Modeling Heterogeneity in Pooled Event History Analysis

Rebecca J. Kreitzer[1] and Frederick J. Boehmke[2]

Abstract

Pooled event history analysis (PEHA) allows researchers to study the effects of variables across multiple policies by stacking the data and estimating the parameters in a single model. Yet this approach to modeling policy diffusion implies assumptions about homogeneity that are often violated in reality, such that the effect of a given variable is constant across policies. We relax this assumption and use Monte Carlo simulations to compare common strategies for modeling heterogeneity, testing these strategies with increasing levels of variance. We find that multilevel models with random coefficients produce the best estimates and are a significant improvement over other models. In addition, we show how modeling similar policies as multilevel structures allows researchers to more precisely explore the theoretical implications of heterogeneity across policies. We provide an empirical example of these modeling approaches with a unique data set of 29 antiabortion policies.

Introduction

Since Berry and Berry's (1990) influential study of state lottery adoptions, state politics researchers have built a significant body of knowledge about the adoption and diffusion of state policies through event history analysis (EHA). The single policy EHA has served as the workhorse of policy diffusion studies within the American

[1]University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[2]University of Iowa, Iowa City, IA, USA

Corresponding Author:
Rebecca J. Kreitzer, University of North Carolina at Chapel Hill, Public Policy Department, Campus Box #3435, UNC-Chapel Hill, NC 27599, USA.
Email: rkreit@email.unc.edu

states or cities as well as across countries. Recently, though, scholars have shifted focus from the study of specific policies in isolation to the study of policy diffusion more broadly by simultaneously looking at the diffusion of many policies or even a single policy with multiple components. Pooled event history analysis (hereafter PEHA) allows researchers to study the effects of variables across multiple policies or components by stacking the data and estimating the parameters in a single model (Boehmke 2009; Makse and Volden 2011; Shipan and Volden 2006). In addition to being more parsimonious than separate estimations of different policies, this technique allows the researcher to evaluate the commonalities between different policies and to leverage information about policies with few adoptions.

With this approach, the level of analysis becomes the policy-state-year, and the researcher estimates a single set of parameters that represent the average effects of the covariates on policy adoption across the range of policies. This has the advantage of allowing researchers to test theories of policy diffusion by focusing on consistent patterns of diffusion over many policies rather than informally aggregating results one policy at a time. At the same time, pooling many policies involves a number of assumptions about homogeneity that might not always hold true. This homogeneity may result in biased estimates of either coefficients or standard errors, which would undermine the value of a parsimonious model. Thus, we need a better sense of how competing models perform with assumptions that are often violated in reality.

In this article, we investigate the issue of heterogeneity in PEHA. If the benefits of pooling policies come at too great a cost, then only in rare circumstances with homogeneous policies will researchers find value in PEHA. Given what we know about the diffusion of policies across the American states or across countries, in fact, complete homogeneity will tend to be a rare case. We therefore explore the ability of various estimators to capture heterogeneity across policies. These approaches include standard fixes in the literature such as clustering, fixed, or random effects. These approaches have typically only been applied to address differences in the mean rate across units, here policies, whereas in the context of policy diffusion data, we strongly suspect strong differences in effects across variables as well. To address this, we suggest the application of multilevel modeling to the PEHA context, treating policies as the second-level grouping and state-years as the units within those groups. Viewing PEHA in this fashion facilitates the inclusion not just of random effects, but also random coefficients to capture heterogeneity in variables' effects across policies. This variation is not simply a statistical nuisance, it may be substantively interesting to researchers as well. For example, some policies within a diverse policy area may be driven by certain factors more than other policies in a theoretically meaningful way. We demonstrate how to recover and substantively illustrate this heterogeneity and discuss various estimation issues for applied researchers.

Using Monte Carlo, we evaluate the ability of these various estimation approaches to capture or correct for heterogeneity across policies. We find that multilevel models with random effects and random coefficients produce coefficient estimates closest to the true value and also have the smallest difference between standard errors of the estimate and the sample standard deviation. We then illustrate the findings of our

Monte Carlo simulations with an application using the case of abortion policy and demonstrate how to test for heterogeneous variable effects as well as how to substantively recover and interpret policy-specific effects for such variables.

PEHA

EHA is the standard approach to modeling policy diffusion because it allows scholars to simultaneously account for internal and external determinants of policy adoption. There is a vast body of research on state policy adoption using EHA on policies such as state lottery adoption (F. S. Berry and Berry 1990), welfare policy (Harris 1993), hate crime legislation (Soule and Earl 2001), health maintenance organization (HMO) reform (Balla 2001), education reform (Mintrom 1997), and smoking bans (Pacheco 2012) that determine which factors contribute to state adoption of specific policies. The ideology of previous adopters (Grossback, Nicholson-Crotty, and Peterson 2004), the perceived success of the policy (Volden 2006), geographic proximity (F. S. Berry and Berry 1990; Walker 1969), public opinion (Pacheco 2012), and policy entrepreneurs (Mintrom 1997) contribute important explanatory variables to the diffusion process. EHA has also been employed widely in both comparative politics and international relations to study the diffusion of policies, democratization, or treaties across countries (Elkins, Guzman, and Simmons 2006; Simmons, Dobbins, and Garrett 2008).

Although the use of EHA to study a specific policy has developed a rich literature with many useful insights into state policy making, there may be "diminishing marginal returns" in continuing this approach (Boehmke 2009, 229). The policy-specific EHA approach emphasizes the unique determinants of a specific policy instead of engaging in a broader discussion of the determinants of policy adoption and diffusion. A scholar looking at only a single policy may find a variable to be significant that is not a significant predictor of other similar policies. Including many policies in a single analysis helps to avoid a situation in which conclusions are drawn from an anomalous policy. This approach to modeling policy diffusion has been used by political scientists studying a variety of questions such as how policy attributes such as complexity and relative advantage influence the diffusion of criminal justice policies (Makse and Volden 2011), how antismoking policies diffuse from cities to states (Shipan and Volden 2006), and the innovativeness of states (Boehmke and Skinner 2012).

This approach to modeling policy diffusion is not without its critiques. Pooling policies almost certainly violates assumptions of homogeneity, and concerns about the violation of this assumption are exacerbated when heterogeneous policies from multiple issue areas are pooled together (as in, for example, Boehmke and Skinner 2012). In this scenario, a variable may have a positive effect for some policies and a negative effect for others (say, Democratic control of the state legislature). The resulting analysis may be biased by the construction of the data set, such as if there are more policies favored by Democrats or Republicans. For this reason, it may be less worrisome to include only policies in a single issue area or policies that have a similar set of predictors, though our results suggest that the method can still be applied to broad sets of policies. In either case, heterogeneous effects can be modeled using the strategies presented in this article.

PEHA builds on the standard EHA method, in which a dichotomous dependent variable is coded as a 1 for the time period in which a policy is adopted; 0 when it is in the risk set of states that have not adopted the specific policy; and missing for years after the policy has been enacted. Specifically, Y_{it} is a indicator variable for whether state i adopts a policy in year t. In a PEHA model, the unit of analysis is the state-policy-year, and the dependent variable includes a subscript for the specific policy k, so that Y_{ikt} is an indicator variable for whether state i adopts a policy k in year t. States remain at risk for adoption of each policy until it is enacted. As such, Y_{ikt} is coded as 0 when the state is at risk to enact policy k but does not adopt it, coded as 1 in the year of adoption, and coded as "missing" for subsequent years. In a standard EHA model, a state is no longer in the data set once it has adopted a policy. Here, a given state remains in the data set until all policies have been adopted by that state.

In this article, we test the implications of assumptions about homogeneity commonly made in PEHA. We focus on the consequences of violating the assumption of coefficient homogeneity across policies on our ability to recover accurate estimates of both mean effects and associated standard errors. We also offer guidance to researchers on specifying and estimating a multilevel PEHA model. Finally, we use an application of abortion policies to demonstrate how multilevel PEHA models contribute to our understanding of the nuances of this policy area.

Heterogeneity in PEHA

Before discussing various estimation alternatives, we begin by writing out the data generating process for a PEHA. This allows us to identify the specific forms of heterogeneity that we wish to account for during estimation. With PEHA, each observation represents a state-year-policy adoption opportunity. If we think of policies as defining the groups, then we will have state-year observations as units within those groups. Because independent variables may vary across states, time, policies, or some combination of these, we label observations by all three pieces with i representing states, k representing policies, and t representing time. This setup has parallels to approaches for repeated events in continuous time duration data in that we pool multiple failures per observation and then account for possible heterogeneity across failures, but rather than treat the failure order as defining the groupings we rely on an exogenous, ex ante identifier of policy type.[1] With policies as the group variable, a discrete PEHA estimated as a multilevel model allows both the intercept and coefficients to possibly vary across policies according to a specified distribution, such as a multivariate normal. A typical PEHA data generating process can thus be represented as follows:

$$Y_{ikt} = \begin{cases} 0 \text{ if } Y_{ikt}^* \leq 0 \\ 1 \text{ if } Y_{ikt}^* > 0, \end{cases} \tag{1}$$

$$Y_{ikt}^* = \beta_{0j} + \beta_{1k} X_{it} + \beta_{2k} X_{ikt} + \beta_{3k} X_{kt} + \varepsilon_{ikt}, \tag{2}$$

$$\beta_{0j} = \beta_{00} + \beta_{01}Z_j + u_{0j}, \tag{3}$$

$$\beta_{1k} = \beta_{10} + u_{1k}, \tag{4}$$

$$\beta_{2k} = \beta_{20} + u_{2k}, \tag{5}$$

$$\beta_{3k} = \beta_{30} + u_{3k}. \tag{6}$$

Here, we follow a standard latent variable setup for a binary outcome and allow for heterogeneity across policies in multiple ways. β_{0j} captures the constant differences in adoption rates across policies by allowing for an overall intercept as well as policy-specific differences in that intercept, some of which might be explained by exogenous variables Z_k measured at the policy level and some of which are not (i.e., u_{0j}). We also include sets of independent variables that capture, respectively, factors measured at the state-time level, the state-policy-time level, and the policy-time level. The first group corresponds to the typical list of internal state factors that would include in single policy EHA models such as ideology or population that capture the tendency for some states to adopt policies sooner than others (this might also include time-invariant differences across states); the second captures the differences in how states view policies over time, such as policy-specific diffusion effects or measures of demand or need for the policy that evolve over time; and the last captures the way policies evolve over time, which could include Federal interventions or policy-specific duration dependence, for example. Within each set of effects, then, we allow for heterogeneity across policies in equations 4 to 6 with u_{1k} through u_{3k}. This allows the effect of any or all of the X variables to differ across policies. For example, state ideology could have a positive effect for some policies and a negative effect for other policies.

These sets of categories are not exhaustive, but should capture a wide variety of effects of interest to diffusion scholars while allowing for heterogeneity in those effects. The specifics of when and where to specify and allow for heterogeneity will depend on the data and theoretical questions being addressed, though we offer some guidelines shortly. Furthermore, while we have not explicitly done so here, scholars might also be interested in specifying cross-level interactions so that, for example, the effects of a state characteristic depend on characteristics of the policy. This can be done with a straightforward extension of the specification above by allowing the coefficient for one or more state characteristics to depend on characteristics of the policy: e.g., $\beta_{1k} = \beta_{10} + \beta_{11}Z_j + u_{1k}$.

We now turn to discussing alternative estimators for estimating a PEHA. State politics researchers have taken different approaches to modeling interdependence and heterogeneity in general, some of which have been applied to PEHA. Below, we discuss three strategies for modeling the heterogeneity in PEHA present in the existing literature: (1) multilevel modeling with random effects; (2) the inclusion of fixed effects, which can be framed through a multilevel modeling framework even if not always conceived that way; and (3) adjusting standard errors by clustering them within policies.

Estimation Strategies for PEHA

Estimating equation 1 via a multilevel model with random effects adds the assumption that the group-level heterogeneity captured by the four u terms follows a joint distribution with unknown variance. A variety of specifications and assumptions can be made; here, we review one common form and refer interested readers to the many excellent resources on multilevel modeling to keep this discussion relatively brief (see, e.g., Franzese 2005; Gelman and Hill 2007; Raudenbush and Bryk 2002; Steenbergen and Jones 2002). Specifically, then, we assume these policy-specific random effects follow a multivariate normal distribution as follows:

$$\begin{pmatrix} u_{0k} \\ u_{1k} \\ u_{2k} \\ u_{3k} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \tau_0^2 & \tau_{01} & \tau_{02} & \tau_{03} \\ \tau_{01} & \tau_1^2 & \tau_{12} & \tau_{13} \\ \tau_{02} & \tau_{12} & \tau_2^2 & \tau_{23} \\ \tau_{02} & \tau_{12} & \tau_{13} & \tau_3^2 \end{bmatrix} \right). \tag{7}$$

Estimating this heterogeneity as random effects imposes additional structure on the heterogeneity across policies as a number of assumptions must be made about the random components of the model. ε and each of the u must be uncorrelated with the included independent variables. Furthermore, ε must be independent from the u. The policy-specific errors, u, however, may be correlated with each other as captured by their joint distribution, which has an arbitrary covariance matrix. We assume that they have a mean of zero. Using a multivariate normal means that many of the policies will have effects similar to the fixed part (i.e., small u) while a few policies will have larger deviations in the tails of the normal distribution, but one can assume other distributions besides the normal here.

The heterogeneity described in equation 1 can also be estimated via multilevel modeling with fixed rather than random effects. In this approach, one includes a set of fixed effects for each policy to capture deviations in the baseline rates of adoption. To capture heterogeneity in the coefficient for a variable, one must then include interactions between the variable and the policy fixed effects. Thus, one adds in $K-1$ fixed effects and $K-1$ interactions between a variable and the policy fixed effects. At the one extreme with no fixed effects, this corresponds to a fully pooled model, while at the other extreme with all variables interacted with policy fixed effects, we have the equivalent of estimating K separate single policy EHA models. In between we have the range of partially pooled specifications (Franzese 2005). One can test for the presence of heterogeneity by a joint test of the null hypothesis that the $K-1$ coefficients or intercepts are zero.

The fixed effect specification has the advantage of fewer assumptions: the heterogeneity does not have to follow a specific distribution, and the u do not have to be independent of the included independent variables. Drawbacks include the introduction of high or perfect collinearity. If a model has slowly evolving state-level variables such as median income, population, or employment growth rates, adding an indicator variable for the states could create a high level of multicollinearity (Cheah 2009).

Fixed effects also preclude the inclusion of policy-invariant variables that will likely be of interest to scholars of state or international politics employing PEHA, a problem that is well known in the context of cross-sectional time-series data (Plümper and Troeger 2007).

A third approach that has been used in previous applications of PEHA involves clustering the standard errors (e.g., Boehmke 2009; Karch et al. 2013; Makse and Volden 2011; Shipan and Volden 2006). This has been described as a "straightforward and practical" approach to account for group-level correlation (Primo, Jacobsmeier, and Milyo 2007). Ignoring differences in the error structure across policies or states or time can lead to biased estimates of the standard errors (Harden 2011; Moulton 1990; Primo, Jacobsmeier, and Milyo 2007), frequently resulting in Type I errors (Franzese 2005; Moulton 1990). While this is a topic that has received a lot of attention in political science and in state politics research in particular, it does not capture the kind of heterogeneity that we describe above. And while one could still employ clustered standard errors to capture yet-unaccounted for differences even after adding in heterogeneity via policy fixed effects and interactions thereof with independent variables, the best approach remains trying to develop a proper and fully specified model (King and Roberts 2015).

Overall, then, we believe multilevel modeling offers the best trade-off between parsimony and accuracy. In addition to its potential advantages for properly estimating measures of uncertainty, multilevel modeling has specific features that fit the PEHA application well since it can help capture differences across policies or states. Pooling many different policies together inevitably leads to the violation of the assumption that the effects of included covariates do not change across policies. By returning an estimate of the variation of the random coefficients across policies, it allows for an explicit evaluation of the amount of heterogeneity as well as the opportunity to recover and substantively evaluate those policy by policy deviations. Furthermore, such deviations can be explored at the same or different levels, which allows researchers to simultaneously capture heterogeneity across policies, states, or, years.

Considerations for Multilevel PEHAs with Random Effects

A few important considerations will be of interest to applied researchers beyond the common ones associated with variable choice, and so on. First, one must choose which coefficients should include possible random effects. Allowing for all of them to have random effects will be computationally impractical with the number of observations that one would have in a typical PEHA. Theory and substantive knowledge should guide this choice, but estimation issues cannot be ignored. For example, theories of policy diffusion suggest that policies with spillover effects may diffuse differently than those without them, which could point the researcher toward including a random coefficient for diffusion between contiguous neighbors. Measures of public opinion also seem like likely candidates for random coefficients because its effect likely varies across policies based on salience or complexity. The nature of the policies included will also help guide

this choice: if the policies tend to be similar, then less heterogeneity may be needed than if they are quite diverse. Estimation issues tend to place some limitations here since our experience indicates that convergence becomes difficult to achieve when specifying more than three or four random coefficients. Choices about which variables might have random effects can be informed by theory and then tested statistically.

Second, and relatedly, one must decide whether to place any restrictions on the covariance matrix for the random effects. This can help ameliorate convergence problems, especially as the number of random coefficients increases. For example, one might restrict the off-diagonal elements by assuming common covariances or move all the way to a diagonal matrix that assumes independence across effects. Researchers should keep in mind that independent random effects can be a strong assumption and in general will want to relax this when possible. Theory may also help here as it may be the case that some variables matter more when others matter less, suggesting a negative correlation between their random coefficients.

Interpretation will also be an important consideration, especially as we view the random effects as capturing theoretically motivated, policy-specific heterogeneity. At the most basic level, we can divide the effect of each variable into its fixed component, which does not vary across policies, and its random component, which does. Researchers will often want to test hypotheses about the former and also whether the variation in the random effects differs from zero. They will often also want to evaluate the substantive magnitude of this variation, which can be ascertained by comparing the magnitude of the fixed component to the standard deviation of the random components. Exploring the random or combined (fixed plus random) effects for each policy can be accomplished by obtaining the estimates of the random effects terms for a given variable and inspecting them numerically or graphically. We illustrate these approaches later in our application to state abortion policy.

Finally, we should note that these specifications can be expanded in a variety of ways. For example, as noted earlier, the equations for the coefficients can be written to allow for cross-level interactions that might capture how the effects of state characteristics interact with those of policies. Or one could add more levels to the specification to capture, for example, the state-year-policy, year-policy, and policy levels. One could also specify nonnested levels such as policy and state, which would allow some variables to have heterogeneous effects across policies or across states.

Monte Carlo

To evaluate different modeling strategies for PEHA, we run Monte Carlo simulations to determine how closely the coefficient estimates match the true values and how closely the standard errors match the standard deviation of the estimates. We created a hypothetical data set with 50 states covering 40 years with three independent variables borrowed from the real state politics literature: citizen ideology (W. D. Berry et al. 1998), state median income (U.S. Census Bureau), and a count of lagged neighbors' adoptions. We created adoption data for 30 different policies, stacked them in the pooled event history format, and ran the five models on each of 500 draws.

Data Generating Process

To generate our data, we start with a standard logistic EHA data generating process for each of 30 policies. The initial equation for generating the latent variables for policy adoption is then

$$Y_{ikt}^* = \beta_0 + \beta_1 \times ideology_{it} + \beta_2 \times income_{it} + \beta_3 \times neighbors_{ikt-1} + \varepsilon_{ikt}. \tag{8}$$

We generate heterogeneity across the 30 simulated policies by adding a random deviation from the average effects for the intercept and the first two covariates. We then vary the amount of heterogeneity to see how the various estimators fare with increasing deviations from the assumption of homogeneous effects. In our smallest deviation case, the constant term has a mean of -4 with a variance of 0.5, the variable based on ideology (X_1) has a mean effect of .5 with a variance of .15, and income (X_2) has a mean effect of 1 with a variance of .15. Here, we assume independent random effects, but future work should explore the consequences of correlated random effects. To increase heterogeneity, we then multiply the variances by a common scale factor, c, of 1, 2, 3, and 4. We chose these values so that with the smallest level of heterogeneity, most of the values would be near the mean effect, but as the variance scale increases, we would start to have some policies with vastly different effects and even differently signed effects. In our empirical example that follows, we estimate levels of heterogeneity with a standard deviation of roughly the same magnitude as the fixed portion of the coefficient that occurs in our simulations for β_{1k} when $c = 2$, but not quite for β_{2k} because its standard deviation reaches a maximum of 0.77, suggesting that our simulation includes levels of heterogeneity appropriate for real-world pooled adoption data. This leads to the following data generating process:

$$Y_{ikt}^* = \beta_{0k} + \beta_{1k} \times ideology_{it} + \beta_{2k} \times income_{it} + 0.25 \times neighbors_{ikt-1} + \varepsilon_{ikt}, \tag{9}$$

$$\beta_{0k} = \beta_{00} + u_0, \tag{10}$$

$$\beta_{1k} = \beta_{10} + u_1, \tag{11}$$

$$\beta_{2k} = \beta_{20} + u_2, \tag{12}$$

$$\begin{pmatrix} u_0 \\ u_1 \\ u_2 \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 0.5c & 0 & 0 \\ 0 & 0.15c & 0 \\ 0 & 0 & 0.15c \end{bmatrix} \right), \tag{13}$$

$$c \in \{1, 2, 3, 4\}. \tag{14}$$

We then generate the corresponding binary outcomes, iterating by year to include the count of lagged neighbors' adoptions. We adjust the dependent variable to the correct event history format by determining the beginning of the risk set for each policy
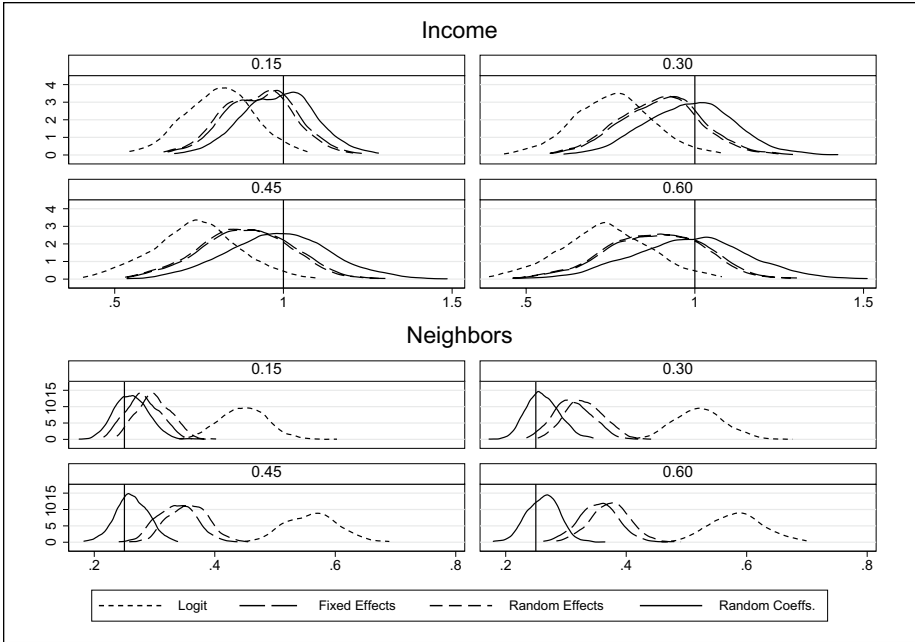
Figure 1. Kernel density plots of estimates of select coefficients, varying magnitude of the random effects.

Note. Results obtained from 500 draws for each value of the variance of the random effects.

by the first observed adoption. We then set the dependent variable to missing before the first adoption, 0 in that year and beyond until a state adopts, 1 in the year a state adopts the policy, and missing after a state adopts.

The Monte Carlo estimates five models with four independent variables: X_1 (ideology), X_2 (income), X_3 (lagged neighbors), and a time trend. The first model is a baseline logistic regression that does not take into account the interdependence of observations. The second model includes fixed effects for each policy. The third model accounts for dependence by clustering the errors by policy. The final two models are multilevel models, with the higher level groups corresponding to each policy. The first version is a random intercept model, and the second one adds random coefficients for both income and ideology.

We now summarize the general trends and conclusions that emerge from our Monte Carlo analysis, highlighting results for specific parameters either as representative or unique. We provide full details of the results in Table 2 of our Supplementary Appendix.

Coefficient and Random Effects Estimates

In Figure 1, we plot the kernel density estimates for two coefficients across the four different values of the random effects variances. We report the results for income,

which is indicative of the results for variables with random coefficients, and lagged neighbors, which is indicative of the results for variables with constant effects. Because the coefficient estimates for the simple logit and clustered logit match exactly, we do not report or discuss them separately. The results for the coefficients with heterogeneity across policies consider only the fixed portion—we examine the estimates of the variation in these effects below.

The general pattern across all parameters and estimators consists of the logit model and logit with clustering producing estimates furthest from the truth, the random effects closer but still off, the fixed effects slightly closer still, and the random coefficients model falling almost exactly around the true value (indicated by the vertical lines). The apparent bias appears to be smaller for the coefficients with random effects than for the two without: for the lagged neighbors variable, the logit model produces estimates more than twice the true value, and for the time variable, the coefficients indicate a negative effect that barely includes zero in the support, despite the fact that time was not included in the data generating process. For all variables, we find that the distributions tend to deviate more from the true value as the variance scale factor increases. These deviations move in opposite directions for the random coefficients than for the fixed coefficients.

Part of the explanation for this apparent bias likely results from a compensatory response to bias in the intercept via the logit functional form. The intercept shows evidence of attenuation with the relative bias in the same order as for the coefficients. Even the random coefficients model shows some slight attenuation for the intercept. Because the logit intercept is generally too large, it would make sense that the effect of income would be underestimated as a smaller change produces a larger increase in the probability of adoption given the relatively low baseline probability of adoption. But that does not explain the bias in the opposite direction for neighbors and time. The small but negative effect of time might occur when we fail to account for random effects via survivorship bias: cases with large random effects tend to adopt sooner and therefore leave the data. When we fail to account for these random effects, the expected value of the combined error term will be negative rather than zero for cases that persist. Yet even though such heterogeneity ought to be captured by fixed effects, we still see some potential bias there, suggesting that more forces may be at work.

While we do not display the results here, we also examine the distribution of the standard deviations of the random effects. The random effects model showed about a 25% to 30% underestimation of the random intercepts on average, while the random coefficients estimator exhibited a roughly 10% underestimation. As the true variance increases, the estimator with random coefficients outperforms the one with only random effects. The standard deviations of the two random coefficients showed a much smaller deviation, if at all, from their true values.

Comparing Standard Errors and Standard Deviations

Figure 2 compares the average standard errors with the standard deviation of the coefficients across the 500 draws for the coefficients on ideology and neighbors'
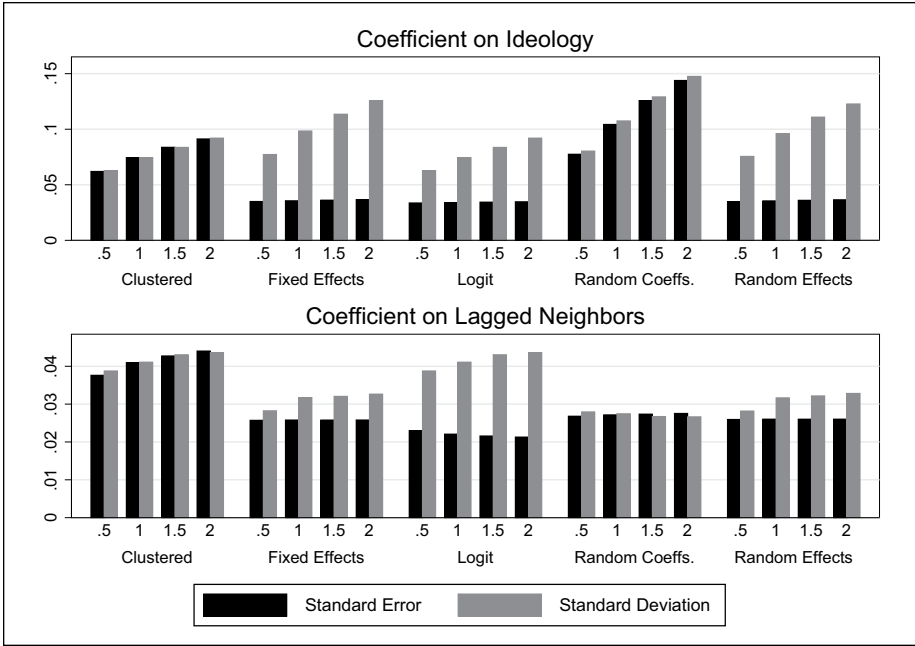
Figure 2. Comparison of standard errors and standard deviation for select coefficients, varying magnitude of random effects.

Note. Results obtained from 500 draws for each value of the scale parameter. Standard deviations calculated from the sampling distribution of the 500 estimated coefficients, while the standard errors represent the average of the 500 standard errors.

adoptions. If the estimator properly captures the fundamental uncertainty in the data generating process, then the two should match. Across all variables, a common pattern emerges, with the logit, fixed effects, and random effects estimators all producing standard errors that reflect too much confidence in the associated coefficient estimate. The degree of overconfidence for the two coefficients with heterogeneous effects exceeds that for the two fixed coefficients, but the pattern persists. Both the clustered and random coefficients estimators produce average standard errors that correspond well to the standard deviations.

What this means substantively is that the standard errors from the clustered, logit, and random effects estimators are too small, and as a result, researchers may incorrectly fail to reject the null hypothesis (Type 1 error). The clustered and random coefficients models more accurately estimate the uncertainty around the estimated statistics, making such an error less likely with these models. Because we know that the random coefficients model matches the true generating process, we can also note that the clustered estimator produces measures that are too small for the coefficients with heterogeneity and too large for those without.

Application: The Diffusion of Antiabortion Rights Policies

To illustrate these various approaches to modeling heterogeneity, we apply these techniques to a political science data set. The number and variety of state policies regulating abortion has grown dramatically over the last few decades. According to the Alan Guttmacher Institute, 2013 saw the second-highest number of new abortion restrictions passed: 24 states enacted 122 new pieces of legislation regarding reproductive health. One-third of these (43) restrict access to abortion services (Guttmacher Institute Report 2013). Previous research has uncovered some of the determinants of abortion policy; however, the focus of study has been exclusively on only a few commonly adopted policies instead of looking at the body of legislation that contributes to a state's abortion climate more broadly. While the most common explanation of abortion policy diffusion is the role of values-oriented factors such as religious constituents and public opinion that are consistent with a morality politics paradigm (Mooney and Lee 1995; Patton 2007), some scholars find that these factors are far less influential compared to partisan or other redistributive policy variables (Medoff and Dennis 2011; Meier and McFarlane 1993).

Although the previous policy-by-policy approach to studying abortion has yielded many useful insights into an important policy area, there is a serious concern with this approach. The contradictory findings of previous scholarship on abortion policy can be explained by the selection of different policies for study. Studies that find morality policy theory to be the most compelling explanation tend to study highly salient abortion policies that attempt to discourage women from obtaining an abortion (such as mandating preabortion counseling or restrictions on minor women seeking abortions). Studies that find redistributive or partisan variables tend to study abortion policies that target abortion clinics or regulate the funding of abortions for low-income women.

Very few studies include a heterogeneous sample of abortion policies (but see Patton 2007). This is problematic, as individual abortion policies may be overdetermined by specific factors that are generally insignificant in predicting the adoption of abortion policy. For this reason, scholars attempting to explain the proliferation of abortion policy by looking at one or only a few policies may reach vastly different conclusions from each other and from scholars looking at the broader abortion policy environment, making the realm of abortion policy a prime candidate for the use of PEHA.

Modeling abortion policy with PEHA allows us to establish the average determinants of abortion policy. Using a near-universe of 29 anti-abortion rights policies from 1973 to 2013, we study the effects of morality, institutional, and state contextual variables on the adoption of state policies that restrict access to legal abortion. These data, created using reports from the Alan Guttmacher Institute, NARAL, and the National Right to Life Committee, include all policies tracked by these three groups that have been adopted by at least five states. In addition, policies deemed unconstitutional by the Supreme Court (such as spousal consent requirements) are excluded as these policies have a truncated diffusion process. For a more detailed discussion of the collection of this data set, including source materials and which policies are excluded, see Kreitzer (2015).

The nature of these policies vary on a number of dimensions, including technical complexity, salience, and whether the policy targets clinics or pregnant women. The abortion policies also diffused at different rates. While some policies were adopted by most states (46 states adopted bans on the use of federal funds for low-income women), others were adopted by very few states (only 6 states require mandatory viability testing). The nuances of specific policies may also mean that certain variables have a heterogeneous effect.

Drawing from the previous literature, we theorize about which factors have a consistent effect across policies. Abortion is often considered a "clash of absolutes," in which advocates on each side of the debate have strongly held and noncompromising beliefs (Cook, Jelen, and Wilcox 1992; Tribe 1990). Opponents of legal abortion are generally supportive of all restrictions on access to abortion. Although measures of public preferences often have heterogeneous effects on policies, public opinion, ideology, and religiosity of constituencies on abortion policy may not. In contrast, partisans are strategic on the issue of abortion. Although Democrats generally oppose restrictions on abortion, they are more supportive of policies that are controversial or have high public support (such as bans on late term abortion procedures, often known by the political term "partial birth abortions"). Thus, the effect of partisans may be heterogeneous. Another variable with potentially heterogeneous effects is the percent of neighboring states that have already adopted a given policy. Some policies may generate "spillover" effects by creating incentives for women to travel to neighboring states to obtain an abortion or increasing public attention on a certain policy. We empirically test these expectations to determine whether there is statistically significant variation around the effect of these variables.

We model state abortion policy using several PEHA models that explicitly account for the heterogeneity that is inherent in the data. We first use a logistic regression with clustered errors for the states.[2] Next, we improve the first model by including policy fixed effects to allow for different baseline hazard rates. We then model policy and state heterogeneity by clustering on policies and states. Finally, we use multilevel models with random effects for policies, with and without random coefficients for Democratic governor and the proportion of neighboring states that previously adopted a policy.

Our dependent variable is the state adoption of legislation that restricts access to legal abortion. Our slate of independent variables measures *public preferences* (public attitudes toward abortion and the state religious adherence rate), *partisan and institutional factors* (unified Democratic control of government, the proportion of the state legislature comprised of female Democrats, Democratic governor, difficulty of the initiative process), *state contextual factors* (such as the median income and state population size), various *time trend variables* to model linear and nonlinear time trends and an indicator variable for years after the influential *Webster v. Reproductive Health Services* (1989) court decision, and the *proportion of neighboring states* that adopted a given policy.

We report the estimates of the five different modeling approaches in Table 1. The statistical significance and size of the coefficients of some of the variables change

Table 1. Pooled EHA Estimates of Diffusion of 29 Antiabortion Policies, 1973–2013, by Estimator (N = 29,663).

	Logit	FEs	Clustered	MLM	MLM-RC
Conservative Abortion Opinion (Norrander)	0.546*	0.290	0.546**	0.522**	0.518**
	(0.317)	(0.410)	(0.217)	(0.219)	(0.226)
Religious Adherence Rate	1.092*	1.433**	1.092***	1.283***	1.538***
	(0.630)	(0.692)	(0.417)	(0.385)	(0.395)
Initiative Difficulty (Bowler, Donovan)	0.073***	0.073***	0.073***	0.076***	0.084***
	(0.026)	(0.028)	(0.020)	(0.018)	(0.018)
Democratic Governor	−0.264***	−0.258**	−0.264***	−0.281***	−0.362***
	(0.096)	(0.103)	(0.083)	(0.085)	(0.129)
Unified Democratic Legislature	−0.174*	−0.183	−0.174*	−0.178**	−0.164*
	(0.104)	(0.113)	(0.091)	(0.090)	(0.091)
Democratic Women	−5.359***	−5.450***	−5.359***	−5.321***	−5.285***
	(1.627)	(1.632)	(1.084)	(1.093)	(1.104)
Neighbors' Adoptions (%)	2.514***	1.692***	2.514***	1.858***	1.455***
	(0.202)	(0.232)	(0.176)	(0.173)	(0.327)
State Median Income (rescaled)	−0.038	−0.305*	−0.038	−0.122	−0.136
	(0.073)	(0.164)	(0.060)	(0.084)	(0.087)
State Population (rescaled)	0.087	0.094	0.087	0.102	0.093
	(0.091)	(0.102)	(0.078)	(0.074)	(0.075)
Time	−0.122***	−0.046*	−0.122***	−0.084***	−0.074***
	(0.016)	(0.026)	(0.014)	(0.018)	(0.019)
Time Squared	0.002***	0.001***	0.002***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Post-Webster Indicator	0.653***	0.530**	0.653***	0.663***	0.603***
	(0.188)	(0.214)	(0.151)	(0.155)	(0.161)
Constant	−5.183***	−3.798**	−5.183***	−5.185***	−5.227***
	(0.965)	(1.620)	(0.663)	(0.717)	(0.737)
var(constant)				0.499***	0.553***
				(0.176)	(0.210)
var(Neighbors)					1.620**
					(0.641)
var(Dem. Governor)					0.203*
					(0.104)
N	29,663	29,663	29,663	29,663	29,663
χ²	357.12	2,275.25	369.19	330.56	225.53
AIC	5,894.06	5,727.27	5,894.06	5,776.21	5,716.90
BIC	6,001.93	6,067.47	6,001.93	5,892.38	5,849.66

Note. EHA = event history analysis; FEs = fixed effects; MLM = multilevel model; MLM-RC = multilevel model–random coefficients; AIC = Akaike information criterion; BIC = Bayesian information criterion. Standard errors are in parentheses.
*p < .01. **p < .05. ***p < .01.

across models. Just as we found in the Monte Carlo simulations, the logit model seems to understate several variables, with the random effects and fixed effects model in between and the model with random coefficients the largest. In addition, the coefficient for the lagged percent of neighboring states is much larger in the logit, and the

size of the coefficient decreases as we move across the table to the random coefficients model. Most significantly, we gain important substantive information in the random coefficients model. The multilevel models also allow us to explicitly model heterogeneity and retrieve separate estimates of the intercept and variance for the group-level variable (here, we use the policies as the second-level group). In the final model, we also include random coefficients for Democratic governor and the proportion of neighboring adopters.

It is clear from our analysis that there exist several types of heterogeneity within the data. The models that fail to incorporate these types of heterogeneity return estimates that are substantially different from those in which the heterogeneity is explicitly modeled through random effects and random coefficients. While we do not know the true data generating process in these data, recall that the multilevel models in the Monte Carlo simulation were the best model in terms of returning accurate estimates and correct standard errors. There is further evidence that the multilevel models outperform other models when we compare the Akaike information criterion (AIC) and Bayesian information criterion (BIC). The multilevel models have the smallest values for both of these goodness of fit tests.

These models provide valuable insights into what shapes abortion policy in the states. Conservative public opinion and the size of the religious population in the states both increase the probability of conservative abortion policy adoption. Democratic partisans have a consistent and negative effect, indicating that partisan control of the executive and legislative branches and the presence of Democratic women in the legislature make conservative abortion policy adoption less likely. There is also evidence that states learn from their neighbors. States are more likely to adopt abortion policy as the percent of their neighboring states that have already adopted that policy increases.

In addition to being the best specified model, taking this multilevel approach also allows us to gain additional information regarding how the effect of a given variable differs across policies. There is statistically significant variation around the effect of neighboring states and Democratic governors. In models not shown here, we find no significant variation around the effect of conservative public opinion or the religious adherence rate in the state. To visualize this heterogeneity, we construct a graph that depicts both the shared effect of a variable across policies as well as the policy-specific deviations from that effect. For each policy, we add the coefficient to the estimated random effect and plot these. We also generate 95% confidence intervals for the combined effects by policy to evaluate significance. These confidence intervals reflect the total uncertainty for each policy by combining the standard errors of the coefficient and the standard errors of the random effects.[3] To facilitate interpretation, we plot cases that exceed the .05 significance level in black, those that exceed the .10 level in medium gray, and those that exceed neither significance level in light gray. The black vertical line and gray shaded region correspond to the coefficient estimate and its 95% confidence interval.

We display these plots for the two random coefficients in Figures 3 and 4. Consider them in turn. The proportion of neighboring adopters is always a significant predictor of policy on average, as evidenced by the fact that the gray shaded region does not
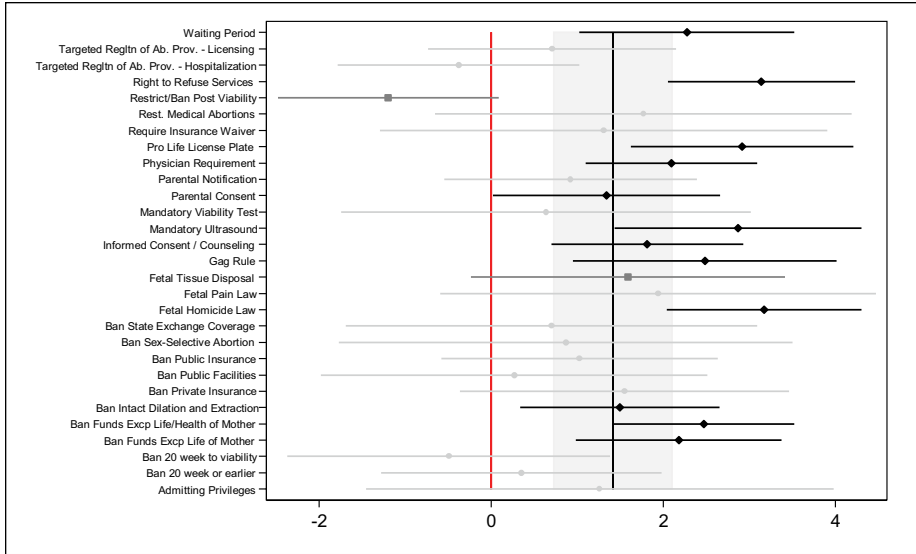
Figure 3. Estimated coefficients for neighboring adopters, by policy.

Note. Points represent the combined fixed and random effect for each variable for each policy. Lines represent a 95% confidence interval based on the combined standard errors of the fixed and random effects. Black cases with diamonds are significantly different from zero at the .05 level, medium gray cases with squares are significantly different at the .10 level, and light gray lines with circles are not. Vertical black line indicates the estimated fixed coefficient for that variable, and the light shaded region gives its 95% confidence interval. Vertical red line indicates zero.

TRAP: Targeted Regulation of Abortion Providers; IDE: integrated development environment

include zero. There is also great heterogeneity across policies, with effects ranging from −1 to more than 3. This reflects the large standard deviation for the random coefficient on neighbors recovered in the multilevel model estimates. The effect of Democratic governor is less heterogeneous and generally less significant. While the mean effect is negative (indicating that Democratic governors make anti-abortion rights policy adoption less likely), it is barely significant. Few of the policy-specific effects have confidence intervals that do not include zero and some even have positive estimated effects.

Graphing the heterogeneity in this way uncovers theoretically and substantively interesting information. The policies for which Democratic governors have an individually significant effect are some of the policies Democratic governors are very likely to veto, such as informed consent policies (often called "Women's Right to Know Acts") and bans on abortions that approach viability. The effect of Democratic governors is positive (albeit not statistically significant) for policies with high levels of public support, such as mandatory parental consent and bans on postviability abortion. The number of policies for which there is an independently significant effect for neighboring adopters is even greater. Geographic diffusion is significant in 15 of the
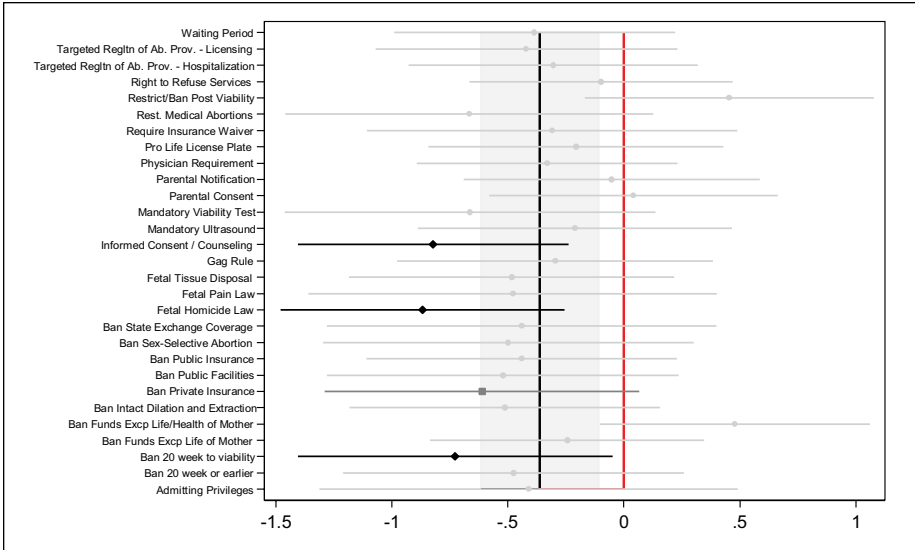
Figure 4. Estimated coefficients for Democratic governor, by policy.
Note. Points represent the combined fixed and random effect for each variable for each policy. Lines represent a 95% confidence interval based on the combined standard errors of the fixed and random effects. Black cases with diamonds are significantly different from zero at the .05 level, medium gray cases with squares are significantly different at the .10 level, and light gray lines with circles are not. Vertical black line indicates the estimated fixed coefficient for that variable, and the light shaded region gives its 95% confidence interval. Vertical red line indicates zero.
TRAP: Targeted Regulation of Abortion Providers; IDE: integrated development environment

29 policies, but there is not a clear pattern for when it is significant. Some of the policies for which neighboring states' adoption is significant include policies likely to generate spillover effects, such as waiting periods and parental consent. Other policies with significant geographic diffusion are technically simple and politically salient, such as "pro-choice" license plates.

Conclusion

Our Monte Carlo simulations of PEHA with varying amounts of heterogeneity around the true coefficients indicate considerable variation in the ability of different modeling strategies to account for this heterogeneity. Using cluster-adjusted standard errors returns equally biased estimates of coefficients compared to the baseline logit that does not attempt to account for heterogeneity, but the standard errors are inflated as expected to account for the group-level dependence. The inclusion of fixed effects does return apparently unbiased estimates, but the standard errors are very large. We find that multilevel models yield the best combination of unbiased estimates and standard errors, and conclude that future research using PEHA should consider using a multilevel pooled event history analysis (MPEHA) model. Even when the research

objectives do not include exploring heterogeneity, researchers may find through the comparison of model fit statistics that a multilevel model is the best fitting model. We find results consistent with the findings in the Monte Carlo simulations through an empirical analysis of a diverse set of 29 antiabortion policies. We demonstrate how to explore the implications of heterogeneity in the data by modeling and graphing how the effect of certain variables differs across policies. This approach yields interesting insights into how political and geographic factors shape state abortion policy.

Which method scholars should use always depends on a variety of factors, and there may need to be some pragmatic sacrifices in explanatory depth. Scholars should think theoretically about how heterogeneity enters the data. For example, using fixed effects for policies is an easy way to change the baseline probability of policy adoption, but does not allow heterogeneous effects of variables across policies. Adjusting standard errors by clustering is an easy way to allow correlation among groups but does not allow the dependence to be explicitly modeled in a way that allows the researcher to include group-specific coefficients or intercepts. Scholars using PEHA should use multilevel models with random effects and random coefficients to retrieve the most accurate estimates, explicitly model heterogeneity, and maximize the information the model provides.

Notes

1. Specifically, the pooled event history analysis (PEHA) looks most like the marginal model of Wei, Lin, and Weissfeld (1989) as it also uses a total time framework; one important difference is that they assume a single event with multiple failures and stratify by the order of failure, whereas we are pooling similar or possible dissimilar events and stratifying by the exogenous policy type. See Kelly and Lim (2000) for an overview of different models for repeated events for continuous time durations.
2. Because no one in practice is naïve enough to not at least cluster, there is no need to continue to look at the baseline logit.
3. Estimates of the random effects and their standard errors come from the empirical Bayes estimates generated during the estimation procedure.

References

Balla, Steven J. 2001. "Interstate Professional Associations and the Diffusion of Policy Innovations." *American Politics Research* 29 (3): 221–45.

Berry, Frances Stokes, and William D. Berry. 1990. "State Lottery Adoptions as Policy Innovations: An Event History Analysis." *American Political Science Review* 84 (2): 395–415.

Berry, William D., Evan J. Ringquist, Richard C. Fording, and Russell L. Hanson. 1998. "Measuring Citizen and Government Ideology in the American States, 1960-93." *American Journal of Political Science* 42 (1): 327–48.

Boehmke, Frederick J. 2009. "Approaches to Modeling the Adoption and Diffusion of Policies with Multiple Components." *State Politics & Policy Quarterly* 9 (2): 229–52.

Boehmke, Frederick J., and Paul Skinner. 2012. "The Determinants of State Policy Innovativeness." Presented at the 2012 State Politics and Policy Conference, Houston, February 16–18.

Cheah, B. C. 2009. *Clustering Standard Errors or Modeling Multilevel Data*. New York: Columbia University.

Cook, E. A., Jelen, T. G., Wilcox, W. C., and Wilcox, C. 1992. *Between Two Absolutes: Public Opinion and the Politics of Abortion*. Boulder, CO: Westview Press.

Elkins, Zachary, Andrew T. Guzman, and Beth A. Simmons. 2006. "Competing for Capital: The Diffusion of Bilateral Investment Treaties, 1960-2000." *International Organization* 60 (4): 811–46.

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

Gelman, Andrew, and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.

Grossback, Lawrence J., Sean Nicholson-Crotty, and David A. Peterson. 2004. "Ideology and Learning in Policy Diffusion." *American Politics Research* 32 (5): 521–45.

Guttmacher Institute Report. 2013. "2012 Saw Second-Highest Number of Abortion Restrictions Ever." http://www.guttmacher.org/media/inthenews/2013/01/02/index.html (accessed June 12, 2015).

Harden, Jeffrey J. 2011. "A Bootstrap Method for Conducting Statistical Inference with Clustered Data." *State Politics & Policy Quarterly* 11 (2): 223–46.

Harris, Kathleen M. 1993. "Work and Welfare among Single Mothers in Poverty." *American Journal of Sociology* 99 (2): 317–52.

Karch, Andrew, Sean Nicholson-Crotty, Neal D. Woods, and Ann O'M. Bowman. 2013. "Policy Diffusion and the Pro-innovation Bias." Presented at the 2013 State Politics and Policy Conference, Iowa City, May 24–25.

Kelly, Patrick J., and Lynette L.-Y. Lim. 2000. "Survival Analysis for Recurrent Event Data: An Application to Childhood Infectious Diseases." *Statistics in Medicine* 19 (1): 13–33.

King, Gary, and Margaret E. Roberts. 2015. "How Robust Standard Errors Expose Methodological Problems They Do Not Fix." *Political Analysis* 23 (2): 159–79.

Kreitzer, Rebecca J. 2015. "Politics and Morality in State Abortion Policy." *State Politics & Policy Quarterly* 15 (1): 41–66.

Makse, Todd, and Craig Volden. 2011. "The Role of Policy Attributes in the Diffusion of Innovations." *Journal of Politics* 73 (1): 108–24.

Medoff, Marshall H., and Christopher Dennis. 2011. "TRAP Abortion Laws and Partisan Political Party Control of State Government." *American Journal of Economics and Sociology* 70 (4): 951–73.

Meier, Kenneth J., and Deborah R. McFarlane. 1993. "The Politics of Funding Abortion State Responses to the Political Environment." *American Politics Research* 21 (1): 81-101.

Mintrom, Michael. 1997. "Policy Entrepreneurs and the Diffusion of Innovation." *American Journal of Political Science* 41 (3): 738–70.

Mooney, Christopher Z., and Mei-Hsien Lee. 1995. "Legislative Morality in the American States: The Case of Pre-Roe Abortion Regulation Reform." *American Journal of Political Science* 39 (3): 599–627.

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

Pacheco, Julianna. 2012. "The Social Contagion Model: Exploring the Role of Public Opinion on the Diffusion of Antismoking Legislation Across the American States." *Journal of Politics* 74 (1): 187–202.

Patton, Dana. 2007. "The Supreme Court and Morality Policy: The Impact of Constitutional Context." *Political Research Quarterly* 60 (3): 468–88.

Plümper, Thomas, and Vera E. Troeger. 2007. "Efficient Estimation of Time-Invariant and Rarely Changing Variables in Finite Sample Panel Analyses with Unit Fixed Effects." *Political Analysis* 15 (2): 124–39.

Primo, David M., Matthew L. Jacobsmeier, and Jeffrey Milyo. 2007. "Estimating the Impact of State Policies and Institutions with Mixed-Level Data." *State Politics & Policy Quarterly* 7 (4): 446–59.

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

Shipan, Charles R., and Craig Volden. 2006. "Bottom-Up Federalism: The Diffusion of Antismoking Policies from U.S. Cities to States." *American Journal of Political Science* 50 (4): 825–43.

Simmons, Beth, Frank Dobbins, and Geoffrey Garrett, eds. 2008. *The Global Diffusion of Markets and Democracy*. Cambridge, MA: Cambridge University Press.

Soule, Sarah A., and Jennifer Earl. 2001. "The Enactment of State-Level Hate Crime Law in the United States: Intrastate and Interstate Factors." *Sociological Perspectives* 44 (3): 281–305.

Steenbergen, M. R., and Jones, B. S. 2002. Modeling Multilevel Data Structures. *American Journal of political Science* 46 (1): 218–37.

Tribe, L. H. 1992. *Abortion: The Clash of Absolutes*. New York, NY: WW Norton & Company.

Volden, Craig. 2006. "States as Policy Laboratories: Emulating Successes in the Children's Health Insurance Program." *American Journal of Political Science* 50 (2): 294–312.

Walker, Jack L. 1969. "The Diffusion of Innovations among the American States." *American Political Science Review* 63 (3): 880–99.

Webster v. Reproductive Health Services 492 U.S. 490 (1989)

Wei, L. J., D. Y. Lin, and L. Weissfeld. 1989. "Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions." *Journal of the American Statistical Association* 84:1065–73.

Author Biographies

Rebecca J. Kreitzer is an assistant professor of Public Policy at the University of North Carolina at Chapel Hill. She received her PhD from the University of Iowa in 2015.

Frederick J. Boehmke is professor of Political Science and Director of the Iowa Social Science Research Center. He received his PhD from California Institute of Technology in 2000 and spent two years as a Robert Wood Johnson Scholar in Health Policy Research at the University of Michigan.