analysts should conduct and report sensitivity analyses vis-a-vis the choice of clustering unit. Taken together, such expectations will raise awareness of the importance of such decisions, while at the same time discouraging fishing expeditions for results most supportive of the authors' hypotheses.

The use of both robust variance estimators and GEEs represent important advances over models that require conditional exchangeability to obtain consistent estimates. To this point, researchers who had reason to believe that their data contained some form of conditional dependence have often turned to methods such as a quick fix to deal with that dependence. While the techniques themselves are both robust and valuable, however, the results here suggest that their application should not be undertaken without substantial thought into the implications of their use.

## APPENDIX
## STATA SYNTAX FOR RESULTS IN TABLE 1

```
. clear
. set obs 25
. gen id = _n
. set seed 7851
. gen X=invnorm(uniform())
. gen Y = 1 + X + invnorm(uniform())
. reg Y X
. expand 100
. reg Y X
. reg Y X, robust cluster(id)
```

## 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 (June): 627-38.

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 (September): 634-47.

_____. 1996. "Nuisance vs. Substance: Specifying and Estimating Time-Series Cross-Section Models." *Political Analysis* 6 (1): 1-36.

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

Eicker, F. 1963. "Asymptotic Normality and Consistency of the Least Squares Estimators for Families of Linear Regressions." *Annals of Mathematical Statistics* 34 (2): 447-56.

Epstein, Lee, and Carol Mershon. 1996. "Measuring Political Preferences." *American Journal of Political Science* 40 (February): 261-94.

Gill, Jeff. 1999. "The Insignificance of Null Hypothesis Significance Testing." *Political Research Quarterly* 52 (September): 647-74.

Hojnacki, Marie, and David C. Kimball. 1998. "Organized Interests and the Decision of Whom to Lobby in Congress." *American Political Science Review* 92 (December): 775-90.

Hsaio, Cheng. 2003. *Analysis of Panel Data*, 2nd ed. New York: Cambridge University Press.

Huber, P. J. 1967. "The Behavior of Maximum Likelihood Estimates under Non-Standard Assumptions." *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* 1 (1): 221-33.

King, Gary. 2001. "Proper Nouns and Methodological Propriety: Pooling Dyads in International Relations Data." *International Organization* 55 (Spring): 497-507.

_____. 1989. *Unifying Politial Methodology: The Likelihood Theory of Statistical Inference*. New York: Cambridge University Press.

Liang, Kung-Yee, and Scott L. Zeger. 1986. "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika* 73 (1): 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* 54 (1): 3-40.

Long, J. Scott, and Laurie H. Ervin. 2000. "Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model." *American Statistician* 54 (August): 217-24.

Mackinnon, J. G., and H. White. 1985. "Some Heteroskedasticity Consistent Covariance Matrix Estimators with Improved Finite Sample Properties." *Journal of Econometrics* 29 (1): 53-57.

McCullagh, Peter. 2004. Exchangeability and Regression Models. Working paper: Department of Statistics, University of Chicago.

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

Mugglestone, M. A., M. G. Kenward, and S. J. Clark. 2002. "Generalized Estimating Equations for Spatially Referenced Binary Data." In Dario Gregori, Gaetano Carmeci, Herwig Friedl, Anuska Ferligoj, and Attilio Wedlin, eds., *Correlated Data Modeling: Proceedings, Trieste 1999,* pp.121-30. Milan: FrancoAngeli S.R.L.

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. 1999. "The Kantian Peace: The Pacific Benefits of Democracy, Interdependence, and International Organizations, 1885-1992." *World Politics* 52 (October): 1-37.

Segal, Jeffrey A. 1986. "Supreme Court Justices as Human Decision Makers: An Individual-Level Analysis of the Search and Seizure Cases." *Journal of Politics* 47 (November): 938-55.

Segal, Jeffrey A., and Albert D. Cover. 1989. "Ideological Values and the Votes of U.S. Supreme Court Justices." *American Political Science Review* 83 (June): 557-65.

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 (March): 70-83.

Stimson, James A. 1985. "Regression in Time and Space: A Statistical Essay." *American Journal of Political Science* 29 (November): 914-47.

Wawro, Gregory. 2001. "A Panel Probit Analysis of Campaign Contributions and Roll Call Votes." *American Journal of Political Science* 45 (July): 563-79.

Wawro, Gregory, and Sean Farhang. 2004. "Institutional Dynamics on the U.S. Court of Appeals: Minority Representation Under Panel Decision-Making." *Journal of Law, Economics, and Organization* 20 (September): 299-330.

White, Halbert. 1980. "A Heteroscedasticity-Consistent Covariance Matrix and a Direct Test for Heteroscedasticity." *Econometrica* 48: 817-38.

Williams, R. L. 2000. "A Note on Robust Variance Estimation for Cluster-Correlated Data." *Biometrics* 56: 645-46.

Zeger, Scott L., and Kung-Yee Liang. 1986. "Longitudinal Data Analysis for Discrete and Continuous Outcomes." *Biometrics* 42 (1): 121-30.

Ziegler, Andreas, Christian Kastner, and Maria Blettner. 1998. "Generalised Estimating Equations: An Annotated Bibliography." *Biometrical Journal* 40 (2): 115-39.

Zorn, Christopher. 2001a. "Estimating Between- and Within-Cluster Covariate Effects, with an Application to Models of International Disputes." 2001. *International Interactions* 27 (4): 433-45.

_____. 2001b. "Generalized Estimating Equation Models for Correlated Data: A Review with Applications." *American Journal of Political Science* 45 (April): 470-90.

Received: August 17, 2004
Accepted for Publication: October 12, 2004
zorn@sc.edu