Predicting House Elections: Who will take the majority?

David Villarreal
Intro to problem

- How can we use ex ante data to predict who will control the majority in the House of Representatives?
- How can we expand upon Prof. Fair’s House Election equations to answer the more politically salient question of who will take the House?
- Vote share is not always accurate measure of House majority (considering gerrymandering).
Past Literature

1. Prof Fair’s Equations
   a. Predicts vote share in congressional elections.

2. What do I add?
   a. Translate vote share into more useful metric
   b. Use different approach, and instead measure percent of seats taken in particular state, not vote share
   c. Use Monte Carlo simulation to predict total #seats taken by Democrats using ex ante data
The Data

1. MIT Election Lab (Tabulated per candidate - district - year, 1978-2020)
   a. R Script to change district-by-district election to state election
2. State growth data and inflation data from St. Louis FED
## Workflow (simplified example)

<table>
<thead>
<tr>
<th>Candidate</th>
<th>State</th>
<th>District</th>
<th>Votes</th>
<th>Party</th>
</tr>
</thead>
<tbody>
<tr>
<td>John Doe</td>
<td>Texas</td>
<td>13</td>
<td>44,000</td>
<td>Republican</td>
</tr>
<tr>
<td>David Villarreal</td>
<td>Texas</td>
<td>13</td>
<td>12,000</td>
<td>Democratic</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>State</th>
<th>Year</th>
<th>%lagged</th>
<th>%DemSeat s</th>
<th>%demInc</th>
<th>%repInc</th>
<th>Inflation</th>
<th>stateGrowt h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Texas</td>
<td>2018</td>
<td>18%</td>
<td>23%</td>
<td>15%</td>
<td>50%</td>
<td>4%</td>
<td>1%</td>
</tr>
<tr>
<td>Vermont</td>
<td>2018</td>
<td>80%</td>
<td>85%</td>
<td>75%</td>
<td>0%</td>
<td>4%</td>
<td>10%</td>
</tr>
</tbody>
</table>
The Regression (SUBJECT TO CHANGE A BIT)

%DemocraticSeats ~

1. Constant
2. Lagged seats (percentage of seats taken by Democrats in previous election)
3. % Candidates that are Democratic Incumbents
4. % Candidates that are Republican Incumbents
5. White House (who’s president, 1 = Dem, 0 = Rep)
6. State Growth Rate * White House
7. Inflation * White House
8. Gerrymandering Index * State Legislature Control (missing data, each state does it differently)
9. National Growth Rate * White House

N = 1,110
Call:
`lm(formula = percentDemocratic ~ laggedSeats + Incumbents + republicanIncumbent + trueGrowth:I + Inflation:I + I, data = subset(electionTable, Year > 1978))`

Residuals:

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>1Q</th>
<th>Median</th>
<th>3Q</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.91948</td>
<td>-0.06492</td>
<td>-0.00043</td>
<td>0.05960</td>
<td>0.94292</td>
</tr>
</tbody>
</table>

Coefficients:

|                        | Estimate | Std. Error | t value | Pr(>|t|) |
|------------------------|----------|------------|---------|---------|
| (Intercept)            | 0.154613 | 0.019455  | 7.947   | 5.02e-15 *** |
| laggedSeats            | 0.623719 | 0.030730  | 20.297  | < 2e-16 *** |
| Incumbents             | 0.187917 | 0.028062  | 6.696   | 3.52e-11 *** |
| republicanIncumbent    | -0.109676| 0.024143  | -4.543  | 6.21e-06 *** |
| I                      | -0.020911| 0.008783  | -2.381  | 0.01746 *  |
| trueGrowth:I          | -0.005525| 0.002049  | -2.697  | 0.00711 ** |
| I:Inflation           | -0.005109| 0.002122  | -2.408  | 0.01623 *  |

---

Signif. codes:  
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1668 on 1021 degrees of freedom
(23 observations deleted due to missingness)
1. Once you have coefficients for regression, you can now create for every “state election” in a future congressional election:
   a. %Democratic Seats (HAT) -> estimate
   b. Standard error
2. Thus, for every state you have an estimated Democratic seat share and a standard error.
3. Monte Carlo simulation:
   a. For every state, pick from distribution with mean = %Democratic Seats (HAT), SD = residual standard error, and multiply by # seats available.
   b. Run a) X times, and return the proportion where the total number of democratic seats > 217 (majority)
Results using ex ante data from previous elections
Things to add/change

1. Better way to handle Gerrymandering
2. Better way of using growth
3. Add boundaries to estimates (you often get estimates that are <0 or >100% Democratic seats, which doesn’t make sense)