Understanding the Impact of Marital Status, Income, and Age on Individuals' Poverty Status*

An Analysis of the 2019 Revised Supplemental Poverty Measure Dataset

Mary Cheng

March 29, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

1 Introduction

You can and should cross-reference sections and sub-sections. We use R Core Team (2023) and Wickham et al. (2019).

The remainder of this paper is structured as follows. Section 2....

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

spm_poor	spm_hmaritalstatus	spm_totval	spm_hage
0	1	127449	63
0	1	127449	63
0	1	64680	64
0	1	64680	64
0	5	40002	54

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

poverty_status	marital_status	income	age
Not in poverty	Divorced	10k-50k	45-54
In poverty	Divorced	below 10k	above 65
In poverty	Divorced	below 10k	above 65
Not in poverty	Married - civilian spouse present	50k-100k	45-54
Not in poverty	Married - civilian spouse present	50k-100k	45-54

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

	poverty_status	marital_status	income	age
	Not in poverty:134359	Married - civilian spouse present :98147	below 10k : 3893	25-34 :24053
	In poverty : 14672	Married - armed forces spouse present : 700	10k-50k :37537	35-44 :39156
	NA	Married - spouse absent (excluding separated): 1961	50k-100k :44384	45-54 :32017
	NA	Widowed : 8130	100k-150k :27865	55-64 :25153
	NA	Divorced :16797	150k-200k :15621	above 65:28652
	NA	Separated : 3184	200k-250k : 8071	NA
	NA	Never Married :20112	above 250k:11660	NA

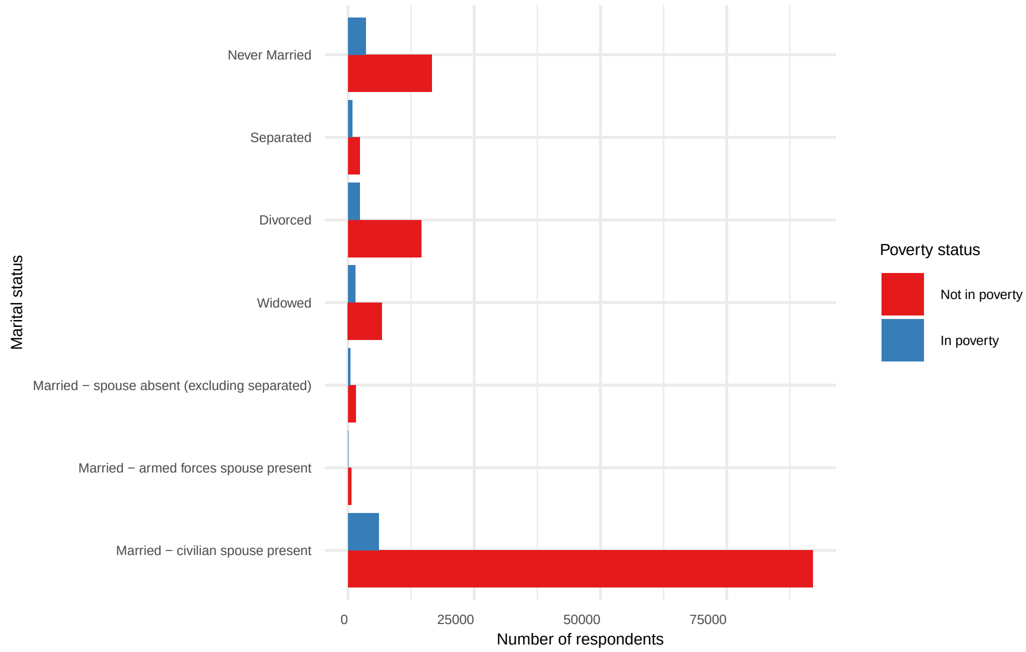


Figure 1: The distribution of poverty status by marital status

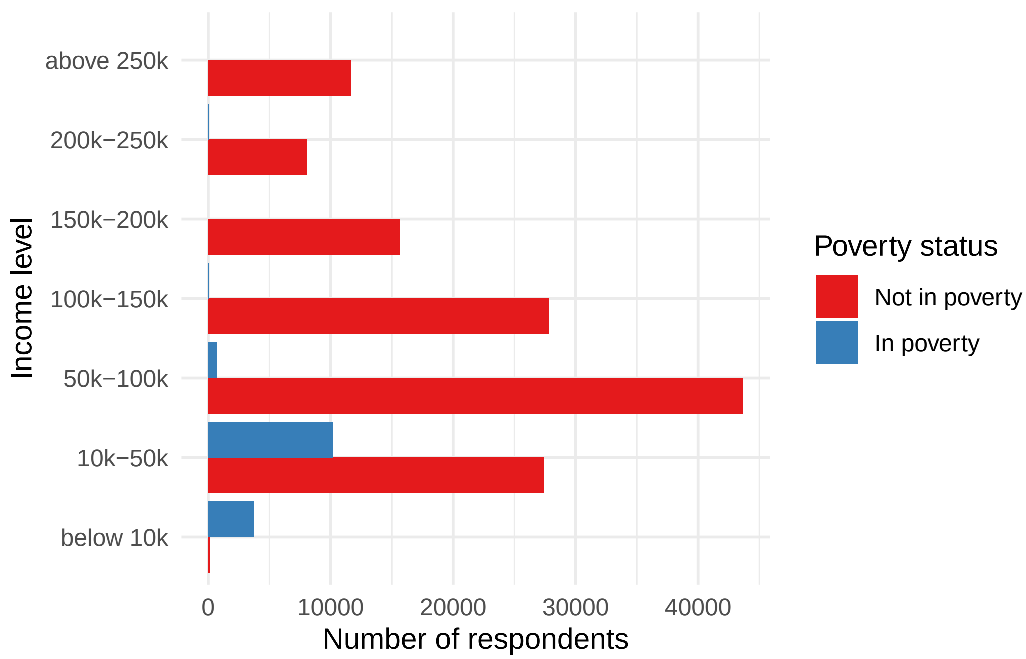


Figure 2: The distribution of poverty status by income level

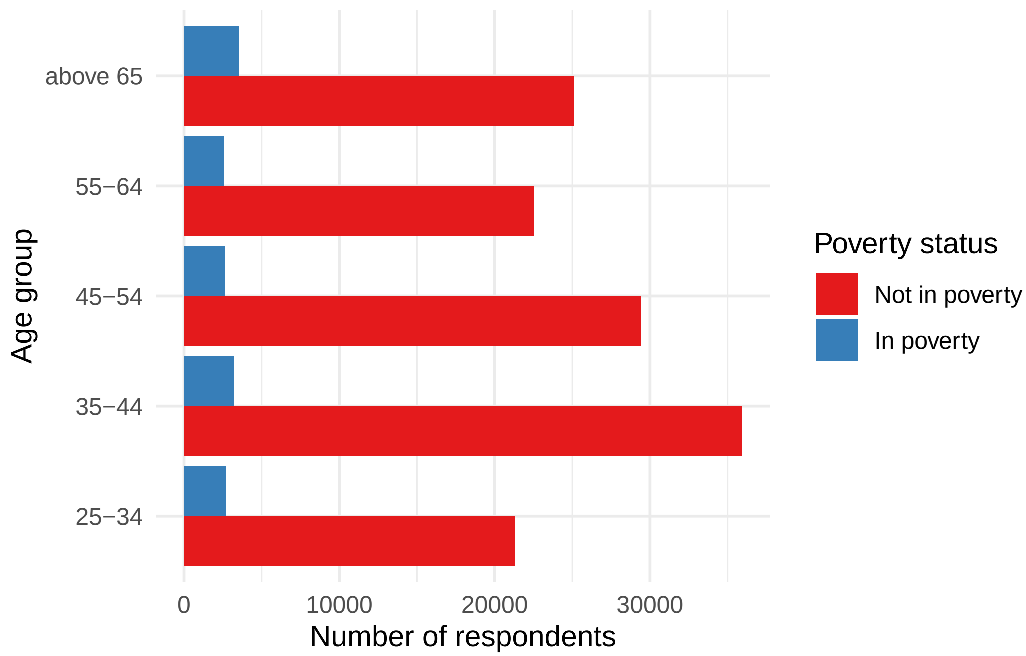


Figure 3: The distribution of poverty status by age group

2 Data

2.1 Data Source

2.2 Features

2.3 Data Measurement

2.4 Methodology

2.5 Data Visualization

3 Model

The goal of our modelling strategy is twofold. Firstly,...

Here we briefly describe the Bayesian analysis model used to investigate... Background details and diagnostics are included in Appendix A.

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

3.1 Model set-up

Define y_i as the number of seconds that the plane remained aloft. Then β_i is the wing width and γ_i is the wing length, both measured in millimeters.

$$
\begin{aligned}
y_i|\pi_i &\sim \text{Bern}(\pi_i) && (1)\\
\text{logit}(\pi_i) &= \alpha + \beta_1 \times \text{mortgage}_i + \beta_2 \times \text{income}_i + \beta_3 \times \text{age}_i && (2)\\
\alpha &\sim \text{Normal}(0, 2.5) && (3)\\
\beta_1 &\sim \text{Normal}(0, 2.5) && (4)\\
\beta_2 &\sim \text{Normal}(0, 2.5) && (5)
\end{aligned}
$$

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

3.1.1 Model justification

We expect a positive relationship between the size of the wings and time spent aloft. In particular…

We can use maths by including latex between dollar signs, for instance θ.

4 Results

Our results are summarized in Table 4.

5 Discussion

5.1 First discussion point

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.

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

	In poverty
(Intercept)	−7.983
	(1.767)
marital_statusMarried - armed forces spouse present	−19.004
	(20.879)
marital_statusMarried - civilian spouse present	0.956
	(0.412)
marital_statusMarried - spouse absent (excluding separated)	0.225
	(0.936)
marital_statusNever Married	1.336
	(0.477)
marital_statusSeparated	1.485
	(0.727)
marital_statusWidowed	1.025
	(0.578)
income10k-50k	6.323
	(1.699)
income150k-200k	−2.334
	(4.195)
income200k-250k	−3.443
	(5.337)
income50k-100k	1.841
	(1.740)
incomeabove 250k	−2.786
	(4.654)
incomebelow 10k	21.414
	(9.381)
age35-44	−0.183
	(0.458)
age45-54	−0.048
	(0.499)
age55-64	−0.110
	(0.490)
ageabove 65	0.208
	(0.464)
Num.Obs.	1000
R2	0.432
Log.Lik.	−160.502
ELPD	−173.0
ELPD s.e.	13.4
LOOIC	345.9
LOOIC s.e.	26.8
WAIC	345.6
RMSE	0.23

5.2 Second discussion point

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Model details

A.1 Posterior predictive check

In Figure 4, 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 (**bayes_post?**), 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.

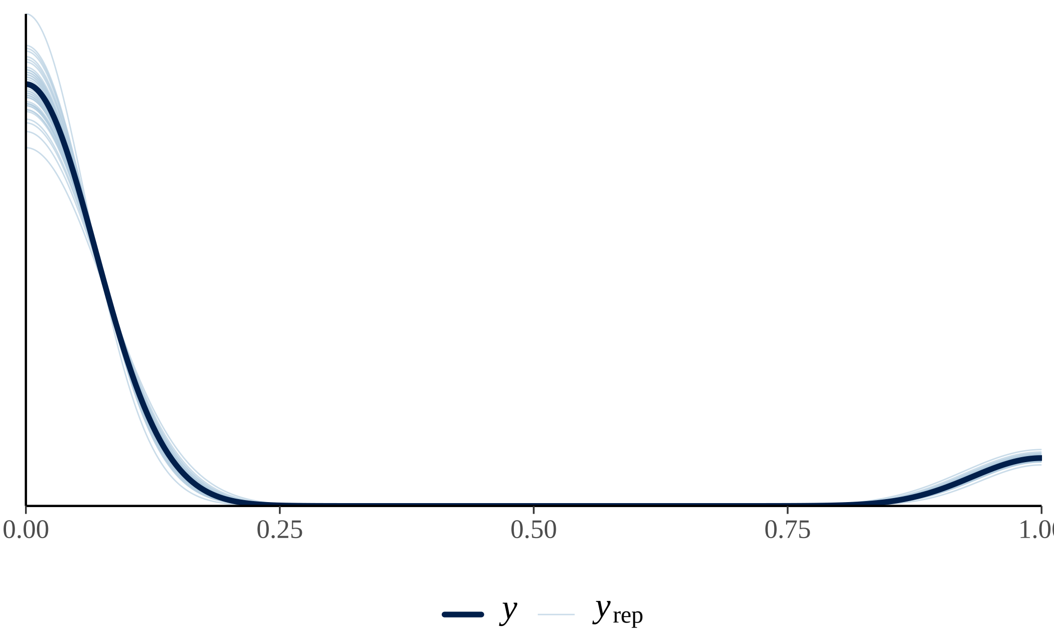


Figure 4: Posterior distribution for logistic regression model

Figure 5 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 10k", the posterior distribution greatly shifts from its prior after we input observed data. This is likely suggesting that the observed data for “below 10k” strongly contradicts our initial belief. This suggests that the majority of people in US who earns less than a total amount of 10k per household annually is in poverty in 2019.

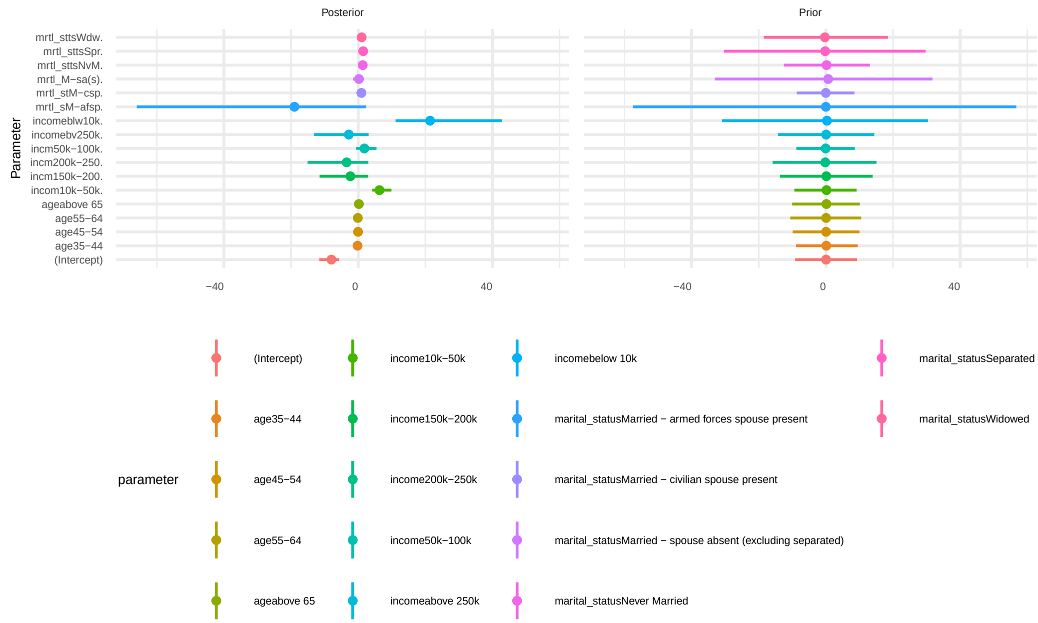
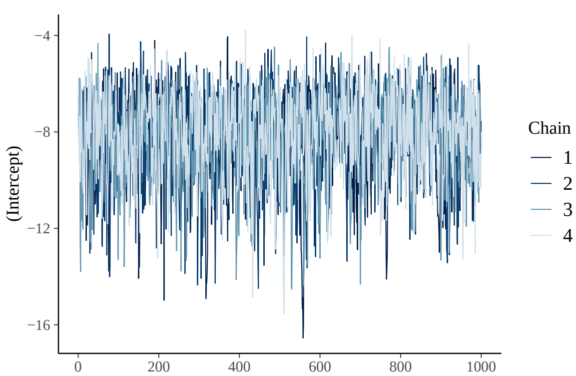


Figure 5: Comparing the posterior with the prior

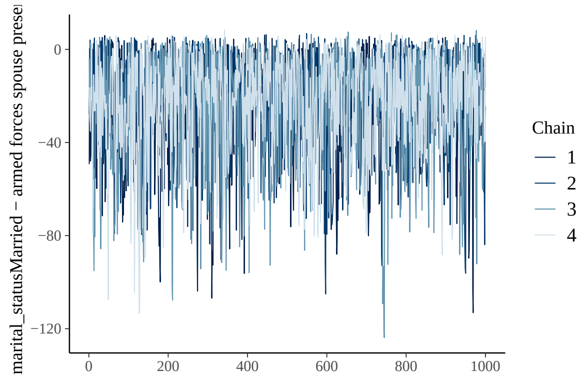
A.2 Markov chain Monte Carlo Convergence Check

Figure 6, Figure 7, Figure 8 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.

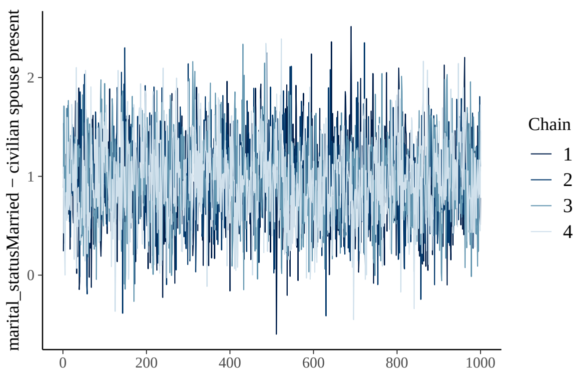
Figure 9 is the Rhat plot of the model. It compares the variability within each chain to the variability between chains in MCMC. We can observe that our Rhat plot are all close to 1, and no more than 1.05. This is a good sign because it suggests that the MCMC algorithm has reached convergence for our model.



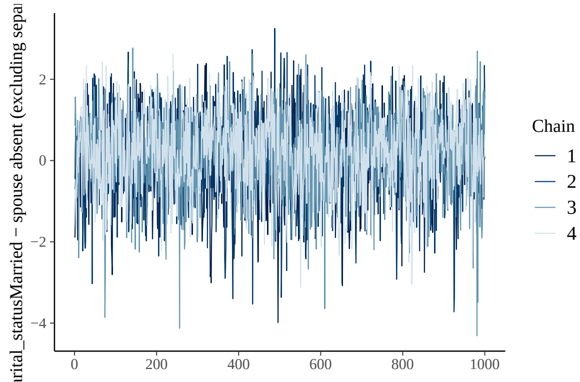
(a) Trace plot of Intercept



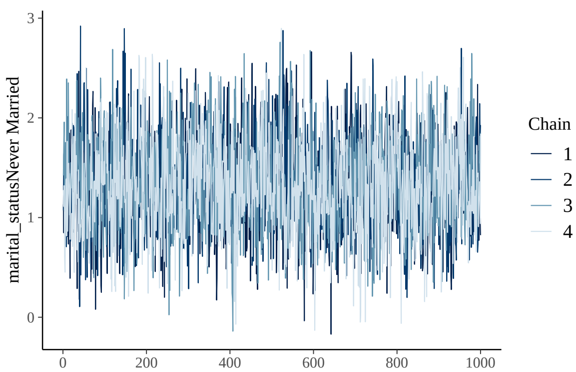
(b) Trace plot of Married - civilian spouse present



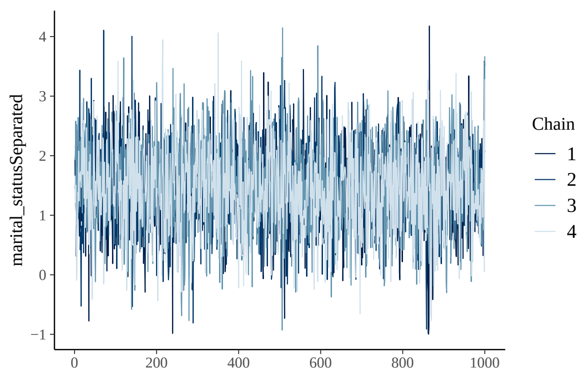
(c) Trace plot of Married - armed forces spouse present



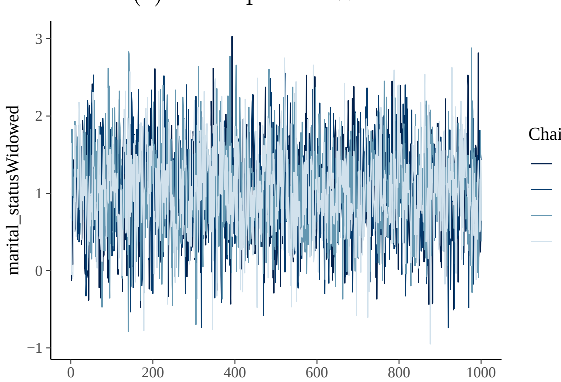
(d) Trace plot of Married - spouse absent (excluding separated)



(e) Trace plot of Widowed

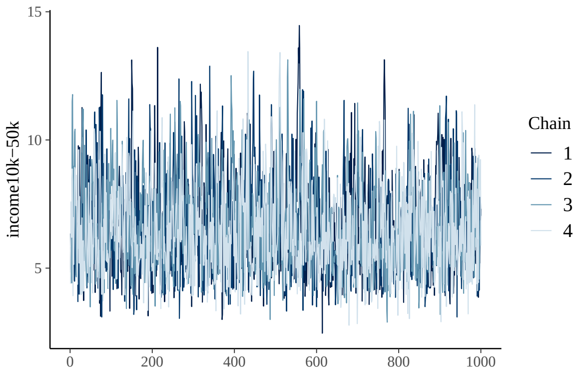


(f) Trace plot of Separated

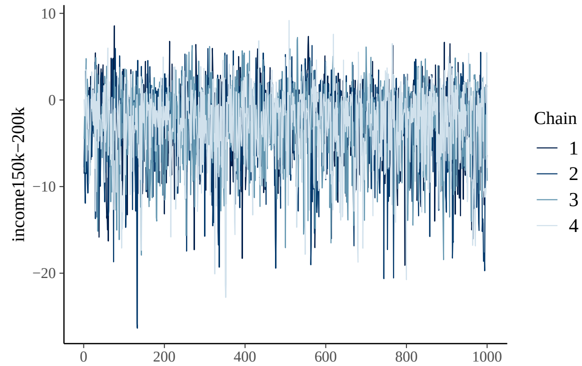


(g) Trace plot of Never Married

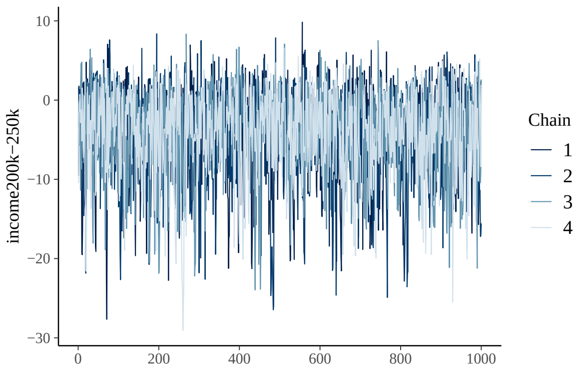
Figure 6: Trace plot of intercept and marital status



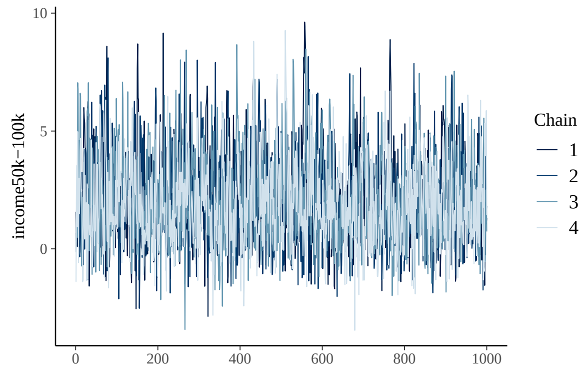
(a) Trace plot of income 10k-50k



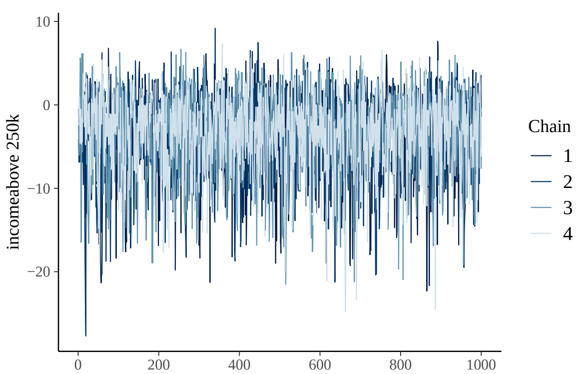
(b) Trace plot of income 150k-200k



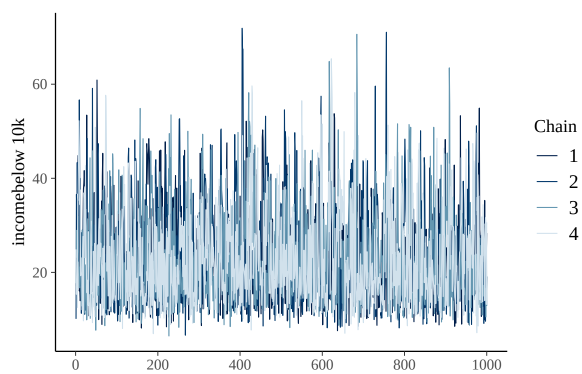
(c) Trace plot of income 200k-250k



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

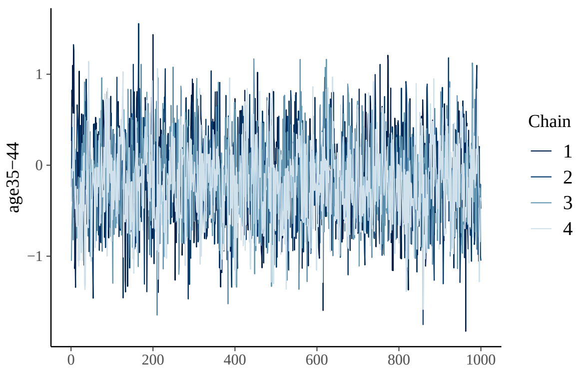


(e) Trace plot of income above 250k

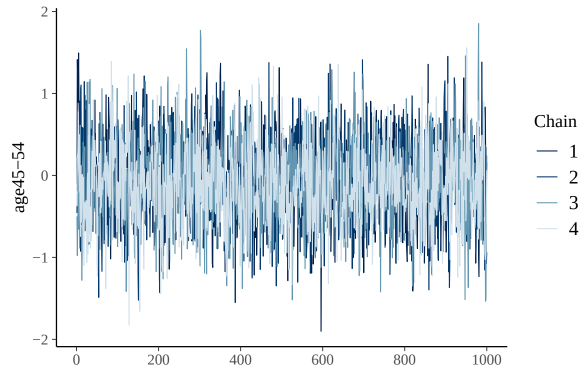


(f) Trace plot of income below 10k

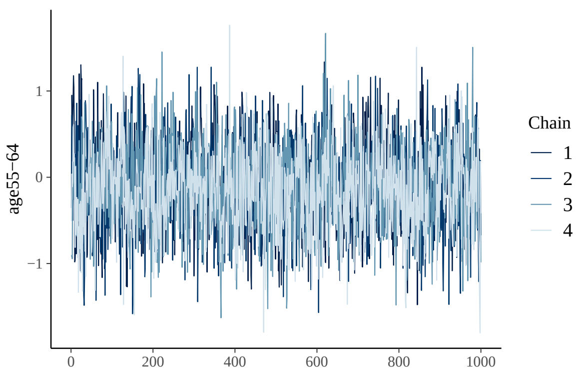
Figure 7: Trace plot of income



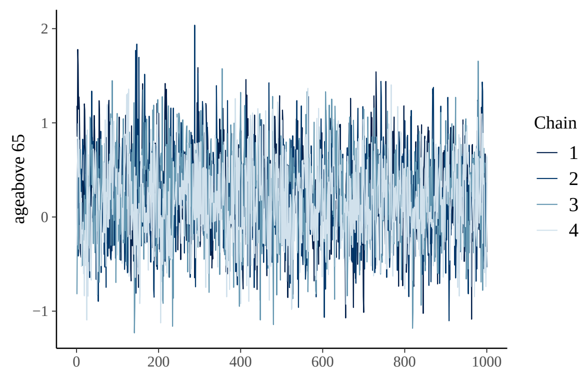
(a) Trace plot of age 35-44



(b) Trace plot of age 45-54



(c) Trace plot of age 55-64



(d) Trace plot of age above 65

Figure 8: Trace plot of employ

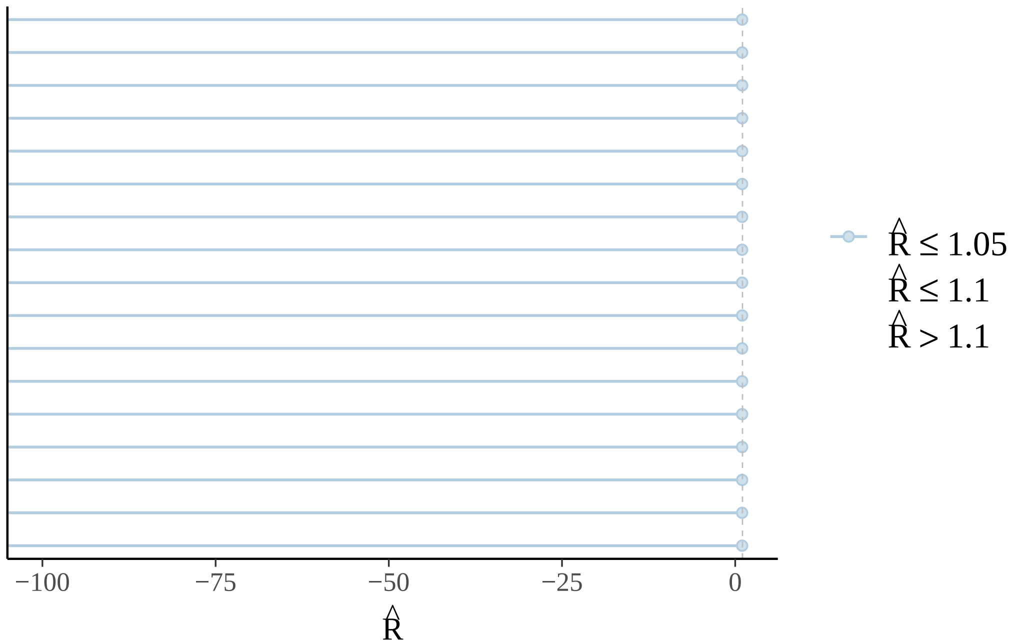


Figure 9: Rhat plot

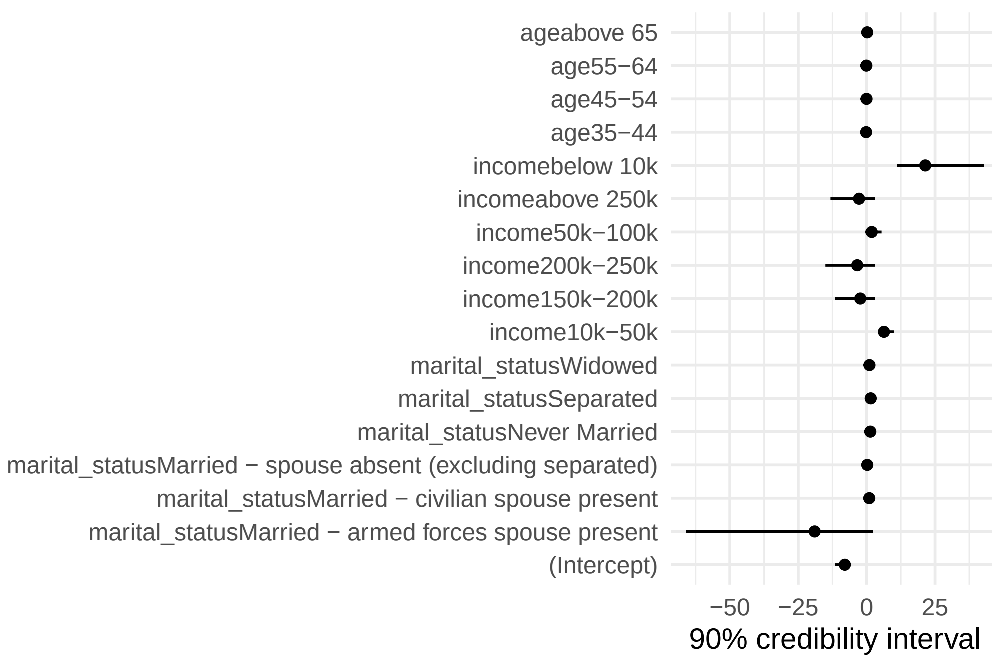


Figure 10: Credible intervals for predictors of support for Biden

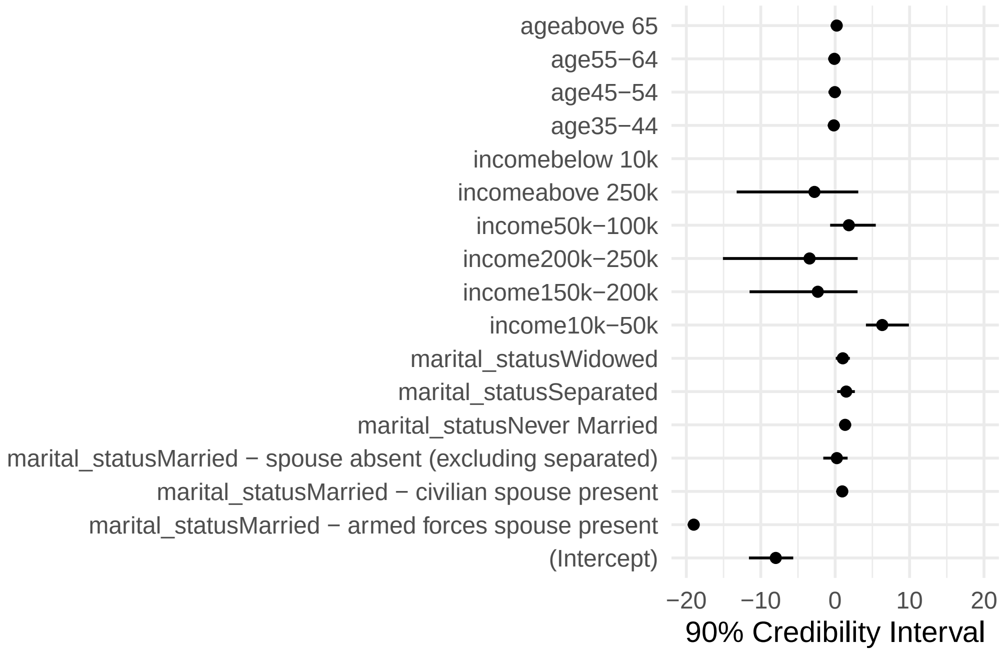


Figure 11: Credible intervals for predictors of support for Biden with x_axis limits

References

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

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