Polling Errors

Graham Tierney

11/6/2020

Assuming unbiased polls

Most of you did something like this. For each poll k of state s at time t, Y_{stk} gives the two-party support share of a candidate.

$$E[Y_{stk}] = \theta_{s,t}$$
$$\theta_{s,t} \sim N(\theta_{s,t-1}, \sigma_\theta)$$

Winner determined by $\theta_{s,1} > 50\%$.

Checking bias

You could do this:

$$
\begin{aligned}
E[Y_{stk}] &= \theta_{s,t} + \alpha_s \\
\theta_{s,t} &\sim N(\theta_{s,t-1}, \sigma_\theta) \\
\alpha_s &\overset{iid}{\sim} N(\mu_\alpha, \sigma_\alpha)
\end{aligned}
$$

Winner determined by $\theta_{s,1} > 50\%$.

How to set hyper parameters?

Could set $\mu_\alpha = \text{mean}(b_1, \ldots, b_S)$ and $\sigma_\alpha = \text{sd}(b_1, \ldots, b_S)$ where b_s is a point estimate of polling error in 2016 in state s (lots of analysts have made these). Could look at longer historical data by state.

Because the polling data cannot separate θ and α, sensitivity checks, particularly in μ_α, are very important. Could make a plot of P(Biden Win) vs μ_α.

How to estimate?

Note lack of information in data is a blessing and a curse. Curse is obvious, you can't learn them easily. But the blessing is you don't necessarily need to re-estimate your model.[1]

If you have your simulated draws of $\theta_{s,1}$, don't need to re-estimate model. For draw of $\theta_{s,1}$, subtract rnorm(1,μ_α,σ_α) and recompute who wins.

[1] The priors do technically give you some identifiability, and a fully Bayesian approac would be to re-estimate the model many times for different priors on α_s.