Understanding the Impact of Mortgage State and Income on Household Poverty Status*

An Analysis of Household Poverty Status in the United States in 2019

Mary Cheng

March 31, 2024

This paper examines household poverty status in the United States in 2019 using the 2019 Revised Supplemental Poverty Measure (SPM) Research Dataset from the United States Census Bureau. Using logistic regression analysis, we investigate the influence of household income level and mortgage state on poverty status. Our findings reveal a higher likelihood of poverty for households' total annual earning less than $50000 and for renters compared to property owners. All households with incomes exceeding $100000 are not in poverty. These insights emphasize the need for targeted interventions to address socioeconomic inequalities and promote financial well-being among US households.

Table of contents

*Code and data are available at: https://github.com/marycx/us_poverty_analysis_2019.git

1 Introduction

Over the last few decades, addressing poverty has remained a persistent issue in the United States, leading to the implementation of multiple policies in an effort to mitigate this problem (Kearney and Harris 2014). In 2019, the United States experienced the culmination of an economic expansion that followed the Great Recession, marking the last full year before the onset of COVID-19 pandemic. The robustness of the U.S. economy during this period was evidenced by low poverty rates, indicating a notable decrease in economic hardship characterized by low income. The poverty rate in the United States was estimated to be 10.5% in 2019, down from 11.8% in 2018, marking the fifth consecutive year of decline (Semega 2022). Despite various interventions and economic policies, understanding the intricate dynamics influencing household poverty status is still difficult. Therefore, it is important for us to analyze why some US household are in poverty in 2019 and how various demographic and financial factors may has an influence on it. Perhaps then we could understand how to reduce the number of households in poverty. We analyze the 2019 Revised Supplemental Poverty Measure (SPM) Research dataset from the United States Census Bureau (USCB) (Bureau 2022). USCB is an agency of the U.S. Federal Statistical System, responsible for producing data about the American people and economy (Bureau 2023). We build a prediction model for households poverty status using 2019 Revised SPM dataset. Supplemental Poverty Measure was developed due to the continued criticism towards the official poverty measure in the US. SPM addresses numerous concerns of official measure critics, and its intent is to provide an improved statistical picture of poverty (Bridges and Robert V.Gesumaria 2015). This allows us to discover trends and address potential economic and societal situations that may cause the poverty status for some US households in 2019.

In this paper, a logistic regression model is used to predict the poverty status of households in the United States in 2019, with data from the 2019 Revised Supplemental Poverty Measure Research dataset. Logistic regression is a great choice since it is used to predict binary outcomes, such as poverty status (in poverty or not in poverty). Our analysis focuses on estimating the likelihood of household being in poverty, based on various demographic and financial factors captured in the SPM dataset. We selected data features: mortgage state and household annual total income. The estimand in this paper is the number of households who are in poverty in reality. However, it is difficult to measure the exact number of households who are in poverty since there are millions of people in the United States and not all of them will be assessed due to various difficulties. Therefore, in this paper, we attempt to estimate the estimand using a logistic regression model which is trained using sample dataset from the 2019 Revised Supplemental Poverty Measure Research Data from USCB.

The logistic regression model shows that in the US in 2019, households with an annual total income less than $50000 are almost all in poverty. For households earning $50000 to $100000, the majority is not in poverty, yet a notable amount of them are still in poverty. For households earning above $100000, all of them are not in poverty. Additionally, renters are more likely to be in poverty compared to owners with or without mortgage.

The remainder of this paper is structured into different sections. Section 2 demonstrates the data used for our report and includes some tables and graphs to illustrate the different groups of people in our data. Section 3 builds the model and discusses its justification and explanation. Section 4 highlights the results of the predictions using tables and graphs. Section 5 contains discussions that conducted based on the findings, which addresses the poverty status results based on mortgage states and income levels. Statistical programming language R (R Core Team 2023) is used in this report, with packages `tidyverse` (Wickham et al. 2019), `here` (Müller 2020), `rstanarm` (Brilleman et al. 2018), `modelsummary` (Arel-Bundock 2022), `ggplot2` (Wickham 2016), `knitr` (Xie 2014), `marginaleffects` (Arel-Bundock 2024), `plotly` (Sievert 2020), `tibble` (Müller and Wickham 2023), `margins` (Leeper 2021), `testthat` (Wickham 2011) and `kableExtra` (Zhu 2021).

1.1 Estimand

The estimand in this paper is the number of households who are in poverty in reality. However, it is difficult to measure the exact number of households who are in poverty since there are millions of people in the United States and not all of them will be assessed due to various difficulties. Therefore, in this paper, we attempt to estimate the estimand using a logistic regression model which is trained using sample dataset from the 2019 Revised Supplemental Poverty Measure Research Data from USCS.

2 Data

2.1 Data Source

This report uses the 2019 Revised Supplemental Poverty Measure (SPM) Research dataset provided by the United States Census Bureau as our main source of data. The Census Bureau provides information about the people and economy of the United States, with a goal to support economic growth, enhance scientific knowledge, and assist in making informed decisions. The official poverty measure has been criticized due to its multiple limitations. For example, it does not consider benefits from most of the largest programs that aid the low-income population. For instance, it uses money income before taxes, so it does not measure the income available for individuals to spend, which for most people is after-tax income. Therefore, any effects of tax credits designed to assist persons with low income are not considered by the official measure. The Supplemental Poverty Measure (SPM), similar to the official poverty measure, tackles the shortcomings of poverty assessment. It determines poverty status by gauging resources against a defined standard of living. This standard considers expenditures on essentials like food, clothing, shelter, and utilities, plus a bit more for additional expenses. The resources assessed include disposable income, accounting for taxes and some non-cash benefits, available to cover these needs (Bridges and Robert V.Gesumaria 2015). In 2019, the dataset contained a total of 157,959 entries.

2.2 Features

The original SPM 2019 dataset, which shows in Table 1, contains 157959 data entries and many variables. Since it is difficult to observe such a large dataset, this report will only explore and analyze through several data features. We first chose these 17 variables: h_seq, spm_poor, spm_tenmortstatus, spm_totval, spm_snapsub, spm_caphousesub, spm_schlunch, spm_engval, spm_wicval, spm_fedtax, spm_eitc, spm_actc, spm_fica, spm_sttax, spm_childsuppd, spm_capwkccxpns, and spm_medxpns. This is because we want to calculate the income after tax, plus all the subsidies and benefits, minus all the expenses, just like how the SPM does it. If we only choose income before tax to analyze, then our analysis will be like the official poverty measure, which requires improvement.

1. h_seq: Household sequence number, unique identifier for each household.
2. spm_poor: the poverty status; 1 representing "In poverty" and 0 represneting "Not in poverty".
3. spm_tenmortstatus: household's tenure/mortgage status; 1 representing "Owner with Mortgage", 2 representing "Owner without Mortgage or rent-free", and 3 representing "Renter".
4. spm_totval: household's cash income, unit in USD.
5. spm_snapsub: Supplemental Nutrition Assistance Program (SNAP) subsidy, unit in USD.
6. spm_caphousesub: capped housing subsidy, unit in USD.
7. spm_schlunch: school lunch subsidy, unit in USD.
8. spm_engval: energy subsidy, unit in USD.
9. spm_wicval: Women, Infants, and Children subsidy, unit in USD.
10. spm_fedtax: Federal tax, unit in USD.
11. spm_eitc: Federal Earned Income Tax Credit, unit in USD.
12. spm_actc: Additional Child Tax Credit, unit in USD.
13. spm_fica: Federal Insurance Contributions Act and federal retirement contribution, unit in USD.
14. spm_sttax: state tax, unit in USD.
15. spm_childsuppd: child support paid, unit in USD.
16. spm_capwkccxpns: capped work and child care expense, unit in USD.
17. spm_medxpns: Medical Out-Of-Pocket (MOOP) and Medicare Part B subsidy, unit in USD.

Table 1: Preview of the raw 2019 US Supplemental Poverty Measure dataset

(a)

h_seq	spm_poor	spm_tenmortstatus	spm_totval	spm_snapsub	spm_caphousesub
1	0	2	127449	0	0
1	0	2	127449	0	0
2	0	2	64680	0	0
2	0	2	64680	0	0
3	0	1	40002	0	0

(b)

spm_schlunch	spm_engval	spm_wicval	spm_fedtax	spm_eitc	spm_actc	spm_fica
0	0	0	13485	0	0	8301
0	0	0	13485	0	0	8301
0	0	0	1459	0	0	2601
0	0	0	1459	0	0	2601
0	0	0	2772	0	0	3060

(c)

spm_fica	spm_sttax	spm_childsuppd	spm_capwkccxpns	spm_medxpns
8301	4624	0	3654	10926
8301	4624	0	3654	10926
2601	0	0	2065	3280
2601	0	0	2065	3280
3060	1152	0	2065	2300

Table 2: Preview of the cleaned 2019 US Supplemental Poverty Measure dataset

poverty_status	mortgage_state	income
Not in poverty	Owner without Mortgage	50k-100k
Not in poverty	Owner without Mortgage	50k-100k
Not in poverty	Owner with Mortgage	5k-50k
In poverty	Renter	5k-50k
Not in poverty	Owner without Mortgage	5k-50k

Table 3: Statistics summary of the cleaned 2019 US Supplemental Poverty Measure dataset

	poverty_status	mortgage_state	income
	Not in poverty:54563	Owner with Mortgage :23587	below 5k : 2273
	In poverty : 8427	Owner without Mortgage:19195	5k-50k :28543
	NA	Renter :20208	50k-100k :19569
	NA	NA	100k-150k : 7326
	NA	NA	150k-200k : 2706
	NA	NA	200k-250k : 1145
	NA	NA	above 250k: 1428

2.3 Data Measurement

2.4 Methodology

The original dataset includes duplicate responses from the same household. As the dataset features are related to the entire household as a whole, this paper will eliminate duplicate entries, using the variable "h_seq" which is an unique household identifier, to ensure data accuracy and consistency.

First, a new column "income after tax" is created by summing up all incomes, credits, and benefits and subtracting all expenses. Then the dataset is cleaned by renaming of column headers, defining column classes, grouping the income variable into different income levels (here the assumed facts are that the income unit is in USD and 'k' stands for thousand, eg. 5k = 10,000), and replacing numerical values in the table with their corresponding descriptions from the data dictionary to improve the readability. After cleaning, 62990 rows of data with 3 data features remain. Table 2 shows a preview of the cleaned dataset.

Table 3 is a summary of the cleaned data, showing detailed statistics about the dataset. As we can see from the table, there are more households which are not in poverty. Respondents cover a wide range of income levels, with "50k-100k" and "100k-150k" being more heavily represented. Also, the mortgage states of the households vary, each of the three mortgage states being fairly evenly represented.

2.5 Data Visualization

Figure 1 illustrates the relationship between mortgage state and poverty status of households. It can be seen that renters are the highest proportion of households experiencing poverty, followed by owners without mortgages, and finally owners with mortgages. Conversely, for households who are not in poverty, owners with mortgages represent the largest segment, followed by owners without mortgages, and then renters. In addition, it can be observed that overall a larger proportion of households are not experiencing poverty. Despite variations in poverty rates across different mortgage states, the cumulative data shows economic stability among the majority of households in the United States in 2019.

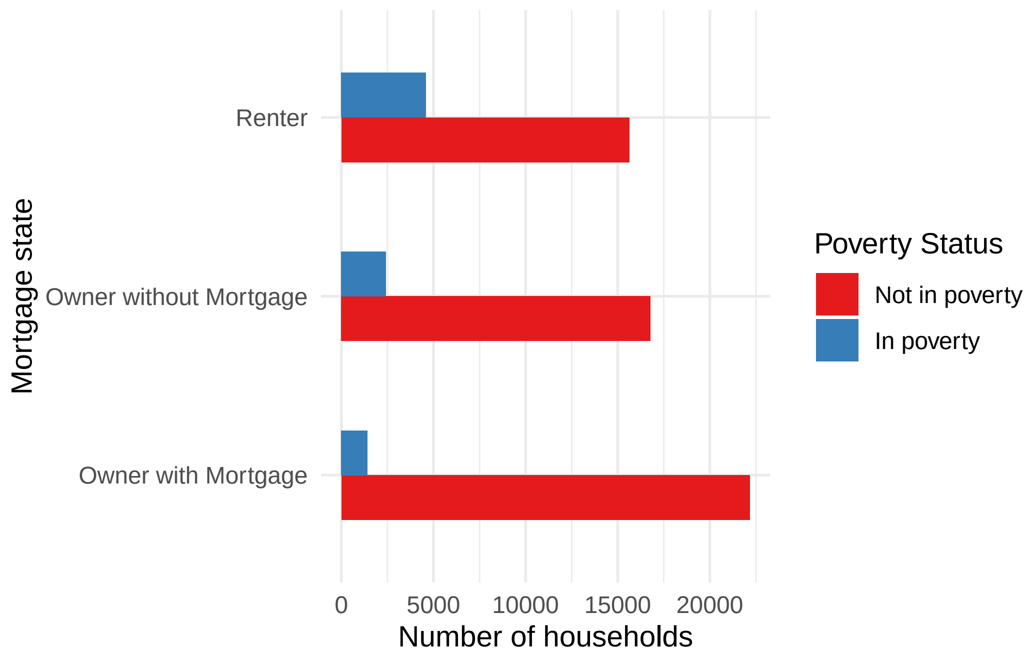


Figure 1: The distribution of poverty status by mortgage state

Figure 2 shows the relationship between household annual total income and their poverty status. We can see that among households earning more than 100k annually, no household is in poverty. For household with income between 50k to 100k, almost all of them are not in poverty with only a very small number of households in poverty. Within the income range of 5k-50k, the majority of households are not in poverty, but there are some families in poverty. However, for households earning below 5k annually, almost all of them are in poverty. This indicates a strong positive correlation between higher income levels and financial stability.

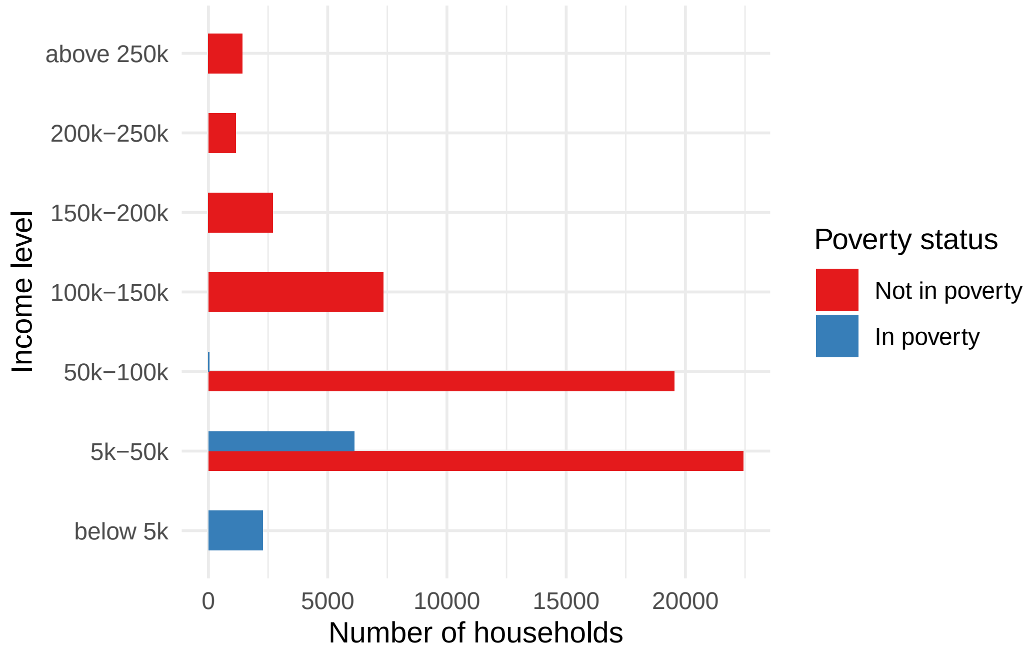


Figure 2: The distribution of poverty status by income level

3 Model

In our analysis, we utilized a Bayesian logistic regression model to examine the relationship between poverty status and two demographic factors, mortgage status and income of the household. Background details and diagnostics are included in Appendix A.

3.1 Model set-up

The model is formulated as follows:

$$y_i|\pi_i \sim \mathrm{Bern}(\pi_i) \tag{1}$$
$$\mathrm{logit}(\pi_i) = \alpha + \beta_1 \times \mathrm{mortgage}_i + \beta_2 \times \mathrm{income}_i \tag{2}$$
$$\alpha \sim \mathrm{Normal}(0, 2.5) \tag{3}$$
$$\beta_1 \sim \mathrm{Normal}(0, 2.5) \tag{4}$$
$$\beta_2 \sim \mathrm{Normal}(0, 2.5) \tag{5}$$

In this model, y_i represents the binary outcome variable indicating whether a household is in poverty (as opposed to not in poverty). The probability of being in poverty (π_i) is modeled using a logistic link function ($\mathrm{logit}(\pi_i)$), which is a linear combination of the intercept (α) and the coefficients (β_1, β_2) corresponding to the predictor variables, mortgage state and income level, respectively. These predictor variables are denoted as income_i and mortgage_state_i, where i indexes the individual households in the dataset.

The intercept (α) and coefficients (β_1, β_2) are assigned informative prior distributions to regularize the model. Specifically, we assume a normal distribution with a mean of 0 and a standard deviation of 2.5 for each parameter.

We chose this modeling approach for several reasons. Firstly, logistic regression is ideal for binary outcome variables, making it appropriate for analyzing poverty status. Also, Bayesian methods enable us to integrate prior knowledge and uncertainty into the analysis, resulting in more reliable estimates of the model parameters.

While alternative modeling approaches like linear regression were considered, we selected Bayesian logistic regression due to the binary nature of our outcome variable, poverty status.

We use the `rstanarm` package (Brilleman et al. 2018) in R (R Core Team 2023) to run the model. Default priors from `rstanarm` is used. Rstanarm employs Markov chain Monte Carlo (MCMC) techniques to estimate the posterior distribution of the parameters. To avoid excessive runtime, 1000 data entries are randomly sampled to fit the model with random seed 215. Model diagnostics, including convergence checks and posterior summaries, are available in the supplementary materials (see Appendix Section A).

3.1.1 Model justification

Regarding the relationship between mortgage state and poverty state, we anticipate that both owners without mortgages and owners with mortgages are less likely to experience poverty. Owners without mortgages typically indicate financial stability, as they have already paid off their homes and can allocate financial resources toward household necessities, other than paying back the mortgage / rent. Also, without monthly mortgage or rent payments, their financial burden is reduced, further decreasing the likelihood of poverty. Similarly, owners with mortgages demonstrate financial responsibility by securing loans from banks. The approval of these loans suggests that the bank is confident in owner's ability to manage repayments, indicating a lower possibility of poverty. On the other hand, renters are more likely to experience poverty due to their lack of property ownership and the hidden implication that they do not earn enough to purchase their own property. Renters typically have fewer savings and may earn insufficient salaries to cover rental and essential expenses. Consequently, the combination of rental payments and other financial obligations increases their vulnerability to poverty.

In terms of income levels, we expect a positive relationship between household's income level and their poverty status. This expectation arises from common sense that higher income typically reduces the likelihood of experiencing poverty. If a family is earning a lot of money, then naturally it would not be in poverty. Households that are earning less than 5k annually in the United States obviously have a larger chance in being in poverty due to the high cost of living. Their income probably barely covers or cannot cover for essential expenses such as housing, transportation, or food.

3.2 Model Implication

For posterior predictive checks, in Figure 3, the great fit of the posterior distribution from our logistic regression model with actual poverty data suggests accurate capture of poverty status patterns. This indicates the accuracy of our model's poverty status prediction to the 2019 SPM data. Also, Figure 4 compares the posterior to the prior, which shows some parameter changes such as "below 5k" and "5k-50k", as well as the intercept (people who are owners with mortgage and earn 100-150k). The discrepancy for the parameter suggest that our prior may not be very accurate in regards to these specific aspects.

The trace plots in Figure 5 and Figure 6 do not suggest anything out of the ordinary. Also, with the Rhat plot (Figure 7), we can observe that everything is close to 1, and no more than 1.05, which shows the great convergence of Markov chain Monte Carlo for our model.

More detailed explanation of each plot can be found in Appendix Section A.

4 Results

Our results are summarized in Table 4. Our results generally matches our expectation. To avoid multicollinearity, the model excludes one variable from each category: mortgage state "Owner with mortgage" and income level "100k-150k". The intercept represents the estimated log-odds of being in poverty when all other predictors are held constant at their reference levels. In this case, the estimated log-odds of being in poverty for people who are owner with mortgage and their total household income level to be 100k-150k annually is -6.896.

The possibility of households with income level below 5k being in poverty is large. The estimated coefficient of 17.352 suggests that, holding all other variables constant, households with income less than 5k are estimated to have a 17.352 unit increase in the log-odds of being in poverty compared to the reference group. Households with income level at "5k-50k" and "50k-100k" on average are also more likely to be in poverty compared to the reference group, with the estimated coefficient to be 5.346 and 1.502 respectively.

The mortgage status of individuals also influences their poverty status. As expected, renters exhibit a higher likelihood of experiencing poverty, as shown by the estimated coefficient of 0.645. However, it's important to note that while the coefficient is positive compared to the reference group, its magnitude is relatively small, suggesting a moderate rather than a substantial difference.

Figure 8 (see Section A.3) shows range of coefficient estimates of our model within the 90% probability. Due to the fact that the credibility interval for mortgage “Renter” and “Owners without mortgage” is quite small, it is hard to observe the trend of the 90% credibility intervals of these two variables. Therefore we created Figure 9 with the x axis limited from -5 to 5.

Combining Figure 8 and Figure 9, we observe statistical significance for the coefficient estimates for household with income below 5k, household income between 5k to 50k, renters, and the intercept, household which owns a property with mortgage and has a income level between 100k to 150k. The estimates are significant because their credibility intervals do not cross 0. The value for the estimates are in log-odds, indicating that if the coefficient is positive, the household is in poverty, if negative, the household is not in poverty.

5 Discussion

5.1 Relationship betwen Mortgage State and Poverty Status

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

5.2 Relationship betwen Income and Poverty Status

Average annual expenditure $63,036

5.3 mortgage interest rate etc...

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Table 4: Explanatory model Poverty Prediction (n = 1000)

	In poverty
(Intercept)	−6.896
	(1.779)
mortgage_stateOwner without Mortgage	−0.248
	(0.351)
mortgage_stateRenter	0.645
	(0.302)
income150k-200k	−5.447
	(7.365)
income200k-250k	−9.312
	(11.319)
income50k-100k	1.502
	(1.938)
income5k-50k	5.346
	(1.795)
incomeabove 250k	−7.004
	(9.012)
incomebelow 5k	17.352
	(5.627)
Num.Obs.	1000
R2	0.403
Log.Lik.	−229.780
ELPD	−234.5
ELPD s.e.	15.5
LOOIC	468.9
LOOIC s.e.	31.1
WAIC	468.9
RMSE	0.27

Appendix

A Model details

A.1 Posterior predictive check

In Figure 3, we implement a posterior distribution. This compares the poverty status of people in reality with the prediction results from the posterior distribution from our logistic regression model. It can be seen that the posterior distribution fits well with the actual data. It suggests that the posterior is able to generate simulated data that closely resembles the actual data (Andrew Gelman and Modrák 2020), because the model accurately captures the observed data patterns. This is good because it indicates that our logistic regression model is a good representation of the actual poverty status in the 2019 poverty data from United States Census Bureau.

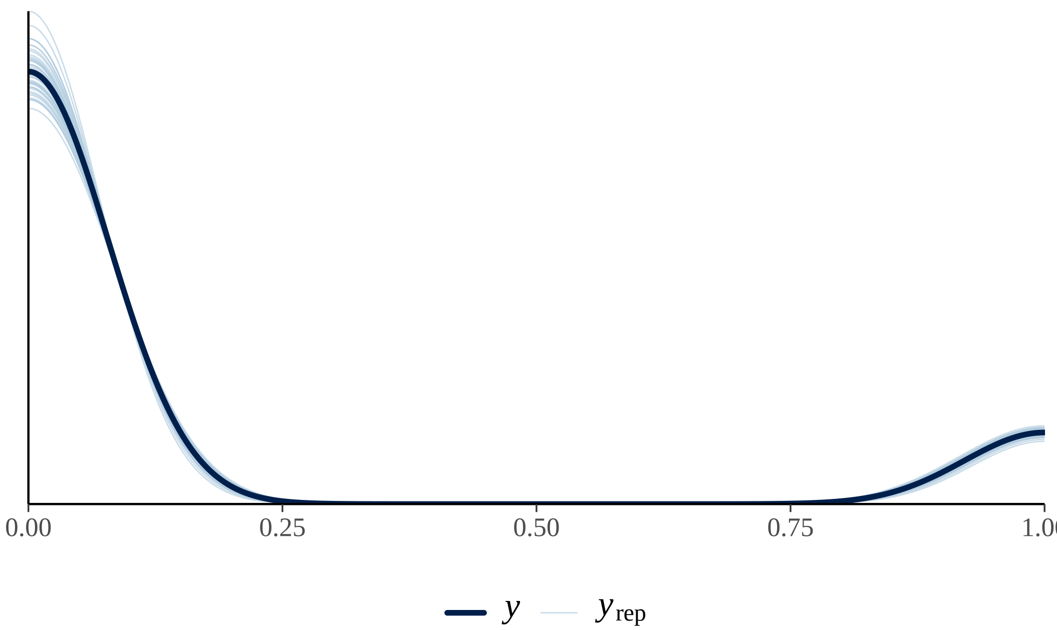


Figure 3: Posterior distribution for logistic regression model

Figure 4 compares the posterior with the prior. This compares the prior distribution of parameters with the posterior distribution of parameters in our logistic regression model. We can see that half of the model parameters do not change after data are taken into account, while some parameters shift slightly. This shows that the observed data partially matches with our initial belief and expectation about the poverty status of people in the United States in 2019. We can see that for people with income level at “below 5k” and “5k-50k” and “Renters”, the posterior distributions shift from their prior after we input observed data; their distribution not crossing 0 at all. This is suggesting that the observed data for "below 5k" and "5k-50k" strongly contradict our initial belief. So the majority of people in US who earns less than a total amount of 50k per household annually was in poverty in 2019. Also, the intercept (people who are owners with mortgage and earn 100-150k) shifts to the left; its distribution not crossing 0. This indicates that the actual observation does not match with our prior belief. People in this category was less likely to be in poverty state in 2019 United States.

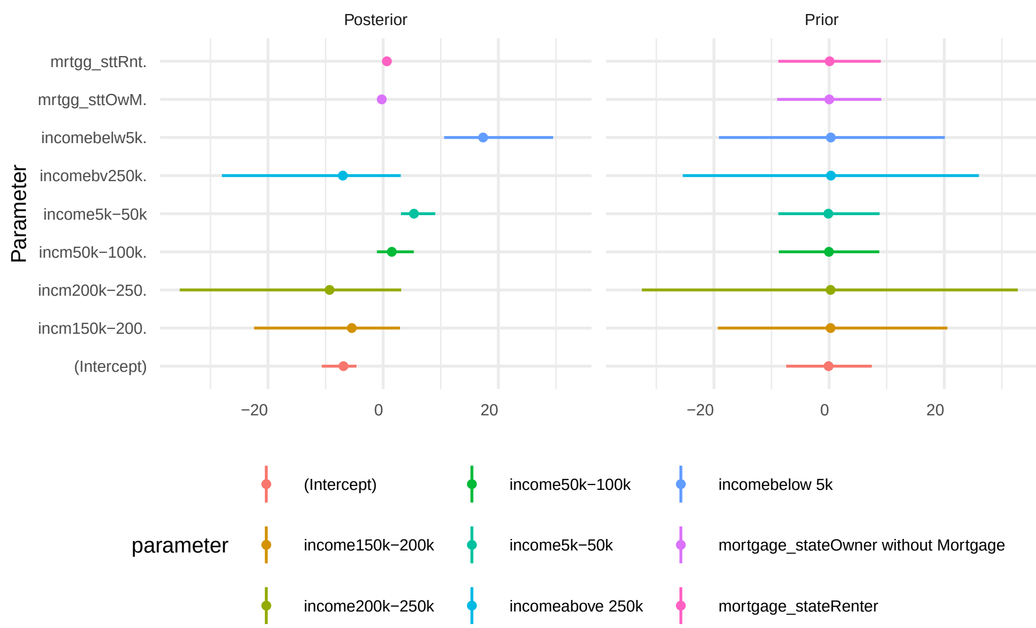
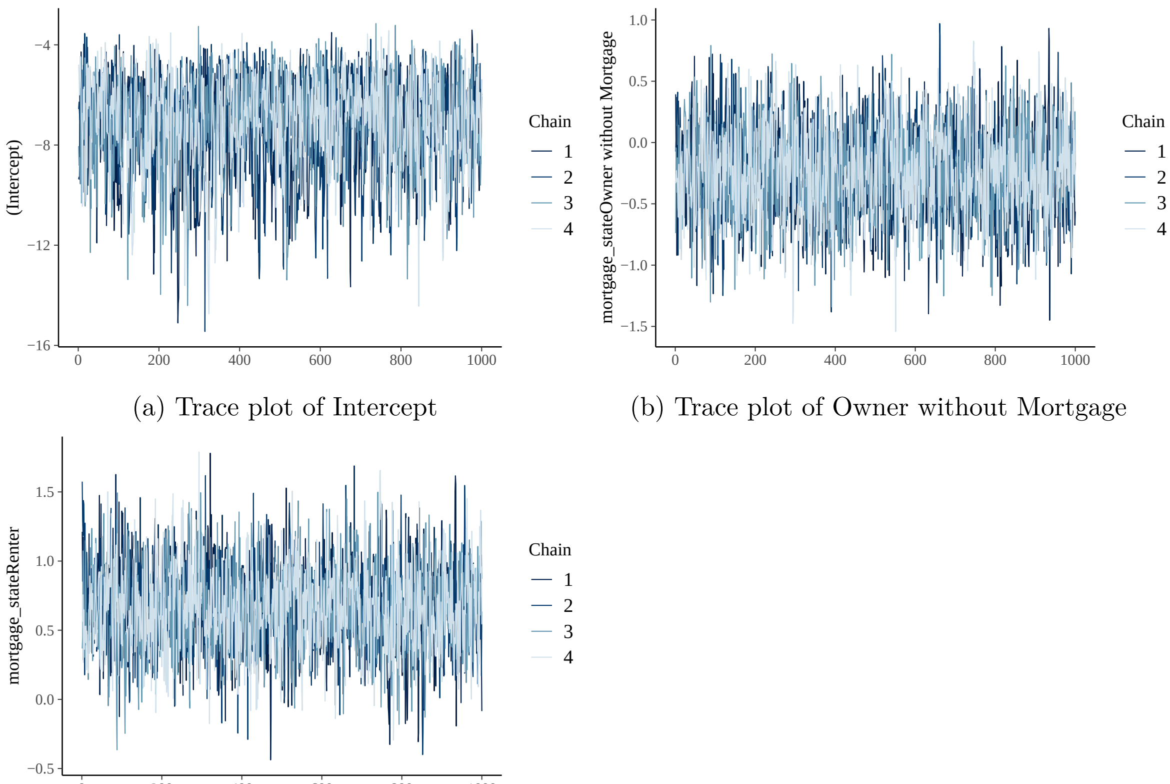


Figure 4: Comparing the posterior with the prior

A.2 Markov chain Monte Carlo Convergence Check

Figure 5 and Figure 6 are the trace plots of the model. It tells us if there is existence of signs that the our model runs into issues. We observe lines in all the trace plots are horizontal and oscillating, and have overlaps between the chains. This suggests that there is nothing strange in this trace plot.

Figure 7 is the Rhat plot of the model. It compares the variability within each chain to the variability between chains in MCMC. We can observe that our Rhat plot are all close to 1, and no more than 1.05. This is a good sign because it suggests that the MCMC algorithm has reached convergence for our model.

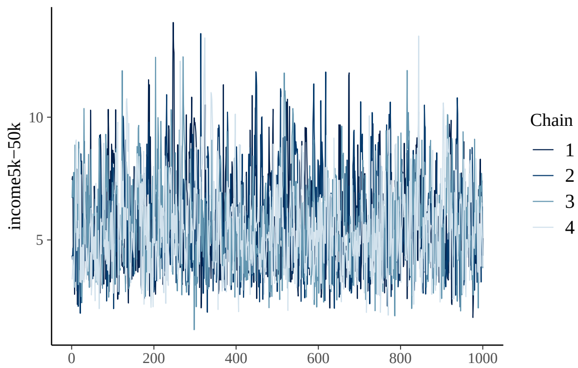


(a) Trace plot of Intercept

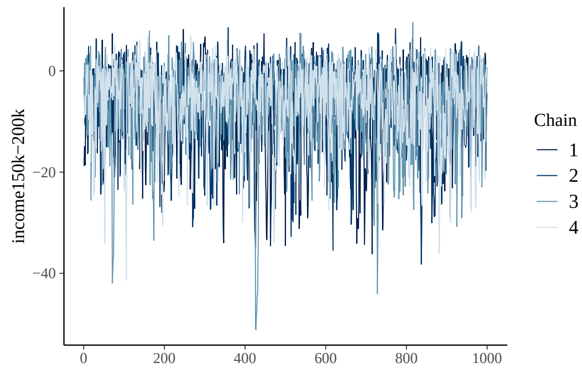
(b) Trace plot of Owner without Mortgage

(c) Trace plot of Renter

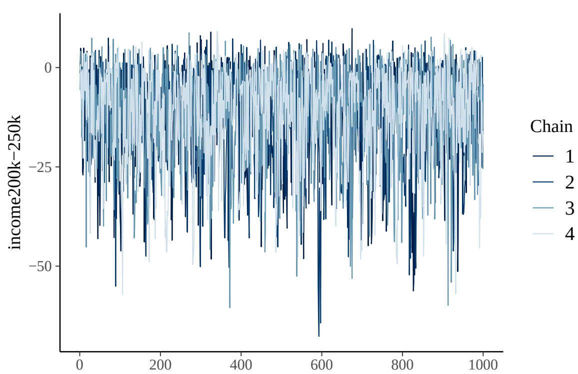
Figure 5: Trace plot of intercept and marital status



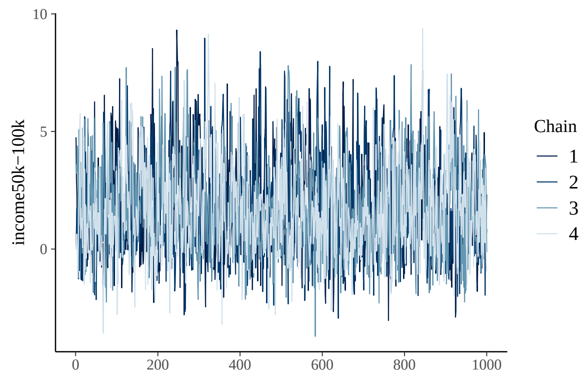
(a) Trace plot of income 5k-50k



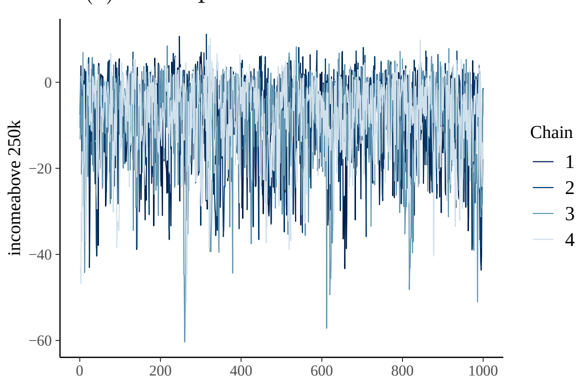
(b) Trace plot of income 150k-200k



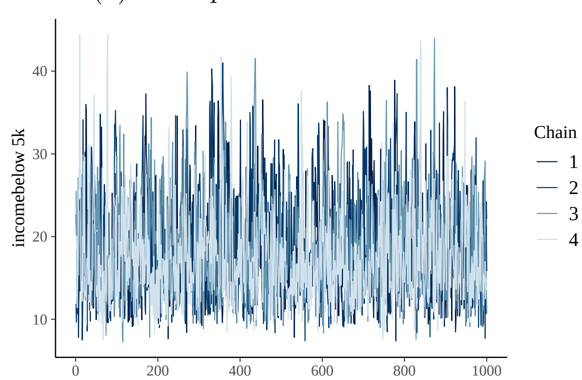
(c) Trace plot of income 200k-250k



(d) Trace plot of income 50k-100k



(e) Trace plot of income above 250k



(f) Trace plot of income below 5k

Figure 6: Trace plot of income

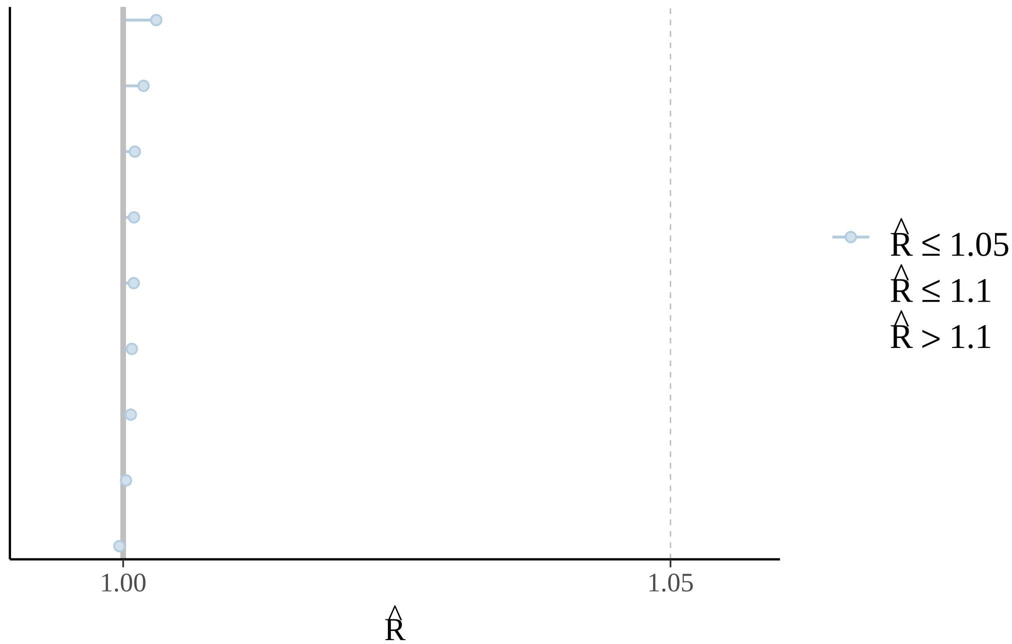


Figure 7: Rhat plot

A.3 90% Credibility Interval

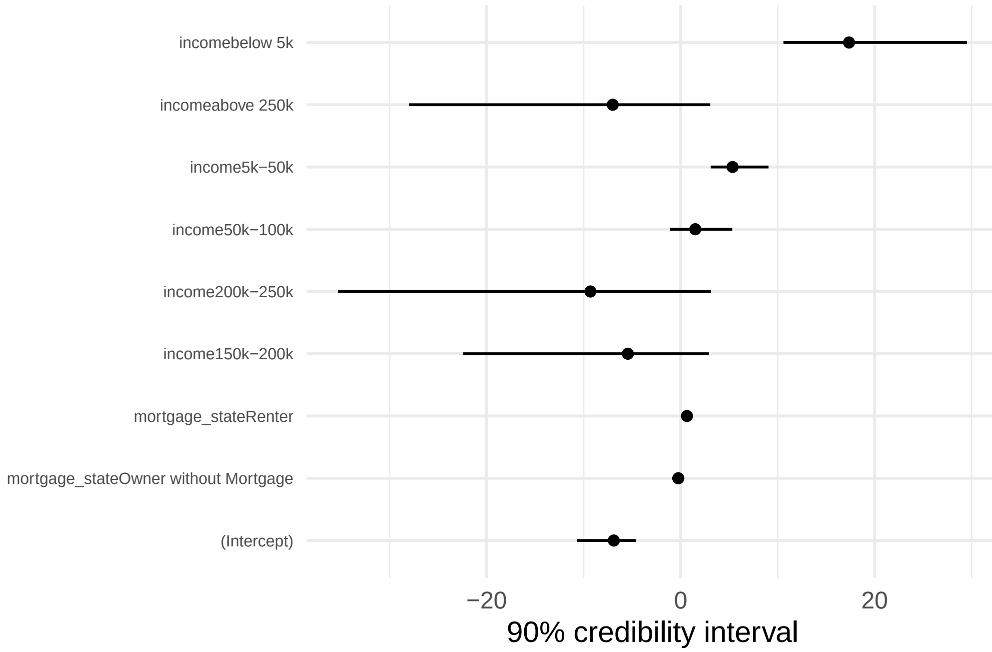


Figure 8: Credible intervals for predictors of positive poverty status

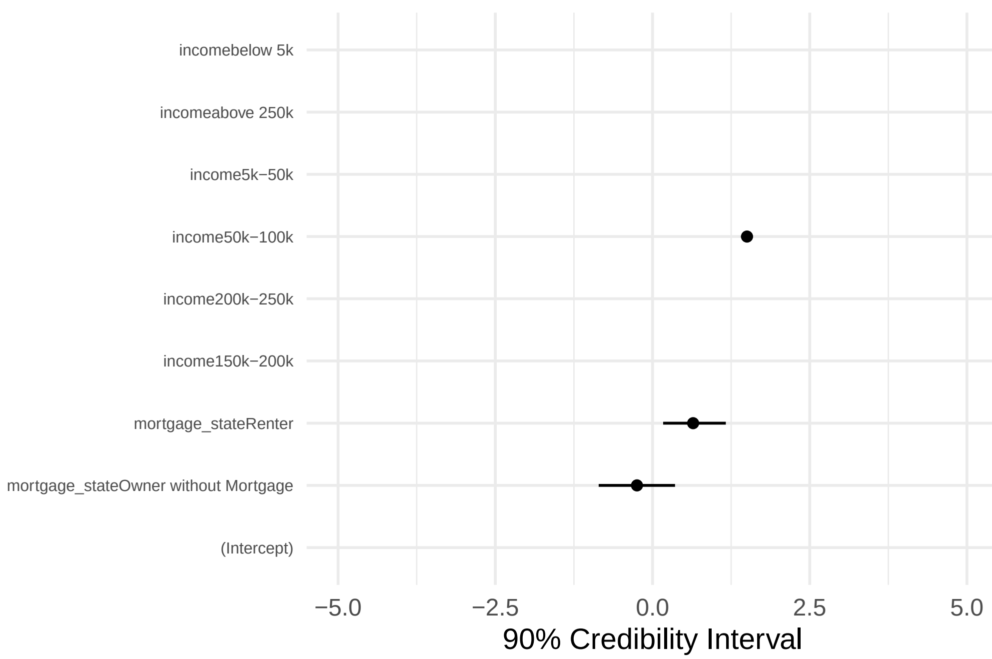


Figure 9: Credible intervals for predictors of positive poverty status with x_axis limits

References

Andrew Gelman, Daniel Simpson, Aki Vehtari, and Martin Modrák. 2020. "Bayesian Workflow." https://doi.org/10.48550/arXiv.2011.01808.

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.

———. 2024. *Marginaleffects: Predictions, Comparisons, Slopes, Marginal Means, and Hypothesis Tests*. https://CRAN.R-project.org/package=marginaleffects.

Bridges, Benjamin, and Robert V.Gesumaria. 2015. "Social Security Administration." *Social Security Administration Research, Statistics, and Policy Analysis*. https://www.ssa.gov/policy/docs/ssb/v75n2/v75n2p57.html.

Brilleman, SL, MJ Crowther, M Moreno-Betancur, J Buros Novik, and R Wolfe. 2018. "Joint Longitudinal and Time-to-Event Models via Stan." https://github.com/stan-dev/stancon_talks/.

Bureau, US Census. 2022. "2019 Revised Supplemental Poverty Measure Research File." *Census.gov*. https://www.census.gov/data/datasets/2019/demo/supplemental-poverty-measure/revised-research.html.

———. 2023. "About the Bureau." *Census.gov*. https://www.census.gov/about.html.

Kearney, Melissa S., and Benjamin H. Harris. 2014. "Policies to Address Poverty in America." *Policies to Address Poverty in America*. https://www.brookings.edu/wp-content/uploads/2016/07/policies_address_poverty_in_america_full_book.pdf.

Leeper, Thomas J. 2021. *Margins: Marginal Effects for Model Objects*.

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/.

Semega, Jessica. 2022. "Income and Poverty in the United States: 2019." *Census.gov*. https://www.census.gov/library/publications/2020/demo/p60-270.html.

Sievert, Carson. 2020. *Interactive Web-Based Data Visualization with r, Plotly, and Shiny*. Chapman; Hall/CRC. https://plotly-r.com.

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, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https:

//CRAN.R-project.org/package=kableExtra.