Estimating Incumbency Advantage and Its Variation, as an Example of a Before–After Study

Andrew GELMAN and Zaiying HUANG

Incumbency advantage is one of the most widely studied features in American legislative elections. In this article we construct and implement an estimate that allows incumbency advantage to vary between individual incumbents. This model predicts that open-seat elections will be less variable than those with incumbents running, an observed empirical pattern that is not explained by previous models. We apply our method to the U.S. House of Representatives in the twentieth century. Our estimate of the overall pattern of incumbency advantage over time is similar to previous estimates (although slightly lower), and we also find a pattern of increasing variation. More generally, our multilevel model represents a new method for estimating effects in before–after studies.

KEY WORDS: Bayesian inference; Before–after study; Congressional election; Gibbs sampler; Incumbency advantage; Metropolis algorithm; Multilevel model.

1. INTRODUCTION

Incumbency advantage is one of the most widely studied features in American legislative elections (see Erikson 1971; Payne 1980; Alford and Hibbing 1981; Alford and Brady 1988; King and Gelman 1991; Cox and Morgenstern 1993; Cox and Katz 1996; Levitt and Wolfram 1997; Jacobson 2000; Campbell 2002; Ansolabehere and Snyder 2002a,b). Our goal in this article is to estimate incumbency advantage in a framework that allows for candidate effects. We seek to do so in the most general manner possible, using only data from two consecutive elections in each legislative district without additional information on the candidates or districts. (By analyzing pairs of elections, we minimize the difficulties that arise with missing data and decennial redistrictings.) Our analysis introduces a new form of multilevel model appropriate for observational studies or experiments based on time series or before–after data—in this case, two successive elections, with the intervening treatment being the decision of whether or not the incumbent runs for reelection.

Our method and results are similar to existing regression estimates but are more general in that we allow the incumbency advantage to vary between incumbents. We achieve this by setting up a multilevel model with three variance components: district-level baseline, candidate-level incumbency advantage, and variation across elections within a district. We check model fit (and demonstrate the flaws of some previous models) by comparing data with simulations of replicated data under the assumed model.

We apply our method to the U.S. House of Representatives in the twentieth century to obtain an estimate of the average incumbency advantage and its variation for each election year. We find that the variation has increased along with the mean level in the second half of the century.

2. REGRESSION–BASED ESTIMATES OF INCUMBENCY ADVANTAGE

In this section we review the advantages and disadvantages of regression models for incumbency effects. We then present our preferred model in Section 3 and fit it to congressional elections in Section 4.

Andrew Gelman is Professor, Department of Statistics and Department of Political Science, Columbia University, New York, NY 40027 (E-mail: *gelman@stat.columbia.edu*). Zaiying Huang is Statistician, Circulation Department, *New York Times* (E-mail: *huang@nytimes.com*). The authors thank Robert Erikson, Gary King, and the editors and reviewers for helpful comments; Gary King, Greg Wawro, and Gary Jacobson for data; and the National Science Foundation for financial support.

2.1 Background and Interpretation as an Observational Study

The advantages of using regression models to estimate incumbency advantage can be seen by comparing them to some simpler approaches. Most directly, we can compute the proportion of incumbents who win reelection; for example, Figure 1 shows the reelection rate for the U.S. House of Representatives for each election year in the twentieth century. This is not a good estimate of incumbency advantage, however, because it does not account for the fact that a high reelection rate could occur in the absence of incumbency effects, simply due to variation among districts. In a highly conservative district, for example, no incumbency advantage is needed to explain that a Republican is likely to be reelected.

Regression methods estimate incumbency advantage by comparing districts with incumbents running to open seats, controlling for district-level measures of partisan strength. Gelman and King (1990) showed that the simple measures of "sophomore surge" and "retirement slump" are biased estimates of incumbency effects, but that the information used in these measures can be put in a regression framework to create unbiased estimates. Their model can be written in terms of v_{it}, the two-party vote share for the Democratic candidate (say), in district i in election t,

$$v_{it} = \beta_0 + \beta_1 v_{i,t-1} + \beta_2 P_{it} + \psi I_{it} + \epsilon_{it}, \qquad (1)$$

where P_{it} represents the incumbent party and I_{it} represents the incumbent candidate (if any),

$$P_{it} = \begin{cases} 1 & \text{if the legislator in district } i \text{ at time } t \text{ is a Democrat} \\ -1 & \text{if the legislator in district } i \text{ at time } t \text{ is a Republican} \end{cases}$$

and

$$I_{it} = \begin{cases} 1 & \text{if a Democrat is running for reelection in district } i \text{ at time } t \\ 0 & \text{if the incumbent is not running for reelection} \\ -1 & \text{if a Republican is running for reelection.} \end{cases}$$

Party is included as well as incumbency, so that ψ captures the effect of the incumbent candidate, after controlling for party.

Journal of the American Statistical Association
June 2008, Vol. 103, No. 482, Applications and Case Studies
DOI 10.1198/016214507000000626

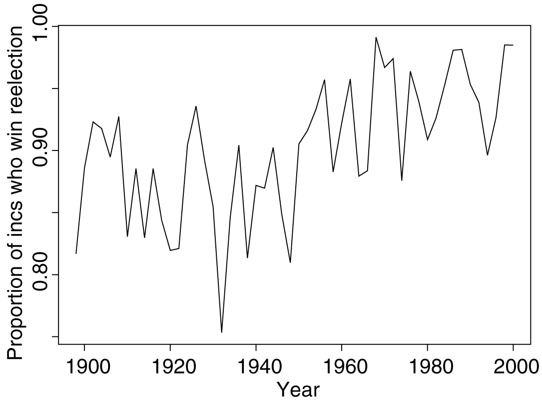

Figure 1. Reelection rate for incumbents in the U.S. House of Representatives over time. We cannot use this as a measure of incumbency advantage, because strongly partisan districts are likely to reelect the incumbent party regardless of the candidate.

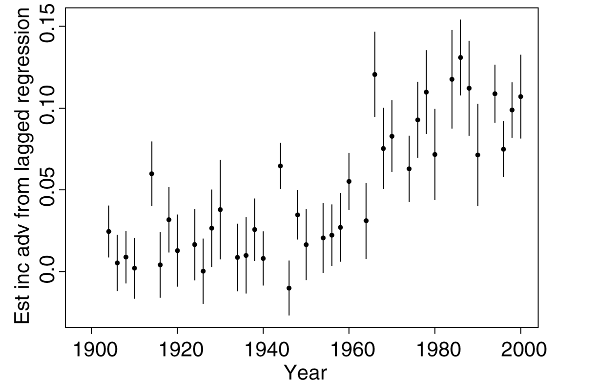

Figure 2. Estimated incumbency advantage over time (with 95% posterior intervals), from regression model (2) that assumes a constant incumbency effect for all districts in any election year, following Gelman and King (1990). These can be compared to the estimates displayed in Figures 9 and 10 for our new model that allows incumbency advantage to vary across districts.

The coefficients in (1) can be estimated separately for each general election t, and ψ is the estimate of incumbency advantage at time t. The other term in the model adjusts for differences among districts. An alternative would be to control in the regression for some measure of baseline or "normal vote" in district i (see Ansolabehere and Snyder 2002a), for example, the Democratic share of the vote in the district in the previous presidential or gubernatorial election. This could improve the estimate slightly but would not change its fundamental form and motivation—a regression-adjusted comparison between districts with different incumbency status.

Estimates of incumbency advantage can be seen as observational studies (see, e.g., Achen 1986), in which the "treatment" is the decision of whether to run a new candidate. Thus incumbents are "controls," and the open seats are "treated" units. This terminology makes sense because the districts with incumbents running are unchanged, while a big intervention is performed in the open seats. Because we are studying general and not primary elections, we view the party, not the individual candidate, as the decision maker. We emphasize this in our notation by defining the treatment indicator, T_{it}, the decision of the incumbent party of whether or not to apply the treatment and run a new candidate:

$$T_{it} = \begin{cases} 0 & \text{if the incumbent legislator is running} \\ & \text{in the general election in district } i \text{ at time } t \\ 1 & \text{if the incumbent is not running for reelection.} \end{cases}$$

Thus $I_{it} = (1 - T_{it})P_{it}$. In any given district, P_{it} is determined by the outcome of the previous election (except in unusual cases, $P_{it} = 1$ if $v_{i,t-1} > .5$), and the "incumbency advantage" is the effect of $T_{it} = 0$ compared with 1. Then regression (1) becomes

$$v_{it} = \beta_0 + \beta_1 v_{i,t-1} + \beta_2 P_{it} + \beta_3 T_{it} - \psi P_{it} T_{it} + \epsilon_{it}, \qquad (2)$$

after adding a main effect for T_{it} (which represents a difference between the incumbency effects for the two parties) to complete the model. The average effect of incumbency is represented by the parameter ψ (coded with a negative sign, because we are considering incumbency as the control condition and open seats as the treatment). In our model the main effect for incumbency appears as an interaction (the coefficient for $P_{it}T_{it}$), whereas the interaction of incumbency and party appears as a main effect for T_{it}. (If we were to reparameterize in terms of the vote for the incumbent party, then ψ would appear as a main effect and β_3 as an interaction.)

Figure 2 displays the estimates of ψ from model (2) for each congressional election between 1900 and 2000, excluding years ending in 2 (for which there was redistricting between elections $t-1$ and t).

The variable T has many of the characteristics of a treatment in a randomized experiment in that in any given election year (except those after a redistricting), the open seats appear to be distributed roughly at random. For example, contrary to what might be expected, there is no correlation between margin of vote and probability of running for reelection in the U.S. House (see Gelman and King 1990, footnote 6). Ansolabehere and Snyder (2002b) studied the issue more thoroughly in the context of strategic retirements and came to the same conclusion, that there is no evidence that open seats generally represent vulnerability of incumbents.

2.2 A Problem With Regression-Based Estimates

The estimates of incumbency advantage ψ in Figure 2 are reasonable, but the underlying model (2) does not quite fit electoral data. Figure 3 illustrates with the data that would be used to estimate the incumbency advantage for the 1988 congressional elections. When plotted on this graph, the regression model would be represented by parallel lines for incumbents and open seats, with the spacing between the lines representing the incumbency effect ψ. The actual data for the incumbents and the open seats are far from parallel, however.

We can study this problem more systematically by adding an interaction term to (2), splitting the coefficient β_1 for the lagged vote into two parts, β_{1a} for incumbents and β_{1b} for open seats,

$$v_{it} = \beta_0 + \beta_{1a} v_{i,t-1} T_{it} + \beta_{1b} v_{i,t-1}(1 - T_{it}) + \beta_2 P_{it} + \beta_3 T_{it} - \psi P_{it} T_{it} + \epsilon_{it}. \qquad (3)$$

We estimate this model for the congresses of the twentieth century and display the estimated slopes β_{1a} and β_{1b} in Figure 4. For the second half of the century, the slope for the lagged vote

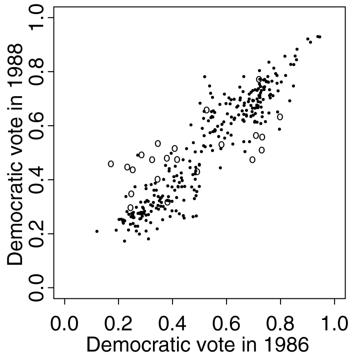

Figure 3. Graph illustrating a problem with the regression model (2). Scatterplot of the Democratic share of the two-party vote for Congress in 1986 and 1988, with each symbol representing a Congressional district. Dots represent districts with incumbents running in 1988, and circles represent open seats in 1988. It is clear that the regression line for the dots is much steeper than that for the circles.

is usually higher for incumbents than for open seats. The pattern is not consistent every year, because only a small fractions of the districts each year are open, and so the slopes for the open-seat elections are estimated imprecisely.

At this point, we could simply fit the interaction model (3) and say that the incumbency advantage is $\psi + (\beta_{1a} - \beta_{2a})v_{i,t-1}$, which depends on the lagged vote $v_{i,t-1}$. But we dislike this interpretation, because this interaction seems entirely motivated by the need to fit a pattern in data, with no direct political interpretation. Fortunately, we can rewrite the interaction model in a more useful and interpretable way, as we discuss next.

2.3 Reformulation of the Interaction as a Variance Component

Why, in the scatterplot in Figure 3, do open seats have a flatter slope compared with those that are contested by incumbents? This can be understood in terms of the interpretation of incumbency as a control condition and open seats as a major intervention. It makes sense that the "before" measure is less predictive of the "after" measure when there has been a disruptive

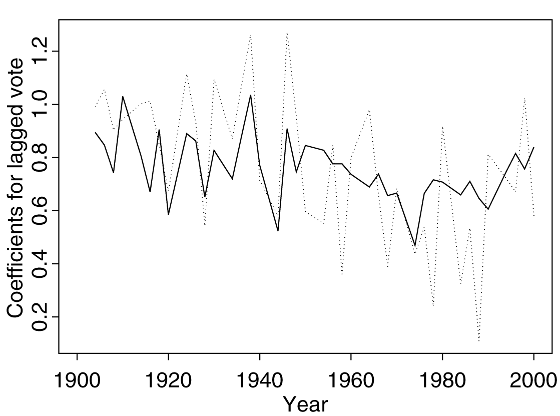

Figure 4. Coefficients for lagged vote in the regression model (3), estimated separately for elections with incumbents (——) and open seats (⋯⋯). The coefficients are consistently different for the two kinds of district elections, indicating a problem with the simpler model (2) that included no interaction.

treatment in between. To put it another way, the previous election is highly predictive of the current election when the same candidate is running but less so when the incumbent has been removed. From this perspective, it is no surprise that the coefficient for lagged vote is close to 1 for incumbents and typically lower for open seats. As Figure 4 shows, this pattern has generally held in congressional elections since the 1960s, when, as all scholars agree, incumbency advantage became substantial.

We would like to interpret difference between the two slopes in (3) as not an *effect* of incumbency but rather as a *consequence* of variation in its effects. If the advantages of incumbency vary among politicians, then the effect of removing an incumbent will be to remove a source of variation that is present in both the lagged and current votes, and the predictive power of lagged vote in that district will be reduced (an example of "regression to the mean"). What is of fundamental importance is the variation in incumbency effects, not the effect of this variation on the lagged regression coefficient.

3. MULTILEVEL MODEL OF INCUMBENCY ADVANTAGE

3.1 The Model

We set up a model that allows incumbents to have their own individual incumbency advantages and estimate the model, as before, using data from two consecutive elections. This time, we set up the full likelihood based on the data from both elections,

$$\text{for } t = 1, 2, \quad v_{it} = .5 + \delta_t + \alpha_i + \phi_{it} I_{it} + \epsilon_{it}. \qquad (4)$$

The parameters in the model are defined as follows:

- δ_t is the national vote (or, for an analysis of state legislatures, the state vote) at each time t, relative to a 50/50 split. We need δ_t in the model to correct for national swings; the centering relative to .5 is purely for convenience.
- α_i is the baseline for district i (relative to the national average); we assign it a normal population distribution with mean 0 and standard deviation σ_α.
- ϕ_{it} is the incumbency advantage in district i at time t; we assign it a normal population distribution with mean ψ and standard deviation σ_ϕ. Thus ψ is the average incumbency advantage and carries over its interpretation from the lagged regression model (2).
- The key feature of the model is its candidate-level incumbency effects. We code these by restricting $\phi_{i2} = \phi_{i1}$ for districts i in which the same incumbent is running for reelection in both years. If this is not the case (typically because the first incumbent lost in election 1), then ϕ_{i1} and ϕ_{i2} are modeled as independent draws from the population distribution of incumbency effects. We assume that all personal candidate effects are contained in ϕ, and we do not model variation in strength among challengers or open-seat candidates (except implicitly as included in the error term ϵ).
- The ϵ_{it}'s are independent error terms assumed to be normally distributed with mean 0 and standard deviation σ_ϵ. We can model the error terms ϵ_{i1} and ϵ_{i2} as independent because any dependence that would have occurred between them is captured by the district-level variable α_i.

If we were analyzing three or more election years simultaneously, then we would need to explicitly model autocorrelation in the ϵ_{it}'s.

Although the vote proportions v_{it} are constrained to fall between 0 and 1, the untransformed linear model (4) is reasonable because the actual data from contested elections almost all fall between .2 and .8 (see, e.g., Fig. 3). A related issue is that the proportion of potential voters who can switch parties between the two elections depends on the vote at time 1 (see Krashinsky and Milne 1993; Ansolabehere, Snyder, and Stewart 2000), suggesting potential nonlinearity at the extremes of the model. Another potential concern is heteroscedasticity, because voter turnout in Congressional elections varies across districts and over time. A study of residuals found no strong connection between residual variance and number of voters, which is consistent with other studies of elections (see Gelman, King, and Boscardin 1998; Mulligan and Hunter 2001).

In estimating the model, we are interested primarily in the individual incumbency effects, ϕ_{it}, their mean, ψ, and their standard deviation, σ_ϕ. The Bayesian approach allows inference for these parameters simultaneously with the district-level parameters. We indicate the complete vector of parameters by $\theta = (\delta, \alpha, \phi, \psi, \sigma_\alpha, \sigma_\phi, \sigma_\epsilon)$; the posterior distribution is then $p(\theta|v, I) \propto p(v|I, \theta)p(\theta)p(I|\theta)$. We describe each of these three factors here and then describe estimation of the parameters in Section 3.2.

The Likelihood. Equation (4) implies the following likelihood for the election data:

$$p(v|I,\theta) = \prod_{i=1}^{n}\prod_{t=1}^{2} \mathrm{N}(v_{it}|.5 + \delta_t + \alpha_i + \phi_{it} I_{it}, \sigma_\epsilon^2),$$

using the $\mathrm{N}(\cdot|\cdot,\cdot)$ notation for the normal density function (as in Gelman, Carlin, Stern, and Rubin 1995).

The Prior Distribution. The models for the district-level baselines and the candidate-specific incumbency effects are

$$p(\alpha|\sigma_\alpha) = \prod_{i=1}^{n} \mathrm{N}(\alpha_i|0, \sigma_\alpha^2)$$

and

$$p(\phi|\psi, \sigma_\phi) = \prod_{i=1}^{n} \mathrm{N}(\phi_{i1}|\psi, \sigma_\phi^2) \prod_{i:\ I_{i1}I_{i2}=-1} \mathrm{N}(\phi_{i2}|\psi, \sigma_\phi^2).$$

This second factor in $p(\phi|\psi, \sigma_\phi)$ counts only the districts in which incumbency status flipped between the two elections. It is necessary to define separate incumbency effects for the two consecutive elections in these districts.

We next assign noninformative hyperprior distributions to the remaining parameters,

$$p(\delta, \psi, \sigma_\alpha, \sigma_\phi, \sigma_\epsilon) \propto 1.$$

In practice, this is equivalent to assigning broad but proper distributions, such as uniform on $[-1, 1]$ for δ and ψ and uniform on [0, 1] for σ_α, σ_ϕ, and σ_ϵ. In contrast, noninformative uniform prior densities on $\log\sigma_\alpha$, $\log\sigma_\phi$, and $\log\sigma_\epsilon$ would lead to improper posterior densities (see, e.g., Gelman et al. 1995; Hobert and Casella 1996).

The Information Provided by the Incumbent Party Indicators. The model is nearly complete, but we must still include $p(I|\theta)$, the information supplied by the incumbency indicators $I_{it} = (1 - T_{it})P_{it}$. We follow previous researchers in ignoring any potential information in the treatment decision T_{it}; this is a reasonable choice given that there is no observed correlation between the decision to run for reelection and the previous year's election outcome.

The incumbent *party* indicators are another matter. P_{i2} provides no additional information in our model, because $P_{i2} = 1$ if and only if $v_{i1} > .5$ (excluding very rare events, such as special elections following death in office). However, P_{i1} provides information about v_{i0}—the previous election result, which is not included in our model—and thus, indirectly, about the baseline α_i. If $P_{i1} = +1$, then α_i is likely to be positive, and if $P_{i1} = -1$, then α_i is likely to be negative. [Recall that in (4), the baseline is defined relative to .5.] The information in P_{i1} comes into the likelihood as the probability of the observed P_{i1} given α_i. For convenience, we use the notation

$$\pi_P(\alpha_i) = \Pr(P_{i1} = 1|\alpha_i) = \Pr(v_{i0} > .5|\alpha_i).$$

The factor in the likelihood arising from the observable I_{i1}'s comprises factors of $\pi_P(\alpha_i)$ for the districts with $I_{i1} = 1$ and $(1 - \pi_P(\alpha_i))$, where $I_{i1} = -1$. We determine these probabilities recursively as follows.

From the model (4), we can write

$$v_{i0} = .5 + \delta_0 + \alpha_i + \epsilon_{i0} + \begin{cases} 0 & \text{if } I_{i0} = 0 \\ \phi_{i0} & \text{if } I_{i0} = 1 \\ -\phi_{i0} & \text{if } I_{i0} = -1. \end{cases}$$

The distribution of v_{i0} is then a mixture of three normals,

$$v_{i0} \sim \begin{cases} \mathrm{N}(.5 + \delta_0 + \alpha_i, \sigma_\epsilon^2) \\ \quad \text{with probability } \pi_{Tt} \\ \mathrm{N}(.5 + \delta_0 + \alpha_i + \psi, \sigma_\epsilon^2 + \sigma_\phi^2) \\ \quad \text{with probability } (1 - \pi_{Tt})\pi_P(\alpha_i) \\ \mathrm{N}(.5 + \delta_0 + \alpha_i - \psi, \sigma_\epsilon^2 + \sigma_\phi^2) \\ \quad \text{with probability } (1 - \pi_{Tt})(1 - \pi_P(\alpha_i)), \end{cases} \tag{5}$$

where $\pi_{Tt} = \Pr(T_{it} = 1)$, the probability that a district will have an open seat. In setting up the second and third of these normal distributions, we have assigned the candidate-specific incumbency effect the $\mathrm{N}(\psi, \sigma_\phi^2)$ distribution from the model, ignoring any information present in ϕ_{i1}. This simplification allows us calculate the probabilities $\pi_P(\alpha_i)$ in closed form. (Another option would be to simply include the data v_{i0} from the previous election, but this would simply push the problem back one step, because we would need to model P_{i0} given $v_{i,-1}$. In addition, including v_{i0} would restrict the applicability of the method, because then data from three consecutive elections would be needed to fit the model.)

We now can determine the probabilities $\pi_P(\alpha_i)$ from the normal distributions in (5):

$$\begin{aligned}\pi_P(\alpha_i) &= \Pr(v_{i0} > .5) \\ &= \pi_{Tt}\Phi\left(\frac{\delta_0 + \alpha_i}{\sigma_\epsilon}\right) \\ &\quad + (1 - \pi_{Tt})\pi_P(\alpha_i)\Phi\left(\frac{\delta_0 + \alpha_i + \psi}{\sqrt{\sigma_\epsilon^2 + \sigma_\phi^2}}\right)\end{aligned}$$

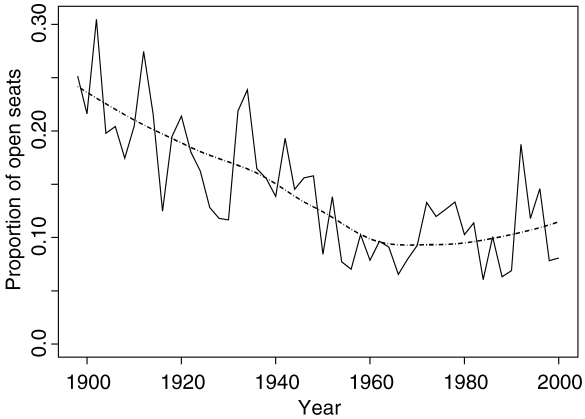

Figure 5. $\overline{T}_t$, the proportion of open seats in each congressional election in the twentieth century. The patterns are similar when considering Democrats and Republicans separately. The dotted line shows a smoothed estimate that we used as an estimate of π_{Tt} in computing (6). The smoothing was done using lowess (Cleveland 1979).

$$+ (1-\pi_{Tt})(1-\pi_P(\alpha_i))\Phi\left(\frac{\delta_0+\alpha_i-\psi}{\sqrt{\sigma_\epsilon^2+\sigma_\phi^2}}\right),$$

where Φ is the normal cumulative distribution function. We can solve for $\pi_P(\alpha_i)$ by

$$\pi_P(\alpha_i) = \left(\pi_{Tt}\Phi\left(\frac{\delta_0+\alpha_i}{\sigma_\epsilon}\right) + (1-\pi_{Tt})\Phi\left(\frac{\delta_0-\alpha_i+\psi}{\sqrt{\sigma_\epsilon^2+\sigma_\phi^2}}\right)\right) \Bigg/ \left(1-(1-\pi_{Tt})\Phi\left(\frac{\delta_0+\alpha_i+\psi}{\sqrt{\sigma_\epsilon^2+\sigma_\phi^2}}\right) + (1-\pi_{Tt})\Phi\left(\frac{\delta_0+\alpha_i-\psi}{\sqrt{\sigma_\epsilon^2+\sigma_\phi^2}}\right)\right). \quad (6)$$

To compute (6) as a function of α_i, we need σ_ϵ and σ_ϕ, which are part of our model, and π_{Tt} and δ_0, which must be computed externally. We estimate π_{Tt} by the historical proportion of open seats (see, e.g., Fig. 5 for congressional elections) and δ_0 by $\bar{v}_0 - .5$, the nationwide average Democratic vote in election 0, relative to .5. The average vote, $\bar{v}_0$, is the only information about election 0 that we use in our analysis. (In the rare scenario in which $\bar{v}_0$ were unavailable, we would simply use $\bar{v}_1$ as an estimate, correcting as best as possible for any national or statewide swings between elections 0 and 1.) We require district-level data only for the two consecutive elections that we have labeled 1 and 2.

The Complete Posterior Distribution. Finally, the joint posterior density is proportional to the product of all of the foregoing pieces,

$$p(\theta|v,I) \propto \prod_{i=1}^{n}\left(\prod_{t=1}^{2}\mathrm{N}(v_{it}|.5+\delta_t+\alpha_i+\phi_{it}I_{it},\sigma_\epsilon^2)\right) \times \mathrm{N}(\alpha_i|0,\sigma_\alpha^2)\mathrm{N}(\phi_{i1}|\psi,\sigma_\phi^2) \times \prod_{i:I_{i1}I_{i2}=-1}\mathrm{N}(\phi_{i2}|\psi,\sigma_\phi^2)\prod_{i:I_{i1}=1}\pi_P(\alpha_i) \times \prod_{i:I_{i1}=-1}(1-\pi_P(\alpha_i)). \quad (7)$$

3.2 Implementation

Given district-level data v, I from any pair of consecutive elections, we fit the model (4) using the Gibbs sampler and Metropolis algorithm, applied to the density function (7). We set up the computation in two steps. First, if we ignore the last two terms—those involving $\pi_P(\alpha_i)$—then the density (7) is a normal linear multilevel model, and all of its parameters can be updated using the Gibbs sampler. The linear parameters δ, α, ϕ, and ψ have a joint normal conditional distribution, and the variance parameters σ_α^2, σ_ϕ^2, and σ_ϵ^2 have independent inverse chi-squared distributions. The Gibbs sampler thus alternates between these two blocks of parameters.

The Gibbs sampler can be slow if parameters are highly correlated (see Gilks, Richardson, and Spiegelhalter 1996), and so we alter the algorithm in two ways to improve the speed of convergence. First, we run the Gibbs sampler for a short time to obtain a preliminary estimate of the variance components and then use them to approximate the posterior covariance matrix of the linear parameters δ, α, ϕ, and ψ conditional on this estimate. We use this approximate covariance matrix to rotate the space of the linear parameters and perform subsequent Gibbs updates on the approximately independent components of this rotated space. Our second improvement is to apply parameter expansion to the scales of the coefficients and the variance parameters, as described by van Dyk and Meng (2001).

However, the Gibbs sampler is just an approximation because it ignores the information in the incumbency indicators I_{i1}. We include this part of the posterior density by adding, after each full step of the Gibbs sampler, a Metropolis accept/reject step, as follows. Suppose that the parameter vector at the previous step was θ^s, which had been altered to a "candidate value" θ^* once all of the parameters had been updated. Under the Gibbs sampler, we would just set the new iteration value θ^{s+1} to the candidate, θ^*. With the Metropolis step, we compute the ratio

$$r = \frac{p(\theta^*|v,I)/g(\theta^*|v,I)}{p(\theta^s|v,I)/g(\theta^s|v,I)},$$

where p is defined in (7) and g is that same expression but omitting the factors of $\pi_P(\alpha_i)$ and $(1-\pi_P(\alpha_i))$; that is, g is the approximate distribution used in the Gibbs sampler updating. The next iteration of θ is then set to either the candidate value or its value at the previous step,

$$\theta^{s+1} = \begin{cases}\theta^* & \text{with probability } \min(r,1)\\ \theta^s & \text{otherwise.}\end{cases}$$

If this step always rejects, then the approximation that led to θ^* is very poor indeed; the acceptance rate is high when the approximation is close to the full model. In the Congressional elections analyses, the acceptance rate of this step ranges from about 50–80%, meaning that the approximation is reasonably close.

The full algorithm, including the improvements to the Gibbs sampler mentioned earlier, converges quickly. For example, a typical pair of congressional elections will have about 700 data points v_{it} (350 contested elections in each of 2 years) and about 800 parameters (α_i and ϕ_{i1} for all districts, ϕ_{i2} for all districts where $I_{i1}I_{i2} = -1$, and δ_1, δ_2, ψ, σ_ϵ, σ_α, and σ_ϕ). Four chains reached approximate convergence [i.e., the Gelman and Rubin (1992) convergence diagnostic $\sqrt{\hat{R}}$ was <1.2] after 4,000 iterations (keeping the last 2,000 iterations of each chain), taking 4 minutes on a PC to fit the model. Computations were performed in the statistical language S–PLUS, using the `apply` function to avoid internal looping.

Once our simulations have converged, we summarize our inferences by posterior medians and interval estimates such as 95% confidence intervals computed from the 2.5% and 97.5% points of the simulations. We illustrate this in Section 4 for the congressional elections analysis.

In fitting our model to pairs of elections, we exclude districts that are uncontested in either election. An alternative would be to model the uncontested elections as missing data, but we avoid the additional complexity that this would bring to our model. Gelman and King (1990) and Ansolabehere and Snyder (2002b) estimated the selection of contested elections to reduce the estimated effect of incumbency by no more than 1%. In addition, we exclude elections after redistricting (i.e., years ending in 2); analyzing these pairs would require additional information on which districts were redrawn, as was provided by Ansolabehere et al. (2000).

3.3 Model Checking

After fitting an elaborate model, it is important to check its fit to data. We do this by simulating replicated data sets conditional on the estimated model parameters and comparing these with the observed data. Systematic discrepancies between data and simulations represent aspects of the data that are not captured by the model (Gelman, Meng, and Stern 1996).

We illustrate this predictive model checking for the 1986–1988 data displayed in Figure 3, fitted by three models: the simple regression (2), the multilevel model (4) fit by the Gibbs sampler ignoring the information provided by I_{i1} [i.e., ignoring the factors of π_P and $(1 - \pi_P)$ in the posterior density (7)], and the multilevel model (4) fit correctly using the full posterior density.

We start with the simple regression model (2) of Gelman and King (1990). As discussed in Section 2.2, election data display an interaction with incumbency that is not captured in this model. Figure 6 shows replicated data from the model, which do not capture the systematic differences between open-seat and incumbent-contested elections. Plots for other election years are similar.

We next consider the multilevel model as fitted by the Gibbs sampler, ignoring the information provided by I_{i1}. This mistaken model is interesting to study for two reasons. First, it was our first try at fitting model (4); we constructed the posterior density too hastily, ignoring those factors with π_P and $(1-\pi_P)$. It is instructive to see whether a model check would catch this omission. Second, all realistic models in the social sciences ignore some potentially relevant information—regressions have omitted variables, predictive relationships are not exactly linear, error distributions are not truly normal, and so forth. Thus it is important to determine whether this particular simplification (whether intentional or unintentional) affects the model's ability to fit the data.

Figure 7 shows several data sets replicated from the incorrectly fit (4) that ignores the information in I_{i1}. Each simulation is created by taking a random draw of the parameters in the model from one of the Gibbs sampler iterations and then, for each district i, sampling new error terms ϵ_{it} to create new data v_{i1} and v_{i2}. The two election outcomes must be simulated in sequence because v_{i1} affects P_{i2}, which in turn affects the predicted value of v_{i2}.

Compared with the real data in Figure 3, the replications in Figure 7 look wrong, with the dots having too low a correlation—an overly puffy appearance. It is good that we looked at these plots, because by doing so we found a serious misfit in our context: By not using the information about α_i derived from knowing the current incumbency status, the model that ignores the information in I_{i1} ends up attributing too much of the variation in the votes v_i to incumbency effects and not enough to variation in the baseline α_i.

The erroneous model would cause us to drastically overestimate incumbency advantage. Although π_P does not contribute much information at each iteration, its aggregate effect is to change the estimates by a fair amount. For example, when we fit the model without π_P, we were estimating the average incumbency effect in recent years as about 15% (compared with the 8% that we actually got). The model checking of Figure 7 was important not because we are particularly concerned about year-to-year correlation or "puffiness" in the graphs, but because the graphical display turned out to be an effective way to reveal a model flaw that we otherwise had not noticed.

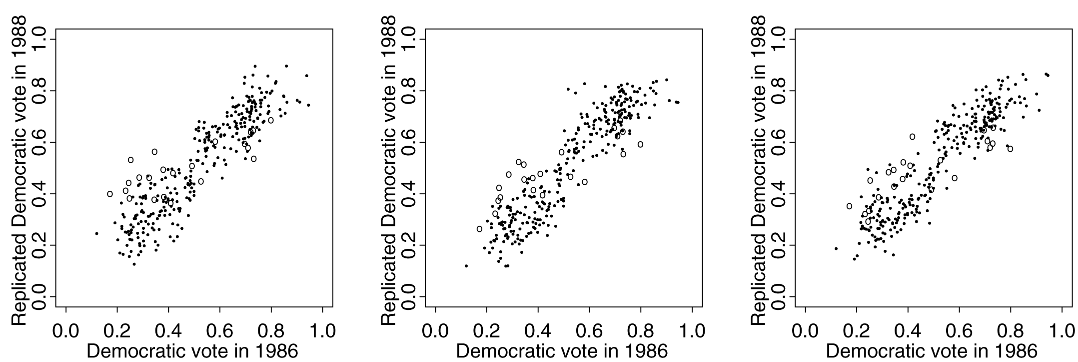

Figure 6. Replicated data sets simulated from the regression model (2) as fit to data from 1986 and 1988. Dots represent elections with incumbents running in 1988 and circles represent open seats. Compare to the actual data in Figure 3. The circles in the actual data have a much flatter slope than in the replications.

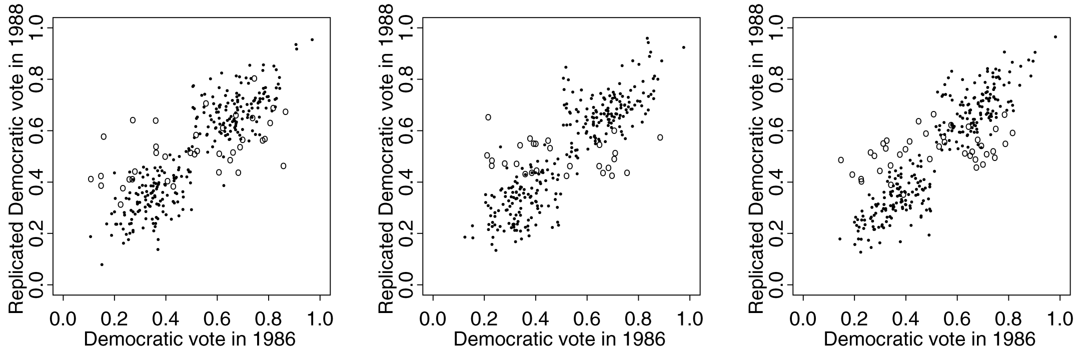

Figure 7. Replicated data sets simulated from the regression model (4) as fit to data from 1986 and 1988, using the Gibbs sampler *ignoring the information in* P_{i0}. Compare with the actual data in Figure 3. The dots are pressed much closer to a 45-degree line in the actual data than in the replications.

Finally, we check the fit of the full model (4) using the full posterior density (7). Figure 8 displays replicated data sets, which look similar to the real data.

Fitting the data is only an intermediate step toward our ultimate goal of estimating incumbency effects. However, the potential problems in estimating the model sloppily, as illustrated in Figure 7, demonstrate why checking the fit is important. We like the model (4) because it captures the important features of electoral data while modeling incumbency in a politically reasonable way. The aspect of the data that motivated the complexity of the model—the interaction between vote and incumbency—allows us to learn about variation between incumbents.

3.4 Connection to Other Estimates of Incumbency Advantage

As discussed earlier, our multilevel model (4) generalizes classical lagged regressions. In particular, if σ_ϕ (the variation in the incumbency advantage) is set to 0, then our estimate is similar to that of Gelman and King (1990) in assuming parallel slopes for incumbents and open seats, also incorporating the improvements of Cox and Katz (1996) and Levitt and Wolfram (1997), who adjusted for incumbency status at election 1. Lee (2003) studied the slightly different problem of estimating how the winner election outcome at time 1 affects the vote share at election 2.

Looked at another way, our model has strong connections to sophomore surge and retirement slump (Alford and Brady 1988), which estimate incumbency advantage by the difference in vote (after adjusting for national swings) between 2 years within districts in which incumbency status changes. Subtracting expression (4) evaluated at $t = 2$ and $t = 1$ yields

$$v_{i2} - v_{i1} = (\delta_2 - \delta_1) + (\phi_{i2} I_{i2} - \phi_{i1} I_{i1}) + (\epsilon_{i2} - \epsilon_{i1}).$$

Given our constraints on ϕ_{it} (as described in Sec. 3.1), we can write this as

$$v_{i2} - v_{i1} = \text{national swing} + \Delta\phi_i + e_i, \tag{8}$$

where $e_i = \epsilon_{i2} - \epsilon_{i1}$ is an independent error term with mean 0 and variance $2\sigma_\epsilon^2$, and $\Delta\phi_i$ is a difference in candidate-specific incumbency advantages, with expectation $\psi(I_{i2} - I_{i1})$. Thus, after correcting for national swing, the change in vote for retirements or for sophomores is a random variable that should equal ψ in expectation.

But sophomore surge and retirement slump are biased measures of the average incumbency advantage ψ because they are based on a nonrandom selection of districts (Gelman and King 1990). In the notation of Section 3.1, the simple estimates based on (8) fail because they ignore the information about ϕ_{i1} that is present in v_{i1} and I_1.

4. APPLICATION TO U.S. CONGRESSIONAL ELECTIONS

We fit the full model (4) to elections in the U.S. House of Representatives in the twentieth century, skipping those pairs of

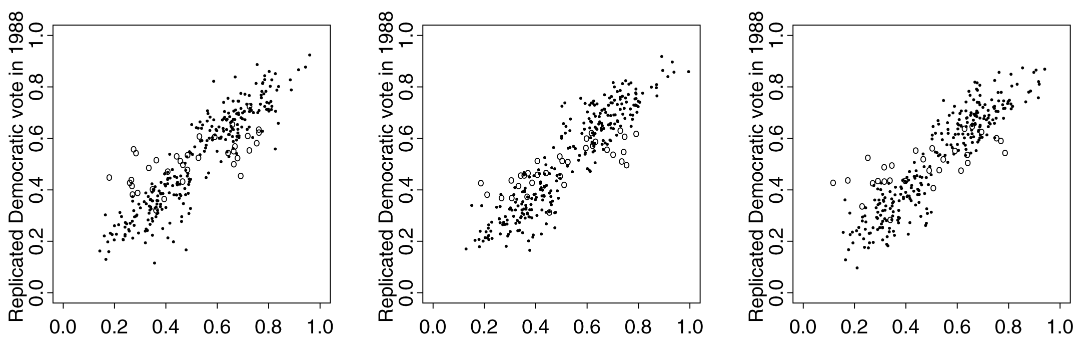

Figure 8. Replicated data sets simulated from the regression model (4) as fit to data from 1986 and 1988, fitting the full model using the Gibbs sampler and the Metropolis algorithm. The replicated data look similar to the actual data in Figure 3.

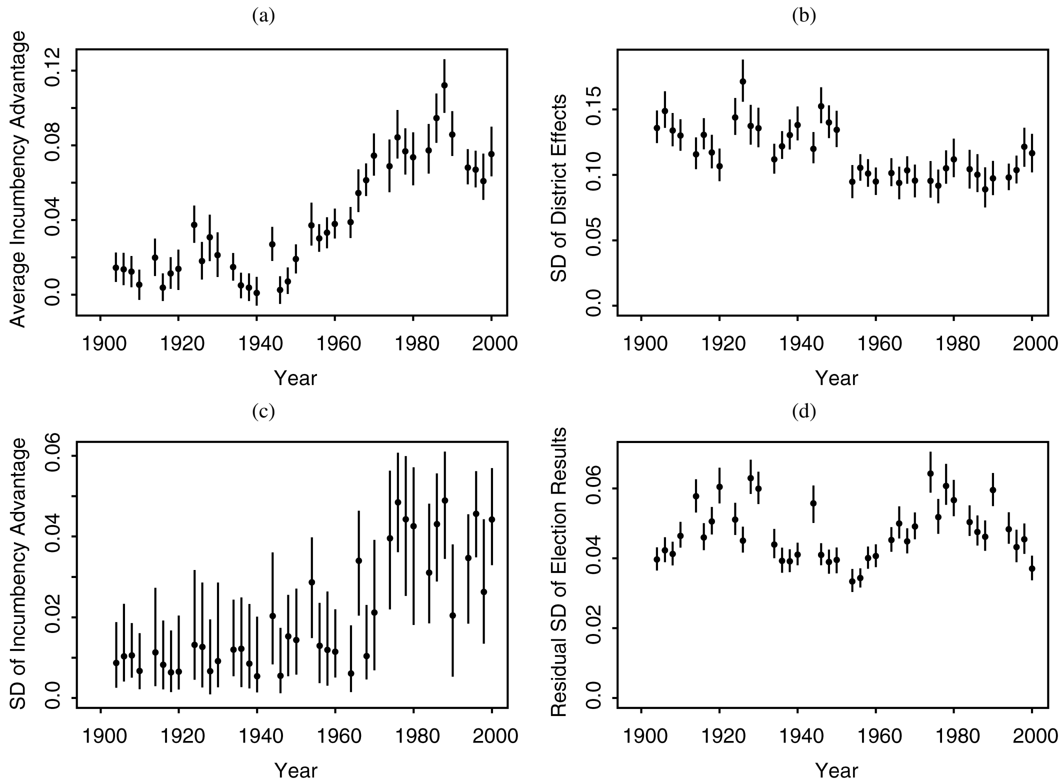

Figure 9. Estimates and 95% posterior intervals for the hyperparameters ψ, σ_α, σ_ϕ, and σ_ϵ in the incumbency model, estimated for each pair of elections for the U.S. House of Representatives in the twentieth century. (a) Average incumbency advantage. (b) Standard deviation of district effects. (c) Standard deviation of incumbency advantage. (d) Residual standard deviation of election results. Election years immediately following redistricting (those ending in "2") are excluded. The vertical axes for the four graphs are on different scales. The estimate of the average incumbency advantage (a) over time is much more stable than from the usual regression estimate (see Fig. 2).

elections that straddled a redistricting (i.e., 1900–1902, 1910–1912, and so forth). For each pair, we fit the model to all districts contested in both years (typically, about 350 out of 435). We are not particularly interested in the parameters for the individual districts (except for the purposes of checking model fit, as described in Sec. 3.3), and so we begin our summary of inferences with estimates of the hyperparameters of the model: the average incumbency advantage ψ and the standard deviations $\sigma_\phi, \sigma_\alpha$, and σ_ϵ, representing the variation in incumbency effects, district baselines, and year-to-year variations within districts.

Figure 9 shows the posterior mean estimates and standard deviations of each of these parameters over time. The estimate of the average incumbency effect ψ is appealingly smooth with low uncertainty, especially compared with previous estimates in the literature (see, e.g., Gelman and King 1990). Our estimates are more precise because we fully use the information from both election years, compared with regression methods that treat the first election merely as a lagged predictor and methods such as sophomore surge and retirement slump that only use a subset of the districts.

The other plots of Figure 9 reveal that σ_ϕ, the standard deviation of the incumbency advantage, is estimated with much less precision than the average effect. The total variance of the vote across all of the districts can be well estimated, but it is more difficult for the model to partition this into district baseline, candidate effects, and year-to-year variation. Thus it is important to display uncertainties in these plots so that we do not overinterpret short-term fluctuations in the estimates.

We are in agreement with previous researchers that the incumbency advantage in Congress was low but positive in the first half of the century and increased rapidly during the 1950s and 1960s, and has remained relatively high for the past 30 years. However, our estimate of the average level of incumbency advantage in recent years is about 8%, compared with the usual estimate from the literature of about 10%. The estimates from our model have much lower uncertainties and are much more stable than the simple regression estimates (compare to Fig. 2), which should be especially important when estimating incumbency advantages in shorter time series or in individual states for which less information is available.

We find that the increase in average incumbency advantage ψ was followed, with about a 15-year lag, by a dramatic increase in σ_ϕ, the variation of incumbency advantage—that is, an increase in the variation of candidate effects. The estimate of the parameter σ_ϕ varies dramatically from year to year, indicating that the data do not supply much information about this parameter in any given year. (This paucity of information also shows up in the wide fluctuations in the estimated lagged regression slope for open seats, as indicated by the dotted line in Fig. 4.) Meanwhile, districts have become slightly more similar to one another in their baselines (i.e., σ_α decreased from about 15% to 12% over the century) and the residual variation σ_ϵ has cycled between about 4% and 6% over the years.

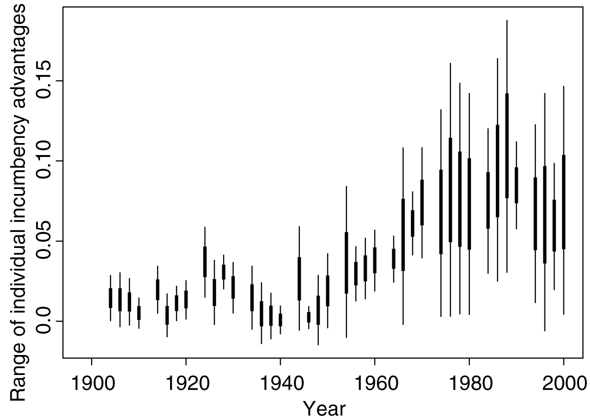

Figure 10. Estimated distributions of incumbency effects ϕ_{i1}, shown by medians, 50% ranges, and 90% ranges, for the twentieth-century U.S. House of Representatives. These ranges represent the estimated variation among incumbents, not uncertainties in the mean level of incumbency. The variation in incumbency effects began to increase about 15 years after the increase in the mean level, with current incumbency effects varying between 0 and 15%.

Finally, we focus on the distribution of incumbency effects by plotting the estimated 5%, 25%, 50%, 75%, and 95% quantiles of the incumbency effects ϕ_{i1} as they vary among districts. We calculate these quantiles for each posterior simulation draw, thus avoiding the problem noted by Louis (1984) that a set of Bayes point estimates tends to be less variable than the set of underlying parameters. Figure 10 displays the posterior mean of each quantile (estimated medians, 50% ranges, and 90% ranges for the incumbency effects across the country) for each estimation of the model (i.e., each pair of consecutive elections). We see a mid-century increase in average incumbency advantage and increased variation starting in the 1960s. In recent years, we estimate that individual incumbency advantages range between 0 and 15% of the vote.

5. DISCUSSION

5.1 Incumbency Advantage

We have set up a full probability model of incumbency effects and applied it to the U.S. House of Representatives in the twentieth century. The new model offers several advantages over previous regression-based estimates: (1) a framework that allows estimation of the variation of incumbency effects as well as their mean level; (2) more precise and stable estimates of the mean level itself; (3) a decomposition of between-district heterogeneity into variation in baselines, incumbency advantage, and year-to-year variability; and (4) better fit to actual election data. These four features are synergistic; expanding the model to add a new component of variation allows it to incorporate more information already present in the data (thus giving more precise estimates) and also to better fitted those aspects of the data. The model is multilevel and was fitted with Bayesian methods using the Gibbs sampler and the Metropolis algorithm. Predictive simulations confirmed the fit.

Now that the model has been programmed (and the program and data are publicly available), it can be used for other electoral systems (e.g., state legislatures). Relatively simple alterations would allow inclusion of additional district-level information, such as other election results and candidate quality, as regression predictors in (4), as in the analyses of Ansolabehere et al. (2000) and Ansolabehere and Snyder (2002a). This could be a powerful tool for research into how differences between candidates lead to different election outcomes and, also should improve the unstable estimates of the variance parameters. These also could be improved by analyzing more than two elections simultaneously, or by smoothing the estimates over time.

5.2 Treatment Interactions in Before–After Studies

Finally, the model developed here is a special case of a more general approach to before–after studies—experiments or observational studies in which a measurement is taken before and after the treatment. Our example is more complicated than the usual before–after setting, because the "treatment" of an open-seat election can happen before election 1 as well; thus it can be more accurately described as a part of a time series in which interventions are possible at each point. However, observational studies commonly allow for the possibility of multiple interventions or choice points while still summarizing measurements before and after some time period of interest.

In various social science applications, before–after correlations have been found to be higher for control units than for treated units (see Gelman 2004), and this interaction can be important—sometimes more important than the main effect of the treatment. (For an example in the effects of legislative redistricting, see figs. 3 and 4 of Gelman and King 1994.) Yang and Tsiatis (2001) showed how joint modeling of pretreatment and posttreatment outcomes can improve efficiency and robustness of estimation. Our incumbency example illustrates how variation in treatment effects can both explain the interaction and also parameterize it more usefully, leading to new insights about the phenomenon under study.

[Received March 2003. Revised August 2003.]

REFERENCES

Achen, C. (1986), *The Statistical Analysis of Quasi-Experiments*, Berkeley, CA: University of California Press.

Alford, J. R., and Brady, D. W. (1988), "Partisan and Incumbent Advantage in U.S. House Elections, 1846–1986," report, Center for the Study of Institutions and Values, Rice University.

Alford, J. R., and Hibbing, J. R. (1981), "Increased Incumbency Advantage in the House," *Journal of Politics*, 43, 1042–1061.

Ansolabehere, S., and Snyder, J. M. (2002a), "The Incumbency Advantage in U.S. Elections: An Analysis of State and Federal Offices, 1942–2000," *Election Law Journal*, 1, 315–338.

——— (2002b), "Using Term Limits to Estimate Incumbency Advantages When Officeholders Retire Strategically," technical report, Massachusetts Institute of Technology, Dept. of Political Science.

Ansolabehere, S., Snyder, J. M., and Stewart, C. (2000), "Old Voters, New Voters, and the Personal Vote: Using Redistricting to Measure the Incumbency Advantage," *American Journal of Political Science*, 44, 17–34.

Campbell, J. E. (2002), "Is the House Incumbency Advantage Mostly a Campaign Finance Advantage?" State University of New York at Buffalo, Dept. of Political Science.

Cleveland, W. S. (1979), "Robust Locally Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association*, 74, 829–836.

Cox, G. W., and Katz, J. (1996), "Why Did the Incumbency Advantage Grow?" *American Journal of Political Science*, 40, 478–497.

Cox, G. W., and Morgenstern, S. (1993), "The Increasing Advantage of Incumbency in U.S. States," *Legislative Studies Quarterly*, 18, 495–514.
Erikson, R. S. (1971), "The Advantage of Incumbency in Congressional Elections," *Polity*, 3, 395–405.
Gelman, A. (2004), "Treatment Effects in Before–After Data," in *Applied Bayesian Modeling and Causal Inference From Incomplete Data Perspectives*, eds. A. Gelman and X. L. Meng, London: Wiley, pp. 195–202.
Gelman, A., and King, G. (1990), "Estimating Incumbency Advantage Without Bias," *American Journal of Political Science*, 34, 1142–1164.
——— (1994), "Enhancing Democracy Through Legislative Redistricting," *American Political Science Review*, 88, 541–559.
Gelman, A., and Rubin, D. B. (1992), "Inference From Iterative Simulation Using Multiple Sequences" (with discussion), *Statistical Science*, 7, 457–511.
Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995), *Bayesian Data Analysis*, London: Chapman & Hall.
Gelman, A., King, G., and Boscardin, W. J. (1998), "Estimating the Probability of Events That Have Never Occurred," *Journal of the American Statistical Association*, 93, 1–9.
Gelman, A., Meng, X. L., and Stern, H. S. (1996), "Posterior Predictive Assessment of Model Fitness via Realized Discrepancies" (with discussion), *Statistica Sinica*, 6, 733–807.
Gilks, W. R., Richardson, S., and Spiegelhalter, D. (1996), *Markov Chain Monte Carlo in Practice*, London: Chapman and Hall.
Hobert, J. P., and Casella, G. (1996), "The Effect of Improper Priors on Gibbs Sampling in Hierarchical Linear Mixed Models," *Journal of the American Statistical Association*, 91, 1461–1473.
Jacobson, G. C. (2000), "Reversal of Fortune: The Transformation of U.S. House Elections in the 1990s," in *Continuity and Change in House Elections*, eds. D. W. Brady, J. F. Cogan, and M. P. Fiorina, Stanford: Stanford University Press, pp. 10–38.
King, G., and Gelman, A. (1991), "Systemic Consequences of Incumbency Advantage in U.S. House Elections," *American Journal of Political Science*, 35, 110–138.
Krashinsky, M., and Milne, W. J. (1993), "The Effects of Incumbency in U.S. House Elections," *Legislative Studies Quarterly*, 18, 321–344.
Lee, D. S. (2003), "Randomized Experiments From Non-Random Selection in U.S. House Elections," technical report, University of California, Berkeley, Dept. of Economics.
Levitt, S. D., and Wolfram, C. D. (1997), "Decomposing the Sources of Incumbency Advantage in the U.S. House," *Legislative Studies Quarterly*, 22, 45–60.
Louis, T. A. (1984), "Estimating a Population of Parameter Values Using Bayes and Empirical Bayes Methods," *Journal of the American Statistical Association*, 78, 393–398.
Mulligan, C. B., and Hunter, C. G. (2001), "The Empirical Frequency of a Pivotal Vote," Working Paper 8590, National Bureau of Economic Research.
Payne, J. L. (1980), "The Personal Electoral Advantage of House Incumbents," *American Politics Quarterly*, 8, 375–398.
van Dyk, D. A., and Meng, X. L. (2001), "The Art of Data Augmentation" (with discussion), *Journal of Computational and Graphical Statistics*, 10, 1–111.
Yang, L., and Tsiatis, A. A. (2001), "Efficiency Study of Estimators for a Treatment Effect in a Pretest–Posttest Trial," *The American Statistician*, 55, 314–321.

Comment

Jonathan N. KATZ

The incumbency advantage is one of the most widely studied phenomena in political science. In fact, it is one of the few quantities of interest in the field where there is relative agreement not only on its directionality, but also on its relative size. Thus I was somewhat dubious that any significant additions remainder to be made to our understanding; however, Gelman and Huang have in fact made two important contributions.

First, they have recast the estimation problem into the more modern causal framework of examining differences between treated and control groups, where the treatment/intervention is an incumbent retiring, thereby producing an open seat in an election. This by itself is not that novel. The earliest estimates of incumbency advantage, the "sophomore surge" and "retirement slump" (Erikson 1971), were naïve estimates of this sort of treatment effect, but these are biased because they use only a selected subsample of the data for estimation (Gelman and King 1990). However, in recasting this problem from the more typical regression framework, Gelman and Huang are able to bring additional information to the estimation problem—the lagged incumbent party indicator—which provides information about v_{i0} that had previously been neglected in estimation. Second, they use a Bayesian multilevel model to allow for random variation in the value of incumbency across individual incumbents. The previous studies all focused only on the average effect, ignoring any variance across districts or candidates.

My comments focus each of these contributions in turn.

Jonathan N. Katz is Professor of Political Science, Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125 (E-mail: *jkatz@caltech.edu*). The author thanks the Center on Advanced Study in the Behavioral Science, where this comment was written.

1. OBSERVATIONAL STUDIES AND NONRANDOM ASSIGNMENT

The bane of any observational study is nonrandom assignment to the treatment and control groups. As is well known, the lack of random assignment can cause confounding factors to contaminate our estimates. The hope is that by conditioning on observables (in this case lagged vote and lagged party incumbency status), we can fix the nonrandom assignment problem. Gelman and Huang are well aware of this, and they claim that the treatment indicator, T_{it}, is essentially random across Congressional districts. Their evidence for this claim is that there is no correlation between margin of vote and the probability of seeking reelection.

Unfortunately, this view is at odds with the generally accepted view that Congressional candidates are strategic in their entry and exit decisions (Jacobson and Kernell 1983; see also Cox and Katz 1996, 2002). In fact, Gelman and Huang's evidence is convincing only if one assumes that the decision to retire is based only on observables available to an outside analyst. But what if candidates have better information than this to forecast their party's prospects in the district in the next election?

In a study with Gary Cox, we examined the loss an incumbent party incurs when an incumbent "retires" (more correctly exits) voluntarily or involuntarily—either by death or primary defeat (Cox and Katz 2002). If there is no strategic retirement,

Journal of the American Statistical Association
June 2008, Vol. 103, No. 482, Applications and Case Studies
DOI 10.1198/016214507000000635

Table 1. Incumbent party's vote loss for voluntary and involuntary retirements in U.S. House elections, 1946–1998

Type of open seat	1946–1964	1966-1998
Voluntary	2.66	7.42
	(.41)	(.40)
Involuntary	.73	2.90
	(.74)	(1.01)

as is claimed by Gelman and Huang, then there should not be a significant difference in the loss that the incumbent party incurs between these two types of open seats. The results of our analysis, which controlled for both lagged vote and year effects, is presented in Table 1. We needed to pool across years because of the small number of involuntary exits, and we chose the periods to correspond to periods in which the incumbency advantage was roughly constant in size. In the first period, when the incumbency advantage is relatively low, there is no significant difference between the two types of open seats. But in the latter period, when the estimated incumbency advantage is large, there is a strong difference between these two types of open seats, which is consistent with strategic exits and nonrandom treatment assignment.

In principle, the information from the involuntary exits by incumbents could be incorporated into the Gelman and Huang framework to correct for this part of nonrandom assignment. Unfortunately, there remains the problem of strategic entry of challengers. I know of no natural experiment that can be used to correct for this effect on the estimation.

2. VARIATION IN INCUMBENCY ADVANTAGE

It is natural to think that the incumbency advantage might vary between incumbents because, for example, some may be better campaigners than others. The Gelman and Huang model, therefore, offers a real improvement over the earlier approaches that assumed a constant incumbency effect. But their analysis misses at least one important source of systematic variability in incumbency advantage: It may vary by party.

Figure 1, based on analysis by Cox and Katz (2002), presents the estimated difference between the Republican and Democratic incumbency advantage using the Gelman and King (1990) regression model but breaking the symmetry constraint that forces incumbents of both parties to have the same size advantage. Positive values indicate that the Republican incumbency advantage is larger, whereas negative values indicate that the Democratic incumbency advantage is larger. The figure suggests that after 1966, there may be significant differences in the incumbency advantage by party, with five of eight elections showing a significant partisan difference. In fact, using the new Gelman and Huang model, which generally gives more precise estimates, the difference may be significant in more elections. It should be relatively straightforward to allow for this systematic difference within the Gelman and Huang framework.

Why should there be a partisan difference in the size of the incumbency advantage? The complete argument was presented by Cox and Katz (2002), but the basics of it relate to the importance of candidate quality. Suppose, following Jacobson (1987) and Cox and Katz (1996), that the incumbency advantage increased because candidate quality became a more important determinant of electoral outcomes beginning in the 1960s. By this line of thinking, the party with the larger incumbency advantage will be the one that has more trouble finding (a) electorally experienced nonincumbents to defend their districts after their incumbents retire and (b) electorally experienced challengers to attack districts vacated by the other party's incumbents. Such a party will more often find, after its incumbents retire, that it has fielded an inexperienced defender, faces an experienced challenger, or both. Its average vote share accordingly will drop rather precipitously, especially if candidate characteristics matter more in determining outcomes. The party that is better at recruiting candidates, in contrast, will more often find that it has fielded an experienced defender, faces an inexperienced challenger, or both after its incumbents retire, and thus its vote share will hold up better.

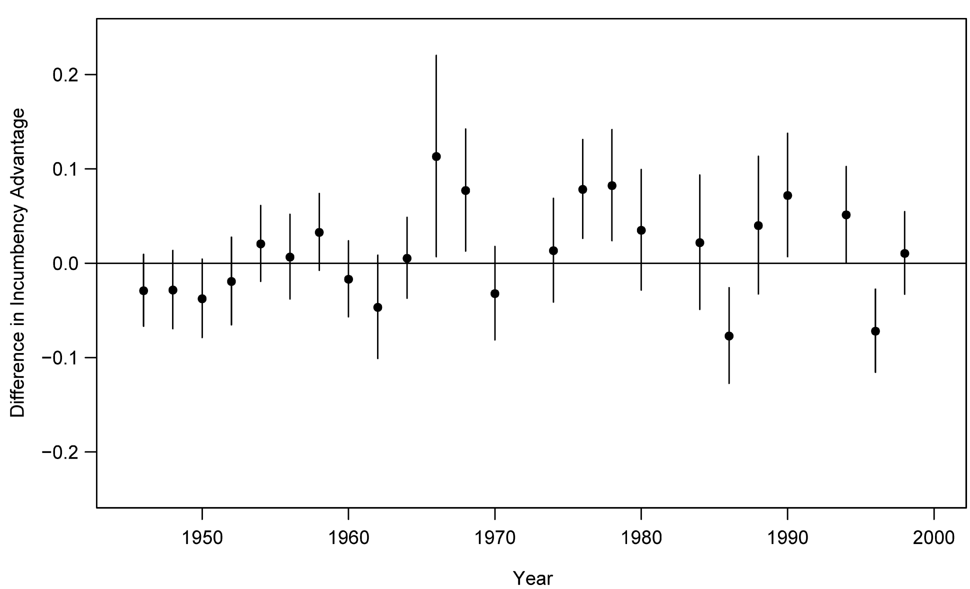

Figure 1. Estimated difference between Republican and Democratic incumbency advantage with 95% confidence intervals from a regression controlling for lagged vote, lagged incumbency status, and year effects.

ADDITIONAL REFERENCES

Cox, G. W., and Katz, J. N. (2002), *Elbridge Gerry's Salamander: The Electoral Consequences of the Reapportionment Revolution*, New York: Cambridge University Press.

Jacobson, G. C. (1987), "The Marginals Never Vanished: Incumbency and Competition in Elections to the U.S. House of Representatives, 1952–82," *American Journal of Political Science*, 31, 126–141.

Jacobson, G. C., and Kernell, S. (1983), *Strategy and Choice in Congressional Elections* (2nd ed.), New Haven, CT: Yale University Press.

Comment

James S. HODGES

I am old enough to remember the last 40 years of elections in Gelman and Huang's (GH's) data set, so it is great fun to pore over their plots and spin stories about why some point is unusually high or low. I am also old enough so that in my first year of graduate school (1981), my mathematical statistics professor could accurately note that although Bayesian theory was very nice, one could not really do anything with it. GH's article is more proof (if we still need any) that those days are gone.

Indeed, this article illustrates nicely how a Bayesian approach can be easier than the alternatives. Bayesian machinery is well suited to individual-specific treatment effects (GH's ϕ_{it}), an idea whose time has come. (So-called "empirical Bayes" analysis is just frequentist analysis of random-effects models and thus is hobbled, like frequentist analyses, by the need to maximize intractable integrals.) Another example of the power of Bayes is GH's treatment of the incumbent party indicators, I_{i0}. This shows what some call the "modularity" of Bayesian analysis, in which a model's unknowns provide "sockets" into which higher-level models for those unknowns can be plugged. In a world with huge, complex data sets, modularity is a natural tactic and is becoming progressively easier with tools like WinBUGS's graphical model specification (*www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml*). This same example also shows that the Bayesian millennium has not yet quite arrived, because GH had to ignore some information in the I_{i0} for the sake of tractability. Presumably such barriers will continue to fall as computing power increases, but for every new chip, someone will always find a bigger problem.

Still, GH's setup suggests—and modern Bayesian computing permits—a wealth of possibilities for adding information and thus power to this analysis. For example, figure 9(a) suggests smoothing the average incumbency advantage ψ_t over time, that is, adding a piece of information that the ψ_ts change fairly smoothly over time and letting the data specify "fairly smoothly" through, say, a pairwise-difference model (Besag, Green, Higdon, and Mengersen 1995). And so on.

However... Although I celebrate the power of Bayes, I must also rain on the Bayesian parade a bit. The deceptive simplicity of Markov chain Monte Carlo (MCMC) can foster a "damn the torpedoes, full steam ahead" confidence that is entirely misplaced. It is misplaced because greater model complexity means more instances in which vague, qualitative information ("the ψ_ts change fairly smoothly over time") must be made specific in an equation or probability distribution. Here I focus on one aspect of complexity, the variance structure (GH's σ_ϕ, σ_α, and σ_ϵ), because it affects many useful models and (surprise!) is my research area.

The difficulty with many-variance models comprises at least three problems tangled together: parameterizing the variance structure, putting priors on the parameters, and drawing robust, efficient MCMC samples. These problems arise because in richly parameterized models like GH's, the data often provide little information about higher-level variances (σ_ϕ and σ_α), so the parameterization and prior are important. Indeed, we do not even know how to determine whether a dataset provides "a lot of" information about the variance parameters. Reich and Hodges (2007) explored this issue for the general two-variance model, and it is a thorny one, even for this simple class of models. Depending on the parameterization and model, it can happen that the variance parameters are only poorly identified, but that identification improves little with increased sample size (Reich, Hodges, and Carlin 2007).

Put plainly, we do not know much about posteriors of variance-structure parameters in richly parameterized models. Even when using gamma priors on precisions, it is not uncommon to be surprised by bimodal posteriors (Wakefield 1998), which arise readily even in simple models with a unimodal likelihood component (Liu and Hodges 2003). It is easy to construct innocuous-looking models and parameterizations with quite strange posteriors (Reich et al. 2007), where the only hint of a problem is an MCMC with high lagged autocorrelations. But 15 years after Gelfand and Smith (1990) opened the modern Bayesian era, a gamma prior on the precisions is still almost a default. A gamma(ϵ, ϵ) prior is a popular "vague" prior because it has mean 1 and variance $1/\epsilon$, although the two most common ϵ's, .01 and .001, give priors with 95th percentiles .34 and 3×10^{-20}, which are anything but vague. We need alternative priors, and we need to know more about how they affect posteriors.

Gelman has proposed at least two alternative priors, both on the standard deviation: a folded noncentral t on the standard deviation (Gelman 2006) and a flat prior, used here. The latter has the advantage that the standard deviation is on the same

James S. Hodges is Associate Professor, Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN 55455 (E-mail: *hodges@ccbr.umn.edu*).

Journal of the American Statistical Association
June 2008, Vol. 103, No. 482, Applications and Case Studies
DOI 10.1198/016214507000000644

scale as the measurements, and the improper version is invariant to scale changes. Its disadvantage is that little is known about it. My doctoral student, Yi He, has compared four parameterizations, each with its own reference prior, according to the bias and mean squared error of the posterior mean and median and the coverage of 95% intervals; the results suggest that for some problems, the *proper* flat prior on standard deviations is highly sensitive to the upper end of the sample space.

Thus, although I very much enjoyed this article, I also would very much like to see what happens with other priors on the variance structure. We just do not know enough about these models and priors. This problem is wide open—calling all graduate students!—and I expect that practice will change radically by the time I retire.

I must note—in case the few remaining Bayes-haters are feeling cocky right now—that these difficulties are not solved by maximizing the likelihood instead of multiplying it by a prior and computing integrals. It is very common for variances to have a likelihood or restricted likelihood that is maximized at 0 but quite flat. In such cases, a point estimate of 0 is the wrong answer, not to mention inconvenient, because in standard software it is accompanied by a zero standard error, if any. In this case, integrating is clearly better than maximizing. The question is not whether to maximize or integrate, but whether general purpose reference prior(s) can be found giving reasonably calibrated posterior intervals and good point estimates.

ADDITIONAL REFERENCES

Besag, J., Green, P., Higdon, D., and Mengersen, K. (1995), "Bayesian Computation and Stochastic Systems" (with discussion), *Statistical Science*, 10, 3–66.

Gelman, A. (2006), "Prior Distributions for Variance Parameters in Hierarchical Models," *Bayesian Analysis*, 1, 515–533.

Gelfand, A. E., and Smith, A. F. M. (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409.

Liu, J., and Hodges, J. S. (2003), "Posterior Bimodality in the Balanced One-Way Random-Effects Model," *Journal of the Royal Statistical Society*, Ser. B, 65, 247–255.

Reich, B. J., and Hodges, J. S. (2007), "Identification of the Variance Components in the General Two-Variance Linear Model," *Journal of Statistical Planning and Inference*, in press.

Reich, B. J., Hodges, J. S., and Carlin, B. P. (2007), "Spatial Analyses of Periodontal Data Using Conditionally Autoregressive Priors Having Two Types of Neighbor Relations," *Journal of the American Statistical Association*, 102, 44–55.

Wakefield, J. (1998), Discussion of "Some Algebra and Geometry for Hierarchical Models, Applied to Diagnostics," by J. S. Hodges, *Journal of the Royal Statistical Society*, Ser. B, 60, 523–527.

Rejoinder

Andrew GELMAN

We published our article in a statistical journal because our main goal was to propose a new method for analyzing before–after data in observational studies. The most common approach is to take simple before–after differences or, one step beyond, to include the before measurements as regression predictors (as was done in Gelman and King (1990) for the incumbency example). The two key points of our article are that (1) in a setting such as incumbency with ongoing "treatments," the before data contain extra information not used in the regression, and, especially, (2) data at two time points allow us to estimate variation, as well as the mean level, of the treatment effect. This is all separate, or in addition to, other concerns about inference from observational studies (and such methods as instrumental variables, matching, regression discontinuity, and selection models), which we do not consider here.

As always in statistics, the general point can be understood only with reference to a real example (not just real data, but also a real problem for which there is outside interest in the answer). Also, as always, the particular example that we study has its own challenges.

We were lucky to get for our article two quite different discussants: Jonathan Katz, a quantitative political scientist who has done important work on understanding the incumbency advantage, and Jim Hodges, a statistician who has done innovative work on constructing and understanding multilevel models.

ADDING COMPLEXITY TO THE INCUMBENCY–ADVANTAGE MODEL

Katz discusses two issues—strategic retirement and interaction of party and incumbency effect—that are not included in our model. In the context of our analysis, each of these can be framed in two successive ways, first as model checks and then as expansions to the model.

We first consider how to model the possibility of different incumbency advantages for Democrats and Republicans. This could—and perhaps should—be easily added into our model by simply replacing our mean incumbency advantage parameter with two parameters, one for each party. They both could be estimated, and the time series of the estimates for each party, or, for more numerical stability, the average of the two (which would presumably be similar to the single parameter for each year estimated in our existing model), could be displayed.

Continuing this thought, we could allow the incumbency advantage to vary systematically by region of the country (e.g., south or nonsouth) or by state (e.g., varying by the strength

Andrew Gelman is Professor, Department of Statistics and Department of Political Science, Columbia University, New York, NY 40027 (E-mail: *gelman@stat.columbia.edu*).

Journal of the American Statistical Association
June 2008, Vol. 103, No. 482, Applications and Case Studies
DOI 10.1198/016214507000000653

of state party organizations, as considered by Ansolabehere, Hansen, Hirano, and Snyder 2005a), and perhaps by features of the incumbents themselves (including sex, ethnicity, and tenure in office) and aspects of the challengers (e.g., the level of previous political experience). At this point, there would be too many coefficients to reliably estimate separately, and the natural next step would be a multilevel model that would essentially explain some of the variation in incumbency advantage given these predictors at the candidate and district levels. Our model is well suited to this sort of expansion, because it already includes variation in the treatment effect.

EVIDENCE FOR STRATEGIC RETIREMENT?

Katz's other concern is strategic retirement. Incumbents who leave office voluntarily may do so because of knowledge about their party's prospects. To the extent that this is occurring, our model is ignoring information in the incumbency status variable.

Katz presents evidence (in his table 1) that the incumbent party's vote loss is much greater for voluntary exits than for involuntary exits and writes, "If there is no strategic retirement, as claimed by Gelman and Huang, then there should not be a significant difference in the loss that the incumbent party incurs between these two types of open seats." We disagree with this statement, because involuntary exit is not itself a randomly assigned treatment; deaths in office are rare and commonly occur in congressmembers who have been in office in safe seats for a long time, and primary losses also are rare events that commonly occur in safe seats (such as in the formerly one-party south; see Ansolabehere et al. 2005b). Given these pretreatment differences between districts with voluntary and involuntary exits, it is hard to make strong conclusions from raw differences in gain scores.

A clever recent systematic study of the effects of different sorts of retirement is that of Ansolabehere and Snyder (2002b), who considered term limits as an externally assigned treatment and concluded that the occurrence of strategic retirement has essentially no effect on estimates of incumbency advantage. Gelman and King (1990) considered the related issue of bounding the bias in estimation arising from uncontested seats.

Although we are not convinced of the substance of Katz's claim about strategic retirement, we agree with his statistical point, which is that additional data can be collected—in this case, a classification of open seats as due to deaths, primary election losses, voluntary retirements, and other causes—for which it would make sense to expand our model. Katz also notes that it would require additional effort to expand our model to include information about challengers. Indeed, much of the variation that we find in incumbency advantage could arise from systematic differences in challenger quality among districts.

CONCERNS WITH BAYES

A favorite saying of Bayesians is that "with great power comes great responsibility." Using hierarchical models and carefully chosen prior distributions, we can estimate many parameters—in our case, a separate incumbency advantage in each district—which is advantageous because it allows us to study issues that we have always worried about as statisticians—in this case, varying treatment effects. But the path from data to inferences is tangled, and there is the worry that we are not really learning from the data so much as spitting out, in a different form, various convenient assumptions from our prior distribution. As Hodges points out in his comment, our inferences are sensitive to assumptions in the prior distribution and in the computation.

Our generic answer to this concern is to do model checking, and indeed our article features an example of a model check—the scatterplots in figure 7 that sent us in a tailspin, leaving our project incomplete for about 2 years until we realized the problem with our model fitting: We had forgotten about the information in the incumbency status at time 1, as we discuss in section 3.3.

A natural worry is that there remain other errors, oversights, or simply inappropriate components of the model. We found one major mistake; why should anyone trust us when we say that we have found no other problems? For that matter, why should we trust ourselves?

As in the old joke about the priest, the doctor, the lawyer, and the bear, we can retreat to the position that we are not trying to get the right answer, but are simply trying to outrun the competition—in this case, the published methods of Gelman and King, Cox and Katz, Levitt and Wolfram, and others, which did not estimate the variation in incumbency advantage and did not make use of certain systematic differences between districts with incumbents and open seats.

It is reassuring that our estimates are roughly consistent with what came before, but smoother [compare fig. 2 with fig. 9(a)]. On the other hand, if that is the point, then why not just take the crude estimate and apply a time series smoother? Smoothing can create its own problems, especially near the end of a time series (often the area in which we are most interested), and in any case we can view our model as a generalization of earlier approaches in which we have added a variance component and are more careful about certain aspects of the likelihood. Ultimately, external validation of predictions can be the most convincing test.

Considering our approach as a statistical method, we agree with Hodges that care is warranted. As with all new technologies, this method will be used first in problems such as incumbency with large sample sizes and a background of substantive understanding that will warn us off of the most extreme errors, and then, as we establish more confidence in the method, it can be extended and used more broadly—as, for example, simple hierarchical Bayes was considered controversial in statistics in the 1970s but is now used routinely in political science (e.g., to estimate state-level opinions from national polls).

To respond slightly more specifically to Hodges's comments about variance component models, I suspect that the best approach is ultimately to extend the model further so that the variance parameters can themselves be modeled hierarchically. This expansion might be performed by using a time series model to link the now-separate inferences from different years, which could fix the implausible and embarrassingly jumpy plot in our figure 9(d). (As we and Hodges note, very little information is available to estimate this particular variance component from the few open seats that occur in any pair of election years.) Expansion raises its own modeling and computational challenges, but with the potential payoff of more precise estimates of year-to-year changes in treatment effects, it is important in many areas of application.

ADDITIONAL REFERENCES

Ansolabehere, S., Hansen, M., Hirano, S., and Snyder, J. M. (2005a), "The Incumbency Advantage in U.S. Primary Elections," technical report, Massachusetts Institute of Technology, Dept. of Political Science.

——— (2005b), "The Decline of Competition in U.S. Primary Elections, 1908–2004," technical report, Massachusetts Institute of Technology, Dept. of Political Science.