STAT 775: Final Project Report

Higher spending leads to poorer education?

Group 7

Team Members:

Hari Krishna Boyapati
Suryaraj Machani
Sheng Huang
Harshita Reddy Edugudi

Introduction:

Do students perform better if a state invests more in education? In the United States, many politicians and academics have had heated debates about this. We might agree with the idea that more money spent on education leads to better school performance. After all, higher financing for education empowers institutions to access better-quality learning resources that improve student learning and, in turn, their performance on standardized examinations like SAT. The National Bureau of Economic Studies discovered that states that give higher education funding to the poorest school districts see more academic improvement in education outcomes over time than those that did not.

However In the paper "Getting What You Pay For: The Debate Over Equity in Public School Expenditures", D. Guber (Guber 1999) found that between 1994 and 1995. It was discovered that, on a state level, higher per capita public school spending is connected with poorer SAT performance. We plan to implement a Bayesian Hierarchical model using STAN to analyze the effects of per capita spending on education and other factors on the total SAT score of the eligible students.

Background and data description:

The dataset first appeared in the “Guber 1999” paper, data were collected to study the relationship between expenditures on public education and test results. Dataset is available as “sat” in “faraway” package in R. The data was chosen because it included information on the educational spending per child and school performance (SAT scores) in different states from 1994 to 1995.

The following table shows the description of the dataset.

Variable	Description
expend	Current expenditure per pupil in average daily attendance in public elementary and secondary schools, 1994-95 (in thousands of dollars)
ratio	Average pupil/teacher ratio in public elementary and secondary schools, Fall 1994
salary	The estimated average annual salary of teachers in public elementary and secondary schools, 1994-95 (in thousands of dollars)
takers	Percentage of all eligible students taking the SAT, 1994-95
verbal	Average verbal SAT score, 1994-95
math	Average math SAT score, 1994-95
total	Average total score on the SAT, 1994-95

Table 1: A description of the Guber (1999) data set

Model:

Inspired by the efficient building of the hierarchical model introduced in the lecture with 8 schools example, our group grew more interest in the flexible usage of the stan programming language. Having knowledge of mathematical theory behind a model but a lack of software expertise can make it cumbersome to apply it in real-world situations. So, we decided to learn more about how the RStan package can be used for Bayesian models and to strengthen our understanding of Bayesian statistics by going one step further. All things considered, a Bayesian hierarchical regression model is by far the best choice for our model.

In this model, we will conduct the Bayesian linear regression of SAT scores with an intercept and five features: expend(expenditure), ratio (student-teacher ratio), salary, and takers (percentage of students eligible for the SAT). Since total is the direct sum of math and verbal columns, it doesn't make sense to include math and verbal columns in the model. In contrast to normal linear regression, which merely accounts for the relationship between predictors and responses, hierarchical linear regression accounts for additional internal dependencies utilizing layered groupings, or hierarchy. Most of the time, these dependencies explain how groups relate to each other in a way that is similar to the hierarchical model we learned about in class. In this case, if the original group is students, the higher-order group is 50 states in the USA.

Hence our dependent variable, total SAT scores, is modelled as a normal distribution with a mean μ and a variance σ^2. μ can, in turn, be described by a linear combination of an intercept, expend, ratio, salary, and takers. The coefficients of each of the features are β.

$$\textbf{Total SAT Score} \sim \mathcal{N}\left(\mu, \sigma^2\right)$$
$$\mu = \beta_0 + \beta_1 \textbf{Expend} + \beta_2 \textbf{Ratio} + \beta_3 \textbf{Salary} + \beta_4 \textbf{Takers}$$

Because we do not have prior information on the effect of each predictor on the SAT score, I used weakly informative priors as defined in 'Gelman 2008'. The prior assumes that the effects of each parameter are centered around zero, with long tails. This can be represented as below:

$$\beta_i \sim \mathcal{N}\left(0, 10^6\right) \text{ for i} = 0 \text{ to } 4$$

σ^2 is assumed to have an inverse gamma distribution.

$$\sigma_\mu^2 \sim \text{Inv. Gamma}\left(\frac{\nu}{2}, \frac{\nu\lambda}{2}\right)$$

For choosing the values of ν and λ, we chose the initial value such that the value of sigma is around 70, which is close to the actual standard deviation of the observed SAT scores, So we pick ν and λ as 1 and 20 respectively. These values allow the parameter to live and even if we change the values, the model is going to work as expected. Hence, having a good idea of what residual standard deviation should be, allows us to tune the prior hyperparameters accordingly. We ran the model using "stan" and Rstudio with the default settings of a stan model of four chains and four thousand iterations, with the first 2000 iterations serving as burn-in. The following are the outcomes of running the model.

Results:

We fit the model using the stan model function from the “RSTAN” package.

```
hierarchical_model <-
  rstan::stan_model(file = "hierarchical.stan")
```

Using sampling function to generate samples for the defined parameters in .stan file.

```
fit1 <-
  rstan::sampling(object = hierarchical_model,
                  data = pooled_args)
```

By default, the rstan runs 4 Markov chains each with 2000 iterations, half of the iterations are discarded as “warm-up”. The numerical output of the hierarchical model for the “sat” sample is shown in below figure:

	n_eff	Rhat	mean	mcse	sd	2.5%	25%	50%	75%	97.5%
beta_0	1,865	1	1046.4	1.3	56	932.9	1010.4	1047.1	1082.9	1156.8
beta_1	1,410	1	4.6	0.3	10.8	-16.7	-2.4	4.7	11.8	25.5
beta_2	1,356	1	-3.6	0.1	3.4	-10.5	-5.7	-3.6	-1.5	3.1
beta_3	1,364	1	1.6	0.1	2.5	-3.4	0.1	1.6	3.1	6.4
beta_4	2,545	1	-2.9	0	0.2	-3.4	-3.1	-2.9	-2.7	-2.4
sigma_squared	2,046	1	1097.7	5.3	238.3	735	926	1068.1	1231.5	1657.1
sigma	2,098	1	32.9	0.1	3.5	27.1	30.4	32.7	35.1	40.7
log-posterior	1,101	1	-203.4	0.1	1.9	-207.9	-204.5	-203.1	-202	-200.9

Table 2: Posterior Summary Statistics

From table 2, Rhat is 1 for all the parameters and from n_eff values effective sample size is above 1000 for all the parameters. To explore the posterior further let us look at the distributions of posterior samples drawn. The below image shows the distributions of posterior samples for all the parameters.

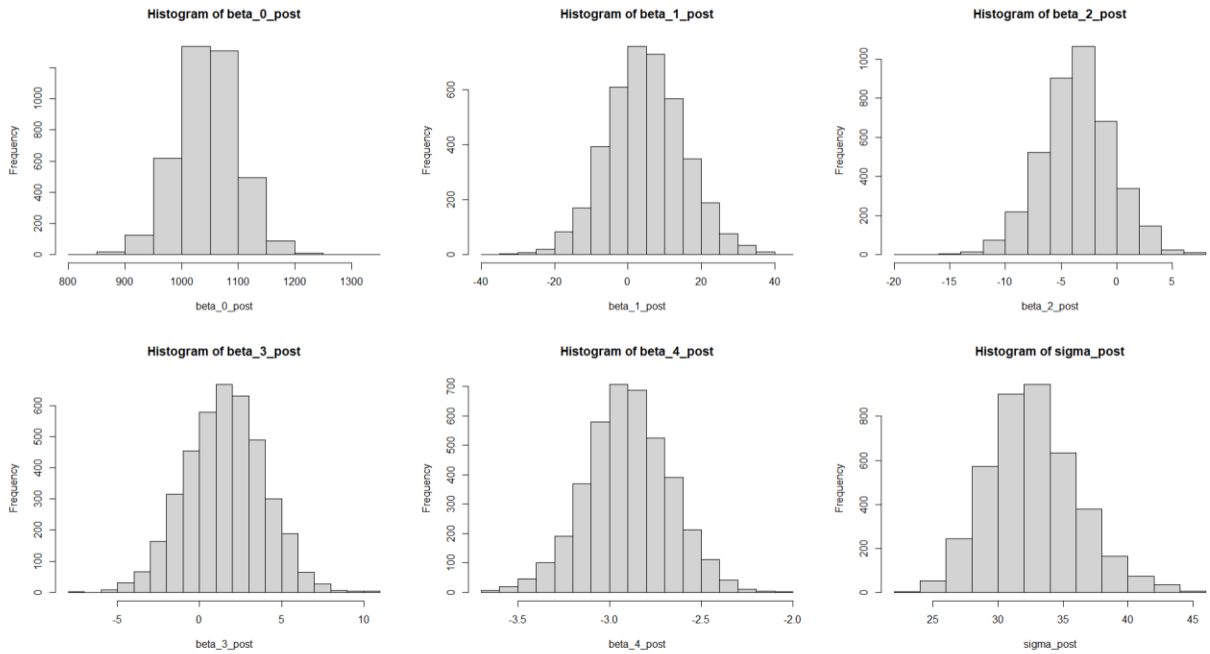

Figure 1: Distributions of posterior samples

The below image shows the convergence diagnostics for all 4 chains for the posterior parameters.

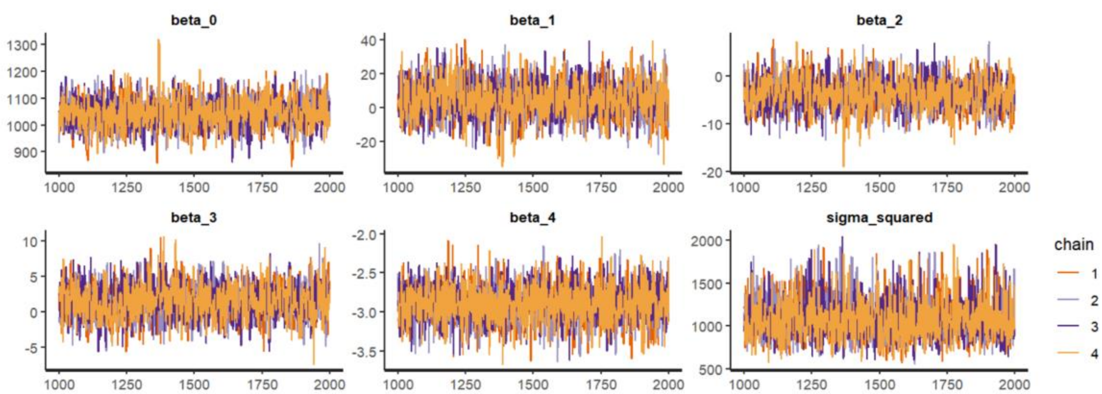

Figure 2: Convergence diagnostics for 4 chains of the posterior parameters

We use posterior intervals (PI) and posterior standard deviation point estimates of the posterior parameters (SD) as the two main ways to look at the data. A measure of central tendency (mean, median) that is chosen to summarize a posterior distribution is a point estimate, which is a single value. In the Bayesian framework, we are interested in the entire posterior distribution rather than the point estimate.

The range of the model parameters is described in probabilistic terms by posterior uncertainty intervals. The center 95% ranges for the posterior probability distribution of the parameter are indicated by the 2.5% and 97.5% columns in Table 1. We have a 95% confidence level that the parameter value falls inside the posterior 95% range. Table 1's SD column denotes the posterior standard deviation, a different way of summarizing the uncertainty surrounding the point estimate.

Using the posterior parameter values, we can generate the SAT scores for given values of expenditure, ratio, salary and takers. The below graph shows the actual observed values and fitted values for the given predictor values.

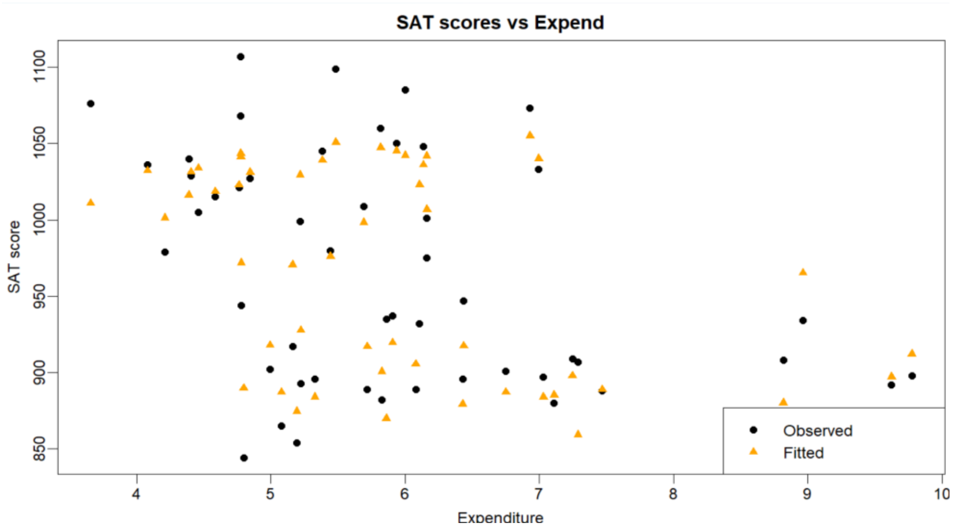

Figure 3: Actual vs Fitted Values

The below graph visualizes the uncertainty of the observed values with respect to the fitted values highlighting the 95% confidence intervals. The below graph shows that 98% of the fitted values are within (Guber) 95% confidence levels of actual values.

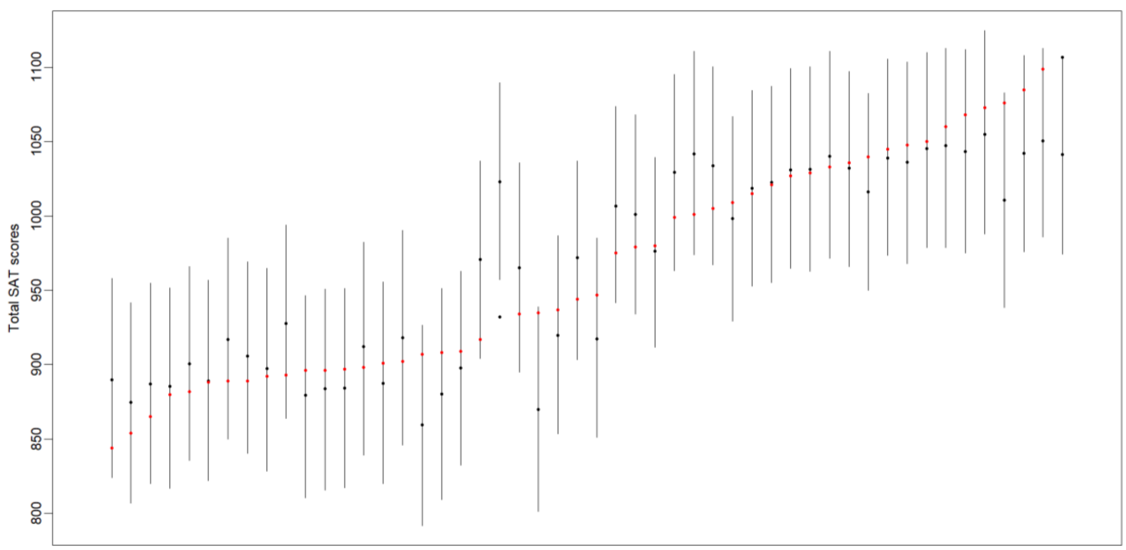

Figure 4: Visualization for uncertainty of Posterior parameters I

The above graph visualizes the uncertainty of the observed values with respect to the fitted values highlighting the 95% confidence intervals. The below graph shows that 98% of the observed values are within 95% confidence levels of the fitted values.

One more visualization on uncertainty around the posterior parameters:

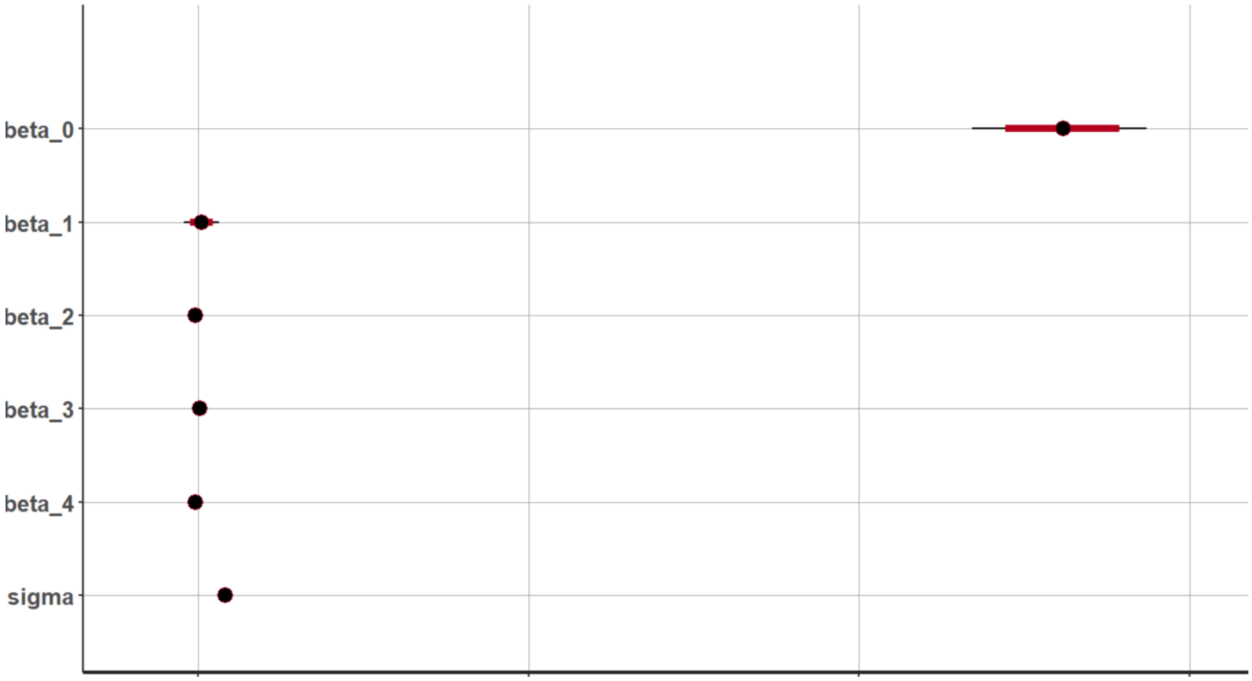

Figure 5: Visualization for uncertainty of Posterior parameters II

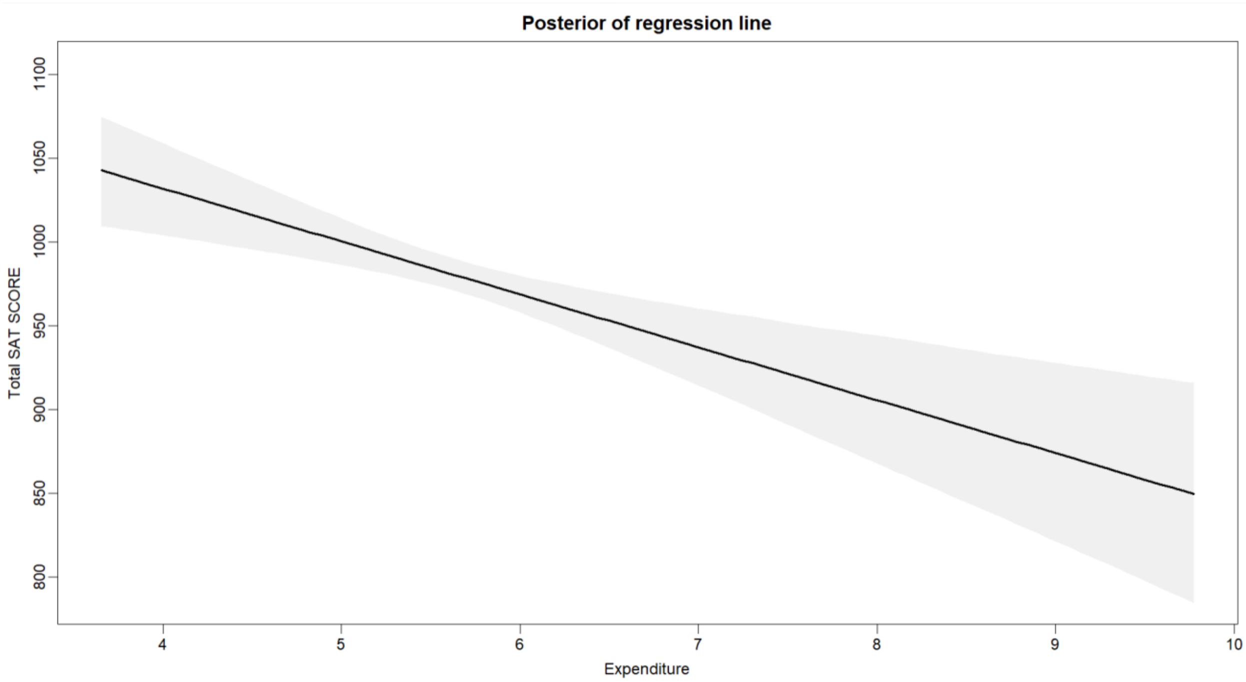

Figure 6: Visualization for uncertainty of Posterior parameters III

The above plot shows the uncertainty around the total SAT scores as the expenditure increases. We can see that the uncertainty is highest when the expenditure is maximum and it is least around the mean.

Discussion:

While a mean value of 4.6 for the coefficient $\beta 2$ might suggest a positive effect on SAT scores with an increase in expenditure, from image 6, we can infer that there is necessarily a negative effect on overall SAT scores with an increase in expenditure. This might not make more sense by itself, but upon examining the other coefficients and correlations. The below image shows the correlation between posterior parameters.

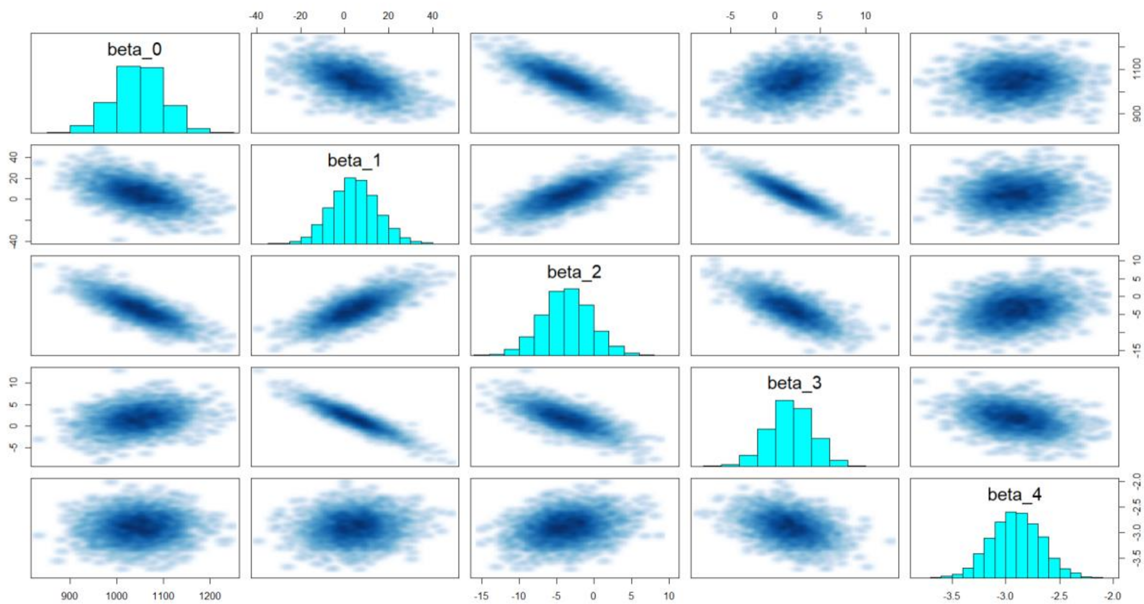

Figure 7: Correlation between posterior parameters

The observations we can make from the above plot are:

Observation 1: Unsurprisingly as expenditure increases, the pupil/teacher ratio increases, and teachers become better compensated as we can see a positive correlation between expenditure and salary.

Observation 2: As expenditure increases, takers increase indicating that more spending on education results in increased accessibility of SAT to a wider population.

The extra expenditure may have provided pupils in the less wealthy school districts with access to educational materials that were previously out of their reach. Thus we can infer that high spending results in an increase in the participation rate, which in turn decreases the overall SAT scores. One probable reason we can think of is when a small percentage of students take the test, only the most prepared students take the test, thus resulting in a higher average SAT score, on the other hand when increased expenditure is allowing a higher percentage of students to appear for SAT, even those who are less prepared for the test, average SAT score drops.

We learned how effective the hierarchical idea is as we explored the hierarchical linear regression model with RStan. This gave us some ideas for future research, such as combining the hierarchical model with more advanced machine-learning techniques to produce results that are more reliable and coherent. As the complexity and depth of the issues increase in an unrestricted manner, more innovative ideas for future applications might include nesting inside models and making compromises across models.

The philosophical idea of compromise between models, in particular, came into light, albeit in a generic and unobtrusive form, and for the time being without demonstrating immediate usefulness to applications. For example, even in the domain of regression, the analogy of Bayesian philosophy (balancing prior belief and real data for the estimation of model parameters) and hierarchical models (in the sense of compromise between aggregate and disaggregate models) has widened our viewpoint.

Meanwhile, the model's seamless implementation and compact representation in RStan exhibited a slew of benefits such as scalability, efficiency, and resilience. Its simple structure also piqued our curiosity in other potential expansions that have yet to be implemented, such as user-specific functions and non-conjugate prior, which might be another area of future investigation.

References:

Guber1999.: https://www.researchgate.net/publication/327474149_Getting_What_You_Pay_For_The_Debate_Over_Equity_in_Public_School_Expenditures

Gelman 2008.: http://www.stat.columbia.edu/~gelman/research/published/priors11.pdf

https://medium.com/analytics-vidhya/higher-spending-leads-to-poorer-education-a-bayesian-statistics-project-using-rjags-b50b213c6961