Bayesian Data Analysis Project

Andrei Iliescu
Miguel Arroyo Marquez

December 15, 2025

Contents

1 Introduction

Education is a major factor that influences the growth and development of a country. The development of the United States is deeply influenced by education. Policy makers frequently advocate for increasing college attainment rates as a universal strategy to combat poverty. However, U.S. cannot be considered as a whole, it is not a uniform entity. A bachelor's degree may result in different economic returns in the industrial Midwest compared to the coastal South. If the relationship between education and poverty varies significantly by geography, a "one-size-fits-all" national policy may be inefficient. This project's motivation is to determine if the relationship between higher education (bachelor's degrees) and poverty is consistent nationwide or geographically distinct.
The central problem of this analysis is to quantify the relationship between educational attainment (specifically bachelor's degrees) and poverty rates at the state level, while accounting for factors such as unemployment and population density. Specifically, we ask: Does the impact of education on poverty reduction remain constant across the country, or does it vary by state? Our project relies on data from the 2010 US Census [2]. We selected the year 2010 because it represents a complete count of the population, providing more reliable data than the sample-based surveys used in other years.

For this project, we analyze the relationship between education and poverty by fitting two models to the same dataset: a pooled regression model and a hierarchical model. Both models control for unemployment and population density, allowing us to isolate the association between educational attainment and poverty rates.
The pooled model assumes one national relationship between education and poverty, treating all states as identical. The hierarchical model does not. It allows for both intercepts and education slopes to vary by state. This way we can prove whether or not there is a higher level of effect of education on poverty state by state.

All models are estimated using Bayesian inference via *brms* and *Stan*, which provide a coherent framework for uncertainty quantification and partial pooling.

2 Data Description

2.1 Data Source

The data used in this project come from the *county_complete* dataset included in the *openintro* [1] R package. These data compile demographic and socioeconomic indicators published by the United States Census Bureau for the year 2010. We specifically rely on county-level measures from the 2010 Decennial Census[2], a once-per decade survey in which the federal government attempts to count every resident. Because Decennial Census values are based on a full population count, rather than statistical sampling, they provide the most reliable and noise-free county-level poverty measurements available.

We believe that this data is more representative than the commonly used American Community Survey (ACS), which produces annual sample-based estimates of poverty, education, and income. ACS data include sampling error and smaller effective sample sizes for many rural counties. These properties motivate the Bayesian Zero-One Inflated Beta models used in Census Bureau internal poverty estimation work (Wieczorek & Hawala, 2011) [3]. However, our dataset avoids these complications entirely: the 2010 Census measures used here represent complete counts, eliminating sampling noise and simplifying the likelihood.

Thus, our analysis is not a small-area estimation problem. Instead, it is an *explanatory*

[1] `https://www.openintro.org`

modeling problem aimed at understanding how the relationship between higher education and poverty varies geographically.

2.2 What the Dataset Measures: Understanding "Poverty"

The variable *poverty_2010* represents the percentage of each county's population classified as living below the Official Poverty Measure (OPM) threshold in 2010. Under the OPM, a person is counted as poor if their family's pre-tax cash income falls below a threshold that depends on family size and composition. Because the threshold is uniform across the United States and does not adjust for regional cost of living differences our outcome variable reflects income poverty in a strict federal sense, but not necessarily material deprivation after adjusting for local prices.

Importantly:

- The measurement is *not* restricted to children (unlike some Census Bureau small-area poverty models, which focus on ages 5–17).
- It is a *headcount rate*: the percentage of the population below the poverty line.
- Because it is derived from a full survey, it contains less sampling variability and less margins of error.

This makes the poverty measure stable and clean for multilevel regression, while also differentiating our work from prior Bayesian studies designed to model ACS uncertainty rather than structural socioeconomic relationships.

2.3 Variables and Data Structure

After removing rows with missing values and ensuring that the log-transformed variables remained finite, our analytical dataset contains 3,139 counties nested within 51 geographic units (the 50 states plus the District of Columbia).

The key variables used in our analysis are:

- **Outcome Variable (y):**
  - *poverty_2010*: Percentage of all individuals in the county whose household income falls below the OPM poverty threshold.
- **Predictor Variables (x):**
  - *bachelors_2010*: Percent of adults with a bachelor's degree or higher.
  - *unemployment_rate_2010*: County-level unemployment rate.
  - *density_2010*: Population density (people per square mile).

This structure is fundamentally hierarchical: counties are the units of observation, but they inherit economic, policy, and cultural characteristics from their states. A model that treats counties as independent ignores this nested dependence. By recognizing the multi-level structure, we are able to investigate not only how education relates to poverty nationwide, but whether the strength of this relationship differs across states.

2.4 Data Transformation

To support efficient Bayesian computation and interpretable priors, all predictor variables and the outcome were standardized using z-score scaling. Population density was first transformed using log(density + 1) to accommodate counties with zero recorded density. Standardization places all variables on a comparable scale and produces posterior geometries that facilitate Hamiltonian Monte Carlo sampling.

```r
df_edu <- df_edu %>%
  mutate(
    log_density_raw = log(density_2010 + 1)   # +1 handles zero densities
  ) %>%
  filter(is.finite(log_density_raw))

df_edu <- df_edu %>%
  mutate(
    poverty_scaled = as.numeric(scale(poverty_2010)),
    edu_scaled     = as.numeric(scale(bachelors_2010)),
    unemp_scaled   = as.numeric(scale(unemployment_rate_2010)),
    dens_scaled    = as.numeric(scale(log_density_raw))
  )
```

Listing 1: Data transformation

2.5 Preliminary Evidence for State-by-State variation

Before fitting any Bayesian models, we performed an exploratory analysis to assess whether education appears to reduce poverty at the same rate in every state. Figure 5 shows regression lines fit within six randomly sampled states. Visually, the slopes vary in magnitude, suggesting that the return to higher education may depend strongly on state level context. This motivates the use of a hierarchical model with state specific education effects.

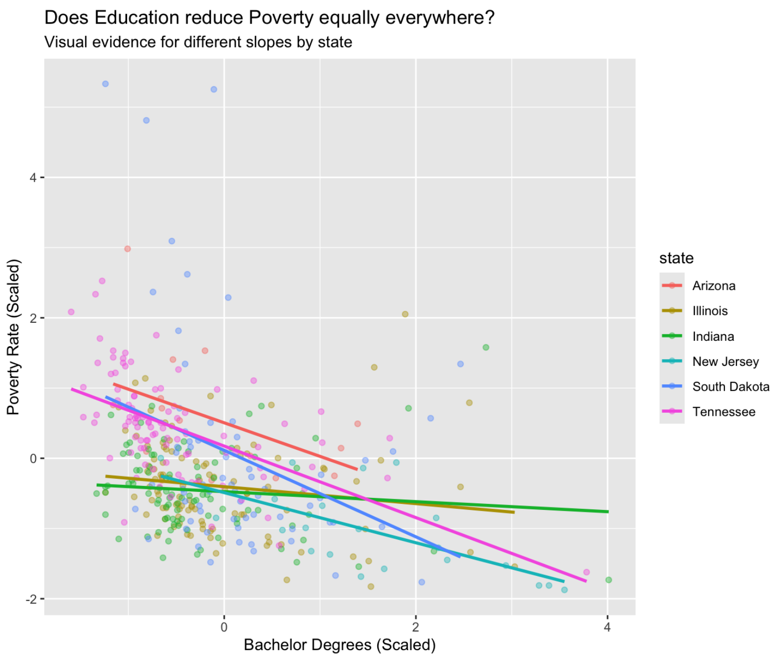

Figure 1: Variation in education–poverty slopes across a random sample of six U.S. states. Lines produced by preliminary state-specific regressions demonstrate clear differences in the magnitude of the education coefficient.

3 Description of Models

Both models assume that the outcome variable, standardized poverty (y_i), follows a Gaussian distribution. The difference lies in how they treat the state feature, the aim of our experiment.

Let y_i be the standardized poverty rate for county i. Let x_{edu}, x_{unemp}, and x_{dens} be the standardized predictors for education, unemployment, and log-density, respectively.

3.1 Model 1: The Pooled Model (Baseline)

The first model is a complete pooling approach. It ignores state boundaries entirely, assuming that the relationship between education and poverty is identical across the entire United States. This serves as a baseline to see if adding geographic complexity is actually necessary.

$$
\begin{aligned}
&y_i \sim \mathcal{N}(\mu_i, \sigma) \\
&\mu_i = \alpha + \beta_{\text{edu}} x_{\text{edu},i} + \beta_{\text{unemp}} x_{\text{unemp},i} + \beta_{\text{dens}} x_{\text{dens},i}
\end{aligned}
$$

For the above:

- α is the global intercept (average poverty when all predictors are average).
- β_{edu}, β_{unemp} and β_{dens} are the slopes describing how education, unemployment and density affects poverty nationwide.
- σ represents the residual variation (noise) in poverty that the model cannot explain.

3.2 Model 2: The Hierarchical Model

The second model is a hierarchical model. We assume that while unemployment and density have fixed effects nationwide, the effect of education and the baseline poverty levels vary by state. This captures the hypothesis that a college degree might be more valuable in some economies than others.

We map each county i to a specific state j using the index $j[i]$.

$$
\begin{aligned}
&y_i \sim \mathcal{N}(\mu_i, \sigma) \\
&\mu_i = \alpha_{j[i]} + \beta_{\text{edu},j[i]} x_{\text{edu},i} + \beta_{\text{unemp}} x_{\text{unemp},i} + \beta_{\text{dens}} x_{\text{dens},i}
\end{aligned}
$$

Notice that the intercept α and the education slope β_{edu} now have the subscript $j[i]$. This means every state gets its own regression line.

The Intercept $\alpha_{j[i]}$ allows baseline poverty to change by state but we left Unemployment and Density fixed to keep the model stable and focused on our goals. This is as adding state by state variation would add overhead and possible divergence issues for states with few counties (data points).

Crucially, we do not estimate these state effects independently (which would be overfitting, especially for small states). Instead, we model them as coming from a joint population distribution:

$$
\begin{pmatrix} \alpha_j \\ \beta_{\text{edu},j} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_\alpha \\ \mu_{\beta_{\text{edu}}} \end{pmatrix}, \mathbf{S} \right)
$$

Here, the vector containing μ_α and $\mu_{\beta_{\text{edu}}}$ represents the global average intercept and slope. The matrix $\mathbf{S}$ is the covariance matrix that determines:

1. How much poverty varies between states.

2. How much the “return on education” varies between states.

3. The correlation between baseline poverty and the education effect.

This structure allows for partial pooling: states with little data borrow strength from the national average, while states with much data are allowed to deviate more from the mean.

4 Priors and Justification

All variables in the model were standardized to mean 0 and standard deviation 1 and we utilized weakly informative priors. These priors are designed to provide regularization, keeping estimates within physically plausible bounds without imposing strong subjective beliefs on the exact values.

4.1 Weakly informative priors

- **Intercept:** We assigned a $\mathcal{N}(0, 1)$ prior. Since the outcome variable is standardized, we expect the global average to be close to 0. A standard deviation of 1 allows the intercept to plausibly range between -2 and +2 standard deviations.

- **Regression Coefficients:** We assigned a $\mathcal{N}(0, 0.5)$ prior to the fixed effects and the global education effect. This implies that a 1 standard deviation increase in a predictor is unlikely to change poverty by more than 1 standard deviation. This acts as regularization to prevent overfitting.

- **Scale Parameters:** For the residual standard deviation and the group-level standard deviations, we used an Exponential(1) prior. This constrains the values to be positive and places more probability mass on smaller standard deviations.

- **Correlation Matrix:** We used an LKJ(2) prior for the correlation between the random intercept and the random slope. By setting $\eta = 2$, we weakly penalize extreme correlations (near -1 or 1), slightly favoring a correlation of 0.

```r
priors_pooled <- c(
  prior(normal(0, 1),    class = "Intercept"),
  prior(normal(0, 0.5), class = "b"),
  prior(exponential(1), class = "sigma")
)

priors_hier <- c(
  prior(normal(0, 1),    class = "Intercept"),
  prior(normal(0, 0.5), class = "b"),
  prior(exponential(1), class = "sigma"),
  prior(exponential(1), class = "sd"),
  prior(lkj(2),          class = "cor")
)
```

Listing 2: Prior definition

4.2 Cynical priors

On top of this choice of priors we have also tested some cynical priors. This is a specific test designed to be skeptical meaning that: "If I assume education does nothing, will the data convince me otherwise?". Therefore, instead of the $\mathcal{N}(0, 0.5)$ distribution assigned to the education regression coefficient, we changed it to $\mathcal{N}(0, 0.05)$ for the cynical prior. The meaning behind this is "I am 95% certain the effect of education is basically zero (between -0.1 and 0.1)."

```r

priors_cynic <- c(
  ... # Same as hierarchical
  prior(normal(0, 0.05), class = "b", coef = "edu_scaled"),
  ... # Same as hierarchical
)
```

Listing 3: Cynical Prior definition

5 Inference and Convergence Diagnostics

5.1 Execution

The models were fitted using the *brms* package in R. The sampling was run with 4 chains, 2000 iterations each (1000 warmup).

```r
fit_pooled <- brm(
  formula   = poverty_scaled ~ 1 + edu_scaled + unemp_scaled + dens_scaled,
  data      = df_edu,
  prior     = priors_pooled,
  family    = gaussian(),
  chains    = 4,
  iter      = 2000,
  warmup    = 1000,
  control   = list(adapt_delta = 0.95),
  seed      = 2025
)

fit_hier <- brm(
  formula   = poverty_scaled ~ 1 + edu_scaled + unemp_scaled + dens_scaled + (1
    + edu_scaled | state),
  data      = df_edu,
  prior     = priors_hier,
  family    = gaussian(),
  chains    = 4,
  iter      = 2000,
  warmup    = 1000,
  control   = list(adapt_delta = 0.95),
  seed      = 2025
)
```

Listing 4: brms Code Example

5.2 Convergence

We assessed convergence for all Bayesian models using the standard diagnostics recommended for Hamiltonian Monte Carlo (HMC) and the No-U Turn Sampler (NUTS). Specifically, we monitored three indicators: the potential scale reduction statistic $\hat{R}$, the effective sample size (ESS), and divergent transitions. Across all models, including the pooled model, the hierarchical model with weakly informative priors, and the hierarchical model with the skeptical ("cynical") prior, we found clear evidence of successful convergence.

5.2.1 R-hat ($\hat{R}$) and Trace Plots

The $\hat{R}$ statistic compares between-chain variance to within-chain variance. Values close to 1.00 indicate that all chains are sampling from the same posterior distribution.

```
> rhat_values <- brms::rhat(fit_hier)
> print(any(rhat_values > 1.01))
[1] FALSE
> rhat_values <- brms::rhat(fit_pooled)
> print(any(rhat_values > 1.01))
[1] FALSE
> rhat_values <- brms::rhat(fit_cynic)
> print(any(rhat_values > 1.01))
[1] FALSE
```

Figure 2: R-hat check for each model

For every parameter in every model, all $\hat{R}$ values were ≤ 1.01, with the vast majority equal to 1.00, indicating excellent mixing and no signs of non-convergence. This includes the fixed effects (e.g., the education slope, unemployment slope, and density slope), as well as the hierarchical variance components such as the state level random intercept standard deviation and the random-slope standard deviation.

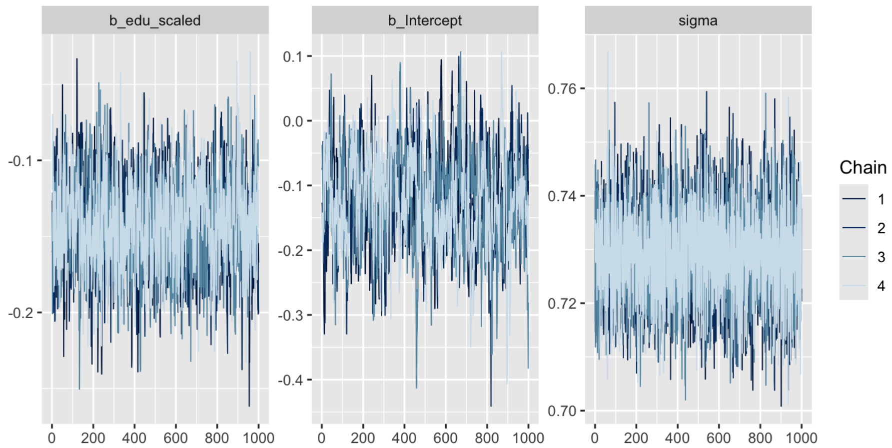

Figure 3: Trace Plots for fit_hier model

For the plots seen above we can observe good mixing between the four chains with tight overlaps, within the same region and similar variability between them. The plots seem stable with frequent jumps and it covers the full range. All which match up with our $\hat{R}$ values.

5.2.2 Effective Sample Size (ESS).

To assess the reliability of the posterior summaries, we evaluated both the Bulk ESS (which reflects the precision of mean and variance estimates) and the Tail ESS (which reflects the precision of quantile estimates). Across all three models.

The pooled model, the hierarchical model with weakly informative priors, and the hierarchical model with the skeptical prior—both ESS metrics comfortably exceeded the commonly recommended minimum of 400. This indicates that the chains explored the posterior distribution efficiently and without pathological autocorrelation.

In the main hierarchical model, for example, the education slope achieved a Bulk ESS of 1313 and a Tail ESS of 2435, while the residual standard deviation σ reached a Bulk ESS of 8245. The group-level standard deviations and the intercept–slope correlation also displayed high ESS values, confirming that the multilevel components were sampled with sufficient precision. The full ESS values for all parameters are summarized in Table 1.

Model	Parameter	Bulk ESS	Tail ESS
Hierarchical Model (Main Prior)			
	sd(Intercept)	756	1455
	sd(edu_scaled)	1490	2577
	cor(Intercept, edu_scaled)	1769	2383
	Intercept	435	989
	edu_scaled	1313	2435
	unemp_scaled	3381	3283
	dens_scaled	3546	2965
	sigma	8245	3026
Hierarchical Model (Cynical Prior)			
	sd(Intercept)	616	1202
	sd(edu_scaled)	1603	2700
	cor(Intercept, edu_scaled)	1423	2183
	Intercept	398	750
	edu_scaled	1543	1851
	unemp_scaled	3127	3150
	dens_scaled	3881	3232
	sigma	7931	2870
Pooled Model			
	Intercept	3803	2790
	edu_scaled	2263	2658
	unemp_scaled	2315	2510
	dens_scaled	2216	2271
	sigma	3727	2960

Table 1: Bulk and tail effective sample sizes (ESS) for all models. All values exceed the commonly used threshold of 400, indicating efficient MCMC sampling.

5.3 Divergences

Divergent transitions are a key indicator of problems in the geometry of the posterior that can invalidate estimates. All models reported *zero* divergent transitions following warmup. We set `adapt_delta = 0.95` to ensure stable sampling, and this proved sufficient for all models, no further increases were necessary.

6 Model Checking

6.1 Posterior Predictive Checks

Posterior predictive checks (PPCs) assess the model's descriptive accuracy by comparing the observed data (y) to synthetic data (y_{rep}) generated by the fitted model. If the model is a good approximation of reality, it should be able to generate synthetic data that follows the same distribution as the observed data.

We utilized density overlay plots to visually compare the distribution of observed poverty rates against 50 simulated datasets drawn from the posterior predictive distribution.

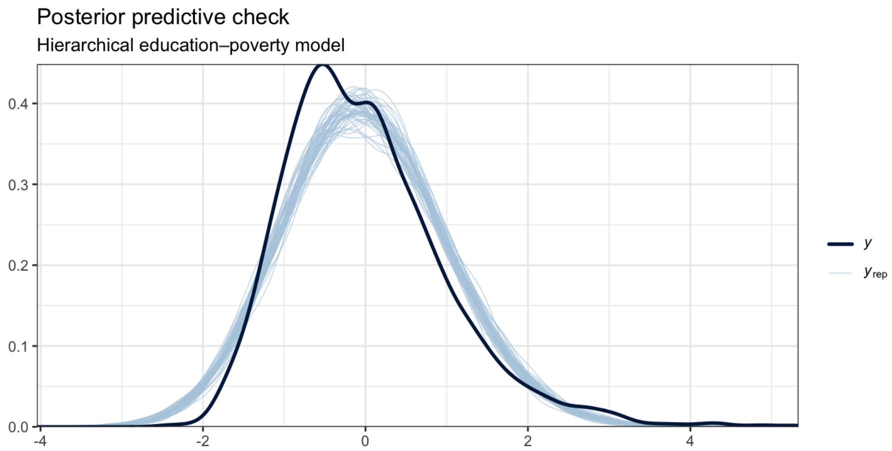

Figure 4: Comparison of the observed data distribution (y) with 50 posterior predictive simulations (y_{rep}).

The visual inspection in Figure 4 indicates a satisfactory fit. The simulated densities (y_{rep}) closely tracked the mode, mean, and tails of the observed poverty data (y). The overlap suggests that the Gaussian likelihood function was an appropriate choice for the standardized response variable.

6.2 Sensitivity Analysis

To ensure our conclusions were not driven solely by subjective prior choices, we performed a sensitivity analysis using a "Cynical Prior." We compared our main model against a model designed to be skeptical of any relationship between education and poverty. The posterior estimates for the global effect of education were:

- **Main Model:** -0.14 (95% CI: $[-0.20, -0.09]$)
- **Cynical Model:** -0.11 (95% CI: $[-0.16, -0.05]$)

While the skeptical prior did shrink the coefficient toward zero (from -0.14 to -0.11), the posterior distribution remained clearly negative and the credible interval excluded zero. This demonstrates that the negative association between education and poverty is significant. The signal in the data is strong enough to overcome even a skeptical prior.

7 Model Comparison

To evaluate whether education truly has a geographically varying effect on poverty, we compared the pooled and hierarchical models using out-of-sample predictive performance. Specifically, we used approximate Leave-One-Out cross-validation (LOO-CV) based on Pareto-smoothed importance sampling.

7.1 LOO-CV Results

Table 2 reports the LOO-CV estimates for both models. Higher elpd_{loo} (Expected Log Predictive Density) indicates better predictive performance.

Model	elpd_loo	SE(elpd_loo)	p_loo	looic
Hierarchical	-3504.8	67.9	81.0	7009.6
Pooled	-3933.1	62.5	7.3	7866.1

Table 2: LOO-CV results for pooled vs. hierarchical models.

The hierarchical model achieves a substantially higher elpd_loo than the pooled model. The difference in expected predictive performance is summarized in Table 3.

Model	elpd_diff	SE(elpd_diff)
Hierarchical	0.0	0.0
Pooled	-428.3	27.1

Table 3: LOO-CV differences in ELPD relative to the hierarchical model (reference).

The pooled model is worse by approximately 428 elpd units with a standard error of 27.1. The ratio

$$\frac{|\text{elpd\_diff}|}{\text{SE(elpd\_diff)}} \approx \frac{428.3}{27.1} \approx 15.8$$

corresponds to a difference of about 16 standard errors, which is overwhelmingly large. In practical terms, this means the hierarchical model is not just slightly better, but decisively superior in terms of out-of-sample predictive accuracy.

All Pareto k diagnostics were below the conventional threshold of 0.7 for both models, indicating that the importance sampling approximation underpinning the LOO-CV is reliable.

7.2 Substantive Interpretation

From a substantive point of view, these results provide strong evidence that the relationship between education and poverty is not adequately described by a single national slope. The hierarchical model's better predictive performance indicates that allowing each state to have its own intercept and education effect is necessary to capture the underlying structure of the data. In other words, a bachelor's degree does not reduce poverty by the same amount in every state.

8 Discussion and Conclusion

Our analysis shows a clear negative association between educational attainment and state level poverty in the United States. In the hierarchical model, a one standard deviation increase in the bachelor's degree rate is associated with roughly a 0.14 standard deviation decrease in poverty, holding unemployment and population density constant. This effect remains negative even under a skeptical prior that strongly weights down any impact of education (posterior mean ≈ -0.11), indicating that the signal is driven by the data rather than by our prior assumptions.

At the same time, the hierarchical model reveals that this relationship is far from uniform across space. Some states, such as Kentucky, South Dakota, Georgia and New Mexico, exhibit strongly negative state specific slopes, indicating that higher education is protective against poverty there. Others, including Indiana, Wyoming and Iowa, have slopes that are close to zero with wide intervals, suggesting that additional education is only weakly associated with lower poverty in those contexts. The strong preference for the hierarchical model in the LOO-CV comparison (an elpd difference of about 428 with a standard error of 27) confirms that allowing state specific intercepts and education effects is statistically essential for describing the data. Substantively, these results support a nuanced version of the "education as opportunity" narrative. On average, states with higher bachelor's attainment have lower poverty, even after

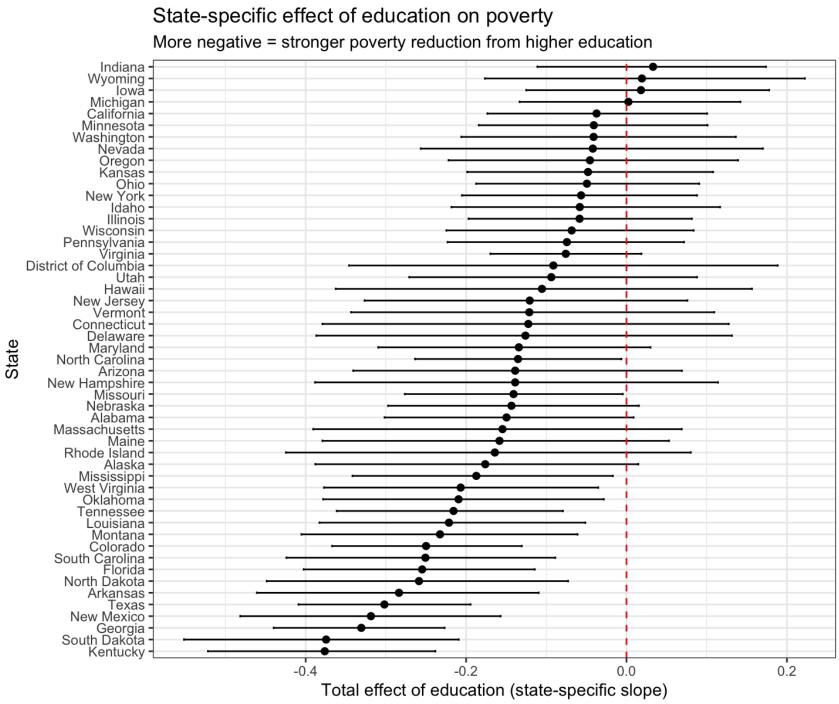

Figure 5: Whisker plot of poverty per state

accounting for unemployment and population density. But the size of that benefit depends heavily on where people live. This is consistent with broader empirical work showing that returns to schooling vary across regions and institutional contexts, and that local labor market conditions shape how strongly education translates into improved economic outcomes [1]. From a policy perspective, our findings suggest that national targets for raising educational attainment should be complemented with state or region specific strategies that address local labor demand, industrial structure, and complementary policies.

8.1 Illustrative Counterfactual: Increasing Education in Alabama

One advantage of the Bayesian hierarchical model is that it can be used to answer policy-style "what if" questions. As a simple illustration, we asked: *What would happen to poverty in Alabama if the county-level bachelor's degree rate were one standard deviation higher, holding all else equal?*

Using the fitted hierarchical model, we constructed a counterfactual dataset in which the predictor `edu_scaled` is increased by 1 for all counties in Alabama, and then generated posterior predictive draws for this hypothetical scenario:

```
# 1. Define the counterfactual (the "intervention"):
#    Current Alabama data, but with higher education
hypothetical_alabama <- df_edu %>%
  filter(state == "Alabama") %>%
  mutate(edu_scaled = edu_scaled + 1)  # add 1 SD of education

# 2. Predict poverty under the counterfactual
preds <- posterior_predict(fit_hier, newdata = hypothetical_alabama)

# 3. Calculate the impact by comparing to current reality
diff <- preds - df_edu$poverty_scaled[df_edu$state == "Alabama"]
mean(diff)   # expected change in poverty (in SD units)
```

Listing 5: Counterfactual simulation for Alabama

The quantity `mean(diff)` summarizes the expected reduction in standardized poverty if Alabama counties were one standard deviation more educated, according to our model. While we do not treat this as a fully causal estimate (because of potential confounding and the cross-sectional design), it illustrates how the hierarchical model can be used to explore and compare hypothetical policy interventions across states.

8.2 Limitations

Several limitations of our study are worth noting. First, the analysis is purely cross-sectional and restricted to 2010. We cannot distinguish long-run causal effects of education because of unobserved factors such as local industry mix, migration, or state policy regimes. Second, our outcome is based on the Official Poverty Measure, which does not adjust for regional cost of living. Third, we assume linear, additive effects and a Gaussian likelihood for the standardized outcome. While posterior predictive checks suggest this is appropriate for the marginal distribution of poverty, more flexible specifications might uncover patterns we miss. Finally, we allow only education to have state-specific slopes, keeping unemployment and density as fixed effects. This choice reflects a desire to focus on our main question, but it imposes structure that may be too restrictive in some contexts.

8.3 Potential Extensions

Our project opens up several natural directions for future work:

- **Non-linear education effects.** We model the education–poverty relationship as strictly linear. However, both economic theory and empirical work suggest that the marginal returns to additional education may decrease once a county reaches high levels of attainment, or may follow an inverted U shape. [4] A straightforward extension would be to include a quadratic term (edu^2) or a spline for education in the hierarchical model. This would allow us to test whether the first few percentage points of bachelor's attainment yield larger poverty reductions than later increments and to visualize how the slope changes across the attainment distribution.

- **Multilevel structure.** We could allow unemployment and log-density to have state-specific slopes as well, or include additional group levels. This would let us ask whether the impact of joblessness on poverty is itself moderated by state institutions, and whether education interacts with local labor market.

In summary, our study demonstrates that education is consistently associated with lower poverty across U.S. counties, but the magnitude of this association varies sharply by state. Hierarchical modeling not only improves predictive performance but also reveals this geography of opportunity, highlighting where additional education appears to be most and least effective as an anti poverty strategy.

9 Group Self-Reflection

Working on this project showed us that the hardest part of Bayesian analysis is often setting up the data correctly before even running the model. We learned that simple steps, like scaling our variables, made a huge difference in whether the code worked or failed. Finally, experimenting with the "cynical prior" gave us confidence that our results were driven by actual data, not just our initial guesses.

AI Usage: We used ChatGPT to help with ideation and dataset searching. On top of that we used it for report formulation and planning.

References

[1] José De Gregorio and Jong–Wha Lee. "Education and Income Inequality: New Evidence from Cross-Country Data". In: *Review of Income and Wealth* 48.3 (2002), pp. 395–416. DOI: 10.1111/1475-4991.00060.

[2] U.S. Census Bureau. *United States Census Bureau*. Accessed: 2025-11-29. 2025. URL: https://www.census.gov/.

[3] Jerzy Wieczorek and Sam Hawala. "A bayesian zero-one inflated beta model for estimating poverty in us counties". In: *Proceedings of the American Statistical Association, Section on Survey Research Methods, Alexandria, VA: American Statistical Association*. 2011, pp. 2812–2815.

[4] Min Xu et al. "Non-linear Links Between Human Capital, Educational Inequality and Income Inequality: Evidence from China". In: *PLOS ONE* 18.8 (2023), e0288966. DOI: 10.1371/journal.pone.0288966.