Economist Model

Graham Tierney

9/29/2020

Modeling Strategy

Ultimate goal is predicting the electoral college result. Strategy: (1) build a model-based prediction for the state-level election outcomes *without* using state-level polls, then (2) use that prediction as a prior for election-day preferences and use state-level polls, similar to Linzer (2013), as noisy measurements of a preferences that evolve over time.

National Popular Vote

First, model the national popular vote with both fundamentals (economic indicators) and national polls.

- "Fundamentals" models have historically been fairly accurate (comparable accuracy to late-campaign polls).
- National polls are conducted *much* more frequently.
- Allows the prior to change over time based on changes in national polls and economic data.

National Popular Vote

Modeling strategy for national popular vote:

1. Select fundamentals variables via leave-one-out cross validation accuracy.[1]
2. Use current national polling averages starting in late June (time also selected by cross validation).
3. For uncertainty quantification, assume the popular vote follows a Beta distribution and both mean and concentration parameters.

[1] Technically, they use elastic net regularization, which combines the lasso and ridge regression penalties.

National to State-Level Outcomes

Next, predict how far above or below the national outcome each state will be *without* using state polls. Features used are past deltas between state and national vote shares, whether a candidate is from the state, and changes in the national electorate.

Using state-level polls

At the core, for poll k of state i at time t, the number of Democrat supporters is

$$Y_{itk} \sim Binom(N_{itk}, p_{itk})$$
$$\text{logit}(p_{itk}) = \mu_{it} + \vec{X}_{itk}^T \vec{\beta} + \sigma_i \epsilon_k \qquad \epsilon_k \sim N(0, 1)$$

Contained in $\vec{X}$ are state, pollster, poll mode (telephone, online), poll population (registered vs likely voters), and whether the poll adjusts for partisan non-response (weighting sample partisanship to match expected partisanship in the election).[2]

[2] Partisan non-response bias and state-level bias are computed slightly differently than the others.

Correlations

A single covariance matrix is passed in as data. State level correlations are inferred from: 2016 presidential vote, racial and educational composition, median age, experienced population density, and the share of white evangelical Christians among voters, all standardized to range from 0 to 1. The covariance matrix is computed from a dataset where each row is one of the above variables and each column a state.[3] Call this matrix **S**.

[3] The computation isn't quite as simple as 'cov(mat)'.

Evolutions

$$\vec{\mu}_{\cdot t} = \vec{\mu}_{\cdot t} + \vec{\epsilon} \qquad \vec{\epsilon} \sim MNV(0, c_{\mu}\mathbf{S})$$

$$\vec{\mu}_{\cdot 1} = \vec{m}_{prior} + \vec{\epsilon} \qquad \vec{\epsilon} \sim MNV(0, c_{prior}\mathbf{S})$$

State-level polling errors are also assumed to have covariance matrix $c_{bias}\mathbf{S}$.

Simulations

All of the above models are implemented in Stan, and posterior distributions for μ_{i1} are used to simulate electoral college outcomes.