Income and Gender: Forecasting the American Voter's Choice in 2018*

Siqi Fei

March 9, 2024

In this study, we examined whether it is possible to predict an individual's voting behavior in the 2018 US elections based solely on their income levels and gender. Utilizing a dataset containing voter information and demographic characteristics, we conducted statistical analyses to identify any patterns or correlations. Our findings suggest that, to a certain extent, a person's income and gender can provide predictive insights into their voting preferences, shedding light on the complex interplay between socio-economic factors and political affiliation. Understanding these relationships can help policymakers and political strategists craft more targeted approaches to engage and mobilize voters across different demographic groups, ultimately contributing to a more informed and inclusive democratic process.

1 Introduction

In the intricate landscape of American politics, unraveling the factors that guide an individual's voting decisions is a crucial endeavor for scholars and political strategists alike. This paper delves into the 2018 U.S. elections—a time characterized by significant voter turnout and pronounced divisions—to investigate the predictive power of income and gender on electoral choices. Despite extensive research into the multifaceted influences on voting behavior, the explicit relationship between these two demographic variables and political preferences remains insufficiently examined. Our focus on the 2018 electoral cycle is intentional, chosen for its illustration of the deep-seated cleavages within the American electorate and its potential to reveal enduring voting behaviors.

Utilizing data from the Cooperative Congressional Election Study (CCES) (Schaffner, Ansolabehere, and Luks 2019), this study employs quantitative methods to explore how variations in income and gender correlate with support for political candidates. The primary

*Code and data are available at: https://github.com/FXXFERMI/Political-support-in-the-United-States.git.

estimand of our analysis is the likelihood of voting for a specific political party or candidate based on an individual's income level and gender. Our findings suggest that these factors are not merely peripheral but central to understanding voter behavior, offering new insights into the demographic contours of political allegiance. This research not only enriches the academic discourse on electoral dynamics but also holds practical implications for crafting more nuanced and effective campaign strategies.

The importance of this study lies in its contribution to a more detailed and informed understanding of the American political landscape. By pinpointing income and gender as significant determinants of voting patterns, we provide a basis for future investigations into how these and other demographic factors shape political identities and allegiances.

The paper is organized into six sections, following this introduction. Section 2 describes the dataset and the rationale behind its selection. Section 3 outlines the statistical methods applied to analyze the data. Section 4 presents the key findings of the study, while Section 5 interprets these results within the broader context of American politics and voting behavior. Through this structure, we aim to offer a comprehensive examination of how income and gender influence voting decisions in the United States, contributing to the broader discourse on democracy and electoral participation.

2 Data

For this paper, we used data from the 2018 Cooperative Congressional Election Study (CCES) (Schaffner, Ansolabehere, and Luks 2019), which we worked with in R (R Core Team 2023), a language for statistical computing. The `tidyverse` suite (Wickham et al. 2019), with its various packages like `ggplot2` (Wickham 2016), `dplyr` (Wickham et al. 2023), `readr` (Wickham, Hester, and Bryan 2024), and `tibble` (Müller and Wickham 2023), made handling the data easier and more precise. We summarized our model results using the `modelsummary` package (Arel-Bundock 2022). The `dataverse` package (Kuriwaki, Beasley, and Leeper 2023) helped us smoothly download the data, and `testthat` (Wickham 2011) made sure our data prep and analysis were solid. The `here` package (Müller 2020) helped keep our files organized and our analysis reproducible.

Our study focuses on three main data points: family income, gender, and individual voting choices in the 2018 election. We organize income into different ranges to examine the potential effect of earnings on political decisions. Gender is categorized simply into male or female, reflecting the dataset's binary format. Regarding voting choices, we concentrate on whether respondents voted for Trump or Clinton, framing it as a binary outcome.

Before analyzing, we thoroughly clean our data. We remove non-essential information and align income brackets with recognized standards. Any problems with data exception or missing values are resolved, ensuring our dataset is reliable.

We present a figure to illustrate our data's narrative. Figure 1 depicts the voting choices broken down by gender and income levels. This bar chart provides a clear visualization of the number of respondents within each income category and their voting behavior.

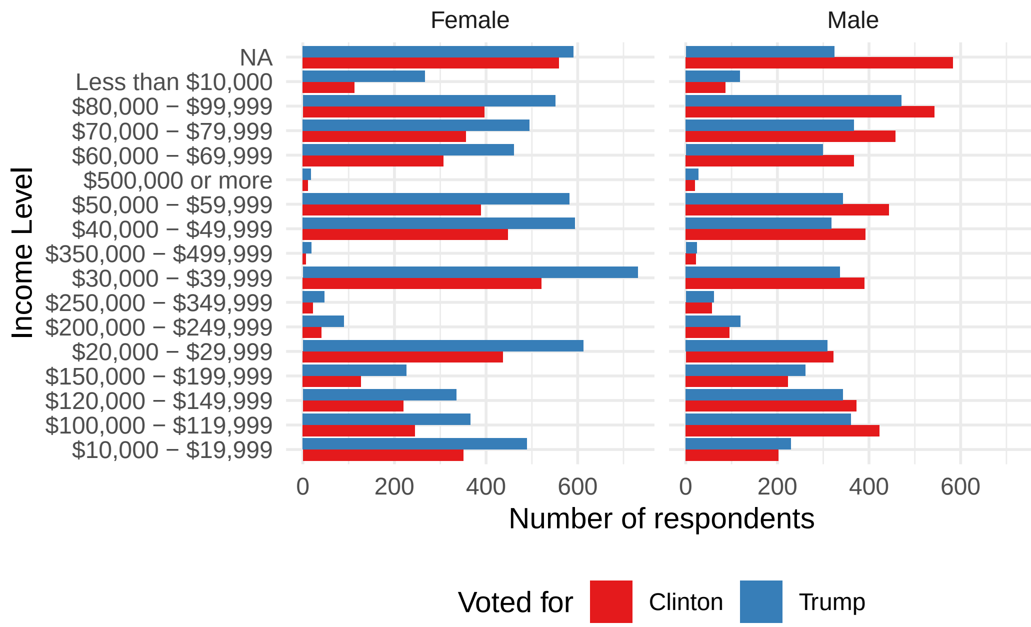

Figure 1: The distribution of presidential preferences, by gender, and income

Reviewing the summary statistics, we notice distinct patterns. For example, higher-income individuals show a tendency to support Trump, while lower-income individuals demonstrate varied support for Clinton.

The next parts of the paper will go into more detail about these findings.

3 Model

The goal of our modeling strategy is to illuminate the underlying dynamics between voters' demographics—specifically, income and gender—and their voting behavior in the 2018 U.S. elections. Additionally, we aim to quantify the strength and direction of these relationships.

To dissect these factors, we apply a Bayesian analysis model, allowing us to incorporate prior knowledge and handle uncertainty in a probabilistic framework. We chose this approach due to its flexibility and the rich interpretability of its results. For those interested in the technical specifics and diagnostic checks of our Bayesian model, these details are thoroughly presented in Appendix A.

3.1 Model set-up

We denote y_i as the binary outcome where 1 represents a vote for Trump and 0 represents a vote for Clinton. In our model, the predictors include gender and income. We use β_1 to represent the effect of gender on the probability of voting for Trump, with gender encoded as a binary variable (e.g., 0 for female and 1 for male). Similarly, β_2 represents the effect of income, categorized into distinct income brackets, on the voting probability.

$$
\begin{aligned}
y_i|\pi_i &\sim \text{Bernoulli}(\pi_i) && (1)\\
\text{logit}(\pi_i) &= \beta_0 + \beta_1 \times \text{gender}_i + \beta_2 \times \text{income}_i && (2)\\
\beta_0 &\sim \text{Normal}(0, 2.5) && (3)\\
\beta_1 &\sim \text{Normal}(0, 2.5) && (4)\\
\beta_2 &\sim \text{Normal}(0, 2.5) && (5)
\end{aligned}
$$

The parameters β_0,β_1,β_2 represent the intercept and the effects of income and gender, respectively, on the log-odds of voting for Trump. A normal prior with mean 0 and standard deviation 2.5 is used for each of these parameters, reflecting a baseline assumption of no effect before observing the data.

We run the model in R (R Core Team 2023) using the `rstanarm` package of (**rstanarm?**). We use the default priors from `rstanarm`.

3.1.1 Model Justification

Given the economic and social context of the 2018 election, it is reasonable to hypothesize that both income and gender would exert influence on voting decisions. We anticipate that income, as a measure of economic status, could have a significant effect on the choice of a candidate known for advocating policies affecting taxations and business. Similarly, gender might play a role given the differing policy stances of the candidates on issues traditionally seen as gendered in public discourse. Therefore, we posit a positive relationship between higher income brackets and support for Trump, while also considering the potential impact of gender on voting behavior.

The subsequent results section will delve into the outcomes of this Bayesian model, revealing the extent to which income and gender can predict support for Trump in the 2018 U.S. elections.

4 Results

The results of our analysis, grounded in the 2018 Cooperative Congressional Election Study (CCES)(Schaffner, Ansolabehere, and Luks 2019), offer a compelling narrative about the interplay between income, gender, and voter preference in the 2018 U.S. elections.

Our results, as summarized in Table 1, explain the relationship between an individual's likelihood to support Trump in the 2018 election and their gender and income level. The intercept, representing the baseline likelihood for the reference group (here presumably female and the omitted income category), is estimated at 0.44. Gender has a notable effect, with males being less likely by 0.50 to support Trump compared to females, holding other factors constant.

The influence of income on voting preference for Trump is more nuanced. For most of the income brackets, such as those earning between $100,000 to $119,999 or $120,000 to $149,999, the coefficients are negative but small in magnitude and not statistically significant, as indicated by their respective standard errors. However, individuals earning between $200,000 to $249,999 show a markedly higher likelihood (an increase of 0.31) of supporting Trump. Notably, the largest income bracket, $500,000 or more, also indicates a higher likelihood of support for Trump, with a coefficient of 0.28. Conversely, individuals in the lowest income bracket, earning less than $10,000, are more likely by 0.40 to support Trump, which is an interesting contrast to the typical association between income level and political preference.

The model's explanatory power, as indicated by the R^2 value of 0.019, suggests that while gender and income do contribute to the prediction of voting behavior, there is a vast amount of variability left unexplained by these factors alone. Other unobserved variables might also play a significant role in an individual's voting decisions.

The fit of the model, evaluated by criteria such as the Widely Applicable Information Criterion (WAIC) and Leave-One-Out Cross-Validation Information Criterion (LOOIC), suggests that the model is relatively well-calibrated given the complexity of human behavior it aims to capture. However, the modest R-squared value indicates that future research could benefit from incorporating additional predictors to provide a more comprehensive understanding of the factors influencing voter behavior.

In summary, the analysis confirms that gender and income are indeed predictive of voter preference, yet the variability indicates that additional factors also play a significant role.

Table 1: Whether a respondent is likely to vote for Trump based on their gender and income

	Support Trump
(Intercept)	0.44
	(0.05)
genderMale	−0.50
	(0.03)
income$100,000 - $119,999	−0.07
	(0.08)
income$120,000 - $149,999	−0.02
	(0.08)
income$150,000 - $199,999	0.19
	(0.09)
income$20,000 - $29,999	−0.05
	(0.08)
income$200,000 - $249,999	0.31
	(0.12)
income$250,000 - $349,999	0.20
	(0.15)
income$30,000 - $39,999	−0.09
	(0.07)
income$350,000 - $499,999	0.28
	(0.25)
income$40,000 - $49,999	−0.15
	(0.07)
income$50,000 - $59,999	−0.10
	(0.07)
income$500,000 or more	0.28
	(0.23)
income$60,000 - $69,999	−0.08
	(0.07)
income$70,000 - $79,999	−0.13
	(0.07)
income$80,000 - $99,999	−0.09
	(0.07)
incomeLess than $10,000	0.40
	(0.11)
Num.Obs.	18 279
R2	0.019
Log.Lik.	−12 442.206
ELPD	−12 459.1
ELPD s.e.	21.2
LOOIC	24 918.1
LOOIC s.e.	42.4
WAIC	24 918.1
RMSE	0.49

5 Discussion

5.1 Discussion

5.1.1 Overview

This paper has navigated through the complex terrain of demographic influence on political preference, particularly in the context of the 2018 U.S. elections. By employing a Bayesian framework to assess the impact of income and gender on voting for Trump, the study enhances our quantitative understanding of electoral behavior, addressing an analytical gap in political science literature.

5.1.2 Gender Dynamics in Political Support

Our research has illuminated the significant role of gender in predicting political preferences. The notable decrease in the likelihood of males voting for Trump, as indicated by the model, adds a quantitative backing to the narrative of gendered political divergence. This gender gap aligns with broader societal discussions on political identity and raises questions about the evolving role of gender in shaping political landscapes.

5.1.3 Economic Status and Voter Preference

A striking aspect of our findings is the complex association between income and support for Trump. The increased likelihood of individuals with lower and higher incomes to vote for Trump could suggest a cross-cutting appeal among disparate economic demographics, challenging conventional wisdom about income-based voting behavior. This nuanced relationship points to the intricate ways economic status intersects with political ideology.

5.1.4 Limitations

The limitations of this study are multifaceted. Primarily, the explanatory power of our model, as indicated by a low R-squared value, suggests that the variables of income and gender, while significant, do not capture the full spectrum of factors influencing voter decisions. Our reliance on binary gender classification and broad income categories may oversimplify the diversity and intricacy of these demographic factors. Additionally, the data's cross-sectional nature restricts our ability to explore causality and changes over time.

5.1.5 Future Research

Moving forward, it is clear that further exploration is needed to deepen our understanding of voter behavior. Incorporating additional demographic variables, such as ethnicity, educational background, and urban versus rural residency, could offer more detailed insights. A longitudinal approach could track changes over multiple election cycles, providing a richer context for understanding voter dynamics.

Integrating qualitative methodologies could offer a layer of depth to the motivations and perceptions behind the quantifiable trends. Surveys, interviews, and focus groups could unveil the personal narratives and societal influences that lie beneath statistical patterns.

To address the limitations around the categorization of gender, future research should consider a more inclusive approach, recognizing and incorporating the full spectrum of gender identities. Similarly, more nuanced economic data could provide a clearer picture of how personal finances translate into political decisions.

5.1.6 Conclusion

This study serves as a stepping stone toward a more comprehensive comprehension of the multifaceted influences on voter behavior. As we peel back the layers of demographic impacts, we pave the way for more informed political discourse and strategy development. The journey of discovery is ongoing, and each successive study builds upon the foundation of understanding laid by its predecessors, contributing to a collective effort to decipher the enigma of voter behavior in the American democratic process.

Appendix

A Model details

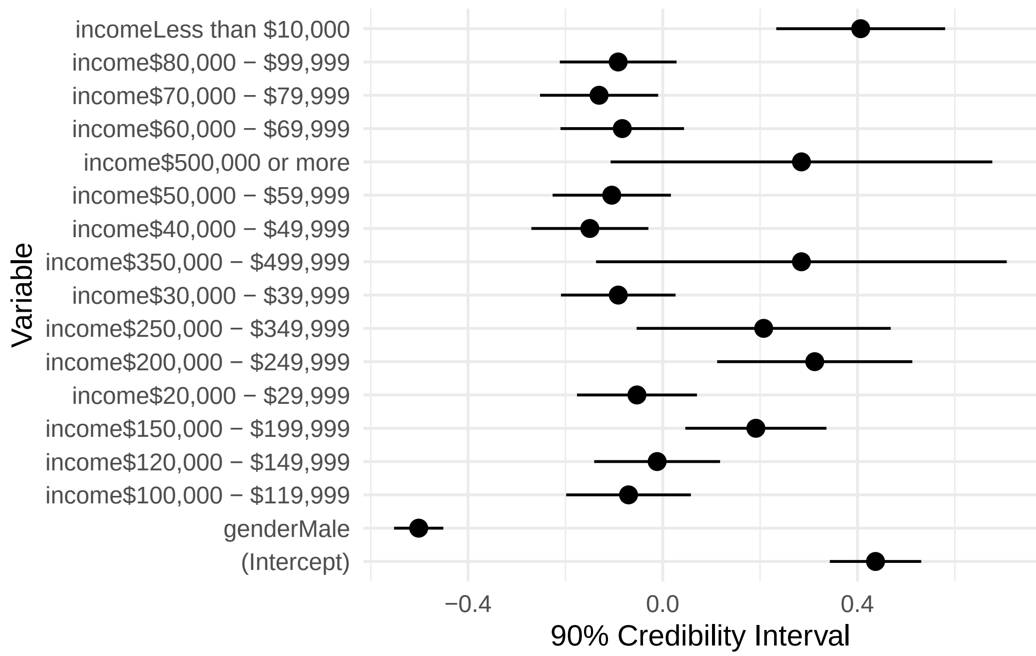

Figure 2: Credible intervals for predictors of support for Trump

References

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Kuriwaki, Shiro, Will Beasley, and Thomas J. Leeper. 2023. *dataverse: R Client for Dataverse 4+ Repositories*.

Müller, Kirill. 2020. *here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

Müller, Kirill, and Hadley Wickham. 2023. *tibble: Simple Data Frames*. https://CRAN.R-project.org/package=tibble.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Schaffner, Brian, Stephen Ansolabehere, and Sam Luks. 2019. "CCES Common Content, 2018." Harvard Dataverse. https://doi.org/10.7910/DVN/ZSBZ7K.

Wickham, Hadley. 2011. "testthat: Get Started with Testing." *The R Journal* 3: 5–10. https://journal.r-project.org/archive/2011-1/RJournal_2011-1_Wickham.pdf.

———. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

——— et al. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2024. *readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.