THE PRACTICAL RESEARCHER

Estimating the Impact of State Policies and Institutions with Mixed-Level Data

David M. Primo, University of Rochester
Matthew L. Jacobsmeier, University of Rochester
Jeffrey Milyo, University of Missouri at Columbia

ABSTRACT

Researchers are often interested in the effects of state policies and institutions on individual behavior or other outcomes in sub-state-level observational units, such as election results in state legislative districts. In this article, we examine the issue of clustered data in state and local politics research and the analytical problems it can cause. Standard estimation methods applied in most regression models do not properly account for the clustering of observations within states, leading analysts to overstate the statistical significance of coefficient estimates, especially of state-level factors. We discuss the theory behind two approaches for dealing with clustering—clustered standard errors and multilevel modeling—and argue that calculating clustered standard errors is a more straightforward and practical approach, especially when working with large datasets or many cross-level interactions. We demonstrate the relevance of this topic by replicating a recent study of the effects of state post-registration laws on voter turnout (Wolfinger, Highton, and Mullin 2005).

RESEARCHERS OFTEN NEED to estimate the effects of state-level policies or institutions on individual or other sub-state-level outcomes.[1] For instance, do state campaign finance laws influence voter perceptions of state government? Do voter registration laws affect turnout? Do legislators in states with term limits behave differently than legislators in states with no term limits? The data used to answer these questions will often include multiple observations from the same state. To the extent that these observations are non-independent by virtue of being linked by state, we say that the observations are in the same "cluster."

Such empirical analyses may be undertaken using either state-level data,

State Politics and Policy Quarterly, Vol. 7, No. 4 (Winter 2007): pp. 446–459

as in a study probing the association between turnout rates and state campaign finance laws, or mixed-level data, as in a study linking state campaign finance laws and individual-level turnout data. These types of studies are known, respectively, as ecological and mixed-level (aka, hierarchical or contextual) analyses.[2] We focus on mixed-level analysis, which is becoming the more prevalent of the two. While both theoretically appropriate and practically possible, this type of analysis is fraught with statistical pitfalls. In particular, standard regression techniques applied to mixed-level data often attribute exaggerated levels of statistical significance to coefficient estimates, especially for state-level variables (Moulton 1990). This outcome occurs, in brief, because mixed-level data typically is comprised of multiple observations per state (for instance, of individuals residing in that state). These observations will often not be independent, thereby violating a standard assumption in regression analysis that the errors are independently and identically distributed (i.i.d.). This problem is closely related to heteroskedasticity, or non-constant error variance, as we will demonstrate.

The goal of this article is to provide a clear illustration of both the ease and importance of correcting regression coefficients' standard errors for clustered observations in the contextual analysis of state politics and policy. This procedure, which estimates what are referred to as clustered (a.k.a. cluster-adjusted or cluster-consistent) standard errors, accounts for the fact that observations within each state are unlikely to be independent, thereby violating a core assumption of many estimation procedures, including ordinary least squares.[3] As a demonstration of this approach, we replicate and extend the analysis in a recent article by Wolfinger, Highton, and Mullin (2005), contrasting clustered standard errors with an alternative approach, the study of contextual relationships using multilevel modeling. The authors of this study (hereafter "WHM") employ individual-level survey data to estimate the effects of state post-registration laws on voter turnout, but WHM do not account for the clustering of survey respondents within states. As we demonstrate, this absence has caused a downward bias in the standard errors associated with estimates of how state post-registration laws affect turnout.

By no means are WHM alone in ignoring the potential problems associated with clustered data. For example, a survey of every issue of *State Politics & Policy Quarterly* revealed a variety of approaches for dealing (or not dealing) with clustered observations in mixed-level studies. Most of these studies simply do not address the issue. These included a study of the effects of state ballot initiatives on individual-level voter turnout and political knowledge (Tolbert, McNeal, and Smith 2003), an examination of the synergistic effects of campaign effort and electoral reforms on turnout in legislative districts (Francia and Herrnson 2004), a study of the effects of state redistricting methods on competition in congressional races (Carson and Crespin 2004), an analysis of the effects of salient state ballot initiatives on voter turnout (Lacy 2005), and a study of the contextual effects of county voting equipment and spoiled ballot rates on individuals' trust in the voting process (Bullock, Hood, and Clark 2005).[4] In contrast, Branton (2004) adjusts for clustering in her study of the effects of racial and ethnic diversity on vote choice. Buckley and Westerland (2004) examine the impact of clustering in discrete event history analysis. And while Bonneau (2005) does not present standard errors corrected for clustering, he notes that his key findings on the determinants of state judicial campaign spending are not sensitive to state-level clustering.

Thus, the frequency with which state politics and policy researchers analyze mixed-level data, and the infrequency with which they account for clustered observations within them, suggest the need for a practical discussion of how best to address clustering. In the following, we provide intuitive and analytical justifications for clustered standard errors, contrasting this method with another popular method of dealing with mixed-level data: multilevel modeling. After this, we use a replication of the WHM study to illustrate the difference that cluster-adjustment can make in testing the effects of state policies on individual behavior. Our goal is to encourage scholars of state politics to take account of clustering issues more consistently by using appropriate statistical techniques.[5]

WHAT IS CLUSTERING?

Clustering arises because the attributes of states in which individuals reside do not vary across individuals within each state. For example, in an analysis on national survey data, every respondent from California will have the same majority party in the state legislature, the same voter registration law, the same type of judicial selection, and so forth. Hence, such clustered observations violate the independence assumption required by most estimation methods. Technical treatises on the problems presented by clustered observations, along with solutions, have been available for some time (e.g., Froot 1989, Moulton 1990, Rogers 1993, Williams 2000, and Wooldridge 2002 and 2003), but such solutions are far from universally implemented, as our review of *SPPQ* articles demonstrated.[6]

What is the effect of such clustering? To illustrate, consider the gold standard for statistical evaluation studies: a double-blind random trial, in which neither the investigator nor the subjects know who is assigned to be in the control or treatment group. Random assignment prevents problems arising from selection bias and endogeneity that might occur if individuals were to self-select into groups or if researchers assigned individuals to groups in a nonrandom manner. Now contrast this with a design where researchers observe the behavior of individuals who are exposed to different state laws by virtue of living in different states. For example, researchers may gather individual-level survey data to evaluate the effects of state voter registration laws on voter turnout. For simplicity, assume that a state either has or does not have "easy" registration, so that the treatment effect is estimated by contrasting turnout in states with and without such laws. Therefore, every respondent that shares the characteristic of "residing in North Dakota" will receive the same treatment (easy registration). In fact, every individual will belong to a group (i.e., a state) that is either all in or all not in the treatment. Because the laws are not assigned to states randomly, and the states individuals live in determine their exposure to the treatment of state-level laws, the independence assumption necessary in most statistical techniques will be violated.[7]

Despite this concern, mixed-level analyses have the great advantage of allowing the researcher to both describe and contrast more confidently compositional (e.g., income level) and contextual (e.g., registration laws) effects on individual behavior. In addition, such studies often have the seeming advantage of having thousands of individual observations available to analyze. In general, additional data allow for increased precision of estimates and, as such, are quite desirable. But if clustering causes a loss of independence among the data, inferences based on these estimates might be misleading.

The main problem here is that, in effect, the number of independent observations is not the number of cases, but rather the number of clusters. In the case of state policy studies, this results in 50 independent observations. Therefore, less information comes from these data than if the individual-level cases were truly independent of one another in terms of the policy or other condition they receive by dint of where they live. The failure to account for this clustering may cause the researcher to understate the standard errors for the estimated regression coefficients, especially for state-level variables (Moulton 1990).

To illustrate this problem, suppose that you are interested in estimating the effects of both state- and individual-level determinants on an individual-level outcome, perhaps the effects of voter registration policy and income on whether a person votes in an election. For ease of exposition, assume that these are only two independent variables, one of each type (individual- and state-level), and you have only a single year's cross-section of observations. (The logic of the following argument applies to situations with both more variables and more years of data.)

Formally, consider the following setup. Let $i = 1, \ldots, N$ index individuals and $j = 1, \ldots, J$ index the cluster to which an individual belongs, where n_j is the number of observations in cluster j. Assume that the state-level (Level 2) variable, w, influences the impact of the individual-level (Level 1) variable, x, on behavior, implying an interaction effect. This assumption can be dropped but is kept in to make the equation slightly more general. Thus, we can write the following equations:[8]

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + \varepsilon_{ij} \tag{1}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}w_j + u_{0j} \tag{2}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}w_j + u_{1j} \tag{3}$$

We can then substitute and generate one equation for our final model:

$$y_{ij} = \gamma_{00} + \gamma_{01}w_j + \gamma_{10}x_{ij} + \gamma_{11}w_jx_{ij} + (u_{0j} + u_{1j}\,x_{ij} + \varepsilon_{ij}) \tag{4}$$

Our resulting Equation 4 looks like a standard regression equation, but with a compound error term due to the fact that the coefficients are assumed to be random. There are typically three ways of estimating the coefficients of theoretical interest (γ_{01}, γ_{10}, γ_{11}) in this model. The first approach runs ordinary least squares on Equation 4, ignoring the clustered nature of the data, and treating the compound error term as if it consisted of only an i.i.d. ε_{ij} term. The second approach uses OLS to estimate the slopes and calculate clustered standard errors. The third approach models the multilevel nature of the data explicitly. We will consider each approach in turn.

The simple OLS approach to estimation and inference is usually the worst option. Recall that to conduct valid statistical inference, OLS requires the assumption that $\varepsilon_{ij} \sim N(0, \sigma^2)$ i.i.d. However, to the extent that observations within a cluster (state) share some common, even if unmeasurable, characteristics, the i.i.d. assumption is violated. As a result, the standard errors are typically underestimated, and the OLS estimator is no longer the best-linear-unbiased estimator (BLUE), even if we assume that coefficients are non-stochastic (i.e., $u_{0j} = u_{1j} = 0$).

The second approach to estimating this model involves using the OLS point estimates of the slopes, but adjusting the estimates of their standard errors to account for non-independence. Cluster-adjustment allows the observations within a cluster to be correlated, but it requires the assumption that observations across clusters are independent. The resulting standard errors are a variant on what are often called Huber-White heteroskedasticity–consistent standard errors, which allow for a general form of heteroskedasticity but do not allow for errors to be correlated across or within units (Huber 1967; White 1980). Clustered standard errors account for both this general form of heteroskedasticity as well as for any intra-cluster correlation.[9]

We can further distinguish between OLS, Huber-White, and clustered standard errors by looking at the variance-covariance matrix for OLS, $E\,[(\mathbf{X'X})^{-1}\mathbf{X'}\varepsilon\varepsilon'X\,(\mathbf{X'X})^{-1}]$. The traditional OLS assumption is that $E\,[\hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\varepsilon}}'] = \sigma^2\mathbf{I}$, where σ^2 is the average of the squared residuals. From this, we obtain the matrix, $\sigma^2\,(\mathbf{X'X})^{-1}$. The OLS standard errors for each beta are then calculated from the diagonals of this variance-covariance matrix. But when heteroskedasticity is present (that is, when the variance of the error term is not constant across observations), $E\,[\hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\varepsilon}}']$ is no longer $\sigma^2\mathbf{I}$. The $n \times n$ matrix $E\,[\hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\varepsilon}}']$ consists of the individual-specific variances along the diagonal and the covariances of the errors on the off-diagonals. White (1980) developed a standard error estimator that addressed heteroskedastic, non-constant variance by maintaining the assumption that the off-diagonal (covariance) terms of the $E\,[\hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\varepsilon}}']$ matrix were still 0 (i.e., that the errors were not correlated across observations), but relaxing the assumption that all the non-diagonal terms were equal.

The clustered standard error approach takes the Huber-White correction one step further by allowing off-diagonal (covariance) elements in the $E[\hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\varepsilon}}']$ matrix from the same cluster to be non-zero. This allows for any arbitrary correlation among the observations within clusters and any arbitrary heteroskedasticity in the error term, but it assumes no correlation among observations across clusters. Thus, if the OLS variance-covariance matrix is $\sigma^2\,(\mathbf{X'X})^{-1}$, the cluster-adjusted matrix is

$$(\mathbf{X'X})^{-1}\sum_{j=1}^{J}\left\{\left(\sum_{i=1}^{n_j}\hat{\varepsilon}_i\mathbf{x}_i\right)\left(\sum_{i=1}^{n_j}\hat{\varepsilon}_i\mathbf{x}_i\right)'\right\}(\mathbf{X'X})^{-1}$$

where $\mathbf{x}_i$ refers to the observations within a given cluster. If J=N and $n_j=1$ (that is, if each cluster is made up of a single case), then the formula reduces to the Huber-White estimator. Clustered standard errors are then calculated by taking the square root of the appropriate element of the variance-covariance matrix (for instance, element (1,1) for the standard error of the constant) and applying the finite-sample adjustment, (J/(J–1))*((N–1)/(N–k)), where J is the number of clusters, N is the total number of observations, and k is the number of regressors.

The clustered standard errors technique works well for a variety of regression methods and estimation procedures, including logit and probit, provided that the number of clusters is large. How many clusters is enough? Fortunately for state politics scholars, Monte Carlo simulations (Kezdi 2003; Bertrand et al. 2004; and Hansen 2005) suggest that 50 clusters are more than sufficient for valid and reliable inference.[10] So when working with individual-level data from a single year across 50 states, one can adjust for clustering at the state level. When working with individual-level data from the 50 states measured over several years, one can cluster either by state or by state-year.[11] The latter is appropriate if one assumes that observations are dependent within a state-year but are independent across years within a state (e.g., vote choice in a state pre- and post-party realignment). But if observations from the same state in different years may be dependent, then clustering should take place by state only. In short, one should cluster on the macro-level features that the researcher believes are causing dependence across observations.

The third approach for dealing with the compound error term (see Equation 4) caused by the clustering of data is to use hierarchical linear modeling (HLM) or related multilevel modeling procedures that explicitly model the error term (Bryk and Raudenbush 1992; Raudenbush and Bryk 2002; Steenbergen and Jones 2002; Franzese 2005; Bowers and Drake 2005). These procedures allow the researcher to estimate how much each level of analysis is contributing to explanation in the model, and how much each level is contributing to the error. In other words, the researcher can assess whether the explanation is primarily macro-level or individual-level. In a nutshell, HLM estimation uses Equations 1–3 and the assumptions below to estimate coefficients, variances, and covariances that maximize the likelihood of observing the data, given the model (Bryk and Raudenbush 1992, 45). Since this is a maximum likelihood estimation (MLE) approach, the misspecification of the error term in HLM models propagates throughout the entire estimation procedure, including the estimation of coefficients. Thus, it important to get the specification correct for HLM, not only for inference but also for point estimation. On the other hand, clustered standard errors, which are calculated after estimation, do not add the additional complexity that estimating variance components does, and this approach does not have the same risk to valid point estimation that the HLM approach does using MLE.

HLM typically requires the following assumptions regarding the components of Equations 1–3:

$$\mathrm{E}\,(\varepsilon_{ij}) = \mathrm{E}\,(u_{0j}) = \mathrm{E}\,(u_{1j}) = 0 \qquad \text{(A1)}$$

$$\mathrm{Var}\,(\varepsilon_{ij}) = \sigma^2 \qquad \text{(A2)}$$

$$\mathrm{Var}\,(u_{0j}) = \dagger_{00} \qquad \text{(A3)}$$

$$\mathrm{Var}\,(u_{1j}) = \dagger_{11} \qquad \text{(A4)}$$

$$\mathrm{Cov}\,(u_{0j}, u_{1j}) = \dagger_{10} \qquad \text{(A5)}$$

$$\mathrm{Cov}\,(u_{0j}, \varepsilon_{ij}) = \mathrm{Cov}\,(u_{1j}, \varepsilon_{ij}) = 0 \qquad \text{(A6)}$$

These assumptions specify the relationship among the error terms; the assumption of constant error variance in the level-1 observations (A2) can be relaxed (see Steenbergen and Jones 2002). Given these constraints, HLM then uses MLE and the data to estimate all of the parameters in Equations 1–3.

How do the approaches compare? First, if there is any relevant correlation among observations within clusters, then the OLS approach will yield biased estimates of standard errors. These standard errors will typically be too small, leading to overly small *p*-values. The intuition is that OLS, by treating every observation as independent, calculates standard errors as if there is more data than actually exists once the dependence of observations is accounted for.

Among the other two approaches, the choice is not so straightforward and will depend on both a researcher's theory and data. The first advantage of the clustered standard errors technique over HLM is that the former requires many fewer assumptions. Because HLM involves estimating all the components of the model using MLE, assumptions must be made about the distribution of all the error terms in Equations 1–3. To the extent that any of those assumptions do not hold, both point estimation and inference will suffer. Simply adjusting standard errors for clustering does not affect the point estimation of coefficients. Third, HLM is data- and computation-intensive. Thus, it will not work if there are too few clusters (Steenbergen and Jones 2002). Furthermore, due to heavy computational demands, HLM may also fail to produce output in analyses with many observations and many cross-level interactions. While clustered standard errors lose some of their excellent large-sample properties if there are too few clusters, even in such a case, they still improve inference over OLS that does not account for clustering. As Steenbergen and Jones (2002, 234), two proponents of HLM, write, "Multilevel models, then, make heavy demands on theory and data. Thus, we caution researchers against 'blindly' using these models in data analysis."

On the other hand, HLM has certain advantages over clustered standard errors. Most obviously, it is always preferable to model a process fully, as suggested by substantive theory. More specifically, HLM allows one to analyze the explanatory power of a model by estimating the variance components directly. For instance, HLM enables the researcher to state what portion of a dependent variable's variance is attributable to state-level versus individual-level variation. Of course, the price of this additional information is that more assumptions, which may be inaccurate, are necessary. All methodological choices entail such trade-offs. For example, White's method for estimating heteroskedasticity-consistent errors is the preferred means for dealing with heteroskedasticity, except when the researcher is confident in assuming the exact form that the heteroskedasticity takes (Kennedy 2003, 153). In such cases, weighted least squares estimation is more efficient. In the next section, we compare the three approaches just described (not accounting for clustering; calculating clustered standard errors; and estimating an HLM).

AN EXAMPLE: THE EFFECTS OF POST-REGISTRATION LAWS ON TURNOUT REVISITED

A question in the turnout literature is whether post-registration laws, such as the presence of polling places that are open early or close late and the information required to be given to voters before election day, increase participation at the polls. Wolfinger, Highton, and Mullin (2005) (WHM) test whether post-registration laws affect turnout in their prize-winning paper, published in this *Quarterly*. To do so, they match individual-level data from the 2000 Voter Supplement of the Current Population Survey with state-level data on relevant state institutions. Thus, they are faced with just the sort of mixed-level data we have discussed here. WHM estimate a logit model of self-reported turnout by respondents who are self-reported to be registered voters (n=44,859). Individual-level covariates in their model include controls for age, education, employment status, ethnicity, family income, race, and residential stability. Their state-level covariates include indicators for the presence of a concurrent statewide election, southern states, and battleground states in the 2000 presidential election. The independent variables of most interest to them are the policy variables, which are the state-level indicators for early voting (polls open before 7:00 am), late voting (polls open after 7:00 pm), whether information about polling places or sample ballots were mailed out to registered voters, and whether state law required that certain workers be given time off from work to vote on election day. Furthermore, several of these state-level policy variables are also interacted with relevant individual-level attributes (e.g., time off work for state employees × state employee). In what follows, we do not alter WHM's variable selection; our goal is not to critique their substantive setup, but rather to demonstrate the impact of alternative approaches to dealing with multilevel data.

First, we replicate WHM's primary empirical analysis, reporting our results in the first two columns of Table 1.[12] Notice that most of the estimated coefficients for the state-level variables are statistically significant at conventional levels.[13] This replication is our baseline, as it does not account for clustering in the data. The third column of Table 1 reports the clustered standard errors for this model, where clusters are the states. This adjustment has the expected effect of increasing the standard errors for the state-level variable coefficient estimates, which demonstrates how failing to account for intra-cluster correlation will bias standard errors downward. In fact, most of the formerly statistically significant estimates in WHM are no longer significant at even the $p < .10$ level.

We also attempted to run an HLM model with these data and in keeping with WHM's initial specification. However, despite repeated attempts using different models (a linear probability model as well as a logit model), the model failed to converge.[14] This result demonstrates that, often, a significant weakness of multilevel models is not methodological but practical. For specifications like WHM's—with many observations and many cross-level interactions and additional variance and covariance elements—modern computing is sometimes insufficient to estimate these models because the likelihood function becomes so complex.

Table 1. A Comparison of Unadjusted and Adjusted Standard Errors in Multilevel, Clustered Data

State-level Independent Variables	Coefficient	Unadjusted Standard Errors	Clustered Standard Errors
Early voting	.14	.03***	.10
Late voting	.08	.04**	.08
Mailed polling place information	.24	.12**	.22
Mailed polling place information × education	–.08	.04**	.04*
Mailed sample ballots	.29	.12**	.18
Mailed sample ballots × education	–.09	.04**	.04**
Mailed sample ballots × age 18–24 and live with parents	.01	.12	.28
Mailed sample ballots × age 18–24 and live without parents	.33	.13**	.16**
Time off work for state employees	.06	.05	.10
Time off work for state employees × state employee	–.02	.19	.16
Time off work for private employees	–.19	.05***	.07***
Time off work for private employees × private employee	.03	.06	.05
Southern state	–.19	.04***	.08**
Battleground state	.08	.03**	.07
Concurrent senatorial or gubernatorial contest	–.09	.04**	.08

*p<.10; **p<.05; *** p<.01; N=44,859

Note: Estimation technique: logit. Other independent variables in the model include age, education, ethnicity, employment status, family income, race, and residential stability. See Wolfinger, Highton, and Mullen (2005) for details.

Finally, note that even though our replication raised questions about the statistical significance of WHM's findings, it does not necessarily mean that their substantive conclusions are incorrect. Perhaps their data are simply insufficient to reveal this relationship. It could be that more years of data and changes in these laws over time might yield findings that are both substantively and statistically significant. In fact, because their number of individual cases is so large, it might seem remarkable that statistical significance was not obtained on all coefficients. However, one must remember that cluster adjustment effectively reduces the information in the data by including the correlation among observations within a cluster in the analysis. But this example at least suggests that caution is in order for policymakers considering future reforms and scholars working with this type of data. This is an important finding in its own right.

CONCLUSION

Our goals in this article were to show the effects of clustered data in analyses of state and local politics and policy and to suggest two ways of dealing with this problem. We focused on a particular type of clustering—by state in cross-sectional survey data—but the same issues we have raised apply to the analysis of other sorts of clustering. While no statistical method is without its limitations, we argue that simply adjusting standard errors for clustering in data is an easy-to-implement methodology that requires fewer assumptions than the alternative technique, hierarchical linear modeling, and that the calculation of these standard errors is not subject to the current computing limitations that HLM is. We hope that our discussion will lead to greater awareness of the clustered nature of much data used in studies of state and local politics and will cause researchers to think carefully about the considerations and trade-offs presented. The researcher will need to evaluate these tradeoffs in light of his or her own theory and data before embarking on a given analysis.

To be sure, both approaches have pitfalls, especially since both require the assumption that the clustering being accounted for is the primary source of non-independence in the data. For example, if cluster adjustment is done by state but there is also a substantial clustering within counties, inference (and, in the case of HLM, even point estimation) may be adversely affected. Therefore, neither methodology is a panacea for the problems related to multilevel data, and we urge readers to use these methods carefully and judiciously and be cognizant of both their drawbacks and their advantages.

ENDNOTES

We thank Jake Bowers, Kevin Clarke, Rob Franzese, and the anonymous reviewers for their helpful comments, and Ben Highton, Megan Mullin, and Ray Wolfinger for making their data available to us.

1. An example of sub-state-level data is local government spending. We focus on the case where the sub-state-level data is at the individual level.

2. While ecological analysis has the advantage of requiring only aggregate data, which is often readily available, its disadvantages loom large. Chief among these is that relationships holding at the aggregate level need not hold at the individual level. For this reason, researchers typically strive to acquire appropriate individual-level data for the purposes of evaluating the effects of state-level phenomena on individual outcomes. King (1997) discusses ecological inference and develops a procedure for improving ecological analyses. For a recent application of King's method, see Tolbert and Grummel's (2003) study of voting on California's Proposition 209 (regarding affirmative action).

3. Clustered standard errors are sometimes also referred to as Rogers standard errors, since Rogers (1993) implemented this technique in STATA.

4. Likewise, Abbe and Herrnson (2003) used data that are likely clustered, but contextual effects are not the focus of their study.

5. To keep our discussion focused, we do not address panel data or time-series-cross-sectional data, which admit many other approaches besides the ones we discuss, including Newey-West standard errors (Newey and West 1987) and panel-corrected standard errors (Beck and Katz 1995). Debate continues over which methodology is best for these types of data, and those issues are beyond the scope of our discussion.

6. The implementation of multilevel models is also relatively recent and rare in both political science and economics (Kedar and Shively 2005; Wooldridge 2003).

7. Furthermore, it may be the case that individual characteristics are related to the probability that a person belongs to either the treatment group (easy registration) or the control group, which might lead to issues of sample selection or endogeneity. For example, if people who are more likely to vote are also more likely either to settle in states with easy registration or to push for legislation that makes registration easy, then unadjusted estimates of the treatment effect of easy registration will be biased upward.

8. The equations in this article draw from Bryk and Raudenbush 1992, Bowers and Drake 2005, and Franzese 2005.

9. Some statistical packages, such as SAS and STATA, now include options for estimating clustered standard errors in most common estimation routines.

10. More statistical work needs to be done to understand the properties of this technique with small samples (Franzese 2005).

11. Practically, one can implement clustering by state-year in the following way: Create an index variable where every state-year combination is given a unique identifier. Suppose that states are indexed 1–50 in a variable called *state* and years run from 1990–2000 in a variable called *year*. Then one can create a unique state-year identifier, *clusterindex*, to be fed to the appropriate command, with the following formula: *clusterindex*= 10,000 *state* + *year*. For example, state 5 in year 1990 would be given the identifier 51990.

12. Our replication exercise yielded one coefficient estimate that did not match exactly

the results reported in Wolfinger, Highton, and Mullin (2005). For the interaction effect, time off work for state employees × state employee, we obtain an estimate of –.02, while WHM report –.03. We suspect this trivial difference is due to rounding, attributable either to humans or to some subtle differences in the routines used by our respective statistical programs.

13. The substantive importance of these estimates is discussed in detail in WHM.

14. The implementation of estimation procedures appropriate for mixed-level data is still a relatively new phenomenon in political science. For example, we used STATA 9.0 and implemented both the *gllamm* commands for the weighted logit and the *xtmixed* command for an unweighted linear probability model. The *xtmixed* command does not admit weights nor does it run logit. However, the linear model without weights should in fact be more likely to converge. Nevertheless, after 1,500 iterations with no movement in the log likelihood, it became clear that the model would not be successfully estimated. Using a subset of three variables from the specification, we verified that the software was working properly.

REFERENCES

Abbe, Owen G., and Paul S. Herrnson. 2003. "Campaign Professionalism in State Legislative Elections." *State Politics and Policy Quarterly* 3(3):223–45.

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(3):634–47.

Bertrand, Marianne, Esther Duflo, and Senhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics* 119(1):249–75.

Bonneau, Chris W. 2005. "What Price Justice(s)? Understanding Campaign Spending in State Supreme Court Elections." *State Politics and Policy Quarterly* 5(2):107–25.

Bowers, Jake, and Katherine W. Drake. 2005. "Applying a Two-Step Strategy to the Analysis of Cross-National Public Opinion Data." *Political Analysis* 13(4):301–26.

Branton, Regina P. 2004. "Voting in Initiative Elections: Does the Context of Racial and Ethnic Diversity Matter?" *State Politics and Policy Quarterly* 4(3):294–317.

Bryk, Stephen W., and Anthony S. Raudenbush. 1992. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks, CA: Sage Press.

Buckley, Jack, and Chad Westerland. 2004. "Duration Dependence, Functional Form, and Corrected Standard Errors: Improving EHA Models of State Policy Diffusion." *State Politics and Policy Quarterly* 4(1):94–113.

Bullock, Charles S., M.V. Hood III, and Richard Clark. 2005. "Punch Cards, Jim Crow, and Al Gore: Explaining Voter Trust in the Electoral System in Georgia, 2000." *State Politics and Policy Quarterly* 5(3):283–94.

Carson, Jamie L., and Michael H. Crespin. 2004. "The Effect of State Redistricting Methods on Electoral Competition in United States House of Representatives Races." *State Politics and Policy Quarterly* 4(4):455–69.

Francia, Peter L., and Paul S. Herrnson. 2004. "The Synergistic Effect of Campaign Effort and Election Reform on Voter Turnout in State Legislative Elections." *State Politics and Policy Quarterly* 4(1):74–91.

Franzese, Robert J., Jr. 2005. "Empirical Strategies for Various Manifestations of Multilevel Data." *Political Analysis* 13(4):430–46.

Froot, Kenneth A. 1989. "Consistent Covariance Matrix Estimation with Cross-Sectional Dependence and Heteroskedasticity in Financial Data." *Journal of Financial and Quantitative Analysis* 24(3):333–55.

Hansen, Christian. 2005. "Asymptotic Properties of a Robust Variance Matrix Estimator for Panel Data When T is Large." Working paper, University of Chicago.

Huber, Peter J. 1967. "The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions." In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley, CA: University of California Press.

Kedar, Orit, and W. Phillips Shively. 2005. "Introduction to the Special Issue." *Political Analysis* 13(4):297–300.

Kennedy, Peter. 2003. *A Guide to Econometrics*, 5th ed. Cambridge, MA: MIT Press.

Kezdi, Gabor. 2003. "Robust Standard Error Estimation in Fixed-Effects Panel Models." Working paper, Central European University.

King, Gary. 1997. *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data.* Princeton, NJ: Princeton University Press.

Lacy, Robert. 2005. "The Electoral Allure of Direct Democracy: The Effect of Initiative Salience on Voting, 1990–96." *State Politics and Policy Quarterly* 5(2):168–81.

Moulton, Brent R. 1990. "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables in Micro Units." *Review of Economics and Statistics* 72(2):334–38.

Newey, Whitney K., and Kenneth D. West. 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55(3):703–8.

Raudenbush, Stephen W., and Anthony S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*. 2nd ed. Thousand Oaks, CA: Sage Press.

Rogers, William. 1993. "sg17: Regression Standard Errors in Clustered Samples." *Stata Technical Bulletin* 13:19–23.

Steenbergen, Marco R., and Bradford S. Jones. 2002. "Modeling Multilevel Data Structures." *American Journal of Political Science* 46(1):218–37.

Tolbert, Caroline J., and John A. Grummel. 2003. "Revisiting the Racial Threat Hypothesis: White Voter Support for California's Proposition 209." *State Politics and Policy Quarterly* 3(2):183–202.

Tolbert, Caroline J., Ramona S. McNeal, and Daniel A. Smith. 2003. "Enhancing Civic Engagement: The Effect of Direct Democracy on Political Participation and Knowledge." *State Politics and Policy Quarterly* 3(1):23–41.

White, Halbert. 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica* 48(4):817–38.

Williams, Rick L. 2000. "A Note on Robust Variance Estimation for Cluster-Correlated Data." *Biometrics* 56:645–6.

Wolfinger, Raymond E., Benjamin Highton, and Megan Mullin. 2005. "How Postregistration Laws Affect the Turnout of Citizens Registered to Vote." *State Politics and Policy Quarterly* 5(1):1–23.

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA: MIT Press.

Wooldridge, Jeffrey M. 2003. "Cluster-Sample Methods in Applied Econometrics." *American Economic Review* 93(2):133–8.