Electoral College Forecasting

Graham Tierney

9/29/2020

US Presidential Elections

US Presidential elections have some unique features among advanced democracies.

1. States have electoral college (EC) votes equal to the number of Congressional representatives.
2. Nearly all states use winner-take-all apportionment of EC votes; the state-level popular vote winner receivers all votes.
3. The candidate who wins at least 270 of the 538 votes in the EC wins the election.

Electoral College

1. Each state has two Senators and at least one Representative, so states have a minimum of 3 EC votes. The states with the most vote are CA (55), TX (38), NY and FL (29).
2. The two exceptions to the winner-take-all are Maine (ME) and Nebraska (NE). They award their EC votes each by congressional district with the two remaining votes. NE-2 (Omaha) and ME-2 (excludes Agusta and Portland) are probably competative in 2020. The state-wide winner must win at least one district.
3. Not all forecasters model NE and ME separately, e.g. the Economist does not.

Swing States

1. Some forecasters focus on so-called “swing states”, where the election outcome is in doubt. There is no consensus definition or list, and forecasters tend to be over rather than under-inclusive.
2. The Economist model rates the following states/districts as at best "likely" (65-85

Forecasting the Presidential Election

Key challenges:

1. Preferences change over time. Polls are a snapshot not a forecast.
2. State-level polls are conducted infrequently and irregularly.
3. **State-level outcomes are correlated.**

State-level correlations

Significant differences in 538's model and others for the 2016 election was modeling correlation in state-level polling misses. If Trump won Iowa, the 538 model would also project him as the likely winner in Wisconsin and Minnesota (nearby, demographically similar states). Clinton was ahead by a small margin in most swing state polls. If states are independent, a Trump win would require multiple unlikely, independent events. If states are dependent, a Trump win only requires a single (or fewer) unlikely events.

State-level correlations

As a simple, example, suppose there are 3 states with equal EC votes and candidate A is has the same probability p of winning each of them. If states are independent, then $P(\text{A wins}) = P(\text{A wins at least two}) = \binom{3}{2} p^2(1-p) + p^3$. In the other extreme of each state's outcome being identical, then $P(\text{A wins}) = p$. For $p > .5$, candidate A has a higher probability of winning in the independent scenario.

Election forecast simplifications

1. Forecasters often focus just on predicting the two-party (Democrat and Republican) share of the vote in each state, ignoring 3rd parties and undecided voters in polls.
2. Polls are conducted over several days, but many models assume that the polls are relevant measures for only a single day (usual the midpoint or end date of the poll). If you do something like this, a good sensitivity analysis is to check how the forecast changes if you change the date used.

Election forecast example - Linzer (2013)

Linzer (2013) presents a Bayesian forecast model for the 2008 election that has become a useful reference and starting point for Bayesian election forecasting. The Economist model builds off of the fundamentals in this model.

Essentially, the model specifies how preferences evolve over time, and how polls are noisy measurements of the underlying preferences.

Linzer (2013)

Each poll k reports y_k, the number of respondents voting for the Democrat, and n_k the number respondents voting for either the Democrat or the Republican. $i[k]$ indexes the state and $j[k]$ the date of poll k. The election happens on day 1, polls from at most J days before the election are used.

$$\begin{aligned}
y_k &\sim \text{Binom}(\pi_{i[k]j[k]}, n_k) && (1)\\
\text{logit}(\pi_{ij}) &= \beta_{ij} + \delta_j && (2)\\
\text{for } j > 1\text{: } \beta_{ij} &\sim N(\beta_{i,j-1}, \sigma^2_\beta) && (3)\\
\delta_j &\sim N(\delta_{j-1}, \sigma^2_\delta) && (4)\\
\text{for } j = 1\text{: } \beta_{i1} &\sim N(\text{logit}(h_i), s_i^2) && (5)\\
\delta_1 &:= 0 && (6)
\end{aligned}$$

Linzer (2013) - Observation model

$$y_k \sim \text{Binom}(\pi_{i[k]j[k]}, n_k) \tag{1}$$
$$\text{logit}(\pi_{ij}) = \beta_{ij} + \delta_j \tag{2}$$

Linzer (2013) - Evolution model

$$\text{for } j > 1\text{: } \beta_{ij} \sim N(\beta_{i,j-1}, \sigma^2_\beta) \tag{3}$$

$$\delta_j \sim N(\delta_{j-1}, \sigma^2_\delta) \tag{4}$$

Linzer (2013) - Anchoring on Election Day

$$\text{for } j = 1\text{: } \beta_{i1} \sim N(\text{logit}(h_i), s_i^2) \tag{5}$$

$$\delta_1 := 0 \tag{6}$$

Linzer (2013) - Interpretations

What is the election forecast?

Linzer (2013) - Interpretations

What is the election forecast?

$$\text{EC Votes Won} = (\text{EC Votes}) * I(\textit{logit}(\beta_{1:K,1}) > .5)$$

$$P(\text{Candidate Wins}) = P(\text{EC Votes Won} >= 270)$$

Linzer (2013) - Interpretations

States are correlated. What is the covariance of $\text{logit}(\pi_{i2})$ and $\text{logit}(\pi_{i'2})$ conditional on the variance terms σ^2_δ and σ^2_β?

Linzer (2013) - Interpretations

States are correlated. What is the covariance of $\text{logit}(\pi_{i2})$ and $\text{logit}(\pi_{i'2})$ conditional on the variance terms σ_δ^2 and σ_β^2?

$$\begin{aligned} \beta_{i2} &= logit(h_i) + e_i + \epsilon_{i2} & \epsilon_{it} \overset{iid}{\sim} N(0, \sigma_\beta^2);\, e_i \overset{iid}{\sim} N(0, s_i^2) \\ \delta_2 &= 0 + \nu_2 & \nu_t \overset{iid}{\sim} N(0, \sigma_\delta^2) \end{aligned}$$

Linzer (2013) - Interpretations

States are correlated. What is the correlation of logit(π_{i2}) and logit($\pi_{i'2}$) conditional on the variance terms σ_δ^2 and σ_β^2?

$$\beta_{i2} = logit(h_i) + e_i + \epsilon_{i2} \qquad \epsilon_{it} \overset{iid}{\sim} N(0, \sigma_\beta^2); e_i \overset{iid}{\sim} N(0, s_i^2)$$

$$\delta_2 = 0 + \nu_2 \qquad \nu_t \overset{iid}{\sim} N(0, \sigma_\delta^2)$$

$$\begin{aligned} Cor(\beta_{i2} + \delta_2, \beta_{i'2} + \delta_2) &= \frac{Cov(\beta_{i2} + \delta_2, \beta_{i'2} + \delta_2)}{sd(\beta_{i2} + \delta_2) sd(\beta_{i2} + \delta_2)} \\ &= \frac{Var(\delta_2)}{\prod_{k=i,i'} sd(\beta_{k2} + \delta_2)} \\ &= \frac{\sigma_\delta^2}{\prod_{k=i,i'} \sqrt{s_k^2 + \sigma_\beta^2 + \sigma_\delta^2}} \end{aligned}$$

Linzer (2013) - Results

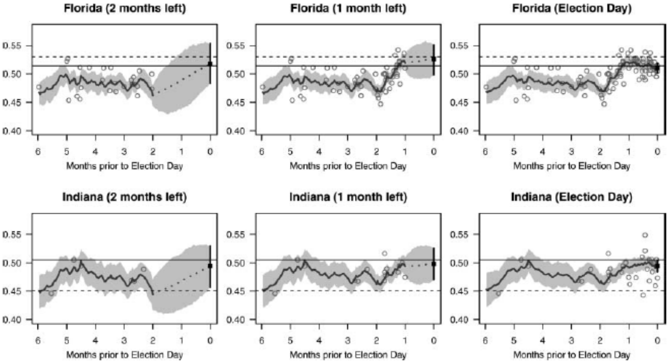

Figure 3. Forecasting the 2008 presidential election in real time. Results are shown for Florida and Indiana. The vertical axis is the percentage supporting Obama; points denote observed poll results. Horizontal lines indicate the late Time-for-Change forecast (dashed) and the actual election outcome (solid). The jagged gray line is the state-level daily estimate of voter preference for Obama, $\hat{\pi}_{ij}$. After the final day of polling, these trends project ahead to the Election Day forecast, $\hat{\pi}_{iJ}$, plotted as ■. Shaded areas, and the vertical bar on Election Day, denote 90% posterior credible intervals.

Linzer (2013) Example Setup

Suppose you want to model L swing states for J days before the election and you have K polls. Prepare the following in a list:

1. y - a K-vector of poll results (number of intended voters for the Democrat, i.e.).
2. n - a K-vector of poll sample sizes.
3. t - a K-vector of the date for each poll.
4. st - a K-vector of indicators for the state of each poll.
5. h - a L-vector of prior guesses for the final result in each state
6. s - a L-vector of prior standard devations for the final result in each state. Recall, for the normal, almost all the mass is on mean $\pm$ 2 standard deviations.

Linzer (2013) Example Code

```
model <- function(){
  for(k in 1:K){
    y[k] ~ dbin(p[k],n[k])
    p[k] = logit(beta[st[k],t[k]] + delta[t[k]])
  }
  for(j in 2:J){
    delta[j] ~ dnorm(delta[j-1],pow(sigma2_delta,-1))
    for(i in 1:L){
      beta[i,j] ~ dnorm(beta[i,j-1],pow(sigma2_beta,-1))
    }
  }
  delta[1] = 0
  for(i in 1:L){
    beta[i,1] ~ dnorm(logit(h[i]),pow(s[i],-2))
  }
}
```

Linzer (2013) - Improvements

The Linzer model did fairly well in 2008, but polls were also fairly accurate that year.

Potential improvements:

1. Adding polling errors, learned from past elections.
2. Allowing for excess polling variance above the implied binomial variance of $\pi_{ij}(1 - \pi_{ij})/n_k$.
3. Removing binomial model entirely and modeling y_k/n_k or $\text{logit}(y_k/n_k)$.
4. Allowing for unique state-level correlations, e.g. not assuming $Cov(NC, NY) = Cov(CA, NY)$.
5. Explicit modeling of h_i and s_i, i.e. using a "fundamentals" prediction to set the prior.