NC STATE UNIVERSITY

ST 533 Final: Modeling Spatiotemporal Trends in Political Poll Bias

Ziyang Jiang, Mingkang Zheng, Tyler Schappe

Introduction

Objectives

- 1. Define the weights used to calculate polling averages
- 2. Test whether there is systematic polling bias under the assumption that the bias is constant over state and election
- 3. Test whether the bias varies by state and/or election and display the estimated bias

Outline

- Methods
- Results
- Conclusions

Weights

- The polling average: $X_{it} = \sum_{j=1}^{N_t} w_{itj} P_{jt}$ $\Longleftarrow$ $w_{itj} = 0.5 \times \boxed{s_{itj}} + 0.5 \times \boxed{d_{itj}}$
  - Based on the poll's sample size: polls that **sample more voters** receive a **larger weight**

$$s_{itj} = \frac{Poll\ j's\ sample\ size\ in\ state\ i\ in\ year\ t}{Sum\ of\ sample\ size\ for\ all\ polls\ in\ state\ i\ in\ year\ t}$$

  - Based on how recently it was conducted: **more emphasis** is placed on **recency**

$$d_{itj} = \frac{F(Poll\ j's\ days\ to\ election\ in\ state\ i\ in\ year\ t)}{Sum\ of\ F(Days\ to\ election)\ for\ all\ polls\ in\ state\ i\ in\ year\ t}$$

*$F(x) = \frac{1}{x}$

(older polls are penalized)

Methods

- Package: CARBayesST
  - Model for capturing the **spatial-temporal autocorrelation** in data via **random effects**
  - Generalized linear mixed model

$$
\begin{aligned}
Y_{kt}|\mu_{kt} &\sim f(y_{kt}|\mu_{kt},\nu^2) \quad \text{for } k=1,\ldots,K, \quad t=1,\ldots,N, \\
g(\mu_{kt}) &= \mathbf{x}_{kt}^\top \boldsymbol{\beta} + O_{kt} + \boxed{\psi_{kt}}, \\
\boldsymbol{\beta} &\sim \mathrm{N}(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta).
\end{aligned} \tag{1}
$$

  - ST.CARar() : one of the models for $\boldsymbol{\psi}$ (Spatio-temporal random effects)

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

Methods

- ST.CARar() : the spatio-temporal random effects follows a **multivariate AR(1) process**
  - Important parameters

    $\boldsymbol{\beta}$: coefficients of covariates ν^2: nugget variance τ^2: spatio-temporal variance parameter

    ρ_S, ρ_T: spatial or temporal dependence parameters
  - Manually change default priors to fit our data: $\nu^2 \sim InvGamma(1, 0.1)$, $\tau^2 \sim InvGamma(0.5, 3)$
  - State adjacency matrix **W:** Border adjacency, $w_{ij} = \begin{cases} 1, if\ state\ i\ and\ j\ share\ a\ common\ border \\ 0, \quad otherwise \end{cases}$

    *Delete states Alaska & Hawaii (no neighbors)
  - State i = 1, 2, … 49; Year t = 2012, 2016, 2020

*In this model, **missing values (NA) are allowed in the response data**, and they can be estimated during fitting model

Covariates

Covariate	Description
Turnout	VEP(voting-eligible population) turnout rate **for all state i and year t**
Income	Household income **for all state i and year t**
Pop.dens	Population density **for all state i and year t**
Age	% 65 years or older (of total population) **for all state i and year t**
Year	**indicator variables** for 2012, 2016, 2020 election years
State	**indicator variables** for 49 states

*Some covariates didn't use data from the election years.

Ex: for *Age* variable, we used the data from 2019 as the data for t = 2020

Models

- **Similar** model setting, **different** covariates

Objective	Model	Covariate	Feature
2	Null model	Turnout, Income, Pop. dens, Age	No fixed effects
3	Full model	Year, State, Turnout, Income, Pop.dens, Age	Fixed election year & state effect
	By Election Year model (no state)	Year, Turnout, Income, Pop. dens, Age	Only fixed election year effect
	By State model (no election year)	State, Turnout, Income, Pop.dens, Age	Only fixed state effect

- Specifically, the **mean term** for each model would be:

$$\begin{cases} \beta_0 + \beta_1 Turnout + \beta_2 Income + \beta_3 Pop.dens + \beta_4 Age \\ \beta_0 + \beta_1 Turnout + \beta_2 Income + \beta_3 Pop.dens + \beta_4 Age + \beta_5 2016 + \beta_6 2020 + \beta_{k+6} State_k \\ \beta_0 + \beta_1 Turnout + \beta_2 Income + \beta_3 Pop.dens + \beta_4 Age + \beta_5 2016 + \beta_6 2020 \\ \beta_0 + \beta_1 Turnout + \beta_2 Income + \beta_3 Pop.dens + \beta_4 Age + \beta_{k+6} State_k \end{cases}$$

Models

Objective	Model	Covariate	Feature
2	Null model	Turnout, Income, Pop. dens, Age	No fixed effects
3	Full model	Year, State, Turnout, Income, Pop.dens, Age	Fixed election year & state effect
	By Election Year model (no state)	Year, Turnout, Income, Pop. dens, Age	Only fixed election year effect
	By State model (no election year)	State, Turnout, Income, Pop.dens, Age	Only fixed state effect

- Objective 2: based on Null model ⇨ **Question: test if β_0 is significantly different from 0**
- Objective 3: compare three models
  - Use DIC/WAIC metrics
  - Analogous to overall F-test

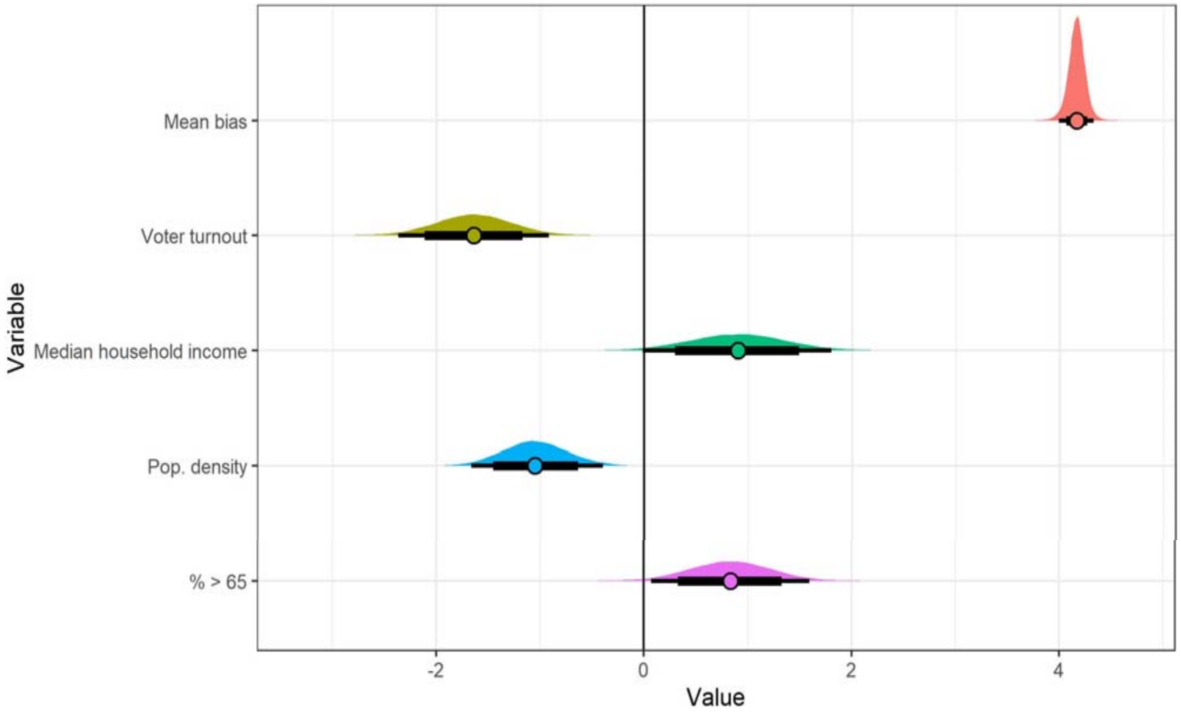
⇨ **Questions:**

(i) Are all coefficients of state predictors equal to 0?

(ii) Are all coefficients of election year predictors equal to 0?

Results for objective 2

Is there a systematic polling bias if assuming bias is constant over state and election?

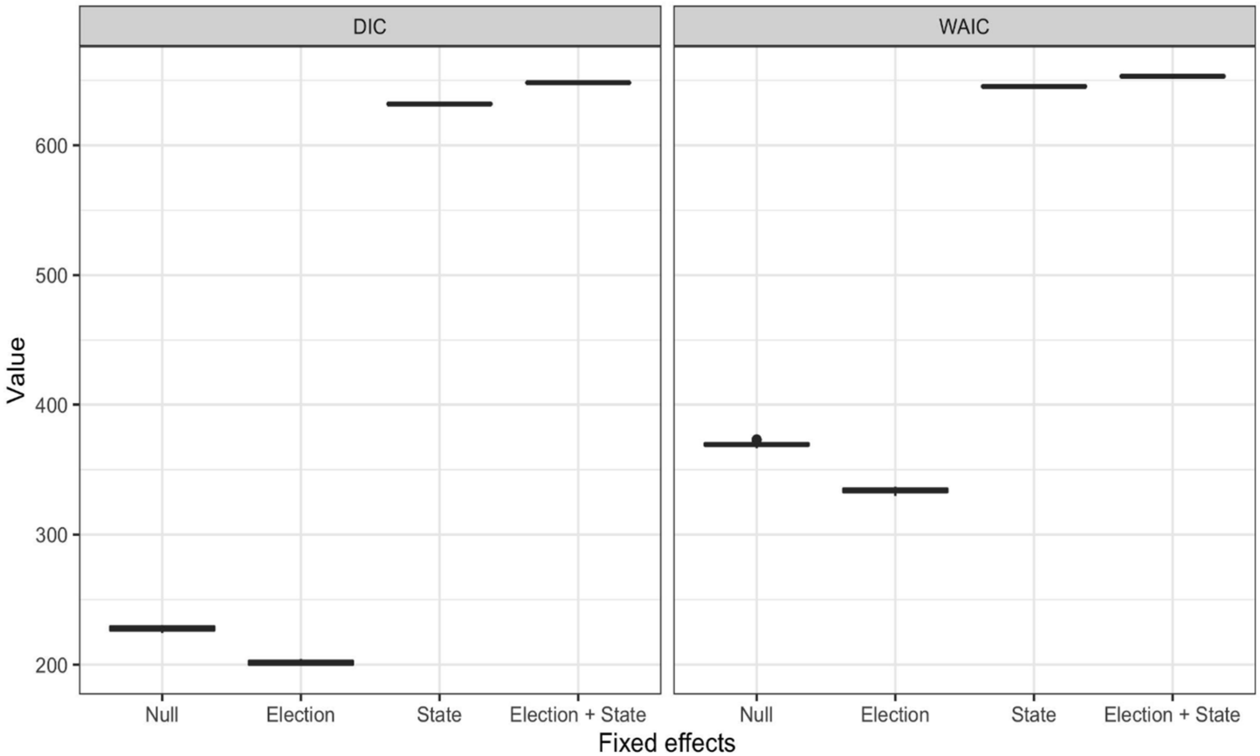


- $B_{it} = \beta_0, \text{for } \forall\, i,t$
- Definitely there is a systematic polling bias
- All the four covariates are significant

NC STATE UNIVERSITY

Results for objective 3

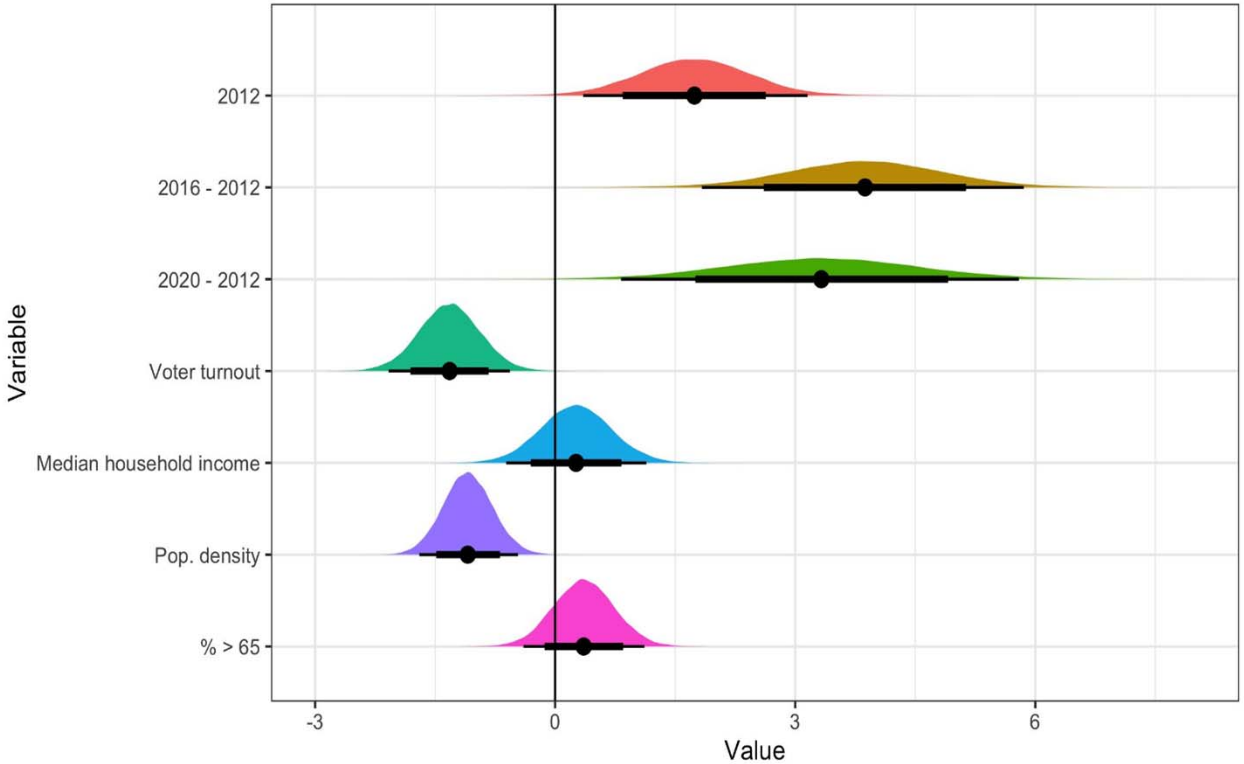
Does the bias vary by state and/or election?



- Results of DIC and WAIC are consistent
- The smallest DIC/WAIC indicates that model only adds 3 elections as extra fixed effects has the best overall performance
- Adding states as fixed effects increases the value of DIC/WAIC
- No evidence that state-level coefficients are non-zero, which doesn't mean the bias doesn't vary by state

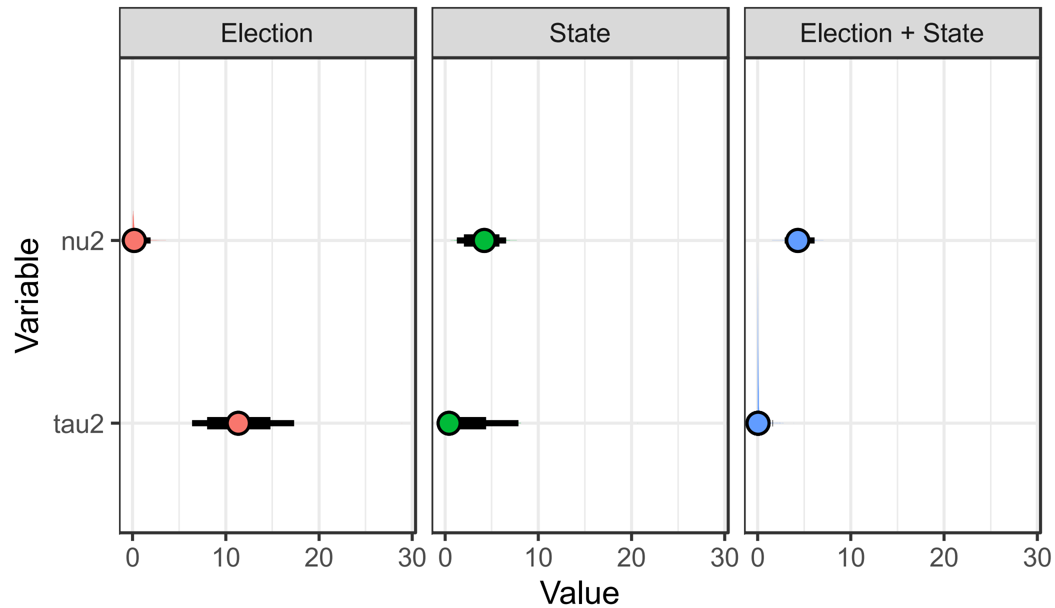
Results for objective 3

Does the bias vary by state and/or election?



- The bias in 2016 & 2020 are calculated based on the bias in 2012
- There is strong evidence that the bias varies by election; the bias that underestimating the GOP support increases these years
- This time median income and age are no longer significant covariates
- Most interesting thing: it seems like the more people vote, the less bias in the election results

Part 3: Closer Look



ψ_{kt}	Election	Election + State
Min	-8.23	-0.2
Max	5.15	0.25

Election Model Covariates (fixed)

- Intercept
- 2016
- 2020
- Turnout
- HH income
- % > 65 yrs

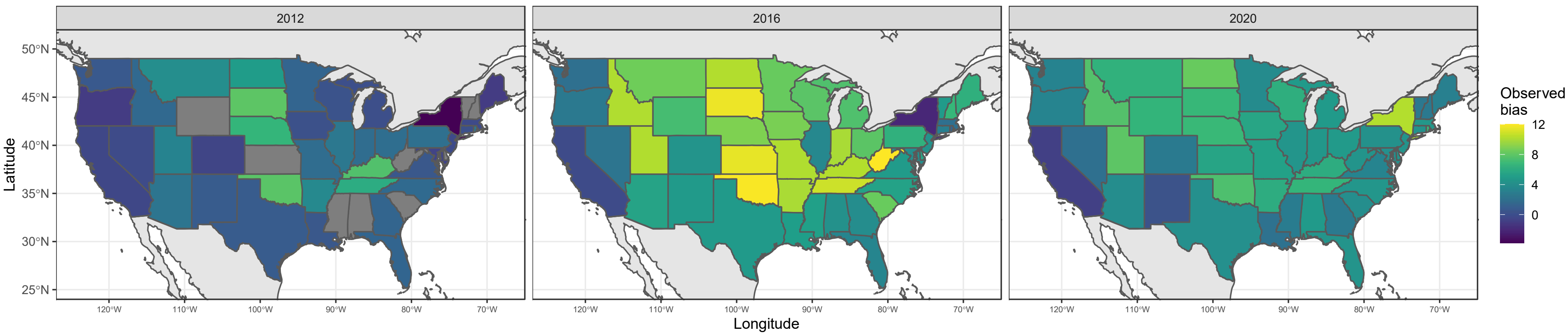
$$g(\mu_{kt}) = x_{kt}^T \beta + \psi_{kt}$$

State/Election + State Model Covariates (fixed)

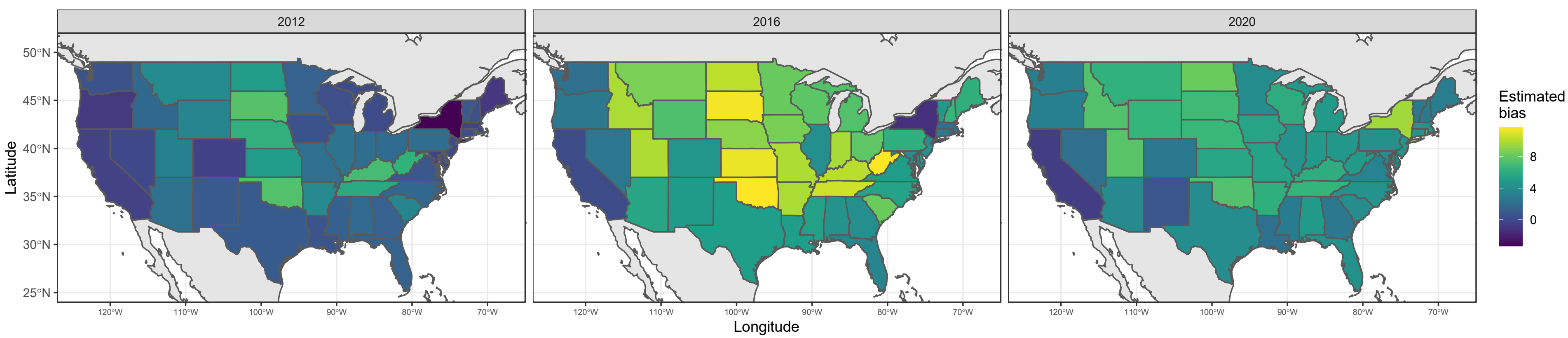
- Intercept
- 2016
- 2020
- Turnout
- HH income
- % > 65 yrs
- Arizona
- . . .
- Wyoming

NC STATE UNIVERSITY

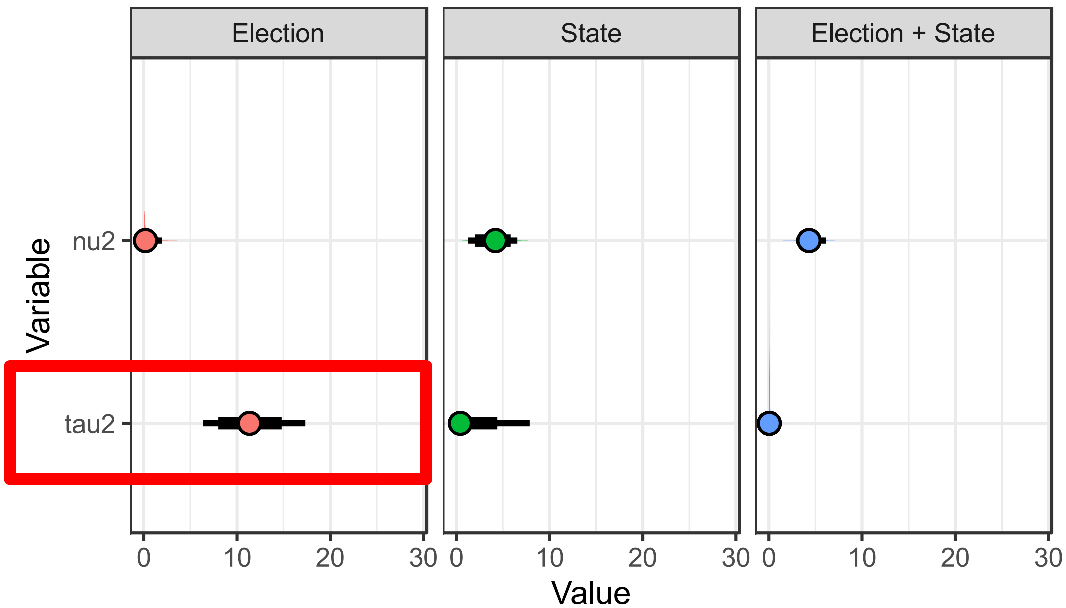
Observed Election Bias



Estimated Election Bias



Part 3: Closer Look



Election Model Covariates (fixed)

- Intercept
- 2016
- 2020
- Turnout
- HH income
- % > 65 yrs

$$g(\mu_{kt}) = x_{kt}^T \beta + \psi_{kt}$$

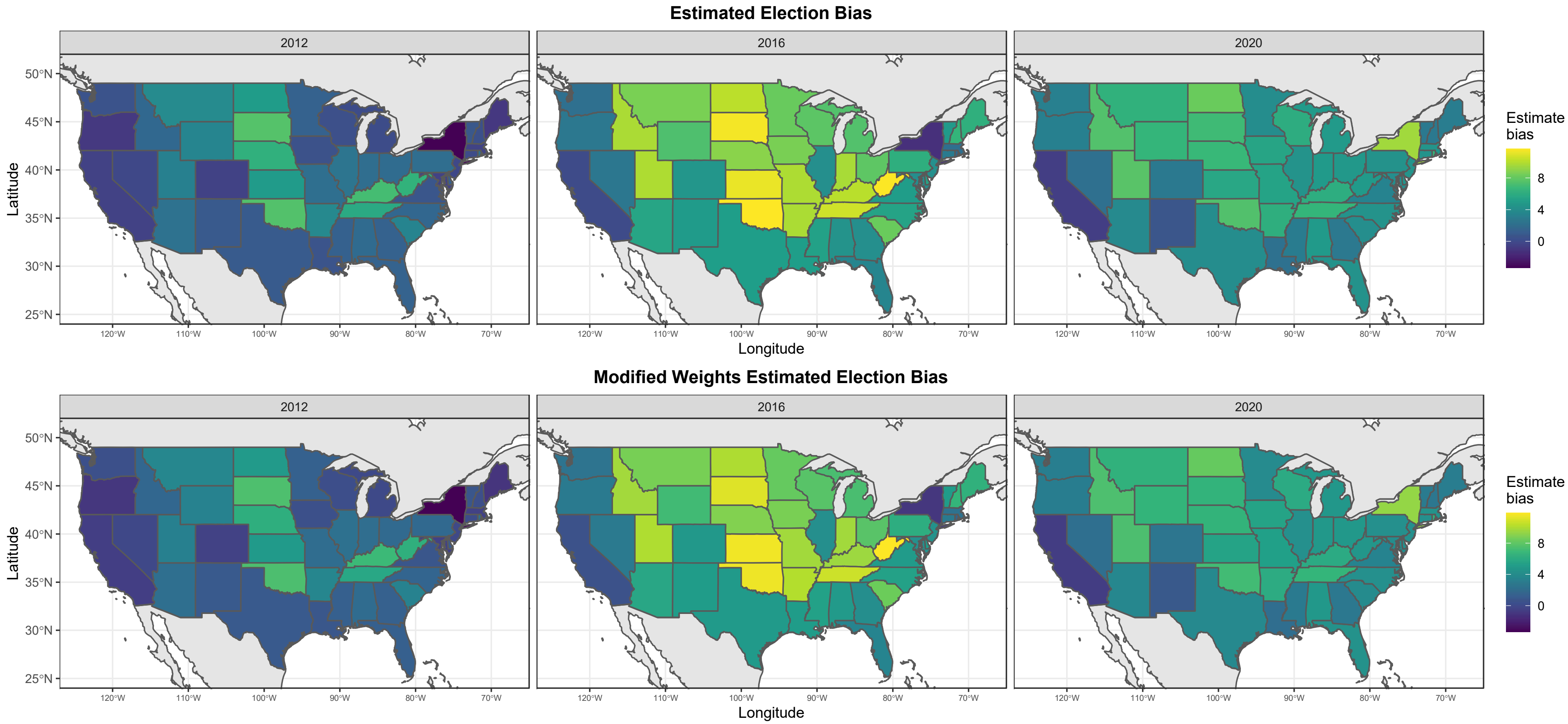
State/Election + State Model Covariates (fixed)

- Intercept
- 2016
- 2020
- Turnout
- HH income
- % > 65 yrs
- Arizona
- . . .
- Wyoming

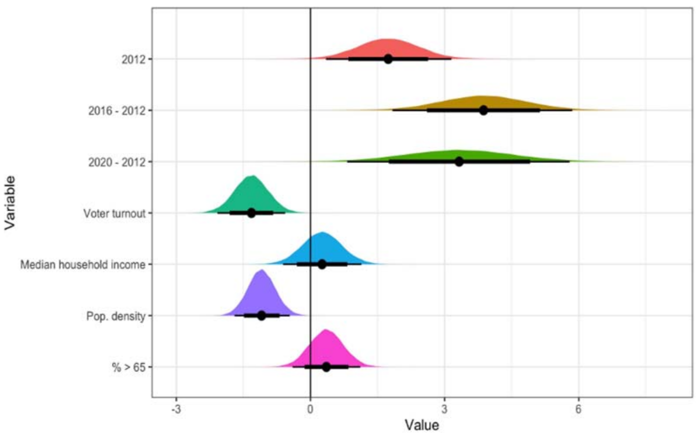
Conclusions

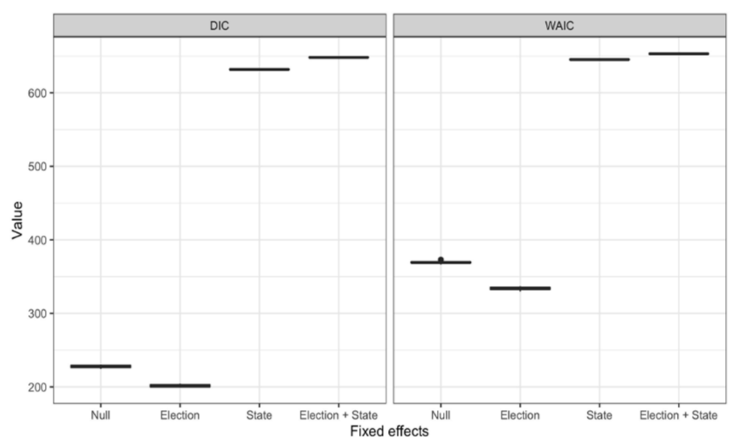
- **Part 2:** Assuming constant bias across all states and elections, consistent underestimation of GOP
- **Part 3:**
  - Underestimation of GOP, magnitude varied by year
  - By state: It's complicated!
    - No evidence of difference among states when considered individually (as fixed effects)
    - BUT we conclude that there ARE differences among states
    - Bias varies among states in a clustered way

Weights Sensitivity

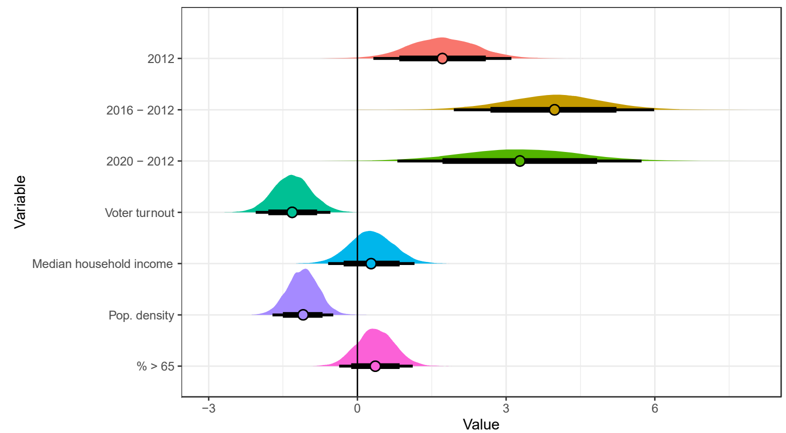


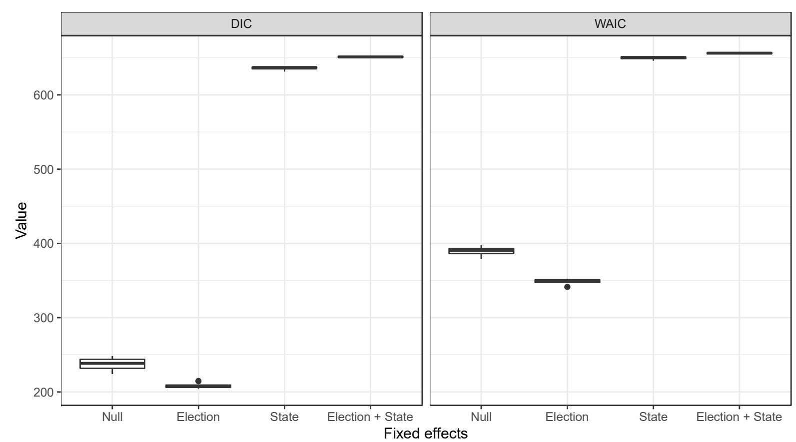
Original Weights (50/50)





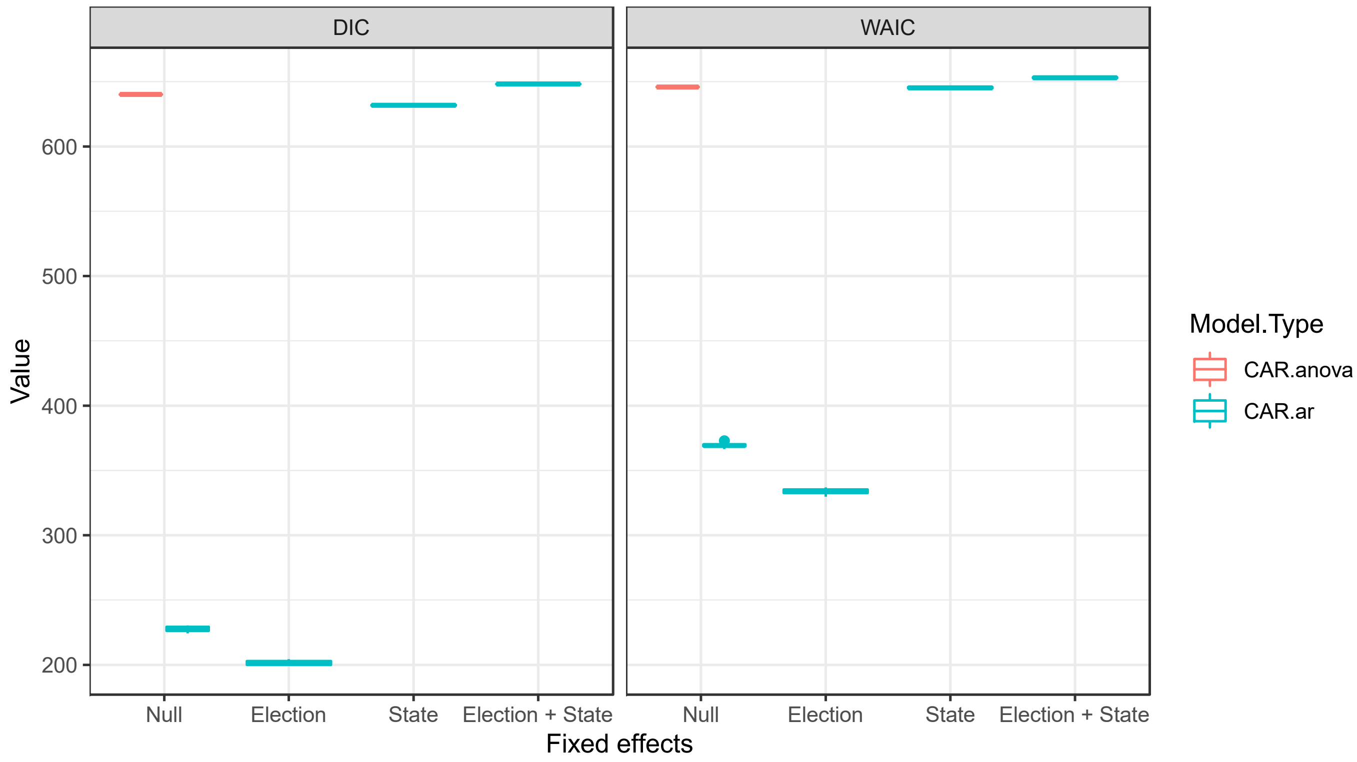
Modified Weights (70/30)





Thank you for listening!
Questions?

CARanova


DIC
WAIC
Value
600
500
400
300
200
Null
Election
State
Election + State
Fixed effects
Model.Type
CAR.anova
CAR.ar