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
<table>
<thead>
<tr>
<th>District</th>
<th>Democrat</th>
<th>Republican</th>
<th>District</th>
<th>Democrat</th>
<th>Republican</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>18,250</td>
<td>22,798</td>
<td>51</td>
<td>22,488</td>
<td>16,951</td>
</tr>
<tr>
<td>2</td>
<td>25,679</td>
<td>17,130</td>
<td>52</td>
<td>24,336</td>
<td>14,083</td>
</tr>
<tr>
<td>3</td>
<td>25,752</td>
<td>33,954</td>
<td>53</td>
<td>25,932</td>
<td>8,997</td>
</tr>
<tr>
<td>4</td>
<td>23,864</td>
<td>10,212</td>
<td>54</td>
<td>22,780</td>
<td>15,229</td>
</tr>
<tr>
<td>5</td>
<td>21,723</td>
<td>16,130</td>
<td>55</td>
<td>20,198</td>
<td>9,583</td>
</tr>
<tr>
<td>6</td>
<td>28,309</td>
<td>0</td>
<td>56</td>
<td>21,603</td>
<td>10,678</td>
</tr>
<tr>
<td>7</td>
<td>20,334</td>
<td>12,675</td>
<td>57</td>
<td>16,533</td>
<td>17,114</td>
</tr>
<tr>
<td>8</td>
<td>16,622</td>
<td>3,656</td>
<td>58</td>
<td>13,587</td>
<td>22,105</td>
</tr>
<tr>
<td>9</td>
<td>11,945</td>
<td>10,396</td>
<td>59</td>
<td>14,877</td>
<td>20,234</td>
</tr>
<tr>
<td>10</td>
<td>12,383</td>
<td>9,576</td>
<td>60</td>
<td>14,556</td>
<td>13,940</td>
</tr>
<tr>
<td>11</td>
<td>20,091</td>
<td>18,539</td>
<td>61</td>
<td>16,507</td>
<td>17,825</td>
</tr>
<tr>
<td>12</td>
<td>18,337</td>
<td>20,561</td>
<td>62</td>
<td>23,668</td>
<td>13,428</td>
</tr>
<tr>
<td>13</td>
<td>16,688</td>
<td>1,970</td>
<td>63</td>
<td>13,886</td>
<td>18,402</td>
</tr>
<tr>
<td>14</td>
<td>22,885</td>
<td>11,218</td>
<td>64</td>
<td>13,984</td>
<td>22,593</td>
</tr>
<tr>
<td>15</td>
<td>21,401</td>
<td>0</td>
<td>65</td>
<td>11,710</td>
<td>29,134</td>
</tr>
<tr>
<td>16</td>
<td>27,783</td>
<td>12,701</td>
<td>66</td>
<td>15,500</td>
<td>30,156</td>
</tr>
<tr>
<td>17</td>
<td>24,511</td>
<td>15,716</td>
<td>67</td>
<td>20,409</td>
<td>17,931</td>
</tr>
<tr>
<td>18</td>
<td>28,805</td>
<td>14,454</td>
<td>68</td>
<td>21,489</td>
<td>15,574</td>
</tr>
<tr>
<td>19</td>
<td>17,687</td>
<td>23,463</td>
<td>69</td>
<td>16,592</td>
<td>21,816</td>
</tr>
<tr>
<td>20</td>
<td>25,225</td>
<td>28,639</td>
<td>70</td>
<td>14,172</td>
<td>21,642</td>
</tr>
<tr>
<td>21</td>
<td>12,392</td>
<td>23,427</td>
<td>71</td>
<td>22,439</td>
<td>20,831</td>
</tr>
<tr>
<td>22</td>
<td>16,635</td>
<td>21,940</td>
<td>72</td>
<td>15,616</td>
<td>19,879</td>
</tr>
<tr>
<td>23</td>
<td>16,986</td>
<td>7,681</td>
<td>73</td>
<td>0</td>
<td>26,079</td>
</tr>
<tr>
<td>24</td>
<td>22,856</td>
<td>12,779</td>
<td>74</td>
<td>22,359</td>
<td>12,626</td>
</tr>
<tr>
<td>25</td>
<td>20,298</td>
<td>12,292</td>
<td>75</td>
<td>14,653</td>
<td>27,063</td>
</tr>
<tr>
<td>26</td>
<td>15,181</td>
<td>30,866</td>
<td>76</td>
<td>16,438</td>
<td>24,947</td>
</tr>
<tr>
<td>27</td>
<td>12,045</td>
<td>35,880</td>
<td>77</td>
<td>14,054</td>
<td>23,185</td>
</tr>
<tr>
<td>28</td>
<td>20,637</td>
<td>27,011</td>
<td>78</td>
<td>18,867</td>
<td>24,829</td>
</tr>
<tr>
<td>29</td>
<td>17,418</td>
<td>13,589</td>
<td>79</td>
<td>15,459</td>
<td>26,221</td>
</tr>
<tr>
<td>30</td>
<td>15,080</td>
<td>9,381</td>
<td>80</td>
<td>24,237</td>
<td>17,392</td>
</tr>
<tr>
<td>31</td>
<td>19,754</td>
<td>12,971</td>
<td>81</td>
<td>14,606</td>
<td>24,845</td>
</tr>
<tr>
<td>32</td>
<td>20,068</td>
<td>13,059</td>
<td>82</td>
<td>18,349</td>
<td>24,436</td>
</tr>
<tr>
<td>33</td>
<td>13,182</td>
<td>22,046</td>
<td>83</td>
<td>12,650</td>
<td>28,287</td>
</tr>
<tr>
<td>34</td>
<td>15,101</td>
<td>14,159</td>
<td>84</td>
<td>23,448</td>
<td>17,882</td>
</tr>
<tr>
<td>35</td>
<td>19,344</td>
<td>10,166</td>
<td>85</td>
<td>15,896</td>
<td>24,702</td>
</tr>
<tr>
<td>36</td>
<td>19,375</td>
<td>7,792</td>
<td>86</td>
<td>18,869</td>
<td>22,815</td>
</tr>
<tr>
<td>37</td>
<td>17,149</td>
<td>11,274</td>
<td>87</td>
<td>21,828</td>
<td>15,253</td>
</tr>
<tr>
<td>38</td>
<td>10,759</td>
<td>30,945</td>
<td>88</td>
<td>20,732</td>
<td>12,816</td>
</tr>
<tr>
<td>39</td>
<td>24,246</td>
<td>18,772</td>
<td>89</td>
<td>27,325</td>
<td>16,336</td>
</tr>
<tr>
<td>40</td>
<td>21,006</td>
<td>20,625</td>
<td>90</td>
<td>25,239</td>
<td>18,272</td>
</tr>
<tr>
<td>41</td>
<td>29,507</td>
<td>11,524</td>
<td>91</td>
<td>19,783</td>
<td>20,492</td>
</tr>
<tr>
<td>42</td>
<td>21,635</td>
<td>17,233</td>
<td>92</td>
<td>20,567</td>
<td>20,749</td>
</tr>
<tr>
<td>43</td>
<td>26,149</td>
<td>9,428</td>
<td>93</td>
<td>11,803</td>
<td>27,093</td>
</tr>
<tr>
<td>44</td>
<td>24,020</td>
<td>17,601</td>
<td>94</td>
<td>16,508</td>
<td>19,097</td>
</tr>
<tr>
<td>45</td>
<td>22,872</td>
<td>0</td>
<td>95</td>
<td>10,642</td>
<td>26,668</td>
</tr>
<tr>
<td>46</td>
<td>23,080</td>
<td>11,743</td>
<td>96</td>
<td>27,270</td>
<td>14,044</td>
</tr>
<tr>
<td>47</td>
<td>20,465</td>
<td>8,920</td>
<td>97</td>
<td>16,859</td>
<td>13,746</td>
</tr>
<tr>
<td>48</td>
<td>18,729</td>
<td>27,079</td>
<td>98</td>
<td>28,067</td>
<td>11,687</td>
</tr>
<tr>
<td>49</td>
<td>19,809</td>
<td>18,632</td>
<td>99</td>
<td>26,945</td>
<td>14,848</td>
</tr>
<tr>
<td>50</td>
<td>18,036</td>
<td>19,734</td>
<td>100</td>
<td>14,994</td>
<td>14,481</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Year</th>
<th>Votes</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1972</td>
<td>1234</td>
<td>5677789999</td>
</tr>
<tr>
<td>1974</td>
<td>0000111122223333344444</td>
<td>56777888888899999999</td>
</tr>
<tr>
<td>0.7</td>
<td>000001111111222222222233333334444444</td>
<td>555555566677788888</td>
</tr>
<tr>
<td>0.8</td>
<td>00011111122333</td>
<td>66799</td>
</tr>
<tr>
<td>0.9</td>
<td>11344</td>
<td>0.5</td>
</tr>
</tbody>
</table>

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.
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 (Tuft 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:

\[
\text{Responsiveness}(V) = \frac{dE(S \mid V)}{dV} \\
\text{Bias}(V) = E(S \mid V) \left[ 1 - E(S \mid 1 - V) \right]. \tag{1}
\]

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 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 |
| 0.4 | 57777788999 |
| 0.5 | 0222233334 |
| 0.6 | 5577899 |
| 0.7 | 0111233 |
| 0.8 | 5566677778899 |
| 0.9 | 000111233344 |
| 1 | 566667788999 |
| 9 | 01134 |

Figure 3. Stem-and-Leaf Plot of the Democratic Proportion of the Two-Party Vote in Contested District Elections in Ohio, 1972.
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 \( u_{it} = \logit(v_{it}) = \ln[v_{it}/(1 - v_{it})] \) 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

\[
u_{it} \sim N(\alpha_i, \sigma^2), \quad \alpha_i = \gamma_i + \delta_i, \tag{2}\]

where \( \gamma_i \) is a district effect, \( \delta_i \) 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 unavali-
able 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:

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

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

We consider the vector \( \gamma = (\gamma_1, \ldots, \gamma_n) \), along with the variance \( \sigma^2 \), 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 \( \gamma \) remain unknown, we model them as random effects by letting the \( \gamma_i \)'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 \( \gamma \) 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 \( \delta_i = 0 \). 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 \( \delta_t \). 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 \( \gamma \) and \( \delta \).

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 \( \sigma^2 \) and uncontested effective vote parameters \( \mu_{in} \) and \( \sigma_{in} \), described in Section 5.

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

3. The Seats–Votes Curve. We average over \( P(\gamma \mid u_i) \) 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 \( \delta_t \) to range from \( -\infty \) to \( \infty \) 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 \( \gamma \):

\[
\text{var}(E(S_i | \mathcal{V}, u_0, \gamma)).
\]

(6)

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

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

\[
\frac{1}{.55 - .45} \int_{.45}^{.55} (E(S | V) - [1 - E(S | 1 - V)]) dV
\]

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

\[
\frac{1}{.55 - .45} [E(S | V = .55) - E(S | V = .45)].
\]

(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 \( u_i \) from the district effects \( \gamma_i \) by adding a constant shift \( \delta \) to the mean in every district. From here, we add the variability in (2); this “unexplained” variance \( \sigma^2 \) 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: \( \gamma_i = u_i \). This assumption of “uniform partisan swing” on the logit scale cannot hope to fit more than a single statewide election.

We estimate \( \sigma^2 \) 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 \( u_{\delta i} - u_{\delta j} \) 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:

\[
s^2_{\delta i} = \frac{1}{n_{\delta i}} \sum_i [u_{\delta i} - u_{\delta j} - (\bar{u}_{\delta i} - \bar{u}_{\delta j})]^2,
\]

where \( n_{\delta i} \) 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 \( s^2_{\delta i} \), 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 \( \sigma \) (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 \( \gamma_i \)

We need to estimate the vector \( \gamma \) of district effects and our uncertainty in it. Embedding \( \gamma \) in a lower-dimensional probabilistic model allows us to estimate these \( n \) district effects from the \( n \) data points \( u_0; \gamma \); 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 \( \theta = (\mu_j, \rho_j^2, \sigma_j^2; \lambda_j; j = 1, 2, 3) \) 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 \( \rho_j^2 \) is the variance of the \( j \)th Normal distribution in the density of observed district vote proportions \( u_{0i} \); \( (\rho_j^2 - \sigma^2) \) is the variance of the \( j \)th Normal distribution in the density of expectations \( \gamma_i \).

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:

\[
\mu_j \sim N(\mu_j, \sigma_j^2), \quad j = 1, 2, 3
\]

\[
\rho_j^2 \sim \Gamma(1/2 a_j, 1/2 b_j), \quad j = 1, 2, 3
\]

\[
(\lambda_1, \lambda_2, \lambda_3) \sim \text{Dirichlet}(a_1, a_2, a_3).
\]

(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 \( \theta \) vague rather than overly restrictive. Note that the prior distribution for \( \gamma_i \) is symmetric about 0, hence treating the political parties equally. We allow the parameters \( \gamma \) and \( \theta \) 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 \( \theta \)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>( j = 1 )</th>
<th>( j = 2 )</th>
<th>( j = 3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu_j )</td>
<td>-.4</td>
<td>.4</td>
<td>0</td>
</tr>
<tr>
<td>( \sigma_j )</td>
<td>.4</td>
<td>.4</td>
<td>3</td>
</tr>
<tr>
<td>( \alpha_j )</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>( \beta_j )</td>
<td>.16</td>
<td>.16</td>
<td>.64</td>
</tr>
<tr>
<td>( a_j )</td>
<td>19</td>
<td>19</td>
<td>4</td>
</tr>
</tbody>
</table>
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 \( i \),

\[
u_i \sim N(\mu_{ui}, \sigma_{ui}^2).
\]

Our data yield the estimates \((\mu_{ui}, \sigma_{ui}) = (7.4, 5.7)\). Assuming this distribution to be independent of \( u_i \) in Equation (2), we get another Normal distribution for the uncontested district effects:

\[
\gamma_i \sim N(\mu_{ui}, \sigma_{ui}^2 - \sigma^2),
\]

(9)

where \( \sigma_{ui}^2 > \sigma^2 \). We then truncate this distribution to be all-positive, so that an uncontested seat will always favor the winning party. We also symmetrically define \( \gamma_i \) for a Republican uncontested district to be distributed as \( N(-\mu_{ui}, \sigma_{ui}^2 - \sigma^2) \), 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. \( P(u_i \mid \delta_i, \gamma, \theta, u_0) = P(u_i \mid \delta_i, \gamma) \).

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 \( \theta \)

The likelihood function \( P(u_i \mid \theta) \) is the product of \( n \) independent densities: \( u_{i0} \sim \text{Normal-mixture}(\mu_j, \rho_j^2, \lambda_j; j = 1, 2, 3) \). 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 \( \theta \) 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 \( (\tau_{i1}, \tau_{i2}, \tau_{i3}) \sim \text{multinomial}(\lambda_{i1}, \lambda_{i2}, \lambda_{i3}; 1) \), 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 \( \tau \).

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:

\[
(\rho_{j}^{-2} \mid \tau, u_0) \sim \Gamma(\begin{pmatrix} \frac{1}{\alpha_{j}}, & n_{j} \end{pmatrix}, \begin{pmatrix} \frac{1}{\beta_{j}}, & SS_{j} \end{pmatrix}),
\]

\[
j = 1, 2, 3,
\]

\[
(\mu_{j} \mid \rho_{j}^{-2}, \tau, u_0) \sim N(\mu_{j}^{*}, \rho_{j}^{*}),
\]

\[
j = 1, 2, 3,
\]

and

\[
(\lambda_{1}, \lambda_{2}, \lambda_{3} \mid \tau, u_0) \sim \text{Dirichlet}(\alpha_{j} + n_{j}; j = 1, 2, 3),
\]

where

\[
n_{j} = \sum_{i} \tau_{ij} \mu_{j}^{*} = \frac{\sigma_{j}^{2} n_{j} u_{ij} + \rho_{j}^{2} \mu_{j}^{*}}{\sigma_{j}^{2} n_{j} + \rho_{j}^{2}},
\]

\[
SS_{j} = \sum_{i} \tau_{ij}(u_{i0} - \bar{u}_{ij})^{2},
\]

\[
\bar{u}_{ij} = \frac{1}{n_{ij}} \sum_{j} \tau_{ij} u_{i0}, \quad \rho_{j}^{*2} = \frac{\sigma_{j}^{2} \rho_{j}^{2}}{n_{ij} \sigma_{ui}^{2} + \rho_{j}^{2}}.
\]

In addition, the values \( \rho_{j}^{2} \) are constrained to be no less than \( \sigma^{2} \). If we simulate too low a value for a \( \rho_{j} \), we just keep repeating the simulation of \( \theta \) until we satisfy the constraint.

Step 1 is intractable as stated but would be easy if \( \theta \) were known, because

\[
(\tau_{i1}, \tau_{i2}, \tau_{i3} \mid \theta, u_0) \sim \text{multinomial}(\lambda_{i1}^{*}, \lambda_{i2}^{*}, \lambda_{i3}^{*}; 1)
\]

for \( i = 1, \ldots, n \),

where

\[
\lambda_{ij}^{*} \propto \frac{1}{\rho_{j}^{*}} \phi \left( \frac{u_{ij} - \mu_{j}^{*}}{\rho_{j}} \right)
\]

for each \( i, j \),

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

1. Choose a reasonable starting point for \( \theta^{*} \). 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 \( \tau \) as unobserved data.

2. Repeat the following steps a number of times: (a) sample \( \tau^{*} \) from \( P(\tau \mid \theta = \theta^{*}, u_0) \) and (b) sample \( \theta^{*} \) from \( P(\theta \mid \tau = \tau^{*}, u_0) \). For our data, the distribution of simulated values \( \theta^{*} \) appears to converge after 10 iterations. Increasing the number of iterations did not noticeably change the distribution of simulated values of \( \theta^{*} \) or our final results.

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

6.2 Averaging Over Uncertainty in \( \gamma \)

We can factor the conditional posterior density as follows:

\[
P(\gamma \mid \theta, u_0) = \prod_{j} P(\gamma_{j} \mid \theta, u_0) \propto \prod_{j} P(u_{j0} \mid \gamma_{j}, \theta)P(\gamma_{j} \mid \theta).
\]
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 $\theta$. Their product yields a new Normal-mixture density with easily calculated parameters $\delta_i$ for each district; we sample from these independent distributions.

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

### 6.3 Averaging Over $u_i$

To estimate the seats–votes curve and its variability, we first approximate the first two conditional moments of the joint distribution $P(V_i, S_i \mid \gamma, \delta_i)$, for several values of $\delta_i$. Figure 5 provides an intuitive sense of our model and sampling procedure by plotting several simulated values $u_i$ for $\delta_i = 0$, 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 $v_{i0}$ and $s_{i0}$, which in turn depend on $\gamma_i$ and $\delta_i$ only through their mean $\alpha_i = \gamma_i + \delta_i$. Thus the desired conditional moments can be expressed in terms of the following expectations:

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

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

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

$$\text{var}(s_i \mid \alpha_i) = E(s_i \mid \alpha_i)[1 - E(s_i \mid \alpha_i)].$$

and

$$\text{cov}(v_i, s_i \mid \alpha_i) = \int_{0}^{\infty} \frac{e^u}{1 + e^u} \frac{1}{\sigma} \phi \left( \frac{u - \alpha_i}{\sigma} \right) du - E(s_i \mid \alpha_i)E(v_i \mid \alpha_i).$$

Some of the foregoing integrals are immediately evaluated through the standard Normal distribution function $\Phi$; 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 $E(S_i \mid \alpha_i)$ versus $E(V_i \mid \alpha_i)$, implicitly parameterized by $\alpha_i$ (or, equivalently, by the scalar $\delta_i$). Similarly, we approximate the variance as follows:

$$\text{var}(S_i \mid V_i) = \text{var}(S_i \mid \alpha_i) - \frac{\text{cov}(V_i, S_i \mid \alpha_i)}{\text{var}(V_i \mid \alpha_i)}.$$

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

### 6.4 Calculating Summaries

Finally, we simulate several vectors $\gamma$ from the posterior density $P(\gamma \mid u_{i0})$. 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 $\gamma$.

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: $E(S \mid V) \pm 2 \text{var}(E(S \mid V, \gamma))^{1/2}$. Instead of presenting seven of these figures for each of three states, we summarize the
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 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
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


