Estimating the Electoral Consequences of Legislative Redistricting

ANDREW GELMAN and GARY KING*

We analyze the effects of redistricting as revealed in the votes received by the Democratic and Republican candidates for state legislature. We develop measures of partisan bias and the responsiveness of the composition of the legislature to changes in statewide votes. Our statistical model incorporates a mixed hierarchical Bayesian and non-Bayesian estimation, requiring simulation along the lines of Tanner and Wong (1987). This model provides reliable estimates of partisan bias and responsiveness along with measures of their variabilities from only a single year of electoral data. This allows one to distinguish systematic changes in the underlying electoral system from typical election-to-election variability.

KEY WORDS: Bayesian estimation; Elections; Political science; Random effects; Simulation.

1. INTRODUCTION

State and national legislators in the United States are largely elected by plurality vote in individual geographic districts, whose boundaries are redrawn after every decennial census. In addition to ensuring equal populations in each district, redistricting affects which candidates are elected, the relative strengths of the two parties in a legislative house, and other features of the electoral system in a state.

Partisans on both sides generally expend considerable political and financial resources trying to control the redistricting process. Because redistricting affects not only immediate political outcomes, but also the fundamental rules of the game, it has always been a highly controversial partisan issue (Cain 1984). When partisans do not receive satisfaction in the legislative arena, they often take their case to the courts. After decades of these cases, the Supreme Court finally declared political gerrymandering justifiable (*Davis v. Bandemer* 1986). The court has not yet settled, however, on an acceptable standard for or measure of an unfair redistricting plan.

In this article, we analyze the effects of redistricting as revealed in the votes received by the Democrats and Republicans in elections for state legislative seats. We also develop measures of partisan bias and the responsiveness of the partisan composition of the legislature to changes in statewide votes. Our conclusions depend on the observed distribution of votes across the legislative districts, as affected by redistricting, and on assumptions about how these district-level votes change as the statewide vote changes. We also explicitly model uncontested district elections.

Related quantitative issues that we do not directly discuss here, but that could be studied with our model, include trends in "marginal seats," the importance of incumbency, the effectiveness of racial gerrymandering, the effect of redistricting on individual districts, and the recent declining responsiveness of the U.S. House of Representatives to vote swings (Gelman and King, in press; King and Gelman, in press).

Our statistical methodology involves a hierarchical random-effects model with a mixture of Bayesian and non-Bayesian estimation, summarized probabilistically. Our Bayesian computation requires simulation along the lines of Tanner and Wong (1987).

2. THE DATA

We analyze the votes received by Democratic and Republican candidates for the lower house of the legislatures of Ohio, Connecticut, and Wisconsin, in the seven elections held in even-numbered years from 1968 through 1980. All elections in these states were by plurality vote in single-member districts, and, except for two districts in Wisconsin in 1980, were won by one of the two major-party candidates. As a result of redistricting in the 1960s, all districts had roughly equal populations. As a sample of our data, Table 1 shows votes in each district election in Ohio in 1972 and 1974. (Our data are available from the Inter-University Consortium for Political and Social Research.)

The Democrats controlled the 1971 Ohio redistricting process and redrew the 99 districts. Connecticut had 177 districts in 1968–1970; during the 1971 redistricting, the number of districts was reduced to 151 and the Republicans controlled where the lines were drawn. Wisconsin's 100 districts were redrawn in 1971 by bipartisan agreement.

For convenience, we will henceforth refer to the Democratic proportion of the two-party vote for a given district election as the *district vote.* We label the average of these proportions, over all districts in a given state and election, as the *average district vote.*

Some district elections feature a single candidate with insignificant opposition or none at all. We refer to such an election as *uncontested* if one candidate gets more than 95% of the two-party vote. The proportion of uncontested elections among all of the district elections varies greatly over the three states and seven election years, with an

* Andrew Gelman is a graduate student, Department of Statistics, Harvard University, Cambridge, MA 02138. Gary King is Professor, Department of Government, Harvard University, Cambridge, MA 02138.
Thanks go to the National Science Foundation for research grants SES-87-22715 and SES-89-09201 to King and for a graduate fellowship to Gelman. Thanks also go to Donald Rubin, Stephen Ansolabehere, Chris Achen, and Arthur Dempster for many helpful suggestions.

Journal of the American Statistical Association
June 1990, Vol. 85, No. 410, Applications and Case Studies

Table 1. Votes Received by Democrats and Republicans in Ohio Legislative House Districts, 1972 and 1974

1972						1974					
District	Democrat	Republican	District	Democrat	Republican	District	Democrat	Republican	District	Democrat	Republican
1	18,250	22,798	51	22,488	16,951	1	20,490	15,107	51	20,952	7,473
2	25,679	17,130	52	24,336	14,083	2	18,669	11,969	52	21,499	7,697
3	0	33,954	53	25,932	8,997	3	12,778	20,272	53	19,522	6,225
4	23,684	10,212	54	22,780	15,229	4	15,765	9,813	54	13,885	15,582
5	21,723	16,130	55	20,198	9,583	5	11,711	9,708	55	19,400	6,538
6	28,309	0	56	21,603	10,678	6	20,584	5,763	56	21,361	9,262
7	20,334	12,675	57	16,533	17,114	7	20,193	9,778	57	11,677	13,944
8	16,622	3,656	58	13,587	22,105	8	11,153	2,261	58	12,286	16,158
9	11,946	10,396	59	14,877	20,234	9	9,566	0	59	13,834	14,211
10	12,383	5,316	60	14,556	13,940	10	8,277	1,890	60	12,550	9.659
11	20,091	18,539	61	16,507	17,825	11	22,398	5,221	61	15,589	13,451
12	18,337	20,561	62	23,668	13,428	12	9,865	19,599	62	18,802	8,178
13	16,688	1,970	63	13,868	18,402	13	10,687	966	63	9,713	9,948
14	22,865	11,218	64	13,984	22,593	14	11,478	8,087	64	10,227	17,747
15	21,401	0	65	11,710	29,134	15	15,905	1,936	65	12,282	21,978
16	27,783	12,701	66	15,500	30,156	16	21,909	10,403	66	11,587	24,978
17	24,511	15,716	67	20,409	17,931	17	22,327	11,274	67	17,556	13,500
18	28,805	14,454	68	21,489	15,574	18	22,416	8,138	68	17,070	12,882
19	17,687	23,463	69	16,592	21,816	19	12,431	19,832	69	12,501	17,328
20	15,225	28,639	70	14,172	21,642	20	17,129	19,927	70	12,708	16,905
21	12,392	23,427	71	22,439	20,831	21	10,732	16,700	71	27,279	0
22	16,635	27,940	72	15,616	19,879	22	13,945	21,762	72	12,734	15,738
23	16,986	7,681	73	0	26,079	23	11,332	0	73	13,178	14,974
24	22,856	12,779	74	22,359	12,626	24	16,270	9,187	74	19,691	9,488
25	20,298	12,292	75	14,653	27,063	25	15,566	7,078	75	15,290	19,913
26	15,181	30,866	76	16,438	24,947	26	13,809	24,345	76	13,940	20,516
27	12,045	35,880	77	14,054	23,185	27	11,655	28,036	77	14,526	18,326
28	20,637	27,011	78	18,867	24,829	28	0	27,907	78	12,307	18,867
29	17,418	13,589	79	15,459	26,221	29	14,001	9,433	79	11,312	19,455
30	15,080	9,381	80	24,237	17,392	30	10,117	3,935	80	23,053	10,137
31	19,754	12,971	81	14,606	24,845	31	16,409	7,302	81	14,778	18,131
32	20,068	13,059	82	18,349	24,436	32	16,402	8,042	82	9,825	23,615
33	13,182	22,046	83	12,650	28,287	33	11,627	16,281	83	11,787	21,775
34	15,101	14,159	84	23,448	17,882	34	12,035	8,516	84	22,858	9,891
35	19,344	10,166	85	15,896	24,792	35	12,146	6,785	85	12,670	19,082
36	19,375	7,792	86	18,969	22,815	36	15,336	2,672	86	12,437	18,466
37	17,149	11,274	87	21,828	15,253	37	13,795	8,310	87	18,484	11,590
38	10,759	30,945	88	20,732	12,816	38	0	23,672	88	20,849	0
39	24,246	18,772	89	27,325	16,336	39	20,149	11,974	89	26,780	9,673
40	21,006	20,625	90	25,239	18,272	40	14,268	14,378	90	23,829	14,405
41	29,507	11,524	91	19,783	20,492	41	22,472	6,734	91	14,733	17,729
42	21,635	17,233	92	20,567	20,749	42	15,888	11,543	92	16,859	15,651
43	26,149	9,428	93	11,803	27,093	43	19,881	5,012	93	11,470	21,709
44	24,020	17,601	94	16,508	19,409	44	15,428	18,232	94	12,036	16,015
45	22,872	0	95	10,642	26,685	45	14,622	4,673	95	8,897	21,921
46	23,080	11,743	96	27,270	14,044	46	19,006	7,538	96	23,133	9,397
47	20,465	8,920	97	16,859	13,746	47	17,031	0	97	21,528	9,742
48	18,756	27,079	98	28,857	11,878	48	18,001	19,673	98	22,598	7,454
49	19,809	18,632	99	26,945	14,848	49	17,406	13,021	99	21,235	10,584
50	18,036	19,734				50	14,994	14,481			

average of 10% of the seats uncontested in any election. No statewide election in our study had more than 23% uncontested seats, except for Wisconsin in 1980, with 32%. Election returns in uncontested districts do not adequately reflect support for the two political parties. Since we are interested in this party support, we define the *effective vote* in the case of uncontested districts to be the (unobserved) proportion of the two-party vote that this candidate would have won in his or her district had the election been contested. We approximate the probability density of the effective vote with a stem-and-leaf plot of the vote proportions received by a party in a contested district, one election *before* an uncontested win by that party in that district. Figure 1 presents this plot, based on data from

0.5	11234
	5667789999
0.6	0000111122223333344444
	55667778888888899999999
0.7	00000011111222222333333333444444
	555555566677778889
0.8	00011111122333
	66799
0.9	11344

Figure 1. Stem-and-Leaf Plot of the Proportion of the Vote Received by a Party in a Contested District Election, Immediately Preceding an Election in Which That Party Was Unopposed in That District.

1968–1980 in the three state legislatures considered in this article.

3. DATA SUMMARIES AND EXPLORATION

Previous work in this field has involved various theoretical constructs and related data summaries, but extremely few statistical models. One early concept is the "swing ratio"—the change in the proportion of legislative seats won by a party (S), divided by the change in the average district vote (V) received (Ansolabehere, Brady, and Fiorina 1988; Kendall and Stuart 1950). This concept was expanded to the "seats–votes curve," which is the fraction of the legislative seats won by a party, as a function of the average district vote (Niemi and Fett 1986; Quandt 1974). This curve can be expressed as the function $S(V)$, where the variables for fraction of seats won and average district vote each vary from 0 to 1. Figure 2 presents two examples of seats–votes curves. One reflects de facto statewide proportional representation, where $S = V$. The other represents a highly responsive electoral system near the middle of the votes scale, where most elections are usually decided. Following King and Browning (1987) and King (1989), we consider these two symmetric seats–votes curves to represent electoral systems that are fair to the political parties. Deviation from bipartisan symmetry is considered partisan bias.

Of course, a party's legislative representation is not a function only of the number of votes it receives; a deterministic seats–votes curve, as defined, cannot be more than a theoretical construct (Tufte 1973). For this reason, we define the seats–votes curve in real electoral systems to be the *expected* value of S, as a function of V, and we will be interested in both this conditional expectation function and variability around it. Responsiveness and bias can be defined more formally as follows:

$$\begin{aligned} \text{Responsiveness}(V) &= dE(S \mid V)/dV \\ \text{Bias}(V) &= E(S \mid V) \\ &\quad - [1 - E(S \mid 1 - V)]. \end{aligned} \tag{1}$$

Past researchers have empirically estimated bias and responsiveness in two ways. The most widely used method uses the statewide Democratic fraction of seats won and the average statewide district vote for a legislature for each of several consecutive elections. One can estimate the seats–votes curve by fitting a nonlinear regression to a scatterplot of these values, and one can calculate summaries of interest from this estimated curve. This method has the disadvantage of ignoring short-term systematic changes in the underlying electoral system, as might result from redistricting. Since only five elections are generally held between redistricting processes, this method is quite limited for present purposes.

The second method, dating back to Butler (1951) [see also Gudgin and Taylor (1979)], creates a "hypothetical" seats–votes curve from the district votes of a single statewide election. This curve plots $S(V)$, under the assumption of "uniform partisan swing"; that is, as the statewide vote V changes, the vote proportion in each district changes

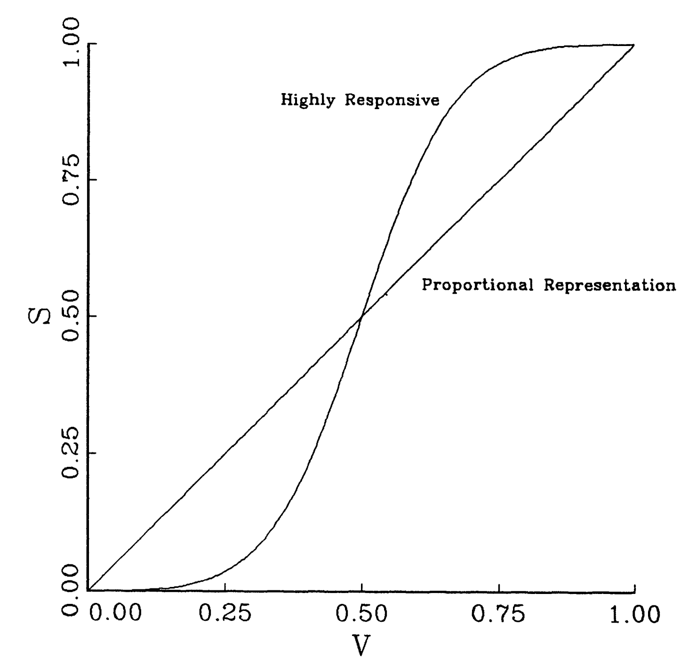

Figure 2. Example of Seats–Votes Curves.

by the same amount. This method breaks down with district votes near 0 or 1 and, in general, is based on an overly strict assumption about voting patterns.

Before describing our stochastic model, we give some exploratory data summaries. We are interested in the distribution of district vote across a state. Figure 3 shows a stem-and-leaf plot of the district votes for the contested elections in Ohio in 1972. This pattern of two main humps with irregular outliers is typical of recent U.S. legislative elections. We identify the two humps with Democratic and Republican "safe seats," and we identify the irregular pattern with the irregular influences of geography on election districts and individual candidates on election results. Sometimes such a plot for an election shows only one main hump in the middle; this corresponds to a competitive system with few safe seats.

0.2	
	5588
0.3	002344
	57777788999
0.4	0222233334
	55778999
0.5	01111233
	55666777788899
0.6	00001112233344
	566667788999
0.7	01134
0.8	1
	9

Figure 3. Stem-and-Leaf Plot of the Democratic Proportion of the Two-Party Vote in Contested District Elections in Ohio, 1972.

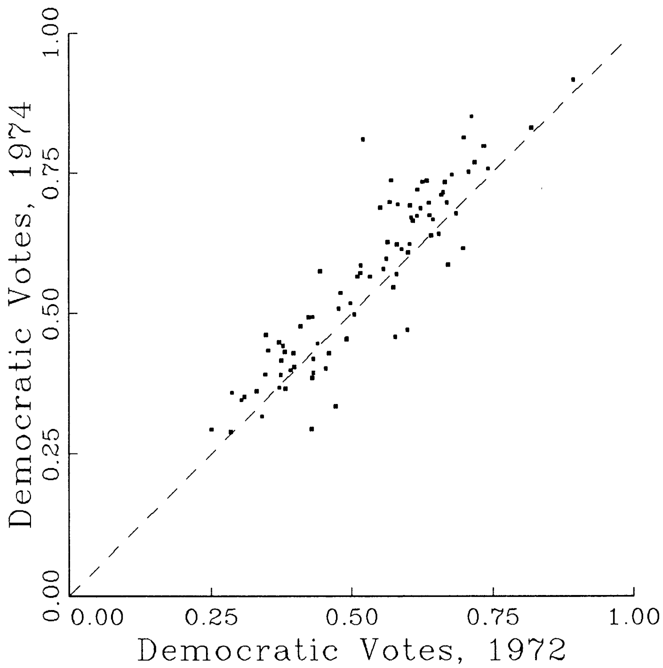

Figure 4. Electoral Swing in Contested Districts, Ohio State House, 1972–1974.

Finally, we would like to know how much partisan voting patterns persist from election to election. As an example of this, Figure 4 shows a scatterplot of district vote proportions for contested elections in Ohio in 1972 and 1974. (Each point on the plot represents one district.) Note that district votes clearly do not move exactly according to "uniform partisan swing"; if they did, all the points would fall precisely on a single line with slope 1. Instead, the points in Figure 4 are scattered around a straight line with slope 1 and intercept equal to the statewide vote swing. We interpret the residual standard deviation in this figure to be within-district random variation about the statewide average vote swing. (A nonuniform shift would be apparent if the points in Fig. 4 fit a clearly nonlinear pattern or no pattern at all.)

4. A PROBABILISTIC MODEL

To avoid problems with vote proportions near 0 or 1, we work with the logit of district votes in contested elections. We label v_{it} as the Democratic vote in district i and election t, and $u_{it} = \text{logit}(v_{it}) = \ln[v_{it}/(1 - v_{it})]$ for contested elections. (For uncontested elections, u_{it} is the logit of the unobserved effective Democratic vote. This will be dealt with in Sec. 5.1.)

Our linear model, fit to a single state, is

$$u_{it} \sim N(\alpha_{it}, \sigma^2), \qquad \alpha_{it} = \gamma_i + \delta_t, \tag{2}$$

where γ_i is a district effect, δ_t is a statewide election effect, and the Normal distributions are independent.

We assume, therefore, that vote swings about the statewide mean are spatially independent across districts. More information about individual districts might enable one to better characterize district-level vote swings. Unfortunately, these data have not been collected, and it would be quite difficult to do so. Modeling districts with additional information such as spatial correlation or covariates, if they were available, would probably yield more accurate estimates of the seats–votes curve. Omitting this unavailable information is unlikely to systematically bias our results.

From the logit effective vote proportions $u_t = (u_{1t}, \ldots, u_{nt})$ for an election t, we define the aggregate Democratic proportions of votes and seats:

$$V_t = \frac{1}{n}\sum_{i=1}^{n} v_{it} = \frac{1}{n}\sum_{i} \text{logit}^{-1}(u_{it})$$

$$S_t = \frac{1}{n}\sum_{i=1}^{n} s_{it} = \frac{1}{n}\sum_{i} 1_{(u_{it}>0)}. \tag{3}$$

We consider the vector $\gamma = (\gamma_1, \ldots, \gamma_n)$, along with the variance σ^2, to identify an "electoral system." We will summarize this system by the seats–votes curve $E(S_t \mid V_t, \gamma)$, its variance $\text{var}(S_t \mid V_t, \gamma)$, and functions of these such as the bias and responsiveness functions. Since the elements of γ remain unknown, we model them as random effects by letting the γ_i's be distributed as a three-point Normal-mixture distribution with a prior distribution, all described in Section 5.2. We then average over our uncertainty in γ as represented by this distribution.

The foregoing model is applied to a single observed statewide election, labeled $t = 0$, with observations u_{i0} ($i = 1, \ldots, n$) and the assignment $\delta_0 = 0$. This assignment is arbitrary and does not affect our estimates of the seats–votes curve. If an arbitrary constant were added to each effective district vote u_{i0}, our results would not change. A family of "hypothetical election" results u_t is defined by the linear model, applied to a range of statewide vote shifts δ_t. This assumption that most electoral districts respond approximately as the statewide total does is widely accepted in the political science literature (Butler 1951; Niemi and Fett 1986), although it has not been formalized statistically. Our data, such as those in Figure 4, are consistent with this pattern. This is also consistent with our assumption in Equation (2) of no interaction between γ_i and δ_t.

We apply this model to our data in four steps.

1. Preliminary Estimation. With data from several consecutive elections, we estimate the global parameters of the model. These include σ^2 and uncontested effective vote parameters μ_{un} and σ_{un}, described in Section 5.

2. Bayesian Estimation for a Single Election. We condition on the data $u_0 = (u_{i0}; i = 1, \ldots, n)$ from a single election to sample from the posterior distribution $P(\gamma \mid u_0)$ of the vector γ. This Bayesian estimation uses the parameters determined in the previous step.

3. The Seats–Votes Curve. We average over $P(\gamma \mid u_0)$ to estimate the posterior seats–votes curve:

$$E(S_t \mid V_t, u_0). \tag{4}$$

(We allow V_t to range from 0 to 1 by allowing δ_t to range from $-\infty$ to ∞ on the logit scale.) We estimate the expected variance of results across hypothetical elections:

$$E(\text{var}(S_t \mid V_t, u_0, \gamma)). \tag{5}$$

We also estimate uncertainty in the seats–votes curve due to our uncertainty in γ:

$$\mathrm{var}(E(S_t \mid V_t, u_0, \gamma)). \quad (6)$$

4. Summaries. From the estimated seats–votes curve (4) and related conditional expectations, we estimate bias and responsiveness summaries of the definitons in (1):

(average bias between $V = .45$ and $V = .55$)

$$= \frac{1}{.55 - .45} \int_{.45}^{.55} (E(S \mid V) - [1 - E(S \mid 1 - V)])\, dV$$

(average responsiveness between $V = .45$ and $V = .55$)

$$= \frac{1}{.55 - .45} [E(S \mid V = .55) - E(S \mid V = .45)]. \quad (7)$$

We define these summaries from $V = .45$ to $V = .55$. This is a convenient range, symmetric about .5, within which most statewide votes fall. We calculate the posterior mean and variance of these summaries.

5. ESTIMATION OF HYPERPARAMETERS

5.1 Election-to-Election Variability

Our linear model creates hypothetical district election results u_{it} from the district effects γ_i by adding a constant shift δ_t to the mean in every district. From here, we add the variability in (2); this "unexplained" variance σ^2 determines the scope of the electoral system identified with the family of hypothetical elections. Setting $\sigma^2 = 0$, for example, causes the district effects to be exactly identified: $\gamma_i = u_{i0}$. This assumption of "uniform partisan swing" on the logit scale cannot hope to fit more than a single statewide election.

We estimate σ^2 from a model of the variances in real district-level election results, across time. We use the following conceptual model:

(variance between two elections, Y years apart)

= (variance due to randomness in individual elections) + (variance due to changes in the underlying electoral system).

In this framework, the first term on the right side of this equality is $2\sigma^2$; we imagine the second quantity to be roughly proportional to Y. Note that, from (2), the difference $u_{it_1} - u_{it_2}$ has variance $2\sigma^2$ if their two Normal distributions are independent.

For each state, we calculate the sample variance of the change in district vote between election years t_1 and t_2, for districts contested in both elections:

$$s^2_{t_1t_2} = \frac{1}{n_{t_1t_2}} \sum_i [u_{it_1} - u_{it_2} - (\bar{u}_{t_1} - \bar{u}_{t_2})]^2,$$

where $n_{t_1t_2}$ is the number of districts in the state contested in both elections t_1 and t_2. We calculate this quantity for all election years (t_1, t_2), $t_1 < t_2$, between 1972 and 1980; that is, we do not track district votes across redistricting.

We then fit a linear regression of the values $s^2_{t_1t_2}$, as a function of the time differences $(t_2 - t_1)$. For each state, our estimate of $2\sigma^2$ is just the estimate of the constant term in this regression, and with an estimate of the regression slope pooled across the three states. This yields estimates of σ (on the logit scale) as .22, .19, and .22 for Ohio, Connecticut, and Wisconsin, respectively, each with a standard error of estimation of .02.

5.2 The Distribution of District Effects γ_i

We need to estimate the vector γ of district effects and our uncertainty in it. Embedding γ in a lower-dimensional probabilistic model allows us to estimate these n district effects from the n data points u_{i0}; we can also then conveniently summarize our results in a posterior distribution.

We consider the district effects to be drawn from a mixture of three Normal distributions, identified by an eight-dimensional parameter $\theta = (\mu_j, \rho_j^2 - \sigma^2, \lambda_j; j = 1, 2, 3)$ of means, variances, and mixture proportions, with the constraint $\lambda_1 + \lambda_2 + \lambda_3 = 1$. These three humps are meant to fit plots like Figure 3, with areas of Democratic strength, areas of Republican strength, and some districts that fit no clear pattern. The parameter ρ_j^2 is the variance of the jth Normal distribution in the density of observed district vote proportions u_{i0}; $(\rho_j^2 - \sigma^2)$ is the variance of the jth Normal distribution in the density of expectations γ_i.

The method of maximum likelihood is inadequate to estimate these eight parameters, since the likelihood function is unbounded. Therefore, we give the eight parameters a prior distribution and move to Bayesian estimation. It is mathematically convenient, and substantively sufficient, to choose a family conjugate to an $N(\gamma_i, \sigma^2)$ distribution:

$$\mu_j \sim N(\mu_{\mu_j}, \sigma^2_{\mu_j}), \qquad j = 1, 2, 3$$

$$\rho_j^{-2} \sim \Gamma(\tfrac{1}{2}\alpha_{\rho_j}, \tfrac{1}{2}\beta_{\rho_j}), \qquad j = 1, 2, 3$$

$$(\lambda_1, \lambda_2, \lambda_3) \sim \mathrm{Dirichlet}(a_{\lambda_1}, a_{\lambda_2}, a_{\lambda_3}). \quad (8)$$

Table 2 specifies these distributions; we have chosen these hyperparameters based on our substantive knowledge, and from inspection of stem-and-leaf plots like Figure 3 and for many statewide elections (King and Gelman in press). When possible, we approximate to make prior assumptions about θ vague rather than overly restrictive. Note that the prior distribution for γ_i is symmetric about 0, hence treating the political parties equally. We allow the parameters γ and θ to change each election year.

Finally, we truncate this distribution so that $(\rho_j^2 - \sigma^2) \geq 0$ for $j = 1, 2, 3$.

Table 2. Specified Hyperparameter Values for the Prior Distribution on θ

Parameter	$j = 1$	$j = 2$	$j = 3$
$\mu_{\mu j}$	−.4	.4	0
$\sigma_{\mu j}$	.4	.4	3
$\alpha_{\rho j}$	4	4	4
$\beta_{\rho j}$	.16	.16	.64
$a_{\lambda j}$	19	19	4

5.3 Uncontested Elections

For an uncontested Democratic district election, we approximate the uncertainty in the effective vote by the information in the stem-and-leaf plot of Figure 1. We then fit this to a Normal density on the logit scale: that is, for each uncontested seat i,

$$u_{i0} \sim N(\mu_{un}, \sigma_{un}^2).$$

Our data yield the estimates $(\hat{\mu}_{un}, \hat{\sigma}_{un}) = (.74, .57)$. Assuming this distribution to be independent of u_{it} in Equation (2), we get another Normal distribution for the uncontested district effects:

$$\gamma_i \sim N(\mu_{un}, \sigma_{un}^2 - \sigma^2), \tag{9}$$

where $\sigma_{un}^2 > \sigma^2$. We then truncate this distribution to be all-positive, so that an uncontested seat will always favor the winning party. We also symmetrically define γ_i for a Republican uncontested district to be distributed as $N(-\mu_{un}, \sigma_{un}^2 - \sigma^2)$, truncated to be negative. (Recall that 0 on the logit scale is .5 on the votes scale.)

6. BAYESIAN ESTIMATION FOR A SINGLE ELECTION

We summarize posterior distributions by sampling from, in the following order:

1. $P(\theta \mid u_0)$
2. $P(\gamma \mid \theta, u_0)$
3. $P(u_t \mid \delta_t, \gamma, \theta, u_0) = P(u_t \mid \delta_t, \gamma)$.

Together, these steps amount to sampling from the desired posterior distribution of election results. (All of these distributions are of course conditional on the parameters specified in Sec. 5.)

6.1 Averaging Over Uncertainty in θ

The likelihood function $P(u_0 \mid \theta)$ is the product of n independent densities: $u_{i0} \sim$ Normal-mixture$(\mu_j, \rho_j^2, \lambda_j; j = 1, 2, 3)$. The posterior density $P(\theta \mid u_0)$ is cumbersome, because of the Normal-mixture terms in the likelihood. Direct sampling or numerical integration over this eight-dimensional distribution seems impossible. With a Normal likelihood, however, simulation of θ would be easy. We exploit this possibility through the data augmentation algorithm of Tanner and Wong (1987).

First, we decompose the Normal mixture through a matrix of unobserved indicator variables $\tau = (\tau_{ij}; i = 1, \ldots, n; j = 1, 2, 3)$. The likelihood $P(\mu_0 \mid \theta)$ can then be factored into independent multinomial distributions for the indicators $(\tau_{i1}, \tau_{i2}, \tau_{i3} \mid \theta) \sim$ multinomial$(\lambda_1, \lambda_2, \lambda_3; 1)$, for $i = 1, \ldots, n$, and a Normal distribution for the data, conditional on the indicators $(u_{i0} \mid \tau_{ij} = 1, \theta) \sim N(\mu_j, \sigma_j^2)$.

Next, we sample from $P(\theta \mid u_0)$, in two steps, using the intermediate variable τ.

1. Sample from $P(\tau \mid u_0)$
2. Sample from $P(\theta \mid \tau, u_0)$.

Step 2, using Bayes's theorem with our conjugate prior distributions (8), is straightforward:

$$(\rho_j^{-2} \mid \tau, u_0) \sim \Gamma(\tfrac{1}{2}(\alpha_{\rho_j} + n_j), \tfrac{1}{2}(\beta_{\rho_j} + SS_j)),$$
$$j = 1, 2, 3,$$
$$(\mu_j \mid \rho_j^2, \tau, u_0) \sim N(\mu_j^*, \rho_j^*), \qquad j = 1, 2, 3,$$

and

$$(\lambda_1, \lambda_2, \lambda_3 \mid \tau, u_0) \sim \text{Dirichlet}(a_{\lambda_j} + n_j; j = 1, 2, 3),$$

where

$$n_j = \sum_i \tau_{ij}, \qquad \mu_j^* = \frac{\sigma_{\mu_j}^2 n_j \bar{u}_j + \rho_j^2 \mu_{\mu_j}}{\sigma_{u_j}^2 n_j + \rho_j^2},$$

$$SS_j = \sum_i \tau_{ij}(u_{i0} - \bar{u}_j)^2,$$

$$\bar{u}_j = \frac{1}{n_j} \sum_i \tau_{ij} u_{i0}, \qquad \rho_j^{*2} = \frac{\sigma_{\mu_j}^2 \rho_j^2}{n_j \sigma_{\mu_j}^2 + \rho_j^2}.$$

In addition, the values ρ_j^2 are constrained to be no less than σ^2. If we simulate too low a value for a ρ_j, we just keep repeating the simulation of θ until we satisfy the constraint.

Step 1 is intractable as stated but would be easy if θ were known, because

$$(\tau_{i1}, \tau_{i2}, \tau_{i3} \mid \theta, u_0) \sim \text{multinomial}(\lambda_{i1}^*, \lambda_{i2}^*, \lambda_{i3}^*; 1) \qquad \text{for } i = 1, \ldots, n,$$

where

$$\lambda_{ij}^* \propto \lambda_j \frac{1}{\rho_j} \phi\left(\frac{u_j - \mu_j}{\rho_j}\right) \quad \text{for each } i, j,$$

and ϕ is the standard Normal density function. In our application of the data augmentation algorithm, we simulate a single random sample θ^* from $P(\theta \mid u_0)$, as follows.

1. Choose a reasonable starting point for θ^*. We use the posterior maximum of $P(\theta \mid u_0)$, which we estimate by the EM algorithm (Dempster, Laird, and Rubin 1977), again treating τ as unobserved data.
2. Repeat the following steps a number of times: (a) sample τ^* from $P(\tau \mid \theta = \theta^*, u_0)$ and (b) sample θ^* from $P(\theta \mid \tau = \tau^*, u_0)$. For our data, the distribution of simulated values θ^* appears to converge after 10 iterations. Increasing the number of iterations did not noticeably change the distribution of simulated values of θ^* or our final results.

Iterations of this procedure yield approximately independent random samples from the posterior distribution of θ. We found that 50 iterations provided sufficient precision.

6.2 Averaging Over Uncertainty in γ

We can factor the conditional posterior density as follows:

$$P(\gamma \mid \theta, u_0) = \prod_i P(\gamma_i \mid \theta, u_{i0})$$
$$\propto \prod_i P(u_{i0} \mid \gamma_i, \theta) P(\gamma_i \mid \theta).$$

The first factor here is just the Normal error density from the model (2), and the second factor is the Normal-mixture density parameterized by θ. Their product yields a new Normal-mixture density with easily calculated parameters $\tilde{\theta}_i$ for each district; we sample from these independent distributions.

For each uncontested district, we simulate γ_i from the truncated Normal distribution (9). We combine these with the simulated values γ_i for contested districts to get a sample vector γ from its posterior distribution.

6.3 Averaging Over u_t

To estimate the seats–votes curve and its variability, we first approximate the first two moments of the joint conditional distribution $P(V_t, S_t \mid \gamma, \delta_t)$, for several values of δ_t. Figure 5 provides an intuitive sense of our model and sampling procedure by plotting several simulated values u_{it} for $\delta_t = 0$, as a function of observed district votes u_{i0}, for Ohio in 1972. Note the assumed distribution of effective votes for the uncontested districts.

The aggregate votes and seats are averages [Eqs. (3)] of their district-level counterparts v_{it} and s_{it}, which in turn depend on γ_i and δ_t only through their mean $\alpha_{it} = \gamma_i + \delta_t$. Thus the desired conditional moments can be expressed in terms of the following expectations:

$$E(v_{it} \mid \alpha_{it}) = \int_{-\infty}^{\infty} \frac{e^u}{1 + e^u} \frac{1}{\sigma} \phi\left(\frac{u - \alpha_{it}}{\sigma}\right) du,$$

$$E(s_{it} \mid \alpha_{it}) = \int_{0}^{\infty} \frac{1}{\sigma} \phi\left(\frac{u - \alpha_{it}}{\sigma}\right) du = \Phi\,(\alpha_{it}/\sigma),$$

$$\operatorname{var}(v_{it} \mid \alpha_{it}) = \int_{-\infty}^{\infty} \left(\frac{e^u}{1 + e^u}\right)^2 \frac{1}{\sigma} \phi\left(\frac{u - \alpha_{it}}{\sigma}\right) du - [E(v_{it} \mid \alpha_{it})]^2,$$

$$\operatorname{var}(s_{it} \mid \alpha_{it}) = E(s_{it} \mid \alpha_{it})[1 - E(s_{it} \mid \alpha_{it})],$$

and

$$\operatorname{cov}(v_{it}, s_{it} \mid \alpha_{it}) = \int_{0}^{\infty} \frac{e^u}{1 + e^u} \frac{1}{\sigma} \phi\left(\frac{u - \alpha_{it}}{\sigma}\right) du - E(s_{it} \mid \alpha_{it}) E(V_{it} \mid \alpha_{it}).$$

Some of the foregoing integrals are immediately evaluated through the standard Normal distribution function Φ; we calculate the rest by approximating the inverse logit function $e^u/(1 + e^u)$ by a third-degree polynomial in u.

We now approximate the seats–votes curve $E(S \mid V)$ versus V by the function defined by $E(S_t \mid \alpha_t)$ versus $E(V_t \mid \alpha_t)$, implicitly parameterized by α_t (or, equivalently, by the scalar δ_t). Similarly, we approximate the variance as follows:

$$\operatorname{var}(S_t \mid V_t) \approx \operatorname{var}(S_t \mid \alpha_t) - \frac{\operatorname{cov}(V_t, S_t \mid \alpha_t)}{\operatorname{var}(V_t \mid \alpha_t)}.$$

This variance depends on V_t and is parameterized by δ_t in the foregoing expression. The formula would be exactly correct if S_t and V_t were jointly Normally distributed, and it is a reasonable approximation for our problem.

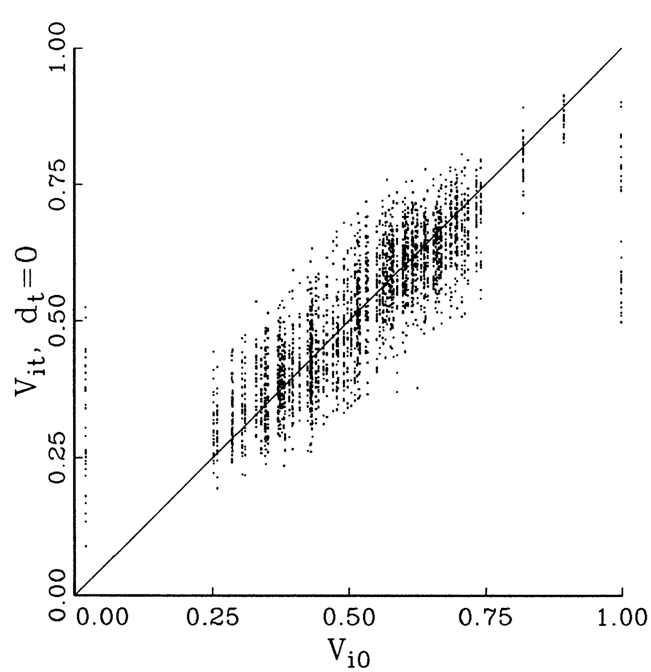

Figure 5. Simulations, Ohio, 1972.

6.4 Calculating Summaries

Finally, we simulate several vectors γ from the posterior density $P(\gamma \mid u_0)$. Each of these samples determines an electoral system, for which we approximate the seats–votes curve and its variance, as described previously. From the seats–votes curve, we calculate the bias and responsiveness of the system between 45% and 55% [Eqs. (7)]. Finally, we estimate the bias and responsiveness of the true electoral system, and our uncertainty in these quantities, with the sample mean and variance of these values, over the many independent samples of γ.

All computations were done in the Gauss computer language on an IBM PS/2.

7. RESULTS

The procedure described in Section 6 produces estimates of an electoral system from the results of a single statewide election. This includes estimates of the seats–votes curve, its variability, and summaries such as the bias and responsiveness functions. Our model assumes that district votes move in an approximate uniform manner as the statewide vote totals change. Because of the lack of information, we assume the absence of spatial correlation. Finally, we assume that the district votes roughly follow a three-hump distribution specified by our family of prior distributions. Within these constraints, our model is quite general and fits recent legislative electoral data quite well.

An example of the complete results appear in Figure 6. The solid line in this figure is the estimated seats–votes curve $E(S \mid V)$ for Ohio in 1972. The dotted lines show plus and minus two standard errors of estimation: $E(S \mid V) \pm 2\operatorname{var}(E(S \mid V, \gamma))^{1/2}$. Instead of presenting seven of these figures for each of three states, we summarize the

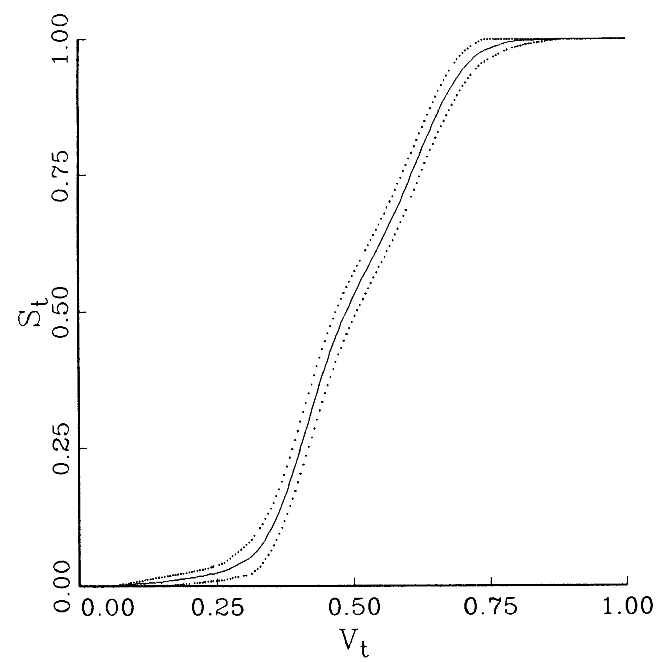

Figure 6. Estimated Seats–Votes Curve, Ohio, 1972.

results for each election from 1968 to 1980, using Formula (7).

The results for all seven years in Ohio appear in Figure 7, where responsiveness is plotted by partisan bias. Pooled standard error estimates appear in the lower left of the figure. The black square marks 1968, a year of moderate responsiveness but with an extreme bias favoring the Republicans. The next square is 1970, which is close to and within two standard errors of 1968. In 1971, the Democrats controlled the redistricting process, dramatically affecting Ohio's electoral system: the dotted line drawn between 1970 and 1972, to indicate redistricting, represents a systematic change from extreme Republican bias to slight Democratic bias—far beyond what one would expect due to mere random variability. The change also appears permanent, as the elections over the course of the rest of the decade remain at or above the initial level of Democratic

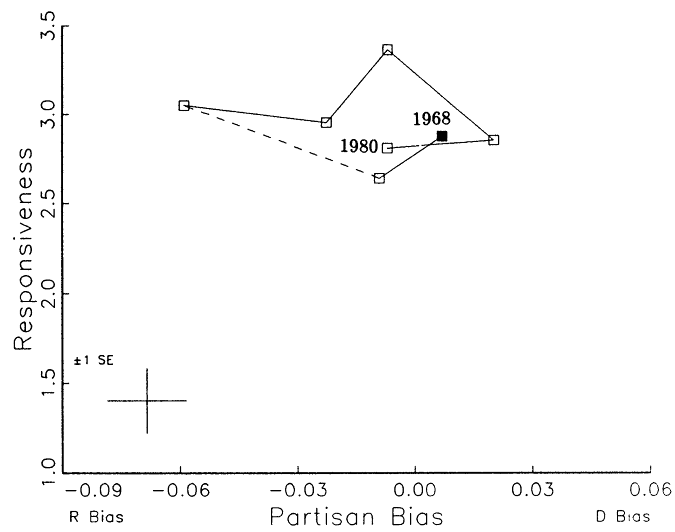

Figure 8. Connecticut House, 1968–1980.

bias. The other change in the figure is a noticeable trend after redistricting toward lower responsiveness.

The changes in Connecticut's electoral system are portrayed in Figure 8. All of the years in Connecticut have electoral systems that are quite responsive, particularly compared with Ohio. In 1968 and 1970, Connecticut had essentially no partisan bias. The 1971 redistricting was controlled by the Republicans, and their effect in biasing the system in their favor seems quite dramatic—again much beyond what one would expect due to random variability. This dramatic effect seems ephemeral, however, since over the course of the rest of the decade the electoral system worked its way back to just about where it began. The Republican gerrymanderers in Connecticut were obviously not as successful as their Democratic counterparts in Ohio. We speculate that the pattern of incumbency retirements accounts for this difference—particularly since the Watergate landslide in 1974 helped to defeat many Republican state legislators.

Figure 9 portrays Wisconsin's electoral system. Because a single party did not elect a governor and a majority of

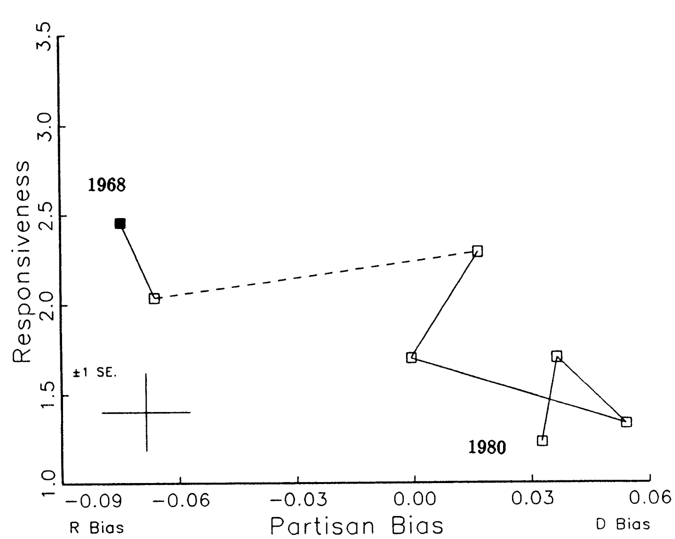

Figure 7. Ohio House, 1968–1980.

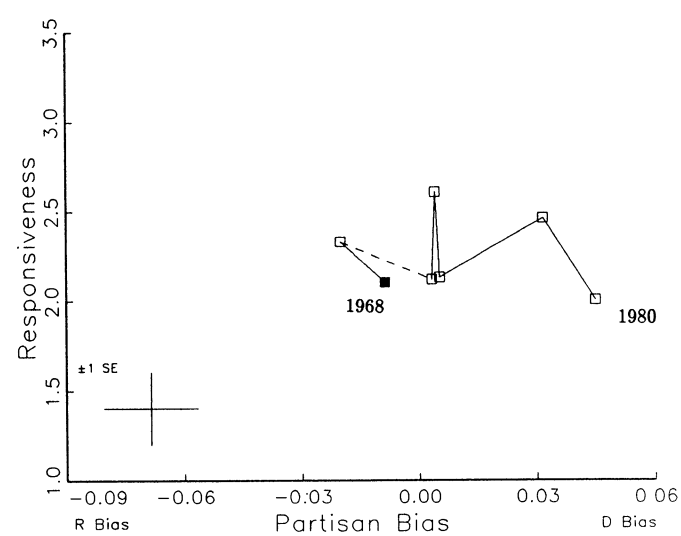

Figure 9. Wisconsin House, 1968–1980.

both houses of the state legislature, Wisconsin was redistricted by a bipartisan agreement between the parties. Redistricting thus has a quite predictable non-effect on the system: the change from 1970 to 1972 is no greater than most other changes between consecutive elections in this graph. Political scientists have speculated that bipartisan redistricters primarily try to protect incumbents; with fewer seats of both parties vulnerable to electoral swings, this would decrease responsiveness (Mayhew 1971). Surprisingly, Wisconsin's responsiveness changes no more across redistricting than between any other two consecutive elections. Of course, responsiveness in Wisconsin started from a low base; perhaps redistricters could not reduce responsiveness any further due to the geographic pattern of voters in the state.

When controlling the redistricting process, partisans have successfully biased the electoral system in their favor, at least in the short term. A glance at Figures 7–9 shows that redistricting had no systematic effect on responsiveness in any of the three states. All previous seats–votes models have been either deterministic, entirely theoretical, or average over many elections. Some have ignored partisan bias and either fit responsiveness or fixed it to the value of 3.0; other models have assumed the electoral system to be constant over several elections. We explicitly model variability and generate estimates and standard errors of bias and responsiveness for each statewide election. A comparison of the changes between elections with the standard errors in Figures 7–9 leads us to reject deterministic models and those with constant bias and responsiveness.

[*Received November 1988. Revised September 1989.*]

REFERENCES

Ansolabehere, S., Brady, D., and Fiorina, M. (1988), "Turnout and the Calculation of Swing Ratios," technical report, Stanford University, Graduate School of Business.

Butler, D. E. (1951), "Appendix," in *The British General Election of 1950,* ed. H. G. Nicholas, London: Macmillan, pp. 306–333.

Cain, B. (1984), *The Reapportionment Puzzle,* Berkeley: University of California Press.

Davis v. Bandemer (1986), 106 U.S. Supreme Court 2797.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood From Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society,* Ser. B, 39, 1–38.

Gelman, A., and King, G. (in press), "Estimating Incumbency Advantage Without Bias," *American Journal of Political Science,* 34.

Gudgin, G., and Taylor, P. J. (1979), *Seats, Votes, and the Spatial Organisation of Elections,* London: Pion.

Kendall, M. G., and Stuart, A. (1950), "The Law of the Cubic Proportion in Election Results," *British Journal of Sociology,* 1, 193–196.

King, G. (1989), "Representation Through Legislative Redistricting: A Stochastic Model," *American Journal of Political Science,* 33, 787–824.

King, G., and Browning, R. X. (1987), "Democratic Representation and Partisan Bias in Congressional Elections," *American Political Science Review,* 81, 1251–1276.

King, G., and Gelman, A. (in press), "Systemic Consequences of Incumbency Advantage in U.S. House Elections," *American Journal of Political Science,* 35.

Mayhew, D. R. (1971), "Congressional Representation: Theory and Practice in Drawing the Districts," in *Reapportionment in the 1970s,* ed. N. W. Polsby, Berkeley: University of California Press, pp. 274–284.

Niemi, R. G., and Fett, P. (1986), "The Swing Ratio: An Explanation and an Assessment," *Legislative Studies Quarterly,* 11, 75–90.

Quandt, R. E. (1974), "A Stochastic Model of Elections in Two-Party Systems," *Journal of the American Statistical Association,* 69, 315–324.

Tanner, M. A., and Wong, W. H. (1987), "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association,* 82, 528–550.

Tufte, E. R. (1973), "The Relationship Between Seats and Votes in Two-Party Systems," *American Political Science Review,* 67, 540–554.