Does Urbanisation Predict Election Outcomes?

A Bayesian's Perspective

Eldaleona Odole Leonor Cunha Anarghya Murthy

Faculty of Statistics

January 26, 2025

technische universität dortmund

Table of Contents

Table of Contents

Introduction

- **Research Question:** How does urbanization of a particular district affect result of an election in the US?
- **Variable of Interest:** Winning party in the House of Representatives 2022 General Election (binary)

Table of Contents

Dataset

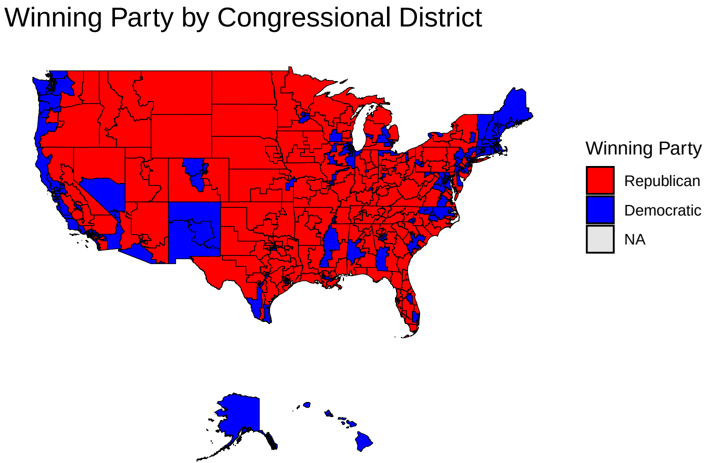
We wanted to consider different factors in the analysis, with our primary focus being the urbanization of each House district. These factors included:

1. Demographic Data (US Census Bureau)
2. Urbanization (FiveThirtyEight)
3. Regional Information (US Census Bureau)
4. Election Results (FiveThirtyEight)

We combined different sources in order to create our data set containing 435 instances of 16 unique covariates.

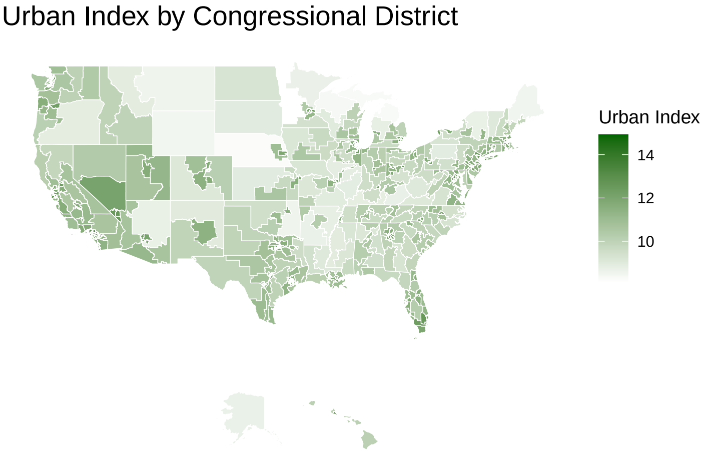
Winning Party

Our independent variable is Winning party in the 2022 Election.

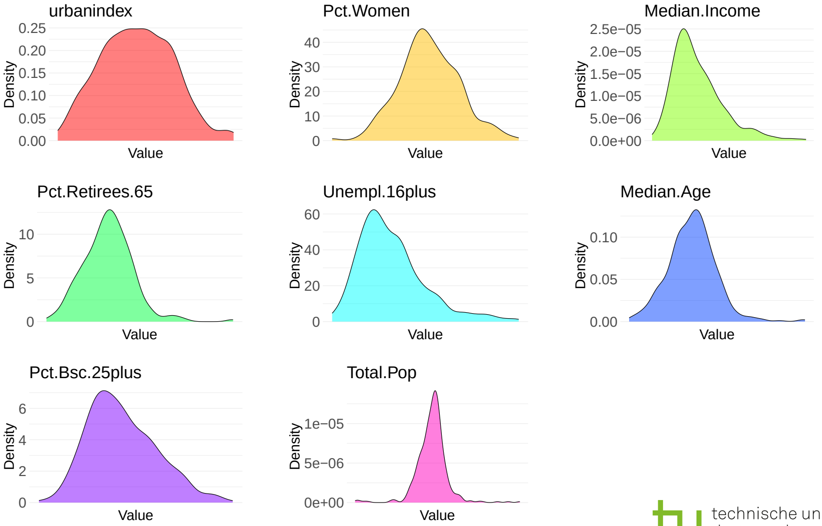


Urban Index

Our dependent variable of interest is the Urban Index from FiveThirtyEight.



Densities


urbanindex
0.25
0.20
0.15
0.10
0.05
0.00
Density
Value
Pct.Women
40
30
20
10
0
Density
Value
Median.Income
2.5e−05
2.0e−05
1.5e−05
1.0e−05
5.0e−06
0.0e+00
Density
Value
Pct.Retirees.65
10
5
0
Density
Value
Unempl.16plus
60
40
20
0
Density
Value
Median.Age
0.10
0.05
0.00
Density
Value
Pct.Bsc.25plus
6
4
2
0
Density
Value
Total.Pop
1e−05
5e−06
0e+00
Density
Value

Correlation Matrix

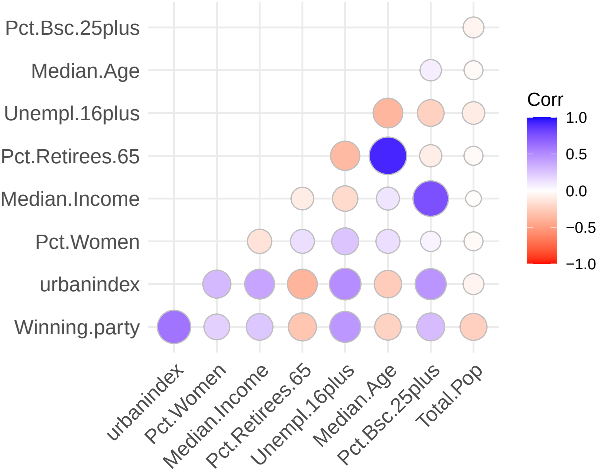


Table of Contents

Model Assumptions

There are many people trying to predict US election outcomes, from the wealth of data available about voters. However we wanted to look at the voters in relation to their geography. In order to do this we assumed

- District voting outcomes can be modeled via logistic regression
- Districts are exchangable within each state and each state is exchangable within its region
- We assume that the variation is explained by out predictors, and were there no additonal predictors $\theta = .5$ across all levels of geography

Table of Contents

Model 1

Let the response variable 'Winning Party' be y, the predictor of interest 'Urban index' be x, and the other covariates be a 15-dimensional vector z. Let i, j, and k be the indices for the district, region, and state respectively.

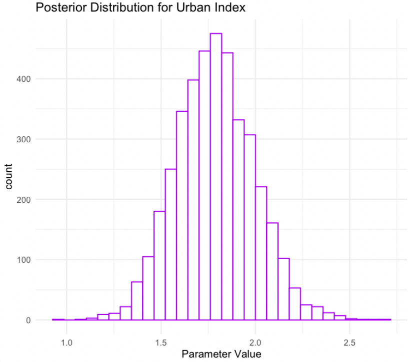
$$y_{i,j} \sim Ber.(logit^{-1}(\theta_j))$$

$$\theta_j := \beta_0, j + x_{i,j} * \beta_{1,j} + z_{i,j}^T * \gamma_{1,j}$$

$$\beta_{1,j} \sim Gam.(1, \tau)$$

$$\tau \sim Normal(0, 1)$$

Model 1: Results



Regression Coefficients:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	-18.00	2.52	-22.94	-13.01	1.00	2713	2434
urbanindex	1.79	0.21	1.40	2.20	1.00	3427	2864
Percentage.Women	0.08	1.01	-1.94	2.08	1.00	7100	2504
pct.bach	0.35	1.67	-2.87	4.51	1.00	3277	1433
Median.Household.Income	-1.37	0.76	-2.90	0.06	1.00	4236	3037
pct.retirees	-3.84	4.41	-15.57	0.97	1.00	2060	1858

Model 2

Let the response variable 'Winning Party' be y, predicted by the variables Urban Index and the Percentage of Retirees. Let i, j, and k represent the indices for district, state, and region, respectively.

$$y_{i,j,k} \sim \text{Ber.}\left(\text{logit}^{-1}(\theta_{j,k})\right)$$

$$\theta_{j,k} = \begin{array}{c} \beta_0 + \beta_1 \cdot \text{Urban\_Index} + \beta_{1,j} \cdot \text{Urban\_Index} \\ + \beta_{1,j:k} \cdot \text{Urban\_Index} + \beta_2 \cdot \text{Pct\_Retirees} \end{array}$$

$$\beta_0 \sim \text{Normal}(0, 0.5)$$

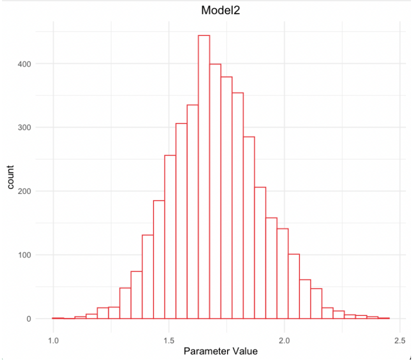
$$\beta_1 \sim \text{Normal}(0, 1)$$

$$\beta_{1,j} \sim \text{Normal}(0, \sigma_j)$$

$$\beta_{1,j:k} \sim \text{Normal}(0, \sigma_{j:k})$$

$$\sigma_j; \sigma_{j:k} \sim \text{Halfcauchy}(10)$$

Model 2: Results




```
Regression Coefficients:
                      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept               -17.98      2.37   -22.70   -13.36 1.00     3212     3406
urbanindex                1.71      0.19     1.34     2.10 1.00     4960     3321
Percentage.Retirees..65..  -4.55      4.18   -16.44    -0.29 1.00     2240     1966
```

Model 3

Let the response variable 'Winning Party' be y, predicted by the variables Urban Index and the Percentage of Retirees. Let i and j represent the indices for district and state respectively.

$$y_{i,j} \sim \text{Ber.}\left(\text{logit}^{-1}(\theta_j)\right)$$

$$\theta_j = \beta_0 + \beta_1 \cdot \text{Urban\_Index} + \beta_{1,j} \cdot \text{Urban\_Index} + \beta_2 \cdot \text{Pct\_Retirees}$$

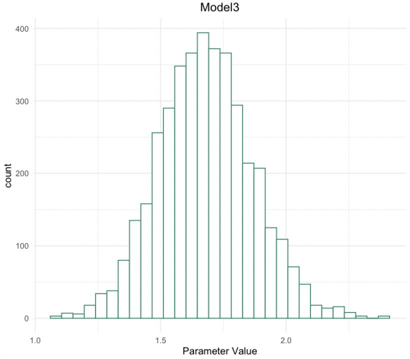
$$\beta_0 \sim \text{Normal}(0, 0.5)$$

$$\beta_1 \sim \text{Normal}(0, 1)$$

$$\beta_{1,j} \sim \text{Normal}(0, \sigma_j)$$

$$\sigma_j \sim \text{Halfcauchy}(10)$$

Model 3: Results



Regression Coefficients:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	-17.77	2.27	-22.20	-13.18	1.00	2303	2442
urbanindex	1.68	0.19	1.32	2.06	1.00	2509	2291
Percentage.Retirees..65..	-4.28	3.82	-15.08	-0.11	1.00	1768	1379

Model 4

Let the response variable 'Winning Party' be y, predicted by the variables Urban Index and the Percentage of Retirees. Let i and k represent the indices for district and region respectively.

$$y_{i,j} \sim \text{Ber.}\left(\text{logit}^{-1}(\theta_j)\right)$$

$$\theta_j = \beta_0 + \beta_1 \cdot \text{Urban\_Index} + \beta_{1,k} \cdot \text{Urban\_Index} + \beta_2 \cdot \text{Pct\_Retirees}$$

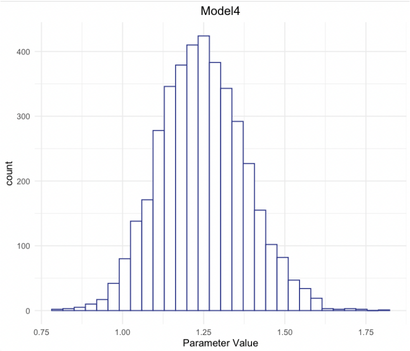
$$\beta_0 \sim \text{Normal}(0, 0.5)$$

$$\beta_1 \sim \text{Normal}(0, 1)$$

$$\beta_{1,k} \sim \text{Normal}(0, \sigma_k)$$

$$\sigma_k \sim \text{Halfcauchy}(10)$$

Model 4: Results



Regression Coefficients:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	-13.24	1.62	-16.53	-10.09	1.00	2028	2228
urbanindex	1.25	0.13	1.00	1.52	1.00	2548	2402
Pct.Retirees.65	-3.79	2.83	-10.98	-0.30	1.01	1460	1075

Table of Contents

Model Comparison: R^2

Model	Estimate	Estimate Error	Q 2.5	Q 97.5
1	0.568	0.0246	0.516	0.612
2	0.534	0.0255	0.479	0.579
3	0.529	0.0255	0.474	0.574
4	0.408	0.0234	0.359	0.449

Model Comparison: RMSE

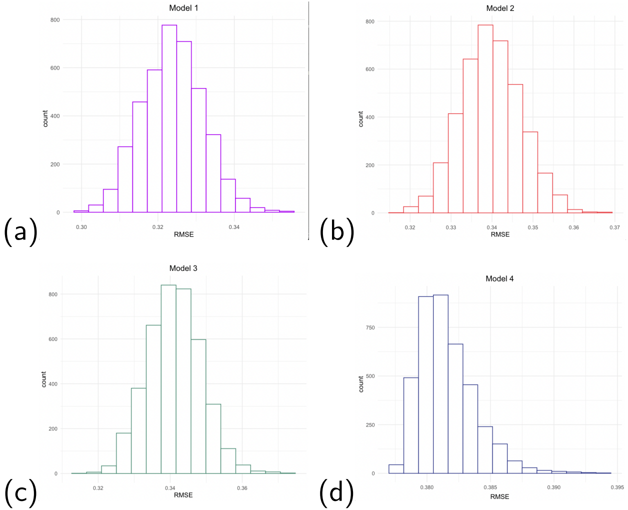


Figure: RMSE Comparison

Model Comparison: Data Log-likelihood

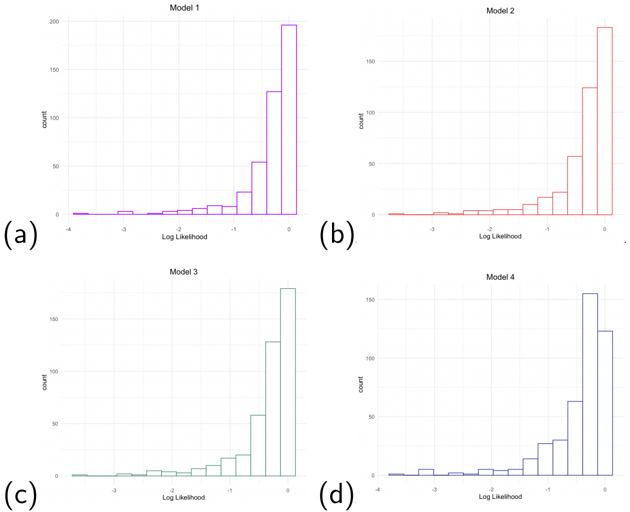


Figure: Log-Likelihood Comparison

Table of Contents

Best Model

Possible Improvemnts

- Scaling all covariates to same scale for better parameter estimation interpretability
- More prior experimentation
- Varying intercept models, as we currently only have varying slopes

Next Steps

Conclusions

In all of our models we found that urbanization has a small positive impact on the logit probility of voting democratically.

Contact:
{eldaleona.odole, leonor.cunha, anarghya.murthy}@tu-dortmund.de

Table of Contents