Estimating incumbency advantage and its variation, as an example of a before/after study*

Andrew Gelman† Zaiying Huang†

August 13, 2003

Abstract

Incumbency advantage is one of the most studied features in American legislative elections. In this paper, 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.

Keywords: Bayesian inference, before-after study, Congressional elections, Gibbs sampler, incumbency advantage, Metropolis algorithm, multilevel model

1 Introduction

Incumbency advantage is one of the most studied features in American legislative elections.[1] Our goal in this paper 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

*We thank Robert Erikson, Gary King, and the editors and reviewers for helpful comments, Gary King, Greg Wawro, and Gary Jacobsen for data, and the National Science Foundation for financial support.

†Department of Statistics and Department of Political Science, Columbia University, New York

[1]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), and Ansolabehere and Snyder (2002a).

elections in each legislative district without additional information on the candidates or districts.[2] Our analysis introduces a new form of multilevel model appropriate for observational studies or experiments based on "before" and "after" data—in this case, two successive elections with the intervening treatment being the decision of whether to the incumbent runs for reelection.

Our method and results are similar to existing regression estimates (e.g., Gelman and King, 1990, Cox and Morgenstern, 1993, Cox and Katz, 1996, Levitt and Wolfram, 1997, and Ansolabehere and Snyder, 2002a, b) 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 to 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

This section reviews 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.

2.1 Background and interpretation as an observational study

The advantages of using regression models to estimate incumbency advantages can be seen by comparing to some simpler approaches. Most directly, one 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

[2] It is standard in this literature to analyze pairs of elections, to minimize the difficulties that arise with missing data and decennial redistrictings.

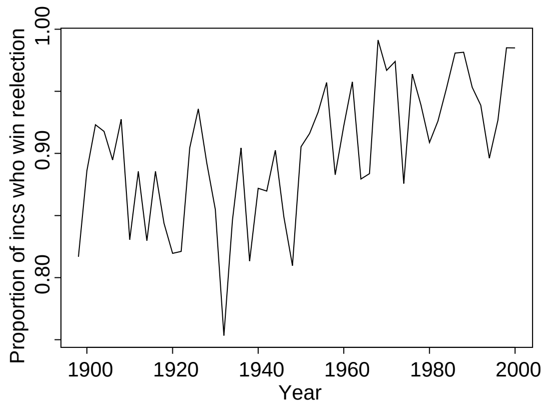

Figure 1: Reelection rate for incumbents in the U.S. House of Representatives over time. We cannot use this as a measure of incumbency advantage since strongly partisan districts are likely to reelect the incumbent party regardless of the candidate.

a highly conservative district, for example, no incumbency advantage is required 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:

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

where P_{it} represents the *incumbent party* and I_{it} 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 ϕ captures the effect of the incumbent *candidate*, after controlling for party. 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 terms in the model adjust 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. Since 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 to apply the treatment and run a new candidate:

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

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 $v_{i,t-1} > 0.5$), and the "incumbency advantage" is the effect of $T_{it} = 0$ compared to 1. Then regression (1) becomes,

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

after adding a main effect for T_{it} (which represents a differential incumbency effect between 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). Figure 2 displays the estimates of ψ (with standard errors) 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 it, in any given election year (except for those following 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

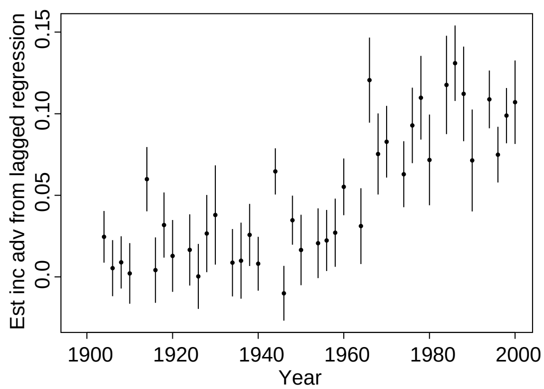

Figure 2: Estimated incumbency advantage over time, from regression model (2) that assumes a constant incumbency effect for all districts in any election year, following Gelman and King (1990). These can be compared to the estimates displayed in Figure 9 and 10 for our new model that allows incumbency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

(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. However, the actual data for the incumbents and the open seats are far from parallel.

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:

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

We estimate this model for the Congresses of the twentieth century and display the estimated slopes β_{1a} and β_{1b} in Figure 4. Figure 3 represents a consistent pattern: for the second half of the century, the slope for the lagged vote is consistently higher for incumbents than for open seats.[3]

[3] A similar point regarding lack of fit of this model (2) was made by Gelman et al. (1995, Section 8.4).

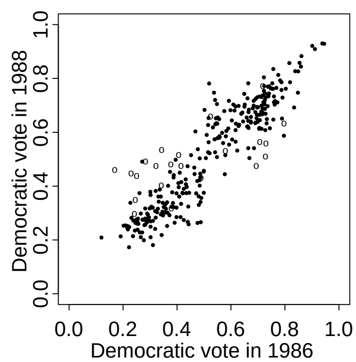

Figure 3: Graph illustrating a problem with the regression model (2). Scatterplot of the Democratic share of the two-party vote for Congress in 1986 and 1988, with each symbol representing a Congressional district. Dots represent districts with incumbents running in 1988, and circles represent open seats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

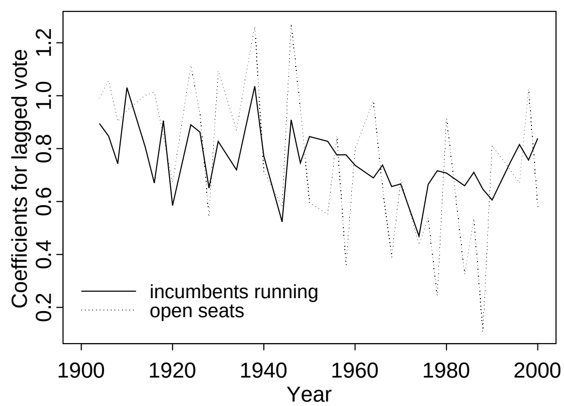

Figure 4: Coefficients for lagged vote in the regression model (3), estimated separately for elections with incumbents (solid line) and open seats (dotted line). The coefficients are consistently different for the two kinds of district elections, indicating a problem with the simpler model (2) that included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

(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 $v_{i,t-1}$. 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 presently.

2.3 Reformulation of the interaction as a variance component

Why, in the scatterplot (3), do open seats have a flatter slope from 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 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) not as an *effect* of incumbency but 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 and thus reduce the predictive power of lagged vote in that district (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 we 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:

$$\text{for } t = 1, 2: \qquad v_{it} = 0.5 + \delta_t + \alpha_i + \phi_{it} I_{it} + \epsilon_{it}, \tag{4}$$

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 0.5 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.
- ϵ_{it}'s are independent error terms assumed 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, 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 0.2 and 0.8 (see Figure 3, for example). 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, and Ansolabehere, Snyder, and Stewart, 2000), suggesting potential nonlinearity at the extremes of the model. Another potential concern is heteroscedasticity, since 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, and Mulligan and Hunter, 2001).

In estimating the model, we are primarily interested 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 $\theta = (\delta, \alpha, \phi, \psi, \sigma_\alpha, \sigma_\phi, \sigma_\epsilon)$, and the posterior distribution is then $p(\theta|v, I)$. We lay out the model here and then in Section 3.2 describe the estimation of the parameters.

The likelihood

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

$$p(v|I, \theta) = \prod_{i=1}^{n} \prod_{t=1}^{2} \mathrm{N}(v_{it} \,|\, 0.5 + \delta_t + \alpha_i + \phi_{it} I_{it},\, \sigma_\epsilon^2),$$

using the $\mathrm{N}(\cdot|\cdot, \cdot)$ notation for the normal density function (as in Gelman et al., 1995).

The prior distribution

The models for the district-level baselines and the candidate-specific incumbency effects are,

$$\begin{aligned} p(\alpha|\sigma_\alpha) &= \prod_{i=1}^{n} \mathrm{N}(\alpha_i|0, \sigma_\alpha^2) \\ p(\phi|\psi, \sigma_\phi) &= \prod_{i=1}^{n} \mathrm{N}(\phi_{i1}|\psi, \sigma_\phi^2) \prod_{i:\ I_{i1} I_{i2} = -1} \mathrm{N}(\phi_{i2}|\psi, \sigma_\phi^2). \end{aligned}$$

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 (for example, uniform on $[-1, 1]$ for δ, ψ, and uniform on $[0, 1]$ for $\sigma_\alpha, \sigma_\phi, \sigma_\epsilon$).[4]

The information provided by the incumbent party indicators

The model is nearly complete, but we must still account for 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.

[4] In contrast, noninformative uniform prior densities on $\log \sigma_\alpha, \log \sigma_\phi, \log \sigma_\epsilon$ would lead to improper posterior densities (see, e.g., Gelman et al., 1995, and Hobert and Casella, 1996).

The incumbent party indicators are another matter. P_{i2} provides no additional information in our model, since $P_{i2} = 1$ if and only if $v_{i1} > 0.5$ (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 0.5.) 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

$$\pi_P(\alpha_i) = \Pr(P_{i1} = 1|\alpha_i) = \Pr(v_{i0} > 0.5|\alpha_i).$$

The likelihood for I 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.[5]

From the model (4), we can write,

$$v_{i0} = 0.5 + \delta_0 + \alpha_i + \epsilon_{i0} + \begin{cases} 0 & \text{if } I_{i0} = 0 \\ \phi_{i0} & \text{if } I_{i0} = 1 \\ -\phi_{i0} & \text{if } I_{i0} = -1. \end{cases}$$

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

$$v_{i0} \sim \begin{cases} \mathrm{N}(0.5 + \delta_0 + \alpha_i,\, \sigma_\epsilon^2) & \text{with probability } \pi_T \\ \mathrm{N}(0.5 + \delta_0 + \alpha_i + \psi,\, \sigma_\epsilon^2 + \sigma_\phi^2) & \text{with probability } (1 - \pi_T)\pi_P(\alpha_i) \\ \mathrm{N}(0.5 + \delta_0 + \alpha_i - \psi,\, \sigma_\epsilon^2 + \sigma_\phi^2) & \text{with probability } (1 - \pi_T)(1 - \pi_P(\alpha_i)), \end{cases} \qquad (5)$$

where $\pi_T = \Pr(T_{it} = 1)$, the probability that a district will have an open seat. In setting up the second and third normal distributions above, we have assigned the candidate-specific incumbency effect the $\mathrm{N}(\psi, \sigma_\phi)$ distribution from the model, ignoring any information present in ϕ_{i1}. This minor simplification allows us calculate the probabilities $\pi_P(\alpha_i)$ in closed form.

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

$$\begin{aligned} \pi_P(\alpha_i) &= \Pr(v_{i0} > 0.5) \\ &= \pi_T \Phi\left(\frac{\delta_0 + \alpha_i}{\sigma_\epsilon}\right) + (1 - \pi_T)\pi_P(\alpha_i)\Phi\left(\frac{\delta_0 + \alpha_i + \psi}{\sqrt{\sigma_\epsilon^2 + \sigma_\phi^2}}\right) + (1 - \pi_T)(1 - \pi_P(\alpha_i))\Phi\left(\frac{\delta_0 - \alpha_i + \psi}{\sqrt{\sigma_\epsilon^2 + \sigma_\phi^2}}\right), \end{aligned}$$

[5] 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.

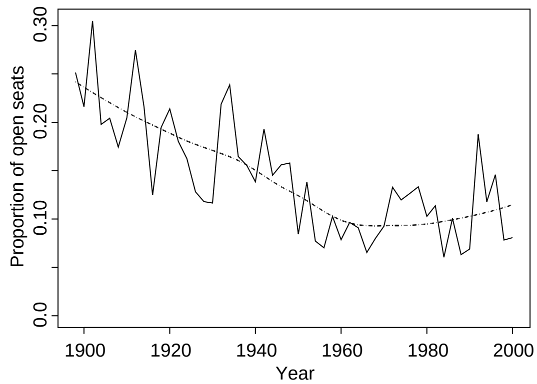

Figure 5: $\overline{T}_t$, the proportion of open seats in each Congressional election in the twentieth century. The dotted line shows a smoothed estimate that we used as an estimate of π_T in computing (6). The smoothing was done using lowess (Cleveland, 1979).

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

$$\pi_P(\alpha_i) = \frac{\pi_T \Phi\left(\frac{\delta_0+\alpha_i}{\sigma_\epsilon}\right) + (1-\pi_T)\Phi\left(\frac{\delta_0-\alpha_i+\psi}{\sqrt{\sigma_\epsilon^2+\sigma_\phi^2}}\right)}{1-(1-\pi_T)\Phi\left(\frac{\delta_0+\alpha_i+\psi}{\sqrt{\sigma_\epsilon^2+\sigma_\phi^2}}\right) + (1-\pi_T)\Phi\left(\frac{\delta_0-\alpha_i+\psi}{\sqrt{\sigma_\epsilon^2+\sigma_\phi^2}}\right)}. \tag{6}$$

To compute (6) as a function of α_i, we need σ_ϵ and σ_ϕ, which are part of our model, and π_T and δ_0, which must be computed externally. We estimate π_T by the historical proportion of open seats (for example, see Figure 5 for Congressional elections) and δ_0 by $\bar{v}_0 - 0.5$, the nationwide average Democratic vote in election 0, relative to 0.5. The average vote, $\bar{v}_0$, is the only information about election 0 that we use in our analysis.[6] We require district-level data only for the two consecutive elections which we have labeled 1 and 2.

The complete posterior distribution

Finally, the joint posterior density is proportional to the product of all the above pieces:

$$\begin{aligned} p(\theta|v,I) \propto & \prod_{i=1}^{n}\left(\prod_{t=1}^{2} \mathrm{N}(v_{it}|0.5+\delta_t+\alpha_i+\phi_{it}I_{it},\sigma_\epsilon^2)\right)\mathrm{N}(\alpha_i|0,\sigma_\alpha^2)\mathrm{N}(\phi_{i1}|\psi,\sigma_\phi^2)\times \\ & \times \prod_{i:I_{i1}I_{i2}=-1}\mathrm{N}(\phi_{i2}|\psi,\sigma_\phi^2)\prod_{i:I_{i1}=1}\pi_P(\alpha_i)\prod_{i:I_{i1}=-1}(1-\pi_P(\alpha_i)). \end{aligned} \tag{7}$$

[6]In the rare scenario in which $\bar{v}_0$ is 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.

3.2 Implementation

Given district-level data v, I from any pair of consecutive elections, we estimate 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 $\pi_P(\alpha_i)$—the density (7) is a normal linear multilevel model, and all of its parameters can be updated using the Gibbs sampler. The linear parameters $\delta, \alpha, \phi, \psi$ have a joint normal conditional distribution, and the variance parameters $\sigma^2_\alpha, \sigma^2_\phi, \sigma^2_\epsilon$ have independent inverse-χ^2 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 a short time to obtain a preliminary estimate of the variance components and then compute the posterior covariance matrix of the linear parameters $\delta, \alpha, \phi, \psi$ conditional on this estimate. We use this 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 in Van Dyk and Meng (2001).

However, the Gibbs sampler is just an approximation since 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 the parameter vector at the previous step was θ^s, which has been altered to a "candidate value" θ^* once all the parameters have 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% to 80%, which means that the approximation is reasonably close.

The full algorithm, including the improvements to the Gibbs sampler mentioned above, converges quickly. For example, a typical pair of Congressional elections will have about 700 data points v_{it} (350 contested elections in each of two 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 $\delta_1, \delta_2, \psi, \sigma_\epsilon, \sigma_\alpha, \sigma_\phi$). Four chains reached approximate convergence (that is, the Gelman and Rubin, 1992, convergence diagnostic $\sqrt{\widehat{R}}$ was less than 1.2) after 4000 iterations, taking four 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 50% intervals computed from the 25% and 75% points of the simulations. We illustrate 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 this would bring to our model.[7] In addition, we exclude elections after redistricting (that is, years ending in "2"); analyzing these pairs would require additional information on which districts were redrawn, as in Ansolabehere, Snyder, and Stewart (2000).

3.3 Model checking

After fitting an elaborate model, it is important to check its fit to data. We do so by simulating replicated data sets conditional on the estimated model parameters, and comparing these to the

[7]Gelman and King (1990) and Ansolabehere and Snyder (2002b) estimate the selection of contested elections to reduce the estimated effect of incumbency by no more than 1%.

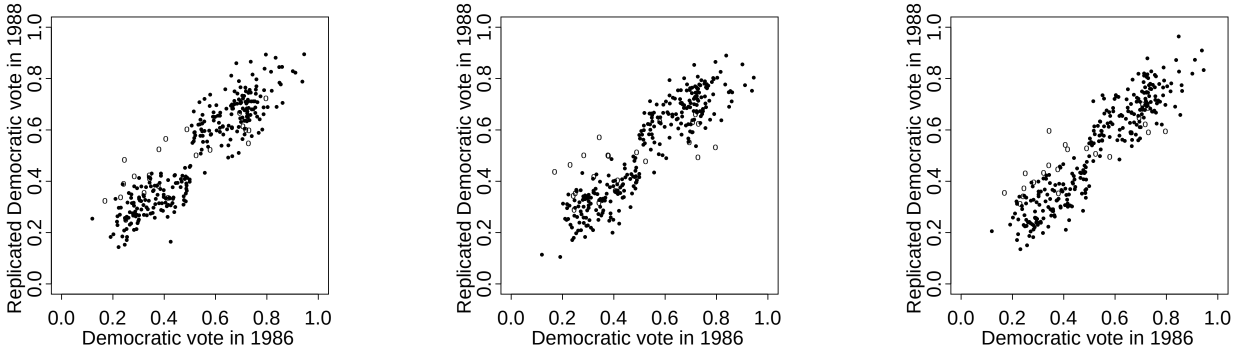

Figure 6: Replicated data sets simulated from the regression model (2) as fit to data from 1986 and 1988. Dots represent elections with incumbents running in 1988 and circles represent open seats. Compare to the actual data in Figure 3. The circles in the actual data have a much flatter slope than in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

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, fit by three models: the simple regression (2); the multilevel model (4) fit by the Gibbs sampler ignoring the information provided by I_{i1} (that is, 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 fit 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 too hastily constructed the posterior density ignoring those factors with π_P and $(1 - \pi_P)$. It is instructive to see if 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. It is thus important to find out if this particular simplification (whether intentional or unintentional) affects the ability of the model to fit the data.

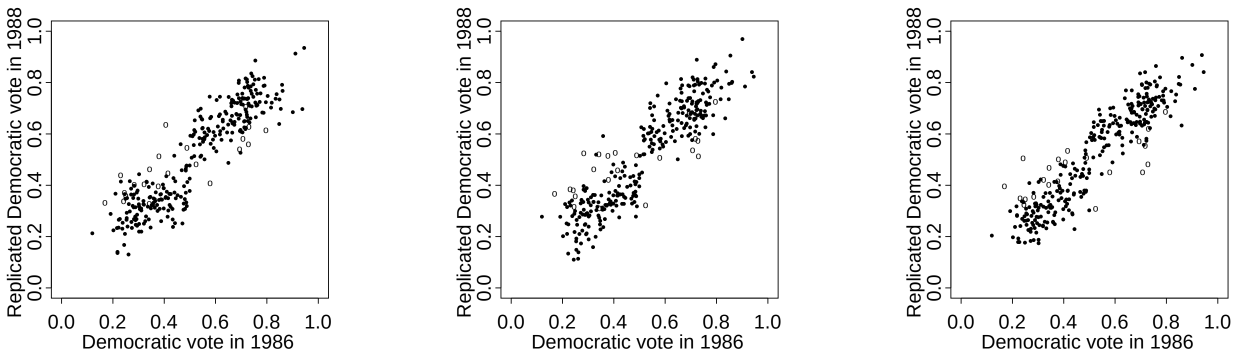

Figure 7: Replicated data sets simulated from the regression model (4) as fit to data from 1986 and 1988, using the Gibbs sampler ignoring the information in P_{i0}. Compare to the actual data in Figure 3. The dots are pressed much closer to a 45° line in the actual data than in the replications.

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}, v_{i2}. The two election outcomes must be simulated in sequence since v_{i1} affects P_{i2}, which in turn affects the predicted value of v_{i2}.

Compared to the real data in Figure 3, the replications in Figure 7 look wrong, with the dots having too low a correlation—too much of a "puffy" appearance. It is good that we looked at these plots, since this is a serious misfit in our context: in underestimating the correlation of the elections with incumbents, 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.

Finally, we check the fit of the full model (4) using the full posterior density (7). Figure 8 displays replicated data sets, which look very 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—show why it is important to check the fit. 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.

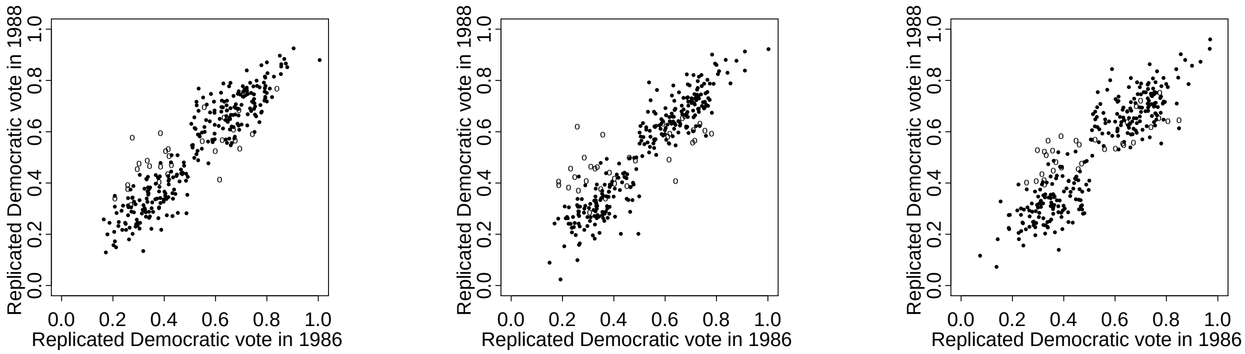

Figure 8: Replicated data sets simulated from the regression model (4) as fit to data from 1986 and 1988, fitting the full model using the Gibbs sampler and the Metropolis algorithm. The replicated data look similar to the actual data in Figure 3.

3.4 Connection to other estimates of incumbency advantage

As discussed above, our multilevel model (4) generalizes classical lagged regressions. In particular, if σ_ϕ, the variation in the incumbency advantage, is set to zero, 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 adjust for incumbency status at election 1.

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 two years within districts where incumbency status changes. Subtracting the 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 Section 3.1), we can write this as,

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

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.

However, 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) fails 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 elections that straddled a redistricting (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 Section 3.3), and so we begin our summarizing 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 to 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 to regression methods that treat the first election merely as a lagged predictor or 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 the districts can be estimated well, 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 rapidly increased during the 1950s and 1960s, and has been relatively high for the past thirty years. However, our estimate of the average level of incumbency advantage in recent years is about 8%, compared to the usual estimate from the literature of about 10%. The estimates from our model have much lower standard errors and

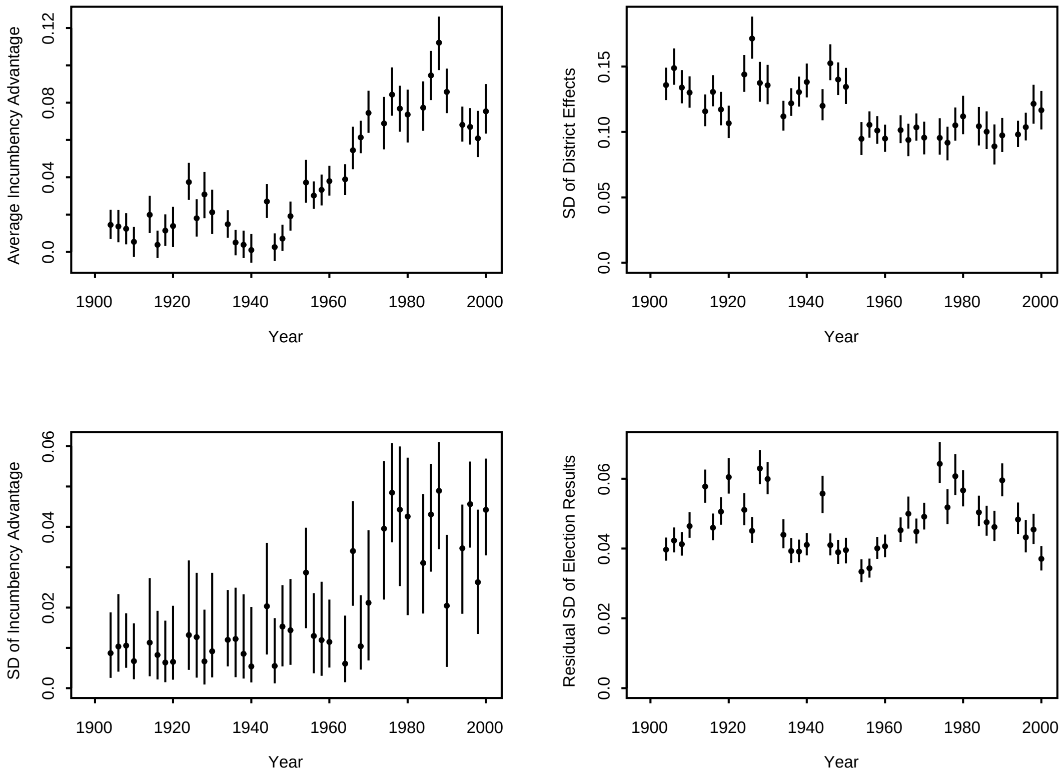

Figure 9: Estimates ± standard errors for the hyperparameters $\psi, \sigma_\alpha, \sigma_\phi, \sigma_\epsilon$ in the incumbency model, estimated for each pair of elections for the U.S. House of Representatives in the twentieth century. Election years immediately following redistricting (those ending in "2") are excluded. The vertical axes for the four graphs are on different scales. The estimate of the average incumbency advantage (the upper-left graph) over time is much more stable than from the usual regression estimate (see Figure 2).

are much more stable than the simple regression estimates (compare to Figure 2), which should be especially important when estimating incumbency advantages in shorter time series or in individual states where less information is available.

We find that the increase in average incumbency advantage ψ was followed, with about a fifteen-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, which we attribute to the lack of information in the model to distinguish between candidate and district effects. (This paucity of information also showed up in the wide fluctuations in the estimated lagged regression slope for open seats as indicated by the dotted line in Figure 4.)

Meanwhile, districts have become slightly more similar to each other in their baselines—σ_α decreased from about 15% to 12% over the century—and the residual variation σ_ϵ has cycled

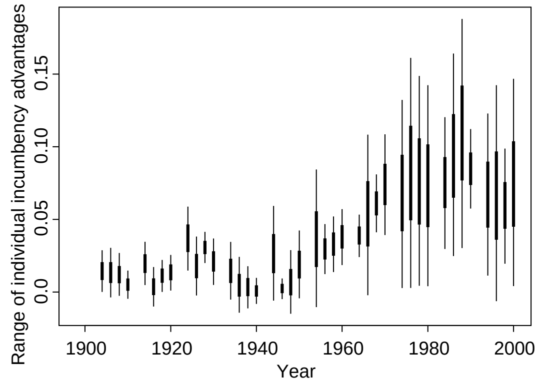

Figure 10: Estimated distributions of incumbency effects ϕ_{i1}, shown by medians, 50% ranges, and 90% ranges, for the twentieth-century House of Representatives. These ranges represent the estimated variation among incumbents, *not* uncertainties in the mean level of incumbency. The variation in incumbency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

between about 4% and 6% over the years.

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), and then in Figure 10 display the posterior mean of each quantile—thus, estimated medians, 50% ranges, and 90% ranges for the incumbency effects across the country—for each estimation of the model (that is, each pair of consecutive elections). We see the midcentury increase in average incumbency advantage and the increased variation starting in the 1960s. In recent years, we estimate individual incumbency advantages to range between 0 and 15% of the vote.

5 Discussion

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 an new component of variation allows it to incorporate more information already present in the data (thus giving more precise estimates) and also to better fit those aspects of the data. The model is multilevel and was fit 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 (for example, state legislatures). Relatively simple alterations would allow additional district-level information, such as other election results and candidate quality, to be included as regression predictors in (4), as in the analyses of Ansolabehere, Snyder, and Stewart (2000) and Ansolabehere and Snyder (2002a). This could be a powerful tool for research on how differences between candidates lead to different election outcomes and, in addition, should improve the unstable estimates of the variance parameters. These could also be improved by analyzing more than two elections simultaneously, or by smoothing the estimates over time.

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. It is common in the social and biological sciences for before-after correlations to be higher for control than for treated units, and this interaction can be important—sometimes more important than the main effect of the treatment.[8] Yang and Tsiatis (2001) show how joint modeling of pre- and post-treatment 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.

References

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

Alford, J. R., and Brady. D. W. (1988). Partisan and incumbent advantage in U.S. House elections, 1846–1986. 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*

[8]For an example in the effects of legislative redistricting, see Figures 3 and 4 of Gelman and King (1994).

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*, to appear.

Ansolabehere, S., and Snyder, J. M. (2002b). Using term limits to estimate incumbency advantages when officeholders retire strategically. Technical report, Department of Political Science, Massachusetts Institute of Technology.

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? Department of Political Science, State University at Buffalo, New York.

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., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995). *Bayesian Data Analysis*. London: Chapman and Hall.

Gelman, A., and King, G. (1990). Estimating incumbency advantage without bias. *American Journal of Political Science* **34**, 1142–1164.

Gelman, A., and King, G. (1994). Enhancing democracy through legislative redistricting. *American Political Science Review* **88**, 541–559.

Gelman, A., King, G., and Boscardin, W. J. (1998). Estimating the probability of events that have never occurred: when does your vote matter? *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.

Gelman, A., and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science* **7**, 457–511.

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*, ed. D. W. Brady, J. F. Cogan, and M. P. Fiorina, 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.

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. National Bureau of Economic Research Working Paper 8590.

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. *American Statistician* **55**, 314–321.