Does Urbanization Predict Elections Outcomes?

— Project Report —
Advanced Bayesian Data Analysis

Eldaleona Odole, Leonor Cunha, and Anarghya Murthy

March 12, 2025

TU Dortmund University

Contents

1 Introduction

In this report we aim to investigate the relationship between urbanization and partisanship voting outcomes in the United States using Bayesian multilevel modeling. Put more concretely we want to answer the question; "How does urbanization of a particular U.S. House district affect the party that is elected?". First we explain our motivation and important background information about the U.S. election system. We then discuss briefly our modeling approach before looking at other literature trying to understand the relationship between urbanization and voting outcomes. Next we discuss our dataset, which is a combination of four different datasets, followed by a brief explanation of our codebase. We then discuss the set up of our four models in detail, followed by a discussion of the priors and our reasoning behind them. We then go into the results of our models and their convergence diagnostics. Next we use Bayesian predictive metrics to compare our models. Finally we perform a prior sensitivity analysis before concluding with limitations, improvements and reflections.

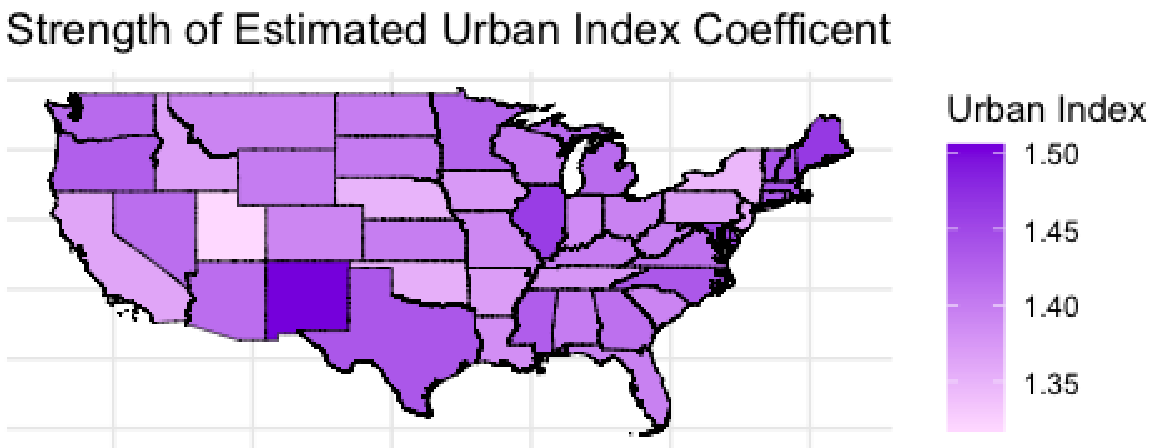


Figure 1: A map of the continental U.S. with the estimated coefficients from Model 4 of *Urban Index* by state.

Background

The association between a voters demographics (gender, age, education etc.) and their propensity to vote for either a democrat or republican candidate is a topic of extensive study. The large political polling organization such as Gallup and Pew Research tend to analyze their polls by breaking respondents into smaller demographic groups (Center, 2020). However, less is known about the relationship between voting outcomes and the voter's environment. Our modeling would like to investigate the conventional wisdom that says cities tend to be more progressive. It is also important to note that the Democratic Party is considered the progressive party and the Republican Party the conservative party, as these terms are at times used interchangeably throughout this report.

To answer our question about the nature of the relationship between urbanization and partisan voting outcomes, we choose to investigate the 2022 U.S. House Election. Every two years the United States elects 435 officials to the House of Representatives. Each state is allocated one of the the 435 House seats with the rest being allocated roughly proportional to the share of total population living in each state.

The 2022 House Election takes places during a non-presidential year, and is the most recent election following the 2020 Census and redistricting, allowing us to use the most recently available district maps, and demographic data. Within this report we combine demographic data and urbanization data into a logistic regression model to predict the district voting outcomes of the 2022 House Election.

From this data we use each of the districts as a observation for our logistic regression model. We then choose to model the effect of urbanization hierarchically. Since urbanization is tied to geography we assume a hierarchy related to the location of the districts at the state and region levels of the United States. We then built four different models to test different aspects of the hierarchical structure, using urbanization and various demographic covariates. By using a Bayesian approach to this multilevel model, we were able to gain a more precise understanding of the relationship between urbanization, geography and voting outcomes. All of our models were fit using the `brms` R package (Bürkner, 2017).

Comparative Literature

Inherent in our analysis is the assumption that Republican and Democrat voters are not evenly distributed, rather we assume that the distribution of voters is informative for our analysis. Much of the current research related to the relationship between urbanization and partisan voting focuses on the practice of gerrymandering. Gerrymandering is the practice of redrawing the voting district boundaries in favor of a particular political party. We ignore the impact of gerrymandering in our analysis, since the method of determining if a district has been gerrymandered is rather unclear and out of the scope of this project.

Since we are using a custom curated dataset, there exists no directly comparable research, however we did find interesting ideas related to urbanization and partisanship. One example of this would be the idea of the inefficient distribution of Democrats in cities, as explored by Eubank and Rodden, 2020. In their analysis they discuss the phenomena of spatial inefficiency wherein higher concentrations of democratic voters in urban areas leads to fewer districts voting for democratic candidates that would be expected. Rather than trying to predict the partisan outcome of a particular district, their analysis focuses on measuring partisan spatial efficiency in order to understand the impact of gerrymandering (Eubank & Rodden, 2020).

Although the Eubank and Rodden look at election outcomes through the lenses of spatial efficiency and we are trying to look at election results of districts with respect to their urbanization, we can take the following insights from their analysis. First the distribution of Democrats and urbanization are inexplicably linked, they found that all over the country Democrats tend to be concentrated in urban areas which indicate that it will likely be a good predictor. We would like to investigate to what extent that is true. They also note that the effect of spatial efficiency is highly dependent on state and the branch of government. When considering that urbanization can be considered a proxy for spatial inefficiency, this further supports our decision to model *Urban Index* hierarchically.

2 Dataset

Our dataset was made by combining four independent datasets related to the 2022 House election. The first dataset is the publicly available urbanization dataset published by FiveThirtyEight from which we incorporate the variables urbanization index (*Urban Index*) and (urban) grouping into our final dataset (Holly Fuong, 2022). The urbanization dataset was put together by FiveThirtyEight as part of their analysis *The Republican Path To a House Majority Goes Through The Suburbs* which gave election predictions leading up to the 2022 U.S. Congressional Election (Skelly, n.d.). From the description of the dataset: *"The urbanization index is calculated as the natural logarithm of the average number of people living within a five-mile radius of every census tract in a given district, based on a weighted average of the population of each census tract. The population of a census tract is according to 2020 census data. This provides a numerical value for how urban or rural a district is."* (Holly Fuong, 2022). Referring to Table 1 we see that *Urban Index* ranges from 8.1 to 15, which are best interpreted in combination with the other variable we included in our curated dataset-urban grouping. Urban grouping is a collapsing of the *Urban Index* into six categories ranging from "Dense Urban" for districts with an *Urban Index* above 13 and "Mostly Rural" for districts with an urbanization index below 9.5. We did not however include urban grouping directly in our modeling.

The second dataset used in our analysis is the Election Results Dataset from FiveThirtyEight (Mehta, n.d.). It is a continuously updated repository of United States Governor, Congressional and Presidential elections. As this dataset includes all elections going back to 1998, we only used a subset of the data relevant to the 2022 House Election. Each state sets its own elections rules which led to additions quirks in the dataset. For example, as of 2020 Alaska uses ranked choice voting, and each stage of the ranked choice voting results is present in this dataset, which made it that much more important to ensure careful preprocessing of the data. From this dataset we used the *State*, and *Winning Party* variables. We also considered using the variable incumbent party, however in addition to being incomplete, incumbent party is highly correlated with our target variable *Winning Party*. Inclusion of incumbent party would lead to a model that places heavy important on incumbent party and much lower importance on other covariates. Since we are more interested in the relationship between our model and our chosen covariates we therefore decided not to include it.

As a small aside, although technically a multiparty system, the U.S. is often called a two party system due to the domination of the major political parties the Democrats and Republicans (Embassy, n.d.). These parties dominate particularly on the federal level because political candidates are required to get a plurality of votes rather than a majority of votes which the two largest parties often reach. This is further reinforced as would be third-party voters, often vote for one of the two major parties to ensure their voice is heard, rather than using their vote on a candidate unlikely to reach a plurality of votes (Embassy, n.d.). Within our dataset, there are no districts represented by a third-party candidate and as such we will refer to the U.S. as being a two party system. Therefore we modeled *Winning Party* as a binary variable, with democrats encoded as (1) and republicans encoded as (0). The third data used in our analysis is a subset of the 2022 American Community Survey Data. The American Community Survey is a yearly survey collecting information about the occupations, education attainment, income and other demographic information carried out by the United States Census Bureau. The United States Census Bureau provides an online tool to access its extensive survey database, which can then be filtered and refined for further analysis. For our analysis we considered the following variables for each House district; *Total Population, Percentage Women*, *Median Household Income*, *Mean Household Income*, *Percentage Retirees*, *Percentage Bachelors* degree holders above the age of 25 years old and *Unemployment Rate* among those above the age of sixteen.

Table 1: Covariate Summary Statistics

Variable	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
Urban Index	11	1.4	8.1	10	12	15
Total Population	764634	44505	543189	744260	784528	1018396
Percentage Women	50	0.93	47	50	51	53
Median Household Income	77782	20652	40532	63352	87918	166489
Mean Household Income	107219	29122	63317	87310	120316	252250
Percentage Retirees	17	3.5	8.7	15	19	35
Percentage Bachelors	22	5.8	7.3	17	25	40

From these covariates we hoped to capture education, income, and, demographic make-up of each district because we thought they might be influential in determining partisan voting outcomes. However not all of these covariates were included in our final models.

The majority of our models are relatively small with only *Urban Index* and one other covariate, *Percentage Retirees*. In our literature review we found that older voters tend to vote more conservatively, which helped inform our prior choice (Brown et al., 2022). We also decided to use percentage retirees instead of median age, as they were highly correlated and median age also includes a part of the population that cannot vote.

In our larger model we included further demographics such as, *Percentage Women*, *Median Household Income*, and *percentage bachelors*. Initially we thought that percentage women would be a poor predictor. As can be seen in Table 1 the difference in minimum and maximum percentage women is only six percentage points, with a mean of 50 percent. However a Pew Research Survey found that women tend to lean more democrat and have higher turnout, which indicates that even a small difference in the percentage women may have a large impact (Center, 2020).

In our largest model we also wanted to capture the impact of income. One idea was to take the difference of mean- and *median household income* as a measure of inequality. However, Gelman et al., 2010 found no strong relationship between income inequality and partisan voting. Instead we included only *median household income* as a measure of income as it is a less skewed measure than mean income as in seen in Table 1.

As previously explained, the districts are drawn in such a way that the total population of each district should be approximately the same (Embassy, n.d.). Although we can see in Table 1 that there is some variability in *total population*, the differences within states are relatively small as can be seen in Table A2 and A3 in the appendix. While a similar population in each district does help support our exchangeability assumption, we determined that total population itself would be an unsuitable covariate. Initially we also thought that unemployment rate may be a good predictor of partisan voting outcomes, however Park and Reeves, 2020 found that unemployment is not a good indicator on its own, rather when unemployment is high the incumbent is more likely to lose regardless of party(Park & Reeves, 2020).

The fourth dataset was the region dataset, which was put together manually by us following the four statistical region designations of United States Census Bureau (Bureau, n.d.).

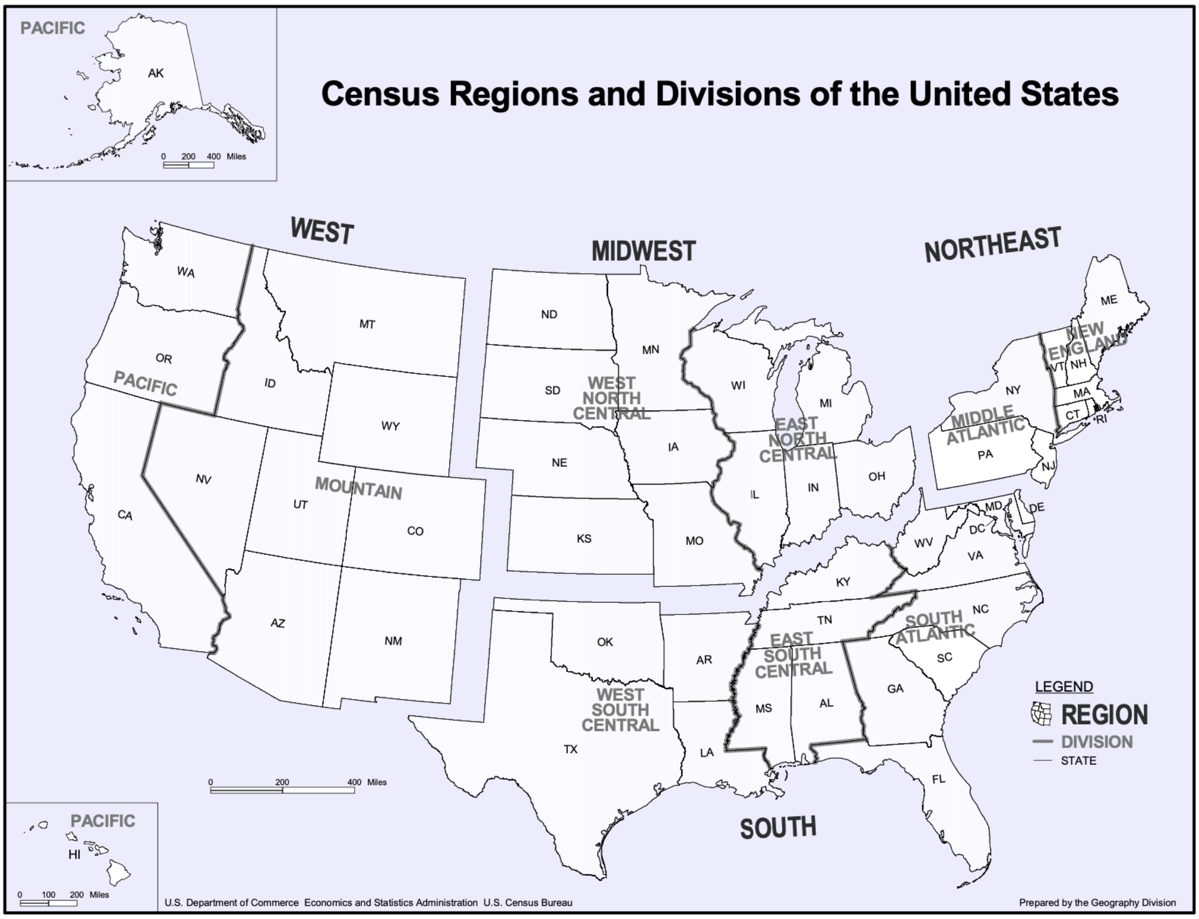


Figure 2: Map showing geographical regions of the United States from the U.S. Census Bureau (Bureau, n.d.)

Data Cleaning

We then merged these four datasets to created our own curated dataset. We did this by merging the different datasets on shared variables. As previously said, each observation represents a particular house district, so for the first three datasets, we simply merged them based on their state and district number. To include the regions we simply used the state variable for each district.

In the election results dataset, we encountered one instance of missing data. For all of the districts in Louisiana, the winning party was not recorded. There was however the incumbent party recorded in the Election Results dataset and by cross referencing this with public record we found for the 2022 House Election only candidates from the incumbent party remained in each of the districts. Therefore as is reflected in our code base we used the data from the incumbent party in place of winning party for the state of Louisiana. For all other states and districts we did not encounter this problem.

In terms of scaling we wanted all the variables to be on roughly the same scale to aid in convergence times. In order to do that we roughly scaled median income and total population by dividing total population by one million and dividing median income by one hundred thousand. This brought each of these to roughly the same scale as the other variables. Similarly we decided to scale all of the percentage variables to be on the scale zero to one hundred rather than zero to one to make their coefficients more interpretable.

3 Code

To fit Bayesian Multilevel Models, one needs to use a probabilistic programming language. We used the R package `brms` (Bürkner, 2017), which allows the user to code entirely in R while running STAN as the backend. There are six R files in the zip file handed in, as well as the trace plots from the prior sensitivity analysis.

The file *data_cleaning.R* shows our process for combining our datasets. The file *modeling.R* contains the `brms` code setting the priors and fitting our four models using the dataset obtained in *data_cleaning.R*. The `brms` package allowed us to efficiently fit many models, as we needed only has to specify the model in a one- or two-line R formula. Each model was then fit with the function `brm()`, using the usual R formula syntax. The priors are set according to what was laid out in Table 2. For the parameter estimates we chose to run 4 chains with 4000 iterations each. We choose to run double the default number of iterations to get more information related to convergence from the trace plots. Further, we set `save_pars = save_pars(all = TRUE)`, so that all parameter draws are saved. This was done to later perform moment matching when computing the LOO-CV statistics for model comparison, although it had no effect on the model estimates.

The file *model_comparison.R* is where we compute and plot the model comparison metrics described in Section 8. The RMSE and LPD are obtained with the method described in Bürkner, 2024, and the LOO cross-validation is performed using the `loo()` and `loo_compare()`, also from the `brms` package.

report_prior_sens_code.r contains the code used to run the prior sensitivity analysis, using the alternative priors described in the corresponding section.

The file *tables_plots.R* contains code for generating tables used in the report (dataset descriptives, model estimates) as well as the marginal effects plots. Which were obtained with the `conditional_effects()` function (setting the option `re_formula = NULL`, to identify the group-level effects).

Finally, the files *report_prior_sens_code.R* and *report_trace_plots_code.R* contain code for the prior sensitivity analysis and trace plots respectively. The former extracts the coefficients from various prior fits and the latter saves the trace plots used in this report.

4 Models

The *Winning Party* in each congressional district race ($y_{i,j,k}$ for district i, state j, region k) can be modeled as the outcome of Bernoulli trial, since this is a binary variable:

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

with probability of a Democrat win $\pi_{j,k}$ modeled as the inverse logit transform of $\theta_{j,k}$, a linear combination of our covariates. The inverse logit function converts real numbers into quantities between 0 and 1, and is therefore a standard way to model probabilities (Hartmann et al., 2023).

We tested four different models for $\theta_{j,k}$, which include different covariates in addition to our variable of interest (Urban index) and incorporate our data's hierarchical structure in different ways. Therefore, all four are Multilevel Bayesian (Logistic) Models, which require particular assumptions (Gelman et al., 2013): first, that a logistic regression accurately represents the relationship between the log-odds of a Democrat win and the explanatory variables, that is, $\theta_{j,k}$ and our covariates are linearly related; second, exchangeability, meaning that each district is exchangeable within each state and each state is exchangeable within each region; and third, that the value of urban index (and other covariates) in a district has a different effect depending on the state/region it belongs to.

The logistic relationship is common choice for modeling binary outcomes. It allows us to model the probability of a Democrat win by a linear predictor which can take any real value, while still having an interpretation for the coefficient estimates (in terms of change of log-odds).

We assume exchangeability because we assume that the districts are drawn in such a way that they are competitive for both parties. Meaning that although the districts may have different characteristics, certain mechanisms can be best captured when thinking of districts as exchangeable parts of a hierarchical model. As an example, for complicated historical reasons, in certain regions of the United States Districts and States are more similar to each other to others outside of the region. For example, the South of the United states tends to be more religious, and religious people tend to vote more conservatively. We can think about this then as a prior telling us about the mix of democratic and republican districts within a particular state. Although the religiosity may increase the number of potential republican districts in each state, whether one district votes republican does not influence the decision of another since each district outcome is determined by thousands of individual votes.

Since each district is part of a state and a region, we can translate the complex geographically determined mechanisms into our model by modeling some of parameters hierarchically.

We chose, however, not to include a varying Intercept in the linear predictor in any of our models. While the varying intercept could potentially improve the accuracy of our predictions, this is not our goal. We are interested primarily in the effect of urbanization on the election outcome in each district and how this effect can be modeled hierarchically. Therefore, we were more concerned with including other covariates in the linear predictor which could interfere with the estimated effect of Urban Index, and opted by leaving the random intercept out to keep the number of parameters to be estimated to a minimum.

Model 1

Our first model includes only our variable of interest, *Urban Index*, plus the *Percentage of Retirees* as covariates to explain θ, plus an intercept. *Percentage of Retirees* was included because it has a relatively high correlation with the *Winning party* and *Urban Index* variables, and there is extensive evidence in the literature of age being related with voting direction (Brown et al., 2022).

Urban index was modeled hierarchically, with the coefficient varying by State, with a prior dependent on common parameters β_{urb} and σ_{urb}, which in turn have (hyper-)priors of their own. The intercept is assumed to be non-variant for all districts, as well as the slope of percentage of retirees.

Equation 2 describes our model conceptually.

$$\theta_j = \beta_0 + \beta_{urb,j}^{uncent} \cdot \text{Urban\_Index} + \beta_{ret} \cdot \text{Pct\_Retirees} \tag{2}$$

To better understand what happens in the backend when we want to fit this model with `brms`, it is helpful to rewrite the equation 3 in terms of 'global' and 'hierarchical' effects. The previously considered coefficient is then decomposed into these effects, i.e. $\beta_{urb,j}^{uncent} = \beta_{urb} + \beta_{urb,j}$ with $\beta_{urb,j}$ centered around zero, which does not alter the meaning of the model (Bürkner, 2024).

$$\begin{aligned} \theta_j =& \beta_0 + \beta_{urb} \cdot \text{Urban\_Index} + \beta_{urb,j} \cdot \text{Urban\_Index} \\ &+ \beta_{ret} \cdot \text{Pct\_Retirees} \end{aligned} \tag{3}$$

Although we assume that there is indeed state-level clustering in the district election outcomes, we have 50 states, and some of them include only one or two districts. This can make the hierarchical estimates unreliable.

Model 2 (region level)

To overcome this problem, we fit another model, with only one difference from the previous one: the hierarchy is at the region level, rather than state. This means the coefficients of *Urban Index* vary by region now, with a common mean and variance which are parameters to be estimated themselves. Equation 4 describes this model, in its specification with separate global and hierarchical effects for *Urban Index*.

$$\begin{aligned} \theta_k =& \beta_0 + \beta_{urb} \cdot \text{Urban\_Index} + \beta_{urb,k} \cdot \text{Urban\_Index} \\ &+ \beta_{ret} \cdot \text{Pct\_Retirees} \end{aligned} \tag{4}$$

Model 3 (nested)

In this model we include the entire geographical hierarchy: a *nested hierarchy* of districts within states within regions. Here the assumption is that the effect of *Urban Index* ($\beta_{urb,j:k}$) depends on state j and region k through a prior with mean parameter $\beta_{urb,k}$, which in turn varies by region and depends on hyper-mean β_{urb} (which has its own prior, with hyper-hyper-parameters). Equation 5 specifies the model, with the centered around zero formulation.

$$\begin{aligned} \theta_{j,k} =& \beta_0 + \beta_{urb} \cdot \text{Urban\_Index} + \beta_{urb,k} \cdot \text{Urban\_Index} \\ &+ \beta_{urb,j:k} \cdot \text{Urban\_Index} + \beta_{ret} \cdot \text{Pct\_Retirees} \end{aligned} \tag{5}$$

Model 4 (big model)

This is our most extensive model. Here we used *Urban Index* and 4 additional covariates plus an intercept to explain $\theta_{j,k}$. It can be seen as an extension of Model 1, as *Urban Index* is modeled hierarchically by state. The region level hierarchy is instead included only in the effect of *Percentage Bachelors degrees*. *Median Household Income* effect is also considered to vary by State, and the intercept and the slopes of *Percentage Women* and *Percentage Retirees* were modeled non-hierarchically.

Equation 6 describes this model, in the `brms` adapted specification.

$$
\begin{aligned}
\theta_{j,k} =& \beta_0 + \beta_{women} \cdot \text{Pct\_Women} \\
&+ \beta_{urbindex} \cdot \text{Urban\_Index} + \beta_{urbindex,j} \cdot \text{Urban\_Index} \\
&+ \beta_{bsc} \cdot \text{Pct\_Bachelor's} + \beta_{bsc,k} \cdot \text{Pct\_Bach.} + \beta_{inc} \cdot \text{Median\_Income} \\
&+ \beta_{inc,k} \cdot \text{Median\_Income} + \beta_{ret} \cdot \text{Pct\_Retirees}
\end{aligned} \tag{6}
$$

5 Priors

Priors represent our initial beliefs about our model parameters' distributions. In each of our models, this means a prior for the Intercept, one for the slope of each covariate that is modeled non-hierarchically (e.g., β_{ret} in all four models) and, in the case of covariates with global and hierarchical effects, priors for the hyper-parameters as well.

Table 2 lists our selected priors for each model, by corresponding covariate.

	Model 1	Model 2	Model 3	Model 4
Intercept	$\beta_0 \sim N(0,10)$	$\beta_0 \sim N(0,10)$	$\beta_0 \sim N(0,10)$	$\beta_0 \sim N(0,10)$
Urban Index	$\beta_{urb} \sim N(0,1)$ $\beta_{urb,j} \sim N(0,\sigma_{urb})$, $\sigma_{urb} \sim Halfcauchy(10)$	$\beta_{urb} \sim N(0,1)$ $\beta_{urb,k} \sim N(0,\sigma_{urb})$, $\sigma_{urb} \sim Halfcauchy(10)$	$\beta_{urb} \sim N(0,1)$ $\beta_{urb,k} \sim N(0,\sigma_{urb,1})$, $\sigma_{urb,1} \sim Halfcauchy(10)$ $\beta_{urb,j:k} \sim N(0,\sigma_{urb,2})$, $\sigma_{urb,2} \sim Halfcauchy(10)$	$\beta_{urb} \sim N(0,1)$ $\beta_{urb,j} \sim N(0,\sigma_{urb})$, $\sigma_{urb} \sim Halfcauchy(10)$
Pct.retirees	$\beta_{ret} \sim t(1,-2,1)$	$\beta_{ret} \sim t(1,-2,1)$	$\beta_{ret} \sim t(1,-2,1)$	$\beta_{ret} \sim t(1,-2,1)$
pct.women				$\beta_{women} \sim N(0,1)$
pct bsc				$\beta_{bsc} \sim t(1,0,1)$ $\beta_{bsc,k} \sim N(0,\sigma_{bsc})$, $\sigma_{bsc} \sim Halfnormal(0,1)$
median income				$\beta_{inc} \sim N(0,1)$ $\beta_{inc,j} \sim N(0,\sigma_{inc})$, $\sigma_{inc} \sim Halfnormal(0,1)$

Table 2: Priors defined for each of our four models, for each parameter; parameters and distributions are listed by their corresponding covariate.

We define such distributions before seeing the data, based on existing literature and our own intuition about the effects of our covariates on the probability of a Democrat win. We assumed the same priors for the same terms included in different models (intercept, percentage of retirees, urban index for the same levels).

For the intercept β_0 we set a Normal prior centered at zero with a large standard deviation. This represents a weakly informative prior, as we had no strong beliefs about the intercept value, nor does it have any straightforward interpretation in our model.

For the population-level component of the *Urban Index* slope we opted for a standard normal prior in all models. We chose not to make assumptions on the sign of the effect of this variable, as it is this variable that we are interested in studying, although we are assuming that its absolute value will be below 1.96 with 95% certainty.

All group-level (zero-centered) priors are Normal, by `brms` specification.

For the standard deviation of the hierarchical effects we opted for a relatively weakly informative prior, a $(half)Cauchy(0, 10)$. We do not want to place strong constraints on the effect of our variable of interest, hence we 'allow the estimates to fluctuate'.

As found in the literature older voters tend to vote more conservatively (Brown et al., 2022), so it makes sense that a higher the *Percentage Retirees* in each district negatively correlates with the probability of a Democrat winning, but we do not know how strong this effect ought to be. Therefore, for the prior on β_{ret} we chose a distribution centered around a negative number, and with relatively heavy tails, reflecting our uncertainty, for all models.

In Model 4, *Percentage Women* is not modeled hierarchically. Although women tend to vote more democratic and have higher voter turnout (Center, 2020), the percentage of women is roughly the same in every district, so we do not expect this covariate to have a strong effect on the probability of either party winning, i.e, we expect β_{women} to be close to zero. So, we set a prior for this slope which is centered around zero and has little variability: a standard normal prior.

The effects of *Percentage of Bachelors* degrees and *Median Household Income* are parameterized in 2 hierarchical levels: an average slope across all districts, and a varying slope by group (State or Region), $\beta_{covariate,j}$ or $\beta_{covariate,k}$, which follows a Normal distribution centered at zero with standard deviation modeled at group level (by a hyperprior). This was done to capture the idea of different costs of living and importance of education respectively, as what is considered rich, poor, or well educated for an individual is highly dependent on the environment.

As we expected the population-level effects of both *Median Household Income* and *Percentage Bachelors* to be positive in some cases and negative in others, we picked symmetric priors for both β_{bsc} and β_{inc}. We are, however, less sure about the on-average-null effect of the *Percentage of Bachelor* degrees, so for this slope parameter we opted for a prior with 'fatter tails', the standard Cauchy distribution rather than the Normal one, representing a higher degree of uncertainty.

For the hyperparameters, we chose a half standard normal prior for both the standard deviations of $\beta_{bsc,k}$ and $\beta_{inc,j}$. This is a narrow distribution, with most values falling between 0 and 1, as we expect to see weak effects for these covariates, and thus small standard deviations (and positive, as any SD is by definition).

6 Results

Each of our models has different estimates for the parameters, even when they are starting from the same prior. As the estimate is given by the mean of the posterior distribution for

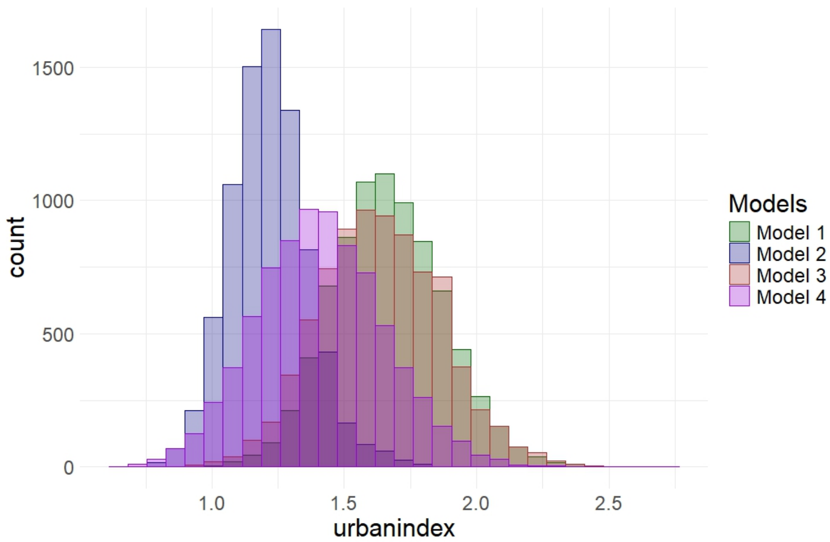


Figure 3: Histogram of posterior draws of Urban Index effect, all models

that parameter, this reflects the different posterior distributions between models, which is only natural given the different structures each model assumes for our data.

The intercept for example, is very different between models, and is particularly large (in absolute terms) in Model 4. Although, as discussed before, the intercept does not have a practical interpretation in our model, it is interesting that this Model's estimate is so far away from the value we initially assumed it would take (between -19.6 and 19.6 with 95% certainty, with a $N(0, 10)$ prior).

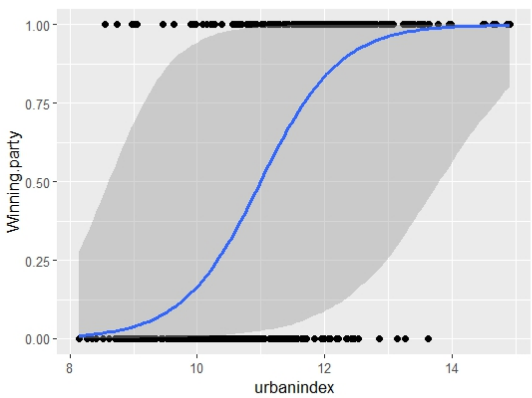
Our variable of interest, *Urban Index*, has a more consistent slope estimate across models, between 1.22 (Model 2) and 1.66 (Model 1). Figure 3 shows the histograms of posterior draw for the *Urban Index* slope parameter, by Model. The results indicate that the variable has, as expected, on average a positive effect on the log-odds of a Democrat Party win. Model 3, for example, estimates that on average a 1 unit increase in *Urban Index* (e.g., 10 to 11) translates into a 1.64 increase in the log-odds (or an $e^{1.64} \approx 5.155$ increase in the odds) of a Democrat being elected in the district. While the estimates are within the range we assumed with the prior for this parameter, the upper bounds of the credible intervals are already close to the boundary of that prior. A prior with more probability mass on the positive range, or a less informative prior, could push the estimates further up.

The standard deviation estimates for the hierarchical effect of *Urban Index* are all on the lower side of the prior, close to zero. This is true both for the cases when the hierarchy on *Urban Index* includes *State* or *Region*.

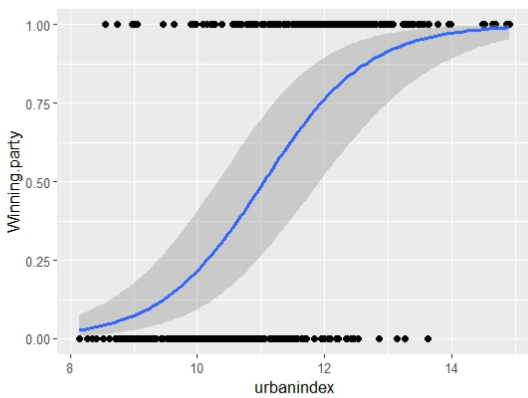
The estimated slope of *Percentage Retirees* is also different between models, particularly in Model 4, where it is larger in absolute value, but always negative as we had assumed. Note that although this estimate is smaller in absolute terms than the slope of *Urban Index*, it does not mean that *Percentage of Retirees* has a weaker effect than urbanization in election outcome. These are different variables on a different scale: in model 3 for example, it is estimated that an increase of 1 percentage point in the percentage of retirees in a district correlates with a 0.16 decrease in the log-odds of a Democrat being elected.

The other covariates included in model 4 also have slope estimates within our expected range. Median Income has a larger (absolute) estimated slope coefficient and standard deviation, but this is likely due to the scale of the variable: 1.43 is the estimated average decrease in log-odds of a Democrat win when the median income of a district increases by 100,000 dollars, which is a big jump. The inclusion of these covariates has some effect on the magnitude of the slope estimate for our variable of interest (*Urban Index*), but not on its sign, and not as much as including a State-specific hierarchical effect vs. Region-specific one.

While the estimated effect is not so different between models, the credible intervals are another story. Figure 4 shows the plotted estimated marginal effect of *Urban Index* on the probability of a Democrat winning according to models 1 and 2, accounting for group-level effects. Model 1 estimates much wider credible intervals than 2. This is due to the number of levels in each model: more levels result in higher uncertainty. and (possibly) to the fact that in model 1 there are several levels with a very small number of observations.



(a) Model 1



(b) Model 2

Figure 4: Conditional effects of *Urban Index*, for Models 1 and 2; mean in blue, with 95% credible intervals in gray, observations in black; group level effects included

7 Convergence Diagnostics

One of the fundamentals of Bayesian analysis is its reliance on MCMC sampling. This ensures we have access to both the posterior samples and the estimates of the posterior regression coefficients. All our data analysis was done using `brms`, which runs on STAN, which itself uses the Hamiltonian Monte Carlo algorithm for the posterior generation.

HMC Convergence diagnostics can be a rather extensive topic, so for this project we mainly consider graphical diagnostics, namely: the MCMC trace plots as provided by the

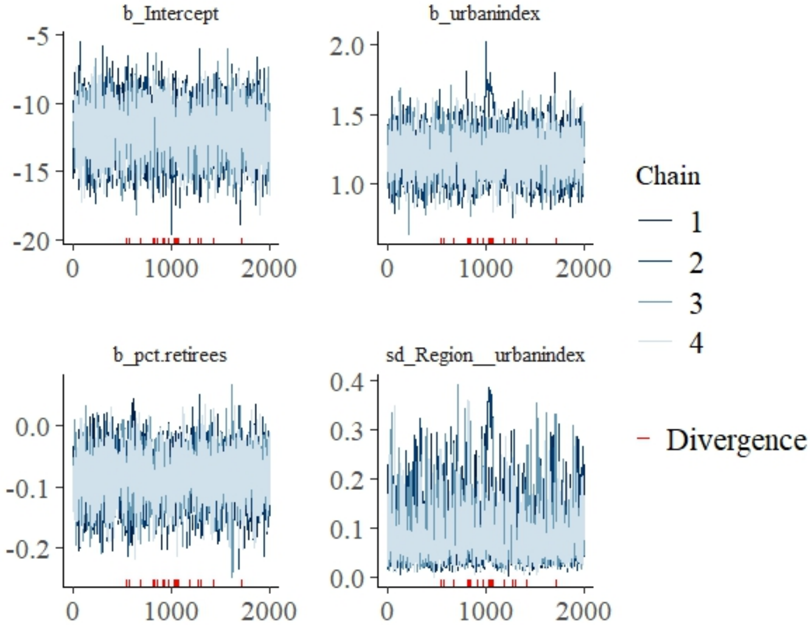


Figure 5: Trace plots for Model 2, all parameters; 2000 sampling iterations

`brms` package.

We present the trace plots for Model 2 (Fig. 5) and most coefficients of Model 3 (Fig. 6). Although we see divergent transitions in both cases, there are two key differences: Firstly, the number of divergent transitions: There are many more of these for Model 3 in comparison to Model 2. Secondly, the place of occurrence: Model 3's divergent transitions occur throughout the chains, while those in Model 2 seem to be concentrated around the 1000th iteration, while noticeably decreasing towards the 2000th iteration. This suggests that increasing iterations further might help with eventually achieving convergence for Model 2, but is unlikely to help with Model 3. The trace plots of the other models and coefficients (in Appendix) either resemble those of Model 2, in that they have divergent transitions near the 1000th iteration and fewer divergent transitions near the 2000th iteration, or have no divergent transitions as is the case with Model 1.

Our key takeaway is the HMC chains have more divergent transitions throughout the chains for Model 3. This is likely due to Model 3 being the only model which takes into account both a Region-wise hierarchy and a State-wise hierarchy for the same coefficient, with the former being nested in the latter. The other models have at most a single hierarchy per coefficient.

8 Model Comparison

Our four models were built based on somewhat different assumptions about the structure of our data, and all produced slightly different results. To better answer our research question, we need to know which of these is *better* prediction-wise. To this end, we

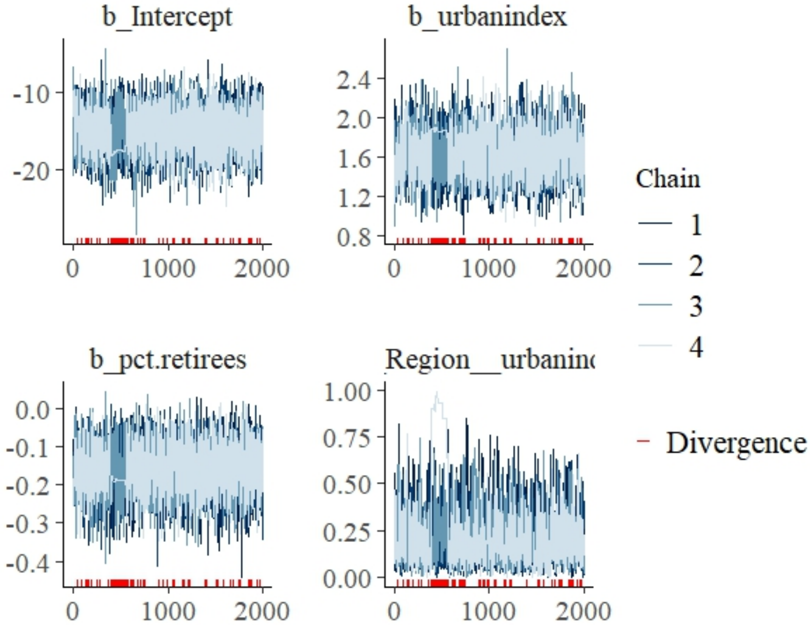


Figure 6: Trace plots for Model 3, intercept, slopes and SD for Region level; SD for the lower level (Region:State) in Appendix; 2000 sampling iterations

measured and compared our models' predictive performance using different metrics.

Absolute predictive performance metrics directly tell us how well the model performs, without looking at other models. To measure absolute predictive performance we used the Root Mean of Squared Error (RMSE). The RMSE for the s-th posterior draw is obtained from the predictive errors, that is, observed outcome y_n minus posterior draw $\hat{y}_n^{(s)}$, squaring those errors and taking the root of their average over all observations, i.e.,

$$RMSE^{(s)} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(y_n - \hat{y}_n^{(s)}\right)^2}$$

Since it is computed for each draw, as opposed to a single point estimate, it takes into account the posterior uncertainty.

The RMSE retains the scale of the response variable. Usually this would mean it has a direct interpretation in the context of the problem, but here we are dealing with a binary response variable, so what the RMSE actually represents is the average distance between the predicted value and 0 or 1, not the true probability of a Democrat win. A model with higher RMSE in this context is not necessarily better at estimating this true value, just on average estimates probability values that are larger for districts where a Democrat won and smaller where a Republican won instead.

Model 4, with higher RMSE, estimates on average closer percentages to the actual outcome which, in a certain sense, can mean it estimates voting outcomes better. This is unsurprising, as this model has many more covariates, and at least some have a large effect on the probability of a Democrat win. Model 2 has the largest average errors, which can mean this model's predictions are more often "wrong" (estimates small probability when a Democrat has won or large otherwise) or that the predictions are in general closer to the center (0.5) rather than the extremes. Model 3 has on slightly smaller RMSE values than Model 1, suggesting the inclusion of the Region hierarchy really does not have a large impact on posterior predictions.

Relative predictive performance measures, contrary to absolute ones, do not have an interpretation in themselves, only as a comparison between models. To assess relative predictive performance, we looked at log-likelihood scores, that is, the average of posterior draws' log-likelihoods for each observation. This is a relative predictive performance measure in the sense that it does not tell us anything about the model's predictive performance alone, rather we need to compare it between different models to establish which is better. So, we examine the differences in log-likelihood scores between models. The sum of these differences corresponds to the difference in Log Predictive Density (LPD) between models, where LPD for a given model is the sum of Log-Likelihood scores of all observations in the model.

Comparison	LPD_Diff	SE_LPD_Diff
Model 1 - Model 2	42.00	7.25
Model 1 - Model 3	-2.32	1.29
Model 1 - Model 4	-20.24	3.91
Model 2 - Model 3	-44.32	7.31
Model 2 - Model 4	-62.24	8.20
Model 3 - Model 4	-17.92	3.66

Table 3: LPD differences between models, with respective standard errors

Table 3 shows the differences in LPD between all four of our models and corresponding standard errors. Model with higher LPD are preferred, as this reflects an overall higher (less negative) log-likelihood score across all observations, meaning the model more accurately predicted each result. The difference between Models 1 and 2 is positive, meaning Model 1 has higher LPD, so it is preferred over 2 using this statistic. In fact, Model 2 is never preferred to any of the other models. Once again, Model 4 provides the best fit and Models 1 and 3 are very close, with 3 performing slightly better. This, together with the comparatively poorer fit of Model 2 points towards Region-level hierarchy possibly not being the correct choice for modeling *Urban Index*.

Both RMSE and LL scores are in-sample predictive performance metrics. In-sample predictive performance measures evaluate only model predictions for the same observations which were used to fit the model in the first place, therefore they only evaluate how well a model predicts the data it was trained on, which means there is a danger of overfitted models, that would not generalize well to new data, performing much better under these metrics (Bürkner, 2024). It is essential to check also out-of-sample predictive performance metrics. These metrics are computed by splitting the dataset into training data and test data, fitting the model on the former and computing the expected LPD (ELPD) of the observations in the latter, given the model estimates with the training set.

The way we choose to split the data into training/test sets naturally impacts the ELPD. So, we rely on cross-validation: we do multiple different splits and average over the results. Our chosen method was Leave-One-Out (LOO) cross-validation, which in theory performs as many splits as observations in the dataset, each time leaving one 'out' as the test data. In practice, a different posterior is not actually computed this many times, but rather an estimate from the full model posterior using importance sampling (Pareto-Smoothed Importance Sampling in this case).

	elpd_diff	se_diff	elpd_loo	se_elpd_loo	p_loo	se_p_loo	looic	se_looic
Model 4	0.00	0.00	-156.49	11.59	32.57	3.25	312.98	23.19
Model 3	-17.99	4.23	-174.48	12.11	33.08	3.33	348.97	24.22
Model 1	-20.35	4.61	-176.84	12.09	33.37	3.43	353.68	24.18
Model 2	-45.87	8.39	-202.36	12.25	6.19	0.58	404.71	24.50

Table 4: LOO statistics for all models

According to the LOO statistics (Table 4), Model 4 (the largest) is the preferred one, followed by Models 3, 1 and 2. The first column of the table shows the difference between each model and the best performing one (in terms of ELPD score, shown in the third column), ranked by best to worst model. This is the same ranking we had seen before, with the more complex models performing better. There is a relatively sizable difference in LOO scores between Model 4 and 3, as well as between 1 and 2, but the difference between 3 and 1 is minimal. The extra level of hierarchy (Region) in the *Urban Index* slope estimate improves model predictions, although not by much. As a small note, when performing the LOO-CV analysis for Model 1, one Pareto-k-estimate was in the 0.7-1 range, which in theory could indicate that the results were not trustworthy, although 1 out of 435 observations is such a small percentage that it is unlikely that it would affect this model's placement in the ranking. In any case, since Model 1 and 3 are very close in ELPD as estimated with LOO, we decided to perform a moment matching correction to the importance sampling for the problematic observation. As expected, the ranking was the same.

9 Prior Sensitivity Analysis

One of the most important parts of Bayesian data analysis is setting the prior distributions. The choice of priors greatly affects the final model estimates, so it is important to test their robustness against different priors (Depaoli et al., 2020). We conducted a prior sensitivity analysis, by refitting our models using alternative priors that also fit our basic model assumptions and assessed the impact on the estimated coefficients.

For easy comparability and scope, we restrict this section to the analysis of Models 1, 2 and 3. We tested 5 different Prior Sets, which are slight variations on the priors we used in our original analysis, all other priors remain as described in Section 5. These changes are as follows :

1. Original priors as per Table 2
2. Change the prior on σ_{urb} from $\sim Halfcauchy(10)$ to $\sim Halfnormal(0,1)$
3. Change the prior on β_{urb} from $\sim N(0,1)$ to $\sim N(0,10)$
4. Change the prior on the intercept from $\sim N(0,10)$ to $\sim N(0,100)$
5. Change the prior on σ_{urb} from $\sim Halfcauchy(10)$ to $\sim Halfnormal(0,1)$ and the prior on β_{urb} from $\sim N(0,1)$ to $\sim N(0,10)$

The first set is just our original set of priors, for comparability. Sets 2-4 represent a single prior that is different relative to the original set, and Set 5 combines two of those changes, to see how estimates behave when more than one prior is altered. The new prior for σ_{urb} in Set 2 was picked because we had chosen a relatively weakly informative prior before, so now we switch it for a more informative one, assuming very small values for the standard deviation. The prior for β_{urb} in Set 3 is less informative than the original, but still symmetric around zero to relax the assumption of the *Urban Index*'s coefficient being 'small' (between -2 and 2 with 95% certainty).

The prior for the intercept in Set 4 was chosen to be *very* weakly informative, to see what happens if we start with the belief that the intercept is nearly as likely to attain very small (or large) values as it is to being close to zero. Prior Set 5 combines the changes in Sets 2 and 3.

Coefficient	Prior Set 1	Prior Set 2	Prior Set 3	Prior Set 4	Prior Set 5
Intercept	-15.7499	-15.7229	-16.4911	-15.6280	-16.4657
Urban Index	1.6647	1.6579	1.7263	1.6502	1.7237
Pct. Retirees	-0.1370	-0.1357	-0.1315	-0.1361	-0.1320
SD Urban	0.09	0.09	0.1	0.09	0.1

Table 5: Prior Sensitivity Results for Model 2

Table 5 displays the (posterior) regression coefficients estimated for Model 2 for the various prior sets. The corresponding results for Model 1 and 3 can be found in the appendix. We see that for the *Urban Index*, our variable of interest, the coefficient changes the most if we set a weakly informative normal prior rather than the standard normal. Interestingly, the intercept seems to respond more to changes (relative to set 1) in the *Urban Index* prior than to changes in its own prior.

The prior sensitivity results for the other two models look rather similar, in that the the coefficient estimate for urban index for prior sets 3 and 5 are the most different from that of prior set 1. The intercept also responds the most in the sets where the urban index prior is changed, as in the results for Model 2.

We also looked at the trace plots arising from the various prior set changes. These can be found in the zip file with the code because there were too many to fit in the appendix. Here is what we found: Model 1 has no divergent transitions for any of the prior sets.

For Model 2, there are more divergent transitions for prior sets 2, 3, and 5 (relative to prior set 1, for all coefficients). As a general rule, however, there are lesser divergent transitions towards the end of the chain.

For Model 3, the situation is reversed: Prior Set 1 has the most number of divergent transitions (for all coefficients). However, just like for prior set 1, there are divergent transitions throughout the chains.

10 Limitations and Improvements

Our analysis and conclusions are limited by our data, the methods and modeling choices made and the scope of the project. We now address these limitations, as well as possible ways our analysis could be expanded upon and improved.

In terms of exploring our research question itself, we assume a linear (in the log-odds) relationship. While our present analysis focuses on the hierarchical nature of the linear model, it would be a natural extension to look at other functional forms of the relationship. Our research question is principally focused around urban index, however it would also be interesting to incorporate the Urban category variable information. For example by examining the behavior of the model at urban category boundaries to see if there is a large change there.

Regarding our specific modeling choices, more models could be fitted using different covariates than the ones we specified. Furthermore, we could fit a model which includes a hierarchically varying Intercept in addition to the random slopes. While modeling the intercept is not the goal of our project, by letting it inform the hierarchical effect of *Urban Index* through a correlation term we could achieve a better fit for our variable of interest effect.

With respect to the priors, although our choices were informed by our knowledge of the data and intuition, there are other distributions that would also meet our beliefs, or make a series of equally valid assumptions. While our Prior Sensitivity Analysis explores this already, more can always be done: different distributions could result in very different estimates or change the convergence of the MCMC chains.

We have also chosen not to do our modeling with STAN directly, but rather to use `brms`. While this package provides a very comprehensive set of instruments for Bayesian Hierarchical Modeling, it is still less flexible than using STAN. For example,with `brm()` all the group-level zero-centered priors are Gaussian. This cannot be changed, we can only affect these parameters' distributions through the prior on the standard deviations. Although a Normal prior is the 'conventional' or 'pragmatic' choice (McElreath, 2016), it would be interesting to experiment with other distributions in our case.

11 Reflection and Discussion

This project has allowed us to better understand how a variable's effect on another can be modeled hierarchically in the context of Bayesian analysis.

By trying different hierarchical structures, we were able to explore how different levels, specifically many levels with few observations in each level (Model 1) vs fewer levels with many observations in each (Model 2) affect the resulting estimates and credible intervals. More levels result in larger credible bands for the estimates, but a better model fit as measured by Leave-One-Out Cross-Validation.

We have discovered that the scale of the variables impacts not just the estimates, but also our prior choice, computation times and the convergence of the chains themselves: in our first attempts at modeling, the variables were not on a similar scale, and while theoretically this would have made no difference for interpretation (we could still adjust the coefficients to read the estimated changes in terms of more sensible units), many of our chains were not converging and the models took a long time to be fitted. Besides, as we were estimating the effects of very large or very small changes in the covariates, the corresponding priors had to be set on very large or very small ranges as well, which affected what would normally be considered a more or a less informative prior. Once we adjusted the scales, computations were faster, we saw better convergence in the trace plots and could reasonably set priors in roughly the same scale.

12 Conclusion

In this report we explored the relationship between urbanization level and election outcomes in the United States using Bayesian multilevel modeling.

We modeled the probability of each party (Democrat or Republican) winning in each congressional district by a logistic regression, where urbanization (measured by an Index) was included as a covariate with a global and a geographical, by State or Region, hierarchical effect. Four different models, which accounted for the geographical hierarchy in different ways were fitted, analyzed in terms of results and convergence, and compared with formal metrics. Furthermore, we conducted a Prior Sensitivity Analysis to check the robustness of our findings.

We used the R package `brms` to fit the models, as well as for the analysis and comparison.

In all models, we found a small positive effect of *Urban Index* on the log-odds of a Democrat winning the district. The estimated effect was smaller when the hierarchical effect or the Index was modeled by Region, and larger when this was done by State, with not much difference between including State-only or a nested hierarchy or Region and State.

All models converged well, except for the nested hierarchy one, which resulted in visible convergence problems and should therefore be considered with caution.

Comparing all our models with Bayesian comparison metrics, we found that Model 4 (the one with the most covariates) performed better, and the Model 2, which included only Region-level hierarchy for *Urban Index*, worse.

The Prior Sensitivity Analysis indicated that our results are robust to different prior choices.

We conclude therefore that the effect of Urbanization on the probability of a Democrat being elected is positive, and modeling this effect by State rather than Region leads to better predictions.

Appendix

Descriptives

Table A1: Correlation matrix for the six variables used in the Models (one response variables, five covariates)

	Winning.party	urbanindex	pct.women	Median.Household.Income	pct.retirees	pct.bach
Winning.party	1.00	0.60	0.20	0.24	-0.30	0.29
urbanindex	0.60	1.00	0.30	0.39	-0.39	0.46
pct.women	0.20	0.30	1.00	-0.15	0.14	0.05
Median.Household.Income	0.24	0.39	-0.15	1.00	-0.10	0.76
pct.retirees	-0.30	-0.39	0.14	-0.10	1.00	-0.09
pct.bach	0.29	0.46	0.05	0.76	-0.09	1.00

Table A2: Total Population Summary Statistics (1)

State	Districts	Average	Standard Deviation	Minimum	Maximum	Difference
Alabama	7	724899	10656	710137	743238	0.045
Alaska	1	733583		733583	733583	0.000
Arizona	9	817689	20000	790643	851459	0.071
Arkansas	4	761409	16746	747672	784904	0.047
California	52	750564	20271	705678	797316	0.115
Colorado	8	729991	9977	718693	748891	0.040
Connecticut	5	725241	10542	710465	735042	0.033
Delaware	1	1018396		1018396	1018396	0.000
Florida	28	794458	24364	740547	839779	0.118
Georgia	14	779491	14149	755007	806637	0.064
Hawaii	2	720098	3338	717738	722458	0.007
Idaho	2	969516	21826	954083	984950	0.031
Illinois	17	740120	14495	708538	766225	0.075
Indiana	9	759226	7898	747577	772783	0.033
Iowa	4	800129	9538	793421	814070	0.025
Kansas	4	734288	6396	727503	741829	0.019
Kentucky	6	752052	9005	739149	762092	0.030
Louisiana	6	765040	22786	727277	796937	0.087
Maine	2	692670	7111	687642	697698	0.014
Maryland	8	770582	18278	744504	792577	0.061
Massachusetts	9	775775	9711	754113	785636	0.040
Michigan	13	771855	8727	757463	782743	0.032
Minnesota	8	714648	11648	700555	731533	0.042
Mississippi	4	735014	21190	704754	750414	0.061
Missouri	8	772245	13199	742101	785669	0.055

Table A3: Total Population Summary Statistics (2)

State	Districts	Average	Standard Deviation	Minimum	Maximum	Difference
Montana	2	561434	11169	553536	569331	0.028
Nebraska	3	655974	5332	649904	659903	0.015
Nevada	4	794443	22073	767891	821679	0.065
New Hampshire	2	697616	8920	691308	703923	0.018
New Jersey	12	771808	13444	746241	791164	0.057
New Mexico	3	704448	8468	696764	713527	0.023
New York	26	756814	24281	699930	786432	0.110
North Carolina	14	764212	9051	751852	779106	0.035
North Dakota	1	779261		779261	779261	0.000
Ohio	15	783737	7429	769169	799350	0.038
Oklahoma	5	803960	4826	796469	807958	0.014
Oregon	6	706690	12034	687278	719249	0.044
Pennsylvania	17	763059	14122	717771	780519	0.080
Rhode Island	2	546867	5201	543189	550545	0.013
South Carolina	7	754662	7586	741110	762713	0.028
South Dakota	1	909824		909824	909824	0.000
Tennessee	9	783482	13836	756975	801730	0.056
Texas	38	790252	25455	732116	846385	0.135
Utah	4	845200	28097	806633	874074	0.077
Vermont	1	647064		647064	647064	0.000
Virginia	11	789420	13074	764684	810541	0.057
Washington	10	778579	11771	751668	789222	0.048
West Virginia	2	887578	15224	876813	898343	0.024
Wisconsin	8	736567	8455	719451	743974	0.033
Wyoming	1	581381		581381	581381	0.000

Trace Plots

Figure A1: Trace plots for Model 1, all parameters; 2000 sampling iterations

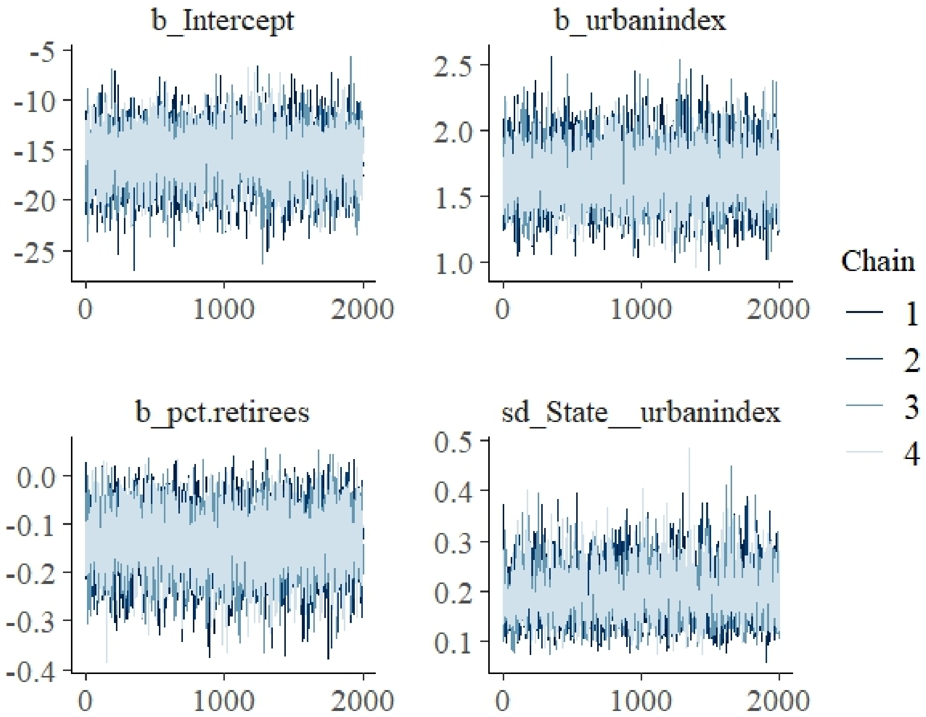


Figure A2: Trace plots for Model 2, all parameters; 2000 sampling iterations

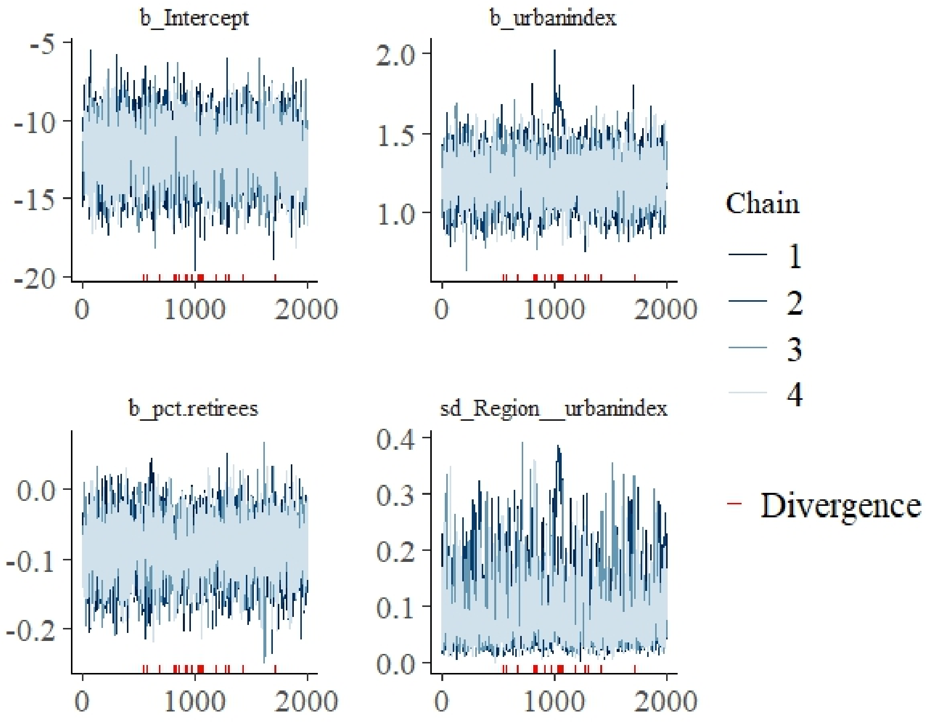


Figure A3: Trace plots for Model 3, all parameters; 2000 sampling iterations

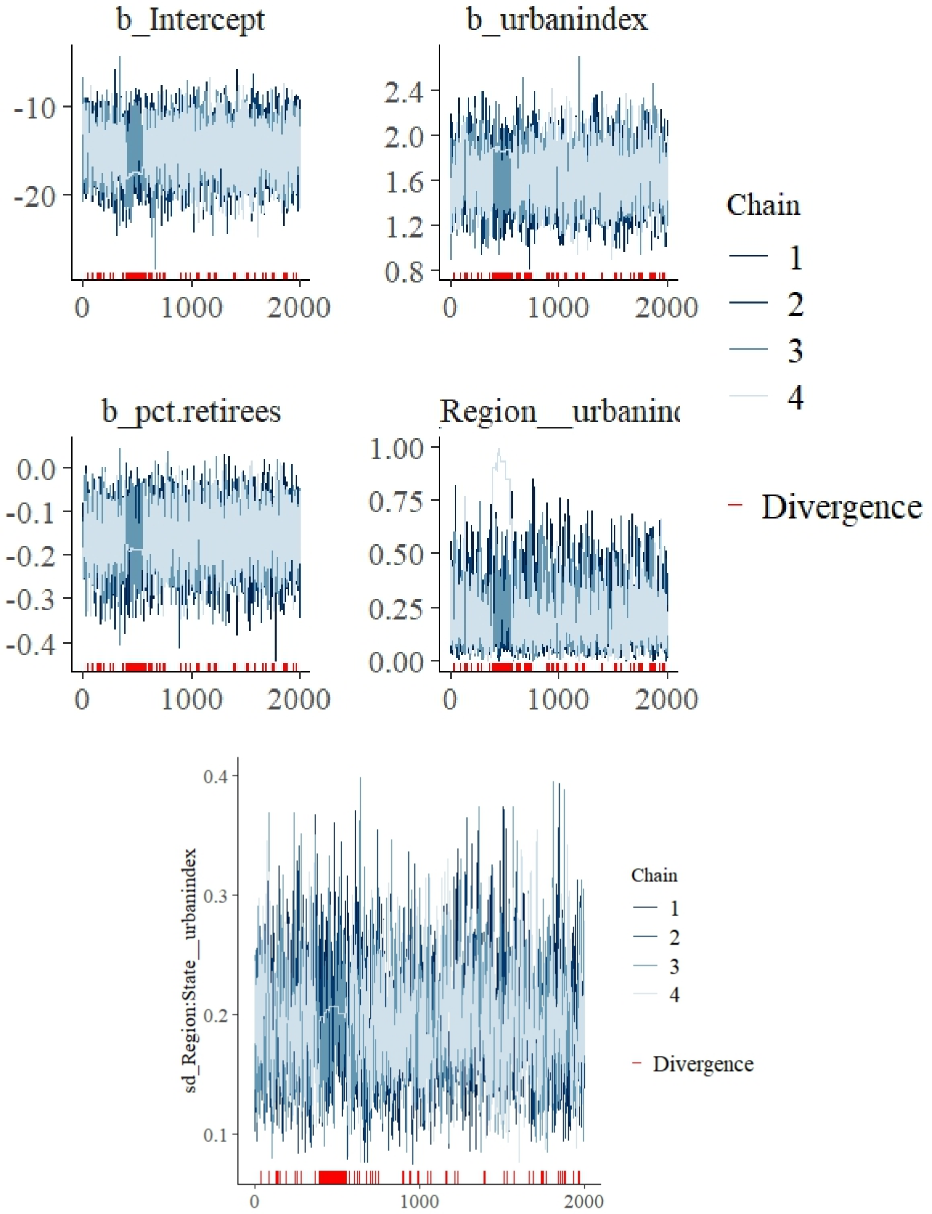
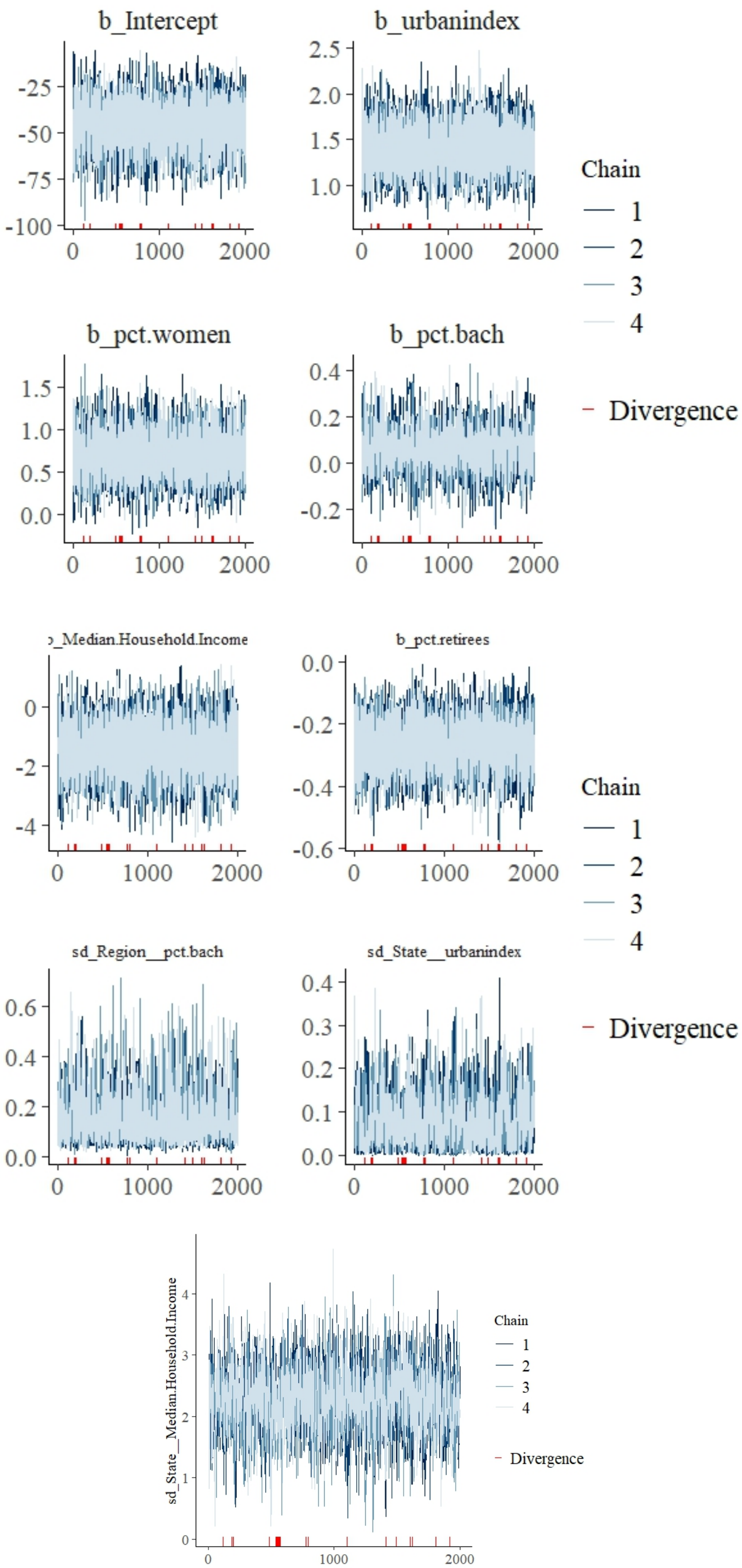


Figure A4: Trace plots for Model 4, all parameters; 2000 sampling iterations



Model Estimates

Table A4: Parameter estimates obtained by Model 1 with all 435 observations, 4 chains, 4000 iterations (2000 warmup, 2000 sampling)

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Regression Coefficients:							
Intercept	-15.75	2.63	-20.96	-10.73	1.00	7397.71	6010.33
urbanindex	1.66	0.21	1.28	2.10	1.00	5191.43	5331.31
pct.retirees	-0.14	0.06	-0.27	-0.02	1.00	6086.48	5724.70
Multilevel Hyperparameters: State (Number of levels: 50)							
sd(urbanindex)	0.19	0.05	0.11	0.30	1.00	1569.53	3108.00

Table A5: Parameter estimates obtained by Model 2 with all 435 observations, 4 chains, 4000 iterations (2000 warmup, 2000 sampling)

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Regression Coefficients:							
Intercept	-12.01	1.78	-15.58	-8.60	1.00	2892.71	3356.25
urbanindex	1.22	0.15	0.95	1.53	1.00	1294.59	473.21
pct.retirees	-0.09	0.04	-0.16	-0.01	1.00	4773.23	4217.12
Multilevel Hyperparameters: Region (Number of levels: 4)							
sd(urbanindex)	0.09	0.06	0.02	0.27	1.00	892.96	465.15

Table A6: Parameter estimates obtained by Model 3 with all 435 observations, 4 chains, 4000 iterations (2000 warmup, 2000 sampling)

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Regression Coefficients:							
Intercept	-15.02	2.69	-20.32	-9.87	1.00	2837.78	5324.03
urbanindex	1.64	0.23	1.21	2.10	1.00	2787.63	2875.43
pct.retirees	-0.16	0.06	-0.29	-0.04	1.00	5424.17	5097.97
Multilevel Hyperparameters: Region (Number of levels: 4)							
sd(urbanindex)	0.21	0.16	0.03	0.67	1.01	348.01	120.67
Region:State (Number of levels: 50)							
sd(urbanindex)1	0.19	0.05	0.11	0.29	1.00	1976.07	3875.21

Table A7: Parameter estimates obtained by Model 4 with all 435 observations, 4 chains, 4000 iterations (2000 warmup, 2000 sampling)

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Regression Coefficients:							
Intercept	-45.84	12.13	-70.40	-22.99	1.00	6668.56	5729.57
urbanindex	1.41	0.24	0.95	1.90	1.00	7707.06	6016.13
pct.women	0.70	0.26	0.21	1.24	1.00	6155.31	4936.03
pct.bach	0.06	0.09	-0.11	0.27	1.00	1593.83	662.43
Median.Household.Income	-1.43	0.87	-3.12	0.25	1.00	6817.58	3411.18
pct.retirees	-0.26	0.08	-0.42	-0.12	1.00	6809.18	5806.15
Multilevel Hyperparameters: Region (Number of levels: 4)							
sd(pct.bach)	0.15	0.09	0.04	0.39	1.00	1855.37	1959.66
Multilevel Hyperparameters: State (Number of levels: 50)							
sd(urbanindex)	0.08	0.06	0.00	0.21	1.00	866.41	1941.43
sd(Median.Household.Income)	2.24	0.51	1.19	3.26	1.00	2627.23	2493.48

Figure A5: Conditional effects of *Urban Index*, for Model 3; mean in blue, with 95% credible intervals in gray, observations in black; group level effects included

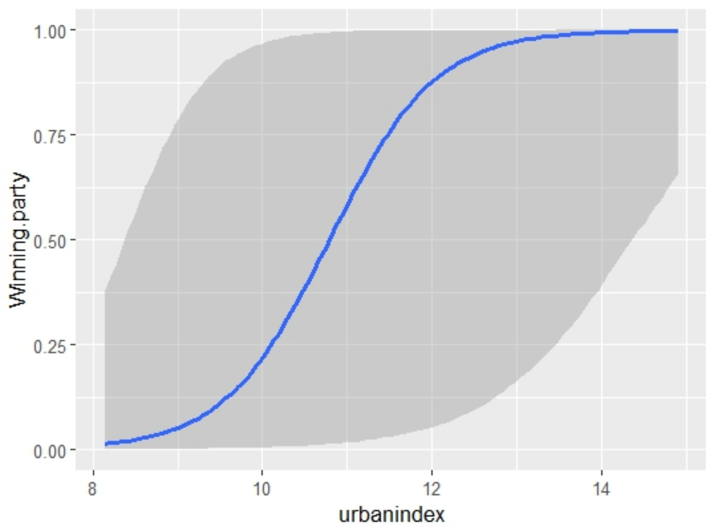
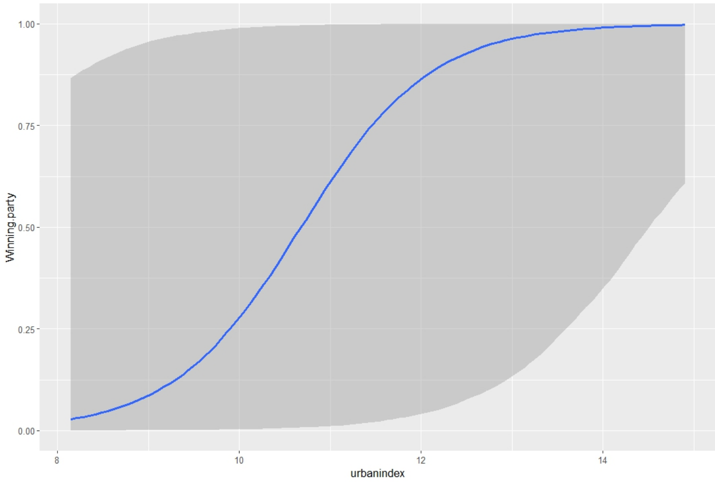


Figure A6: Conditional effects of *Urban Index*, for Model 4; mean in blue, with 95% credible intervals in gray, observations in black; group level effects included



Model Comparison Statistics

Figure A7: Histograms of RMSE draws for all models

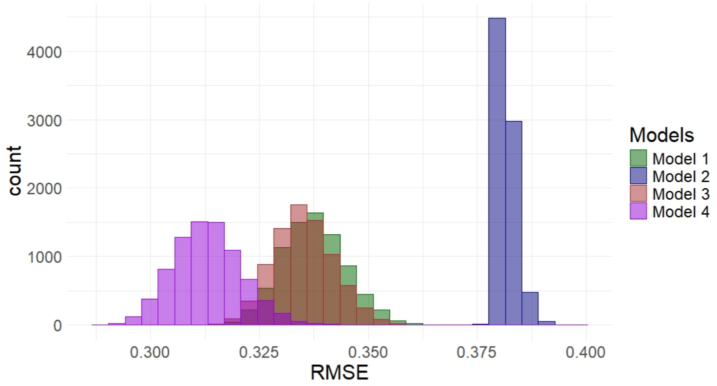


Figure A8: Histograms of differences in Log-Likelihood Scores for each observation between all 4 models; black dotted line marks zero; red vertical line marks the mean of the difference between the corresponding models

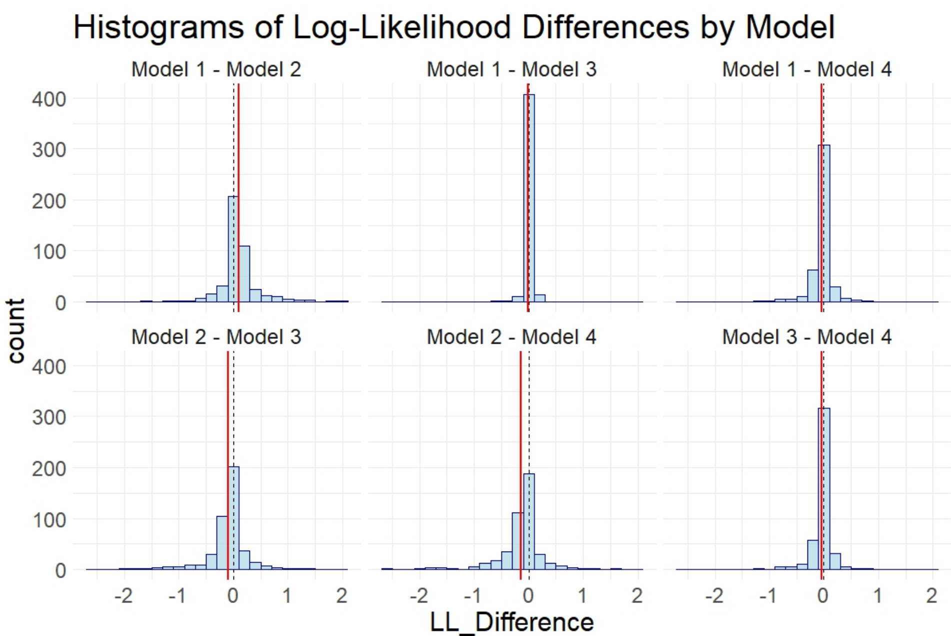


Table A8: LOO statistics with moment matching

	elpd_diff	se_diff	elpd_loo	se_elpd_loo	p_loo	se_p_loo	looic	se_looic
model4.final	0.00	0.00	-156.49	11.59	32.57	3.25	312.98	23.19
model3.final	-17.99	4.23	-174.48	12.11	33.08	3.33	348.97	24.22
model1.final	-20.32	4.60	-176.81	12.08	33.34	3.42	353.61	24.17
model2.final	-45.87	8.39	-202.36	12.25	6.19	0.58	404.71	24.50

Prior Sensitivity Analysis

Table A9: Prior Sensitivity Results for Model 1

Coefficient	Prior Set 1	Prior Set 2	Prior Set 3	Prior Set 4	Prior Set 5
Intercept	-12.0110	-11.8230	-12.1806	-11.9011	-12.1806
Urban Index	1.2167	1.2005	1.2311	1.2047	1.2311
Pct. Retirees	-0.0861	-0.0882	-0.0843	-0.0873	-0.0843
SD Urban	0.19	0.19	0.2	0.19	0.2

Table A10: Prior Sensitivity Results for Model 3

Coefficient	Prior Set 1	Prior Set 2	Prior Set 3	Prior Set 4	Prior Set 5
Intercept	-15.0192	-14.9687	-15.7396	-14.9113	-15.7476
Urban Index	1.6368	1.6327	1.7139	1.6312	1.7189
Pct. Retirees	-0.1599	-0.1596	-0.1577	-0.1606	-0.1568
SD Urban	0.1	0.1	0.1	0.1	0.1

References

Gelman, A., Kenworthy, L., & Su, Y.-S. (2010). Income inequality and partisan voting in the united states. *Social science quarterly*, 1203–1219.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013, November). *Bayesian data analysis* (0th ed.). Chapman; Hall/CRC. https://doi.org/10.1201/b16018

McElreath, R. (2016). *Statistical rethinking: A bayesian course with examples in r and stan*. CRC Press/Taylor & Francis Group.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Center, P. R. (2020). In changing u.s. electorate, race and education remain stark dividing lines. https://www.pewresearch.org/politics/wp-content/uploads/sites/4/2020/06/PP_2020.06.02_Party-ID_FINAL.pdf

Depaoli, S., Winter, S. D., & Visser, M. (2020). The importance of prior sensitivity analysis in bayesian statistics: Demonstrations using an interactive shiny app. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.608045

Eubank, N., & Rodden, J. (2020). Who is my neighbor? the spatial efficiency of partisanship. *Statistics and Public Policy*, *7*(1), 87–100.

Park, T., & Reeves, A. (2020). Local unemployment and voting for president: Uncovering causal mechanisms. *Political Behavior*, *42*(2), 443–463.

Brown, J. R., Cantoni, E., Enos, R. D., Pons, V., & Sartre, E. (2022). *The increase in partisan segregation in the united states* (tech. rep.). Working paper.

Holly Fuong, G. S. (2022). District urbanization index 2022. https://github.com/fivethirtyeight/data/tree/master/district-urbanization-index-2022

Hartmann, K., Krois, J., & Rudolph, A. (2023). *Statistics and geodata analysis using r (soga-r)*.

Bürkner, P.-C. (2024). *The brms book: Applied bayesian regression modelling using r and stan (early draft)* [Early Draft]. https://paulbuerkner.com/software/brms-book

Bureau, U. S. C. (n.d.). *Geographic levels*. https://www.census.gov/programs-surveys/economic-census/guidance-geographies/levels.html

Embassy, U. (n.d.). *Presential elections and the american political system*. https://dk.usembassy.gov/usa-i-skolen/presidential-elections-and-the-american-political-system/

Mehta, D. (n.d.). Election results. https://github.com/fivethirtyeight/election-results/blob/main/README.md

Skelly, G. (n.d.). *The republican path to a house majority goes through the suburbs*. https://fivethirtyeight.com/features/the-republican-path-to-a-house-majority-goes-through-the-suburbs/