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

February 2006

Abstract

Researchers often seek to understand the effects of state policies or institutions on individual behavior or other outcomes in sub-state-level observational units (e.g., election results in state legislative districts). However, standard estimation methods applied to such models do not properly account for the clustering of observations within states and may lead researchers to overstate the statistical significance of state-level factors. We discuss the theory behind two approaches to dealing with clustering—clustered standard errors and multilevel modeling. We then 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). While we view clustered standard errors as a more straightforward, feasible approach, especially when working with large datasets or many cross-level interactions, our purpose in this Practical Researcher piece is to draw attention to the issue of clustering in state and local politics research.

Researchers are often concerned with estimating the effects of state-level policies or institutions on individual or other sub-state-level outcomes.[1] Such empirical analyses may be undertaken using either state-level data, 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 contextual (a.k.a. hierarchical or mixed-level) analyses. While ecological analysis has the advantage of requiring only aggregate data, which is often readily available, its disadvantages loom large.[2] 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. However, contextual analyses also are fraught with potential pitfalls. For example, in standard regression techniques state-level variables often achieve exaggerated levels of statistical significance (Moulton 1990).

This paper provides a clear illustration of both the ease and importance of correcting standard errors for clustered observations in contextual analyses of state politics and policy. These standard errors, known as clustered (or cluster-adjusted) standard errors, adjust for the fact that observations within each state are unlikely to be independent, thereby violating a core assumption of many estimation procedures (including OLS).[3] We accomplish this by replicating and extending the analysis in a recent article by Wolfinger, Highton, and Mullin (2005) and contrasting this correction with the modeling of contextual relationships via multilevel modeling.

---

[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] 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 referred to as Rogers standard errors, since Rogers (1993) implemented this technique in STATA.

We selected this study in part because it is particularly interesting and prominent, but also because the authors were most gracious in making available to us all of their data for the purpose of this exercise. The authors (hereafter, WHM) employ individual-level survey data to estimate the effects of state post-registration laws on voter turnout. In doing so, their study improves upon previous ecological analyses of such laws (and so deservedly won the State Politics and Policy Quarterly Best Paper Award for 2002). However, in their empirical analysis, WHM do not make any adjustments for the clustering of individual survey respondents within states. This may cause them to understate the standard errors associated with their estimates of the effects of state post-registration laws. As we demonstrate, this is in fact the case.

By no means are WHM alone in ignoring the potential problems associated with grouped data. We examined every issue of State Politics and Policy Quarterly and found a variety of approaches for dealing (or not dealing) with clustered observations in mixed-level studies. Most studies simply do not address the issue. These include: Tolbert, McNeal, and Smith (2003), who investigate the effects of state ballot initiatives on individual-level voter turnout and political knowledge; Francia and Herrnson (2004), who examine the synergistic effects of campaign effort and electoral reforms on turnout in legislative districts; Carson and Crespin (2004), who study the effects of state redistricting methods on competition in Congressional races; Lacy (2005), who analyzes the effects of salient state ballot initiatives on voter turnout; and Bullock, Hood, and Clark (2005), who estimate the contextual effects of county voting equipment and spoiled ballot rates on individuals' trust in the voting process.[4]

In contrast, Branton (2004), who employs exit poll data to estimate the effects of effects of racial and ethnic diversity within districts on vote choice by whites, does adjust standard

---

[4] Likewise, Abbe and Herrnson (2003) do not discuss clustering, although contextual effects are not the focus of their study.

errors for the clustering of individual exit poll respondents within districts. Buckley and

Westerland (2004), in a Practical Researcher piece addressing discrete event history analysis,

also examine the impact of clustering in their work. Finally, Bonneau (2005) does not present

estimates that have been corrected for clustering, but does note that his key findings on the

determinants of campaign spending in state judicial elections are not sensitive to correcting for

clustering of judicial elections within states.

The frequency with which state politics and policy researchers employ mixed-level data,

and the infrequency of accounting for grouped observations, suggests the need for a practical

discussion of the importance of accounting for clustering. In what follows, we first provide an

intuitive as well as analytical explanation of the justification for cluster-adjusted standard errors

and contrast it with another popular method, hierarchical linear modeling. Next, we describe the

WHM study and present our replication results. Finally, we illustrate the importance of cluster-

adjustment for understanding the effects of state post-registration laws on turnout of citizens

registered to vote. Our goal is to encourage scholars in the state politics arena to more

consistently take account of clustering issues using appropriate techniques.

To keep the paper focused, we do not address panel data or time-series-cross-sectional

data, which admit many other approaches besides the ones discussed in this paper, including

Newey-West standard errors (Newey and West 1987) and panel-corrected (aka Beck-Katz)

standard errors (Beck and Katz 1995). Debate continues over which methodology is best for

these types of data, and these issues are beyond the scope of our paper.

WHAT IS CLUSTERING?

Simply put, the attributes of states in which individuals reside do not vary across individuals

within a particular state. Hence, such clustered, or grouped, observations violate the assumption

of independence implied 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, Wooldridge 2002, and

Wooldridge 2003), but such solutions are far from universally implemented, as our review of

SPPQ articles has demonstrated.[5]

Before discussing how to address clustering, we demonstrate why it is an issue. 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

group or the treatment group. Random assignment prevents problems of selection bias and

endogeneity that might otherwise occur if individuals were to self-select into one group or

another. For example, a test subject in a well-designed study evaluating the efficacy of a new

medicine will be selected for the control or treatment group independently of his or her

characteristics. In other words, a test subject may or may not have a college degree, but this

characteristic will not influence whether that subject is given a placebo or the investigational

drug. This makes the estimated impact of the drug much more reliable.

Now contrast this setup with a situation in which a researcher employs 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 an "easy registration" law, so that the

treatment effect is estimated by contrasting individual turnout in states with and without such

laws.[6] It is now the case that every individual that shares the characteristic of "residing in a

---

[5] The implementation of multilevel models in both political science and economics is still a relatively new phenomenon. In a recent special issue of *Political Analysis* devoted to multilevel modeling, the editors note, "The use of multilevel models…has blossomed recently in political science" (Kedar and Shively 2005). And as Wooldridge writes, in discussing the state of economics, ``While adjusting for clustering is much more common than it was 10 years ago, inference methods robust to cluster correlation are not used routinely across all relevant settings" (2003, 133).

[6] As an aside, in this example, it is no longer necessarily true that all individual characteristics are unrelated to the

group called 'North Dakota'" will be in the treatment group (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. This is in contrast to other types of groupings we might imagine. For example, some college-educated subjects will reside in easy registration states, and others will not. For this reason we are less concerned about the grouping of individuals by education level as a violation of independence. However, the grouping of observations at the state level is a violation of the assumption that observations are independent.

Despite this concern, we should not lose sight of the fact that mixed-level analyses have the advantage of allowing the researcher to more confidently describe compositional (e.g., income level) versus contextual (e.g., registration laws) determinants of voter turnout. In addition, because such studies may include thousands of individual observations, the researcher is more apt to find statistically significant estimates, especially for any state-level variables. In general, the increased precision of estimates resulting from additional data is desirable, but in this case such results may be misleading with regard to the statistical significance of state-level factors on turnout.

The key problem with this hypothetical study is that in effect there are still only 50 independent state-level observations. In other words, there is less information in this data than if the individual-level observations were truly independent, rather than systematically placed into either the treatment or control group according to the individual's state of residence. The failure to account for clustering may cause the researcher to greatly understate the standard errors on the estimated coefficients for the state-level variables (Moulton 1990).

---

probability that individual belongs to either the treatment group (easy registration) or the control group, for the simple reason that people are not randomly placed in states. This may lead to issues of sample selection or endogeneity; for example, if people who have an unobserved tendency to vote are more likely to either 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.

To illustrate, suppose that you are interested in estimating the effects of both state- and individual-level determinants on an individual-level outcome. For ease of exposition, assume that there are only two independent variables, one of each type (individual-level and state-level), and only a single year's cross-section of observations. The logic of the following argument applies to situations both with more variables or 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 $i = 1, \ldots, n_j$ is the number of observations in cluster j. $N = Jn_j$ if the number of observations per cluster is the same, which we will assume here. Further assume that the state-level (or formally, Level 2) variable, w, influences the impact of the individual-level (Level 1) variable, x, on behavior. Then we can write the following equations:[7]

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + \varepsilon_{ij} \qquad (1)$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}w_j + u_{0j} \qquad (2)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}w_j + u_{1j} \qquad (3)$$

We can substitute and generate one equation:

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

Note that (4) looks like a standard regression equation, but with a compound error term. There are typically three ways of estimating the coefficients of interest in this model. The first is to run ordinary least squares on (4) and ignore the clustered nature of the data, treating the compound error term as if it consisted of just the ε_{ij} term. The second is to implement clustered standard errors. The third is to explicitly model the multilevel nature of the data. We will consider each in turn.

---

[7] The equations in this paper draw from Bryk and Raudenbush (1992) and Franzese (2005).

To see why doing nothing is usually the worst option, note that a standard assumption of OLS for conducting inference is that $\varepsilon_{ij} \sim N(0, \sigma^2)$, where the errors are also distributed independently and identically (i.i.d.). However, to the extent that observations within a state share some common (often unmeasurable) characteristics, the i.i.d. assumption is clearly violated, standard errors are no longer accurate, and the OLS estimator is no longer the best-linear-unbiased-estimator (BLUE). In other words, even if $\varepsilon_{ij} \sim N(0, \sigma^2)$, the compound error term violates the assumptions of OLS unless $u_{0j} = u_{1j} = 0$.

So what is the researcher to do? Cluster-adjustment assumes that observations within groups are correlated but that observations across clusters are independent. These standard errors are a variant on what are often termed 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 a general form of heteroskedasticity as well as for any intra-cluster correlation, and are therefore termed cluster-consistent standard errors.[8]

We distinguish among these standard errors by referencing the variance-covariance matrix for OLS. The typical variance-covariance matrix for OLS is $\mathrm{E}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon\varepsilon'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}]$. Since by assumption $\mathrm{E}[\hat{\varepsilon}\hat{\varepsilon}'] = \sigma^2\mathbf{I}$, where σ^2 is the average of the squared residuals, we obtain a matrix consisting of $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$. The standard errors for each beta are calculated from the diagonals of this variance-covariance matrix. When heteroskedasticity is present, the variance of the error term is not constant across observations, so $\mathrm{E}[\hat{\varepsilon}\hat{\varepsilon}']$ is no longer $\sigma^2\mathbf{I}$. The $n \times n$ matrix

---

[8] Some statistical packages, such as SAS and STATA, now include user-friendly options for estimating cluster-adjusted standard errors in most common estimation routines.

E[$\hat{\varepsilon}\hat{\varepsilon}'$] consists of the squared errors of the individual-specific errors terms along the diagonal and the covariances of error terms on the off-diagonals.

White (1980) developed an estimator that addressed non-constant variance by assuming that the off-diagonal terms of this *n x n* matrix were still 0—i.e., errors were not correlated across observations—but allowed the diagonals to vary. Because fully estimating different variances would be impossible—we would have to estimate *n* variances from *n* observations—White came up with a solution of focusing just on the averages of expectations. Cluster-adjustment takes this correction one step further by allowing off-diagonal elements from the same cluster to be non-zero. By doing this, one allows for any arbitrary correlation of the observations within clusters, any arbitrary heteroskedasticity in the error term, but no correlation across clusters. If the standard variance-covariance matrix is written as $\sigma^2(\mathbf{X'X})^{-1}$, we can write the cluster-adjusted

matrix as $(\mathbf{X'X})^{-1}\sum_{j=1}^{J}\left\{\left(\sum_{i=1}^{nj}\hat{\varepsilon}_i\mathbf{x_i}\right)\left(\sum_{i=1}^{nj}\hat{\varepsilon}_i\mathbf{x_i}\right)'\right\}(\mathbf{X'X})^{-1}$, where $\mathbf{x_i}$ refers to the observations within a

given cluster. If J=N and n_j=1, then we obtain White's estimator. Clustered standard errors are calculated by taking the square root of the appropriate term of the matrix (for instance, element (1,1) for the standard error of the constant) and applying a finite sample adjustment (J/(J-1))*((N-1)/(N-k)), where J is the number of clusters, N is the total number of observations, n is the number of observations per cluster, and k is the number of regressors.

Cluster-adjustment works well for a variety of estimation procedures, including logit and probit, provided that the number of clusters is large. This begs the question "how large"? Fortunately, Monte Carlo simulations (e.g., Kezdi 2003, Bertrand et al. 2004, and Hansen 2005) suggest that 50 clusters are more than sufficient.[9] For example, when working with individual-

---

[9] More work remains to be done on the properties of this technique with small samples (Franzese 2005).

level data from a single year and across 50 states, one adjusts for clustering at the state level. When working with individual-level data from the 50 states measured over several years, it is possible to cluster by either state or (state x year). The latter is appropriate if we believe that observations are not independent within a state-year, but are independent across years within a state (e.g., vote choice in a state pre-and-post party realignment). If we think that observations from the same state but different years are unlikely to be independent, then clustering should take place by state only. In short, one should cluster on the macro-level features that the researcher believes is causing dependence across observations.

One popular alternative to clustered standard errors is to use hierarchical linear modeling (HLM) or related multilevel modeling procedures which explicitly model the error terms in equations (1) – (3).[10] HLM therefore allows the researcher to estimate how much each level is contributing to explanation in the model. In other words, is the explanation primarily a macro-level one or an individual-level one? HLM estimation, in a nutshell, uses equations (1) - (3), along with the assumptions below, to find coefficient, variance, and covariance estimates such that the likelihood of observing the actual values of y is at a maximum (Bryk and Raudenbush 1992, 45). This typically requires the tools of maximum likelihood estimation, which means that misspecification of the error term propagates throughout the entire estimation, including the estimation of coefficients.[11]

---

[10] For an introduction to HLM, see Steenbergen and Jones (2002). For a discussion of other multilevel models, see the Autumn 2005 issue of *Political Analysis*, especially Franzese (2005) and Bowers and Drake (2005). The seminal work in this area is Bryk and Raudenbush (1992), since updated as Raudenbush and Bryk (2002).
[11] Granted, this is also the case with a logit or a probit, also estimated via MLE. However, cluster-adjusted standard errors, which are calculated after estimation, do not add the additional complexity that estimating variance components does.

HLM typically requires the following assumptions:

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

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

$$\text{Var}(u_{0j}) = \tau_{00} \qquad (A3)$$

$$\text{Var}(u_{1j}) = \tau_{11} \qquad (A4)$$

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

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

In general, HLM methods estimate all of the variance and covariance elements.

So what are the trade-offs between the two approaches? The advantage of clustered standard errors is that they require many fewer assumptions to be made as part of the estimation. Because HLM is estimating the components of the error term, by necessity assumptions must be made about the distribution of those error terms for the purposes of estimation. To the extent that those assumptions are inaccurate, those inaccuracies will influence the entire analysis. In contrast, cluster-adjusted standard errors do not impact coefficient estimates. Second, measurement error looms large in HLM models (Steenbergen and Jones 2002). Third, HLM is data- and computationally-intensive; it often will not run if there are too few clusters (Steenbergen and Jones 2002). As we see below, the HLM model also may fail to produce output in analyses with many observations and many cross-level interactions. While clustered standard errors lose some of their large-sample properties if the number of clusters is too small, they still provide an advantage over an estimation that does not account for clustering. As Steenbergen and Jones, 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" (2002, 234).

10

On the other hand, HLM has some advantages. HLM allows one to make claims about what is driving explanation in a model by estimating the variance components directly. For instance, HLM enables the researcher to state what portion of the variance in a dependent variable is attributable to state-level versus individual-level variation. The price of this information is a model with more (possibly inaccurate) assumptions. Like all methodological choices, there are trade-offs.[12] For the present context we ultimately fall on the side of clustered standard errors, because they require fewer assumptions and are feasible for large datasets and many cross-level interactions. Still, we can imagine many scenarios where HLM would be the preferred method.

To be sure, both approaches have pitfalls, in that they both assume that the clustering being accounted for is the primary source of non-independence in the observations. For example, if cluster-adjustment is done by state but we also believe that observations within counties are also clustered, inference may be affected. In other words, neither methodology is a panacea, and we urge readers to use these methods carefully and judiciously, being aware of their drawbacks as well as their advantages. In the next section, we compare the three approaches just described (not accounting for clustering; running clustered standard errors; and running HLM).

THE EFFECTS OF POST-REGISTRATION LAWS ON TURNOUT REVISITED

Wolfinger, Highton, and Mullin (2005) examine the effects of state post-registration laws on the turnout of registered voters. They employ individual-level data from the 2000 Voter

---

[12] An analogy exists between heteroskedasticity-consistent estimation methods and cluster-adjustment. In general, White's method for estimating heteroskedasticity–consistent errors is the preferred means for dealing with heteroskedasticity, except when the researcher is confident in the exact form that the heteroskedasticity takes—in which case, an appropriate weighted estimation will be more efficient.

Supplement of the Current Population Survey and match it to state-level data on relevant state institutions. WHM estimate a logit model of self-reported turnout among those respondents that also self-reported that they are registered to vote (n=44,859). Individual covariates include controls for age, education, employment status, ethnicity, family income, race, and residential stability. State-level covariates include indicators for the presence of a concurrent statewide election, southern states, and battleground states in the 2000 presidential election. The key independent variables of interest are state-level indicators for early voting (polls open before 7:00am), late voting (polls open after 7:00pm), whether information about polling places or sample ballots were mailed out to registered voters, and whether state law requires that workers be given time off to vote on election day. Several of these state-level independent variables are also interacted with relevant individual attributes (e.g., "time off work for state employees X state employee"). In what follows we do not alter WHM's variable selection in any way; our goal here is not to critique their substantive setup, but rather to demonstrate the impact of alternative methodological techniques, given a particular selection of variables.

We were able to replicate WHM's primary empirical analysis almost exactly, and we report our results in column (1) of Table 1.[13] Notice that most of the state-level coefficients have a statistically significant effect on turnout.[14] This replication is our baseline, as it does not account for clustering in the data. Having successfully replicated the key findings in WHM, we now turn to the issue of adjusting standard errors for clustering. In column (2) of Table 2, we report the results of a logit estimation with clustered standard errors, where clusters correspond to states. This has the expected effect of "blowing up" the standard errors on the state-level

---

[13] Our replication exercise yielded only one coefficient estimate that didn't match exactly the results reported by WHM. For "time off work for state employees X state employee" we get an estimate of -.02, while WHM report -.03. We suspect this trivial difference is due to rounding (either attributable to humans or to some subtle differences in the routines used by our respective statistical programs).

[14] The substantive importance of these estimates is discussed in detail in WHM.

variable coefficient estimates. In fact, most of the formerly statistically significant estimates in WHM are no longer significant at even the $p < .10$ level.

We attempted to also run a multilevel model in keeping with WHM's initial specification. However, despite repeated attempts using different models (a linear probability model as well as a logit), the model failed to converge.[15] We verified, using a set of 3 variables from the specification, that the software was working properly. Ultimately, a significant weakness of multilevel models is not methodological but practical: for specifications like WHM's—with many observations and many cross-level interactions—modern computing is sometimes insufficient to estimate these models.

Finally, note that even though our replication raised questions about the statistical significance of the findings, inquiry should not stop there. It may be that collecting more years of data and analyzing changes in laws over time may lead to findings that are both substantively and statistically significant. In fact, because N is so large, one might find it remarkable that statistical significance was not obtained on all coefficients. However, one must keep in mind that the cluster-adjustment effectively reduces N by noting the correlation among observations within a cluster. For now, however, the revised results suggest caution is in order for policymakers considering future reforms. This is an important finding in its own right.

CONCLUSION

The goal of this paper is to encourage greater attention to the clustered nature of data present in many studies of state and local politics. We focused here on a particular type of

---

[15] We used Stata 9.0 and implemented both the gllamm commands for the weighted logit, as well as 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 easier to run. Nevertheless, after 1500 iterations with no movement in the log likelihood, it became clear that the model would not be successfully estimated.

clustering—by state or state-year in non-panel survey data—but the same issues extend to other

sorts of analyses. While no statistical method is without drawbacks, we argue that using

standard errors that account for clustering in data is an easy-to-implement methodology that

requires fewer assumptions than an alternative technique, hierarchical linear modeling, and that

the calculation of these standard errors is not subject to current computing limitations, as HLM

is. The researcher, of course, will want to evaluate these tradeoffs before embarking on a given

data analysis. This paper, then, is not advocating the blind adoption of clustered standard errors.

Rather, we hope that our discussion will lead to greater attention to the clustered nature of data in

studies of state and local politics and that the technique ultimately selected reflects a

consideration of the trade-offs we discussed here.

References

Abbe, Owen G., and Paul S. Herrnson. 2003. "Campaign Professionalism in State Legislative
    Elections." *State Politics and Policy Quarterly* 3(3):223-245.

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):634-647.

Bertrand, Marianne, Esther Duflo, and Senhil Mullainathan. 2004. "How Much Should We
    Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics*
    119(1):249-275.

Bonneau, Chris W. 2005. "What Price Justice(s)? Understanding Campaign Spending in State
    Supreme Court Elections." *State Politics and Policy Quarterly* 5(2):107-125.

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):301-326.

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.

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):283-294.

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):455-469.

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):430-446.

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):333-355.

Hansen, Christian. 2005. "Asymptotic Properties of a Robust Variance Matrix Estimator For Panel Data When T is Large." Working Paper.

Huber, Peter J. 1967. "The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions." *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* 1. Berkeley: University of California Press, 221-233.

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

Kezdi, Gabor. 2003. "Robust Standard Error Estimation in Fixed-Effects Panel Models." Working Paper.

King, Gary. 1997. *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data.* Princeton: 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):168-181.

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):334-338.

Newey, Whitney K., and Kenneth D. West. 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55(3):703-708.

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

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):218-237.

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):817-838.

Williams, Rick L. 2000. "A Note on Robust Variance Estimation for Cluster-Correlated Data." *Biometrics* 56:645-646.

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: MIT Press.

Wooldridge, Jeffrey M. 2003. "Cluster-Sample Methods in Applied Econometrics." *American Economic Review* 93(2):133-138.

Table 1. Effect of Robust and Clustered Standard Errors on Logit Estimates

State-level independent variables	Our Replication	Clustered Standard Errors
Early voting	.14**	.14
	(.03)	(.10)
Late voting	.08**	.08
	(.04)	(.08)
Mailed polling place information	.24**	.24
	(.12)	(.22)
Mailed polling place information X Education	-.08**	-.08*
	(.04)	(.04)
Mailed sample ballots	.29**	.29
	(.12)	(.18)
Mailed sample ballots X Education	-.09**	-.09**
	(.04)	(.04)
Mailed sample ballots X Age 18-24 and live with parents	.01	.01
	(.12)	(.28)
Mailed sample ballots X Age 18-24 and live without parents	.33**	.33**
	(.13)	(.16)
Time off work for state employees	.06	.06
	(.05)	(.10)
Time off work for state employees X State employee	-.02	-.02
	(.19)	(.16)
Time off work for private employees	-.19**	-.19**
	(.05)	(.07)
Time off work for private employees X Private employee	.03	.03
	(.06)	(.05)
Southern state	-.19**	-.19**
	(.04)	(.08)
Battleground state	.08**	.08
	(.03)	(.07)
Concurrent senatorial or gubernatorial contest	-.09**	-.09
	(.04)	(.08)

Notes: *p<.10, **p<.05; N=44,859. Other independent variables include age, education, ethnicity, employment status, family income, race, residential stability, and southern region; see WHM for details.