Spatiotemporal Election Bias Model

Nate Wiecha, Hongjian Yang, Shiyue Yao

Outline:

- Poll weighting
- Test for constant bias
- Test for different state/year bias

Poll weighting

- Must weight polls closer to date of election higher
- Several different weighting methods to test sensitivity of later analysis

General weighting scheme

For a particular state in a particular year, with *p* a parameter, the weight given to a poll *d* days before the election is:

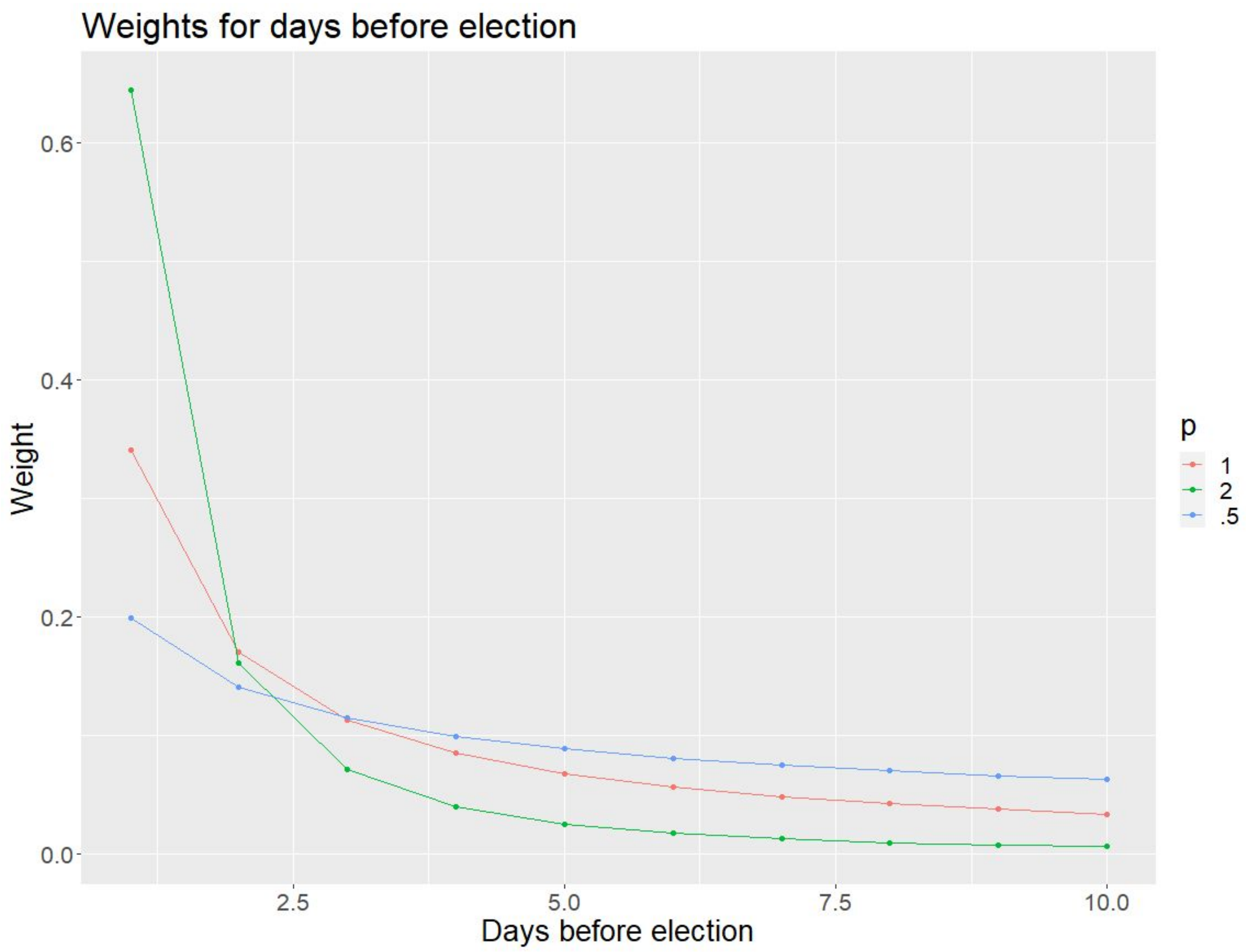
$$w(d;p) = \frac{1}{d^p}\frac{1}{c}$$

Where *c* is a normalizing constant so the weights sum to 1 for a particular state in a particular year.

Weighting specifics

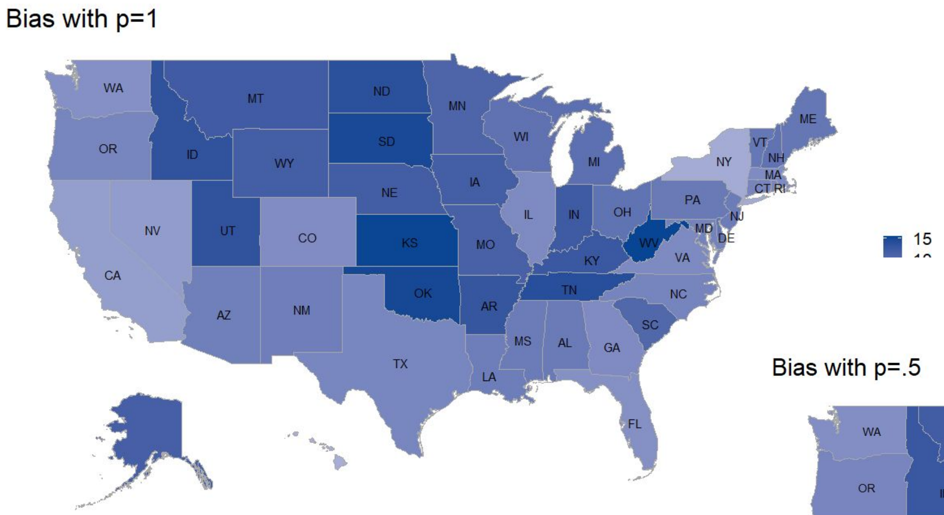
- We used $p = 1, \frac{1}{2}, 2$.
- $p = 1$ was our standard weight that we used as the default for our analysis.
- $p = 2$ weights polls closer to the election date more heavily compared to $p = 1$.
- $p = \frac{1}{2}$ weights polls closer to the election date less heavily compared to $p = 1$.

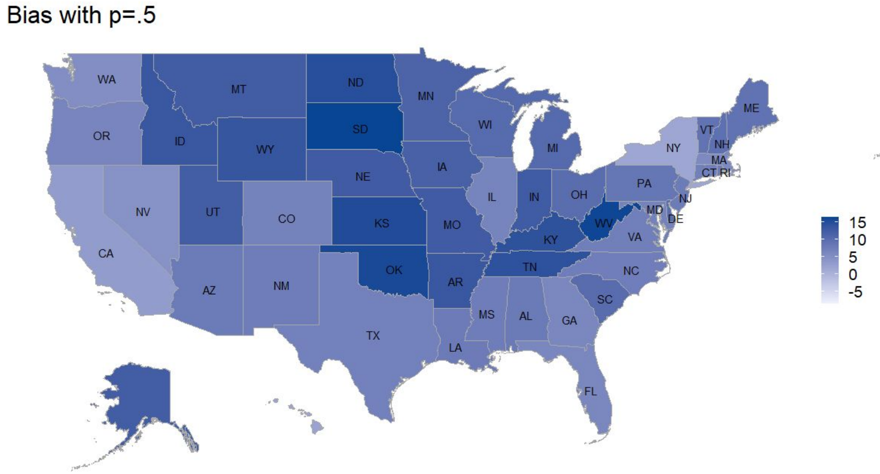
Visualization of the weights

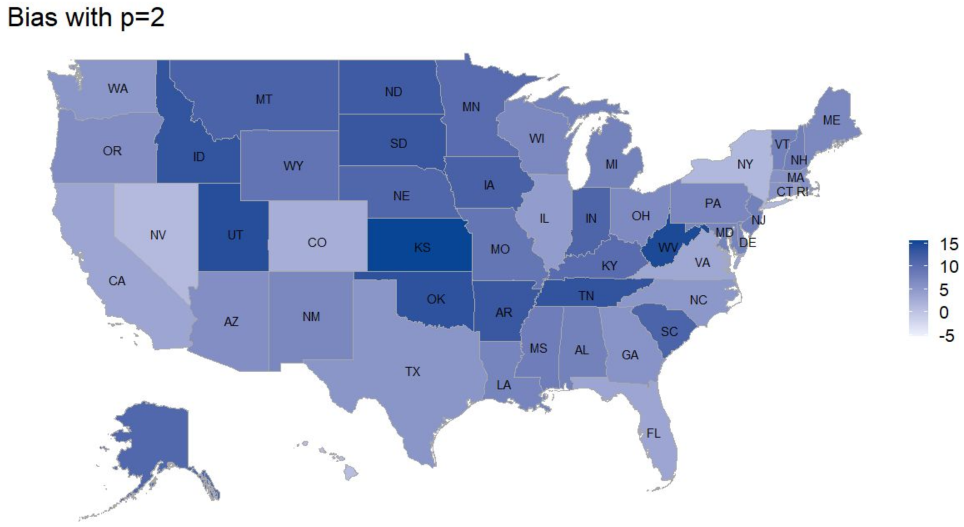


- Hypothetical example where a state has 10 polls, spaced a day apart leading up to the election.

Visualization of the weights using 2016







Test for systematic polling bias

- Spatial temporal CAR model
- ST.CARar() in *CARBayesST* package
- Usage is similar to spatial CAR model, with a few exceptions
  - **formula:** response ~ covariates. Response and each covariate should be vectors of length (KN) * 1 where k is the number of spatial units and N is the number of time periods. All vectors are ordered so that 2012 data comes first, and then 2016 and 2020
  - **family, W, burnin, and n.sample** arguments usages are the same as spatial CAR model

Formula:

$$\begin{aligned} \psi_{kt} &= \phi_{kt}, \\ \boldsymbol{\phi}_t | \boldsymbol{\phi}_{t-1} &\sim \mathrm{N}\left(\rho_T \boldsymbol{\phi}_{t-1}, \tau^2 \mathbf{Q}(\mathbf{W}, \rho_S)^{-1}\right) \qquad t = 2, \ldots, N, \\ \boldsymbol{\phi}_1 &\sim \mathrm{N}\left(\mathbf{0}, \tau^2 \mathbf{Q}(\mathbf{W}, \rho_S)^{-1}\right), \\ \tau^2 &\sim \text{Inverse-Gamma}(a, b), \\ \rho_S, \rho_T &\sim \text{Uniform}(0, 1). \end{aligned}$$

where $\mathbf{Q}(\mathbf{W}, \rho_S) = \rho_S[\mathrm{diag}(\mathbf{W}\mathbf{1}) - \mathbf{W}] + (1 - \rho_S)\mathbf{I}$

(Leroux parameterization)

Result and comparison

(Regress bias ~ 1)

	p=1(standard)	p=1/2	p=2
Intercept	5.23	5.47	4.76
95% CI	[5.16,5.31]	[5.42,5.51]	[4.54, 4.97]
tau2	28.4	29.47	28.55
nu2	0.016	0.016	0.03
DIC	-287	-217	-222

Conclusion

- There is systematic polling bias. With standard weighting method, averaged poll underestimates actual GOP support by 5.23 percent.
- When $p=\frac{1}{2}$, averaged poll has a higher bias; When $p=2$, averaged poll has a smaller bias.
    - If we give recent polls a higher weight, bias is smaller, which indicates that some voters switch to GOP in the last minute.

Spatial Covariates explanation:

State: contains 48 state covariates, eg: North Carolina

Color:

- Red states: gop wins in 2012, 2016, 2020; assigned Color = 1
- Blue states: gop loses in 2012, 2016, 2020; assigned Color = -1
- Swing states: others; assigned Color = 0

Agriculture:

- Farm related income of each state; scaled
- Data source: USDA agriculture census in 2017

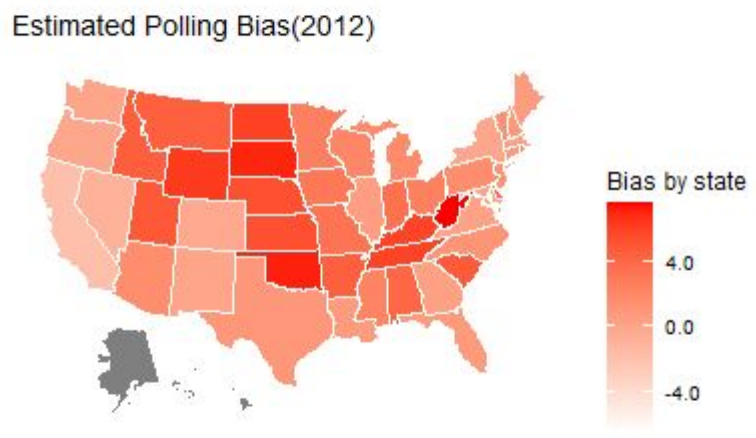
Model output and comparison

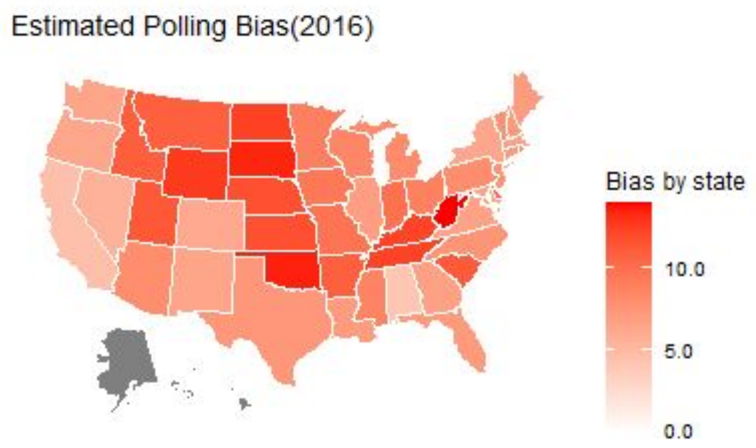
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(Intercept)	5.23	2.357	4.051	7.477	5.210	2.198	2.160	5.201
year2016	-	6.303	6.362	-	-	5.585	5.612	-
year2020	-	2.186	2.168	-	-	2.173	2.237	-
State	-	-	Yes	Yes	-	-	-	-
Color	-	-	-	-	0.689	0.479	-	-
Agriculture	-	-	-	-	-	-	-0.683	-0.805
tau2	28.212	10.345	0.009	11.776	27.335	11.421	11.070	26.999
nu2	0.015	0.074	5.220	1.906	0.018	0.016	0.015	0.018
rho_s	0.876	0.229	0.372	0.965	0.886	0.101	0.093	0.888
rho_t	0.458	0.488	0.376	0.240	0.446	0.468	0.477	0.452
DIC	-288.626	-109.495	664.525	77.383	-275.268	-229.924	-219.113	-287.607

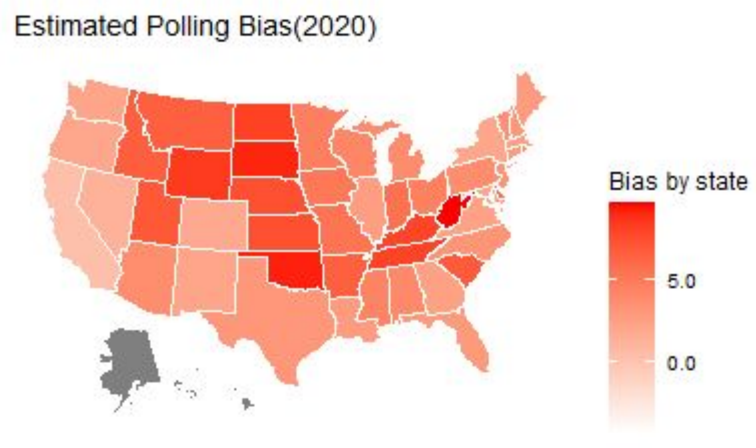
Conclusion

- Generally we see there are positive polling bias, which indicates that the polls understate GOP support rate.
- Looking at year coefficients, we can see that the bias in 2020 is less than 2016, which means that the polls did better in this election.
- When we divide the states into red, blue and swing state, we find the GOP support rates are understated in red states, and overstated in the blue states.
- Specifically, the bias goes up by 0.689 if it’s a red state, and falls by 0.689 if it’s a blue state.
- The polls generate less bias for GOP support in farm states. If the farm-related income goes up by one unit, the polling bias falls by 0.805.

Estimated Polling Bias







We can conclude that the polling bias varies by state and election.

- Spatial pattern: shown in the graphs
- Temporal: 2016 election have the largest bias

Sensitivity

	(5)			(8)		
	p=1	p=1/2	p=2	p=1	p=1/2	p=2
(Intercept)	5.210	4.797	4.794	5.201	5.202	4.783
State	-	-	-	-	-	-
Color	0.689	0.749	0.703	-	-	-
Agriculture	-	-	-	-0.805	-0.841	-0.780
tau2	27.335	27.257	27.851	26.999	26.972	26.511
nu2	0.018	0.046	0.034	0.018	0.021	0.073
rho_s	0.886	0.810	0.792	0.888	0.888	0.826
rho_t	0.446	0.429	0.417	0.452	0.454	0.451
DIC	-275.268	-126.355	-183.896	-287.607	-256.678	-75.004

Conclusion:

When examining spatial bias patterns, the result is not significantly different among different weight methods.

Thanks for listening!