Prediction of Dutch People's Trust in Parliament

Bayesian Statistics

E. P. Swens (0902462)

1. Introduction

The European Social Survey (ESS) is a cross-national survey. It has been administered across Europe, covering more than 35 countries. The target population is all residents within private households in each participating country aged 15 and over, regardless of nationality, citizenship, or language. The survey is administered by face-to-face interviews. Since 2001, the ESS has been conducted every two years. The ESS uses cross-sectional sampling, meaning each round a new sample is taken from the population. In addition, countries use random probability sampling, therefore everyone who belongs to the target population has a chance to be selected.

This study aims to model the trust in parliament in the Netherlands based on the trust in politicians, police, and legal systems. It will use the information from the 2016 and 2018 survey. The information of 2016 will serve to create informative priors. There were four hypotheses formulated. The first hypothesis states that trust in the legal systems can predict trust in parliament. Evidence for this hypothesis will consist of the parameter estimates (i.e. 95% credible interval) and a model comparison using the deviance information criterion (DIC). The other hypotheses involve the ordering of importance of the predictors. These hypotheses will be evaluated using the Bayes factor provided by bayesian informative hypothesis evaluation (bain).

The regression estimates will be obtained using Gibbs sampling and a Metropolis–Hastings algorithm. The model convergence will be evaluated using history plots, autocorrelations, and Monte Carlo (MC) error. Further, model assumptions of homoscedasticity, independence of observations, and normality of residuals will be investigated using posterior predictive checking using discrepancy measures. After presenting the model result, the hypotheses will be evaluated and the Bayesian approach will be compared with a frequentist approach.

1.2. Hypotheses

Table 1 presents the study hypotheses. I am wondering if people's trust in the legal systems helps to predict trust in parliament. Furthermore, I am interested in informative hypothesis testing. I suspect all predictors to be positively associated with the outcome and *trust in politicians* to be the most important predictor.

Table 1 **Study Hypotheses**

H_1	Trust in the legal systems can predict trust in parliament	$\|\beta_3\| > 0$
H_2	Trust in politicians is the most important predictor and trust in police and legal systems are equally important	$\beta_1 > \beta_2 = \beta_3 > 0$
H_3	Trust in politicians is a more important predictor than trust in police and trust in police is a more important predictor than trust in legal systems	$\beta_1 > \beta_2 > \beta_3 > 0$
H_4	Trust in politicians is a more important predictor than trust in legal systems and trust in legal systems is a more important predictor than trust in police	$\beta_1 > \beta_3 > \beta_2 > 0$

2. Method

This study aims to model the trust in parliament in the Netherlands based on the trust in politicians, police, and legal systems. The model was fitted to the data from the European Social Survey of 2016 and 2018.

2.1. Data

The data can be accessed from the ESS Data Portal [1]. The information from 2018 formed the density of the data. Whereas the responses from 2016 formed the basis for the informative priors. People's trust was measured with an 11-point Likert scale from 0 to 10. In this study, all measurements were considered continuous variables. The data was grand mean-centered and missing data were handled using listwise deletion. The sample size for rounds 8 and 9 was 1.681 and 1.673, respectively.

2.2. Model

The Dutch people's trust in parliament (y) is assumed to follow a normal distribution, $y_i \sim N(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i}, \sigma^2_\epsilon)$, where $x_1$represent *trust in politicians*, $x_2$represent *trust in police*, and $x_3$represent *trust in legal systems*, and $\epsilon \sim N(0, \sigma^2_\epsilon)$.

The regression estimates were obtained using a Markov chain Monte Carlo (MCMC) method. Specifically, the MCMC method combined Gibbs Sampling and a Metropolis–Hastings algorithm. Gibbs Sampling approximated the joint distribution by taking 10.000 samples per chain from each conditional posterior distribution. For the regression estimates β_0, β_2, β_3, the conditional posterior distribution was a normal distribution composed of a prior which was also normal. The hyperparameters of the priors were estimated from a regression model using the data of 2016. The mean parameter was set to the regression estimate and the variance parameter to the squared standard error of that estimate. For the regression estimate of *trust in politicians*, the prior was a Student's t-distribution. Because this prior is not conjugate, the conditional posterior could not be identified. Therefore, a random walk Metropolis-Hastings algorithm was implemented in the Gibbs sampler. To be completely honest, the decision for a Student's t-distribution as a prior was more an academic challenge than an informative decision. Still, the hyperparameters were estimated in a similar fashion. The mean parameter was set equal to the regression estimate, the precision parameter to the reciprocal of the squared standard error, and the degrees of freedom equal to the degrees of freedom of the regression model. Although the Gibbs sampler and MH-algorithm are presented as separate methods, one is a special case of the other. Namely, the Gibbs Sampler is a Metropolis-Hastings algorithm wherein proposals are always accepted.

3.3. Model Diagnostics

The convergence diagnostics for the MCMC algorithm included history plots, autocorrelations, and Monte Carlo (MC) error. Firstly, the history plot shows the parameter value at iteration i against the iteration number. A sign that the model converged with the history plot is that the parameter values will move around the mode of the distribution. Secondly, the autocorrelations of the parameter values were plotted for a lag from 0 to 40. An absolute autocorrelation less than 0.10 would indicate the model converged. Finally, the MC error should not be larger than 5% of the sample standard deviation.

3.4. Model Assumptions

There are several assumptions associated with a linear regression model, e.g., linearity, homoscedasticity, independence of observations, normality of residuals, and no extreme multicollinearity. In this study, we focused on testing the assumption of homoscedasticity, independence of observations, and normality of residuals. The assumptions were investigated by posterior predictive checking using discrepancy measures. The following paragraph details how data was simulated from the posterior distribution.

First, assume that the null model is true, defined as $y_i \sim N(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i}, \sigma^2_\epsilon)$. Second, sample t sets of parameter estimates from the posterior distribution of the model, denoted by $\theta^t = [\beta^t_0, \beta^t_1, \beta^t_2, \beta^t_3, \sigma^{2,t}]$. Here, we specified t to be 10000. Third, for each set t we generated data according to the specified null model. Resulting in t simulated datasets $[Y^t_{sim}, X]$. Where Y^t_{sim} represents the simulated values of the dependent variable of n persons and X the ancillary observed scores for the predictor variables. From the simulated data and parameter estimates, it is possible to compute discrepancy measures.

For each assumption, we created a discrepancy measure. First, the discrepancy measure for the homoscedasticity combines the information of the predicted values and residuals. The residuals were partitioned based on the corresponding fitted value. Consider a set of residuals R. Then, let R_{less} be residual values where the corresponding predicted values are less than the median of predicted values and let $R_{greater} = R \backslash R_{less}$. Then, to test for homoscedasticity we compare the ratio between the variance of the residual groups, $D_1 = var(R_{greater}) \,/\, var(R_{less})$. For the normality assumption, we computed the theoretical and standardized residuals, denoted by $R_{theoretical}$ and $R_{standardized}$, respectively. The mean absolute difference was computed to test normality, $D_2 = mean(|R_{standardized} - R_{theoretical}|)$. For the assumption of independence, the autocorrelation of the residuals was computed with a lag of two, $D_3 = \frac{1}{n \cdot var(R)} \sum_{i=1}^{n-2} (R_i - \overline{R})(R_{i+2} - \overline{R})$.

Finally, for each discrepancy measure the posterior predictive p-value (PPP) was derived by $P(D([Y^t_{sim}, X], \theta^t) > D([Y, X], \theta^t) \mid [Y, X] \wedge H_0)$. A p-value centered around 0.5 would indicate that a model assumption holds, as the discrepancy measure from the observed data is centered around the corresponding posterior predictive distribution. In contrast, a strong deviation from 0.5 would indicate a violation of a model assumption.

3.5. Testing Hypotheses

For the first hypothesis, we evaluated the support for the predictor *trust in the legal systems* in two ways. First, we estimated if the 95% credible interval includes the value zero. Second, by comparing DIC values of the null model with a model excluding *trust in legal systems*. The support for the informative hypotheses, involving constraints among the parameters of interest, was twofold. First, we compared the Expected A Posteriori (EAP) of the parameter estimates.

Second, we compared the constrained models to their corresponding complements using a Bayes factor (BF) provided by the BAIN package [2]. A BF of 2 indicates that there is twice as much support for the hypothesis at hand versus its complement, while a BF of 0.5 indicates that there is twice as much support for the hypothesis's complement than for the hypothesis.

3.6. Bayesian versus Frequentist Analysis

In conjunction with the Bayesian approach, we modeled *trust in parliament* with an ordinary least squares linear regression model. This model only included the information of 2018. We evaluated the first hypothesis with a two-sided *t*-test and a comparison of Akaike information criterion (AIC) values of the null model against a model excluding *trust in legal systems*. For the normality assumption, we performed a Shapiro-Wilk normality test. For the homoscedasticity assumption, we used a studentized Breusch-Pagan test. Finally, for the informative hypotheses, we evaluated the three *t*-tests of the coefficients and ran three F-tests. These F-tests consisted of a Wald-test-based comparison between a model and a linearly restricted model. The results of this analysis will solely be used in the discussion section.

4. Results

The diagnostic results indicate that the model converged. Figure 1 presents the trace plot and autocorrelation plot for the regression parameter *trust in legal systems*. The parameter values appear to cover the entire distribution and the absolute autocorrelation less than 0.10 given the lag is greater than 2. The diagnostics plots are similar for the other parameters values from the Gibbs sampler (Appendix A and B). The results for the parameter values obtained by the MH-algorithm also support that the model converged. The MH-algorithm yielded an acceptance rate of 56.3%. Table 1 presents the MC error. For each parameter, the MC error is not larger than 5% of the posterior standard deviation.

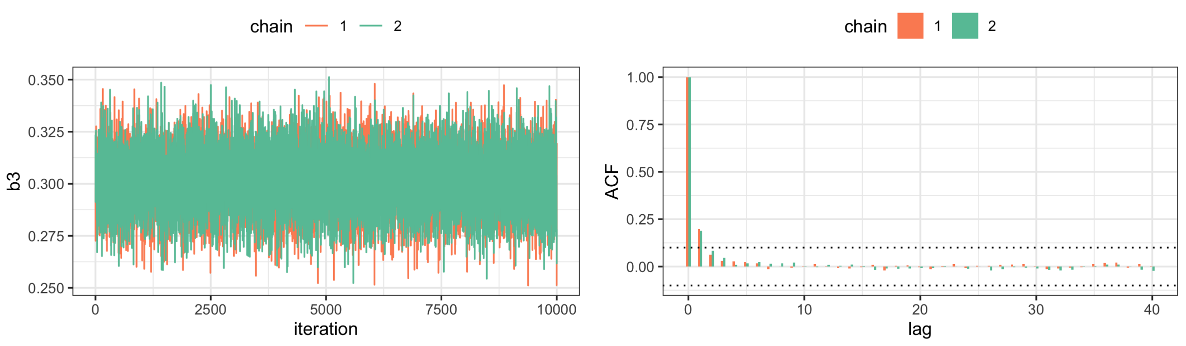


Figure 1 **Diagnostics plots left) Trace Plot right) Autocorrelation Plot**

The posterior mean for the intercept parameter is 0.000. The EAP for the parameter *trust in politicians* is 0.444, while the EAP for *trust in police* is 0.075. The posterior mean for *trust in legal systems* is 0.301. The mean of the residual error variance is 1.296. The 95% credible interval for all predictors does not contain the value zero. The plots of the conditional posterior distributions are provided in Appendix C.

The PPP-value for the assumption of homoscedasticity and normality are round zero, implying both assumptions are violated. In contrast, the PPP-value for the assumption of independence is 0.54. This evidence supports that the assumption of independence of observation holds.

The DIC value of the model including all predictors is 5011. Whereas, the DIC value of the model excluding the parameter *trust in legal systems* is 5236. Furthermore, the BF for the second and third hypothesis was roughly zero. The BF for the fourth hypothesis was 176438.

Table 2 **Sample Statistics for the Parameter Posterior Distributions**

	EAP	SD	95% CI	MC error
β_0	0.000	0.020	-0.040 - 0.040	0.00014
β_1	0.444	0.014	0.417 - 0.472	0.00010
β_2	0.076	0.015	0.047 - 0.105	0.00010
β_3	0.301	0.013	0.275 - 0.327	0.00009
σ^2	1.296	0.047	1.206 - 1.390	0.00033

5. Discussion

There is strong evidence that the model converged from the model diagnostics. However, the PPP suggests a violation of the assumption of normality and homoscedasticity. Still, a visual inspection of the standardized vs theoretical residuals and residuals vs fitted values suggest that we should be too worried (Appendix D). Given that the variables were measured using a 11-point Likert scale, it might be unreasonable to assume that normality holds. Given the scope of the assignment, we did not investigate other assumptions. In any case, for now I will assume there is no issue in interpreting the model results and I will continue to discuss the evaluation of the hypotheses.

5.1 Hypotheses Evaluation

There is evidence supporting that *trust in legal systems* can predict *trust in parliament* (H_1). First, the 95% credible interval of *trust in legal systems* does not contain the value zero. Second, based on the DIC comparison, the null model is prefered over the model excluding *trust in legal systems*.

For the second and third hypotheses, there was virtually no support. To elaborate, the BF for the second and third hypotheses was roughly zero. This Bayes factor value suggests there is almost infinitely more support for the hypothesis complement than for the hypothesis itself. Further, both hypothesized constraints appear to be unlikely according to the EAP of the parameters.

There is substantial support for the fourth hypothesis. First, the BF for the fourth hypothesis was 176438. This result indicates there is almost infinitely more support for H_4 than for its complement. Second, the hypothesized orderings appear to be in accordance with the EAP of the parameters.

To conclude, based on the evidence I accept H_1 and H_4, and reject H_2 and H_3.

5.2 Bayesian versus Frequentist Analysis

Table 3 presents the results of the ordinary least squares regression model. In contrast to the Bayesian approach, the model could not include the information of 2016. Instead, the obtained estimates are similar to the Bayesian model using uninformative priors. Still, the effect estimates do not deviate more than 0.14 from each other. According to the Shapiro-Wilk normality test and studentized Breusch-Pagan test, both the assumption of normality and homoscedasticity were violated. This conclusion is in accordance with the PPP results.

We would accept the first hypothesis that trust in legal systems can predict trust in parliament, $t(1613) = 31$, $p < 0.001$. For the trust in legal systems coefficient, both the 95 % credible interval and 95% confidence interval do not contain zero. Yet, the interpretation of both intervals is quite different. To elaborate, the confidence interval is based on repeated sampling, where 95% of the computed confidence intervals would contain the true value. Whereas the credible interval is an interval within which an unobserved parameter value falls with a particular probability given the observed data. Additional support for the first hypothesis included the model comparisons using the AIC. The model with all predictors (AIC = 4956) is favored over the model excluding trust in legal systems (AIC = 5127). This conclusion is in accordance with the DIC results.

The most striking difference between Bayesian and frequentist analysis occurred during the informative hypothesis testing. Using a frequentist approach, informative hypothesis testing became very cumbersome, as it required six tests (3 t-tests to test if the predictors are positively associated with outcome and 3 F-tests to compare an unconstrained and constrained model). Using a Bonferroni correction to account for multiple testing, the α was set to 0.008. For this alpha, the positive association between trust in police and trust in parliament does not hold. From the frequentist analysis, we have to reject H_4, which is not in accordance with our conclusion from the Bayesian analysis. This result highlights, in my opinion, an important disadvantage of the frequentist approach. Namely, multiple testing and having to correct the alpha level. It is most likely, that in this case, the correction for multiple testing inflates the false negative rate and we now wrongfully reject that $\beta_2 > 0$.

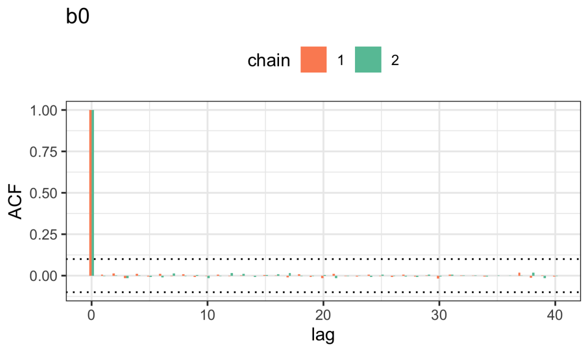
Table 3 **Parameter Estimates of the OLS regression model**

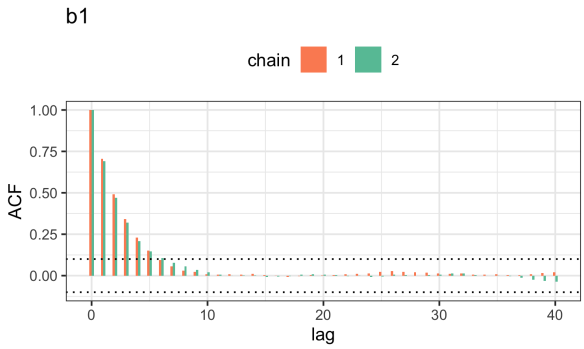
	Estimate	SE	95% Confidence Interval	P(>|*t*|)
β_0	0.000	0.028	-0.055 - 0.055	1.00
β_1	0.580	0.019	0.544 - 0.617	< 0.001
β_2	0.067	0.021	0.026 - 0.109	< 0.01
β_3	0.259	0.019	0.222 - 0.297	< 0.001

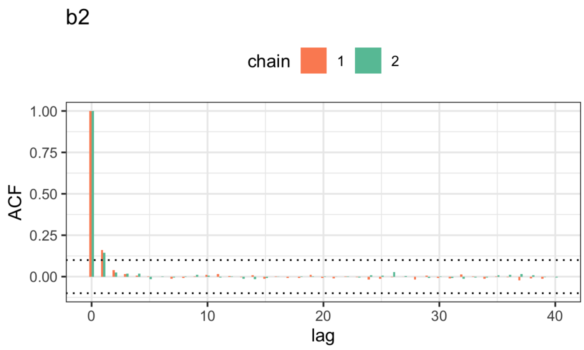
References

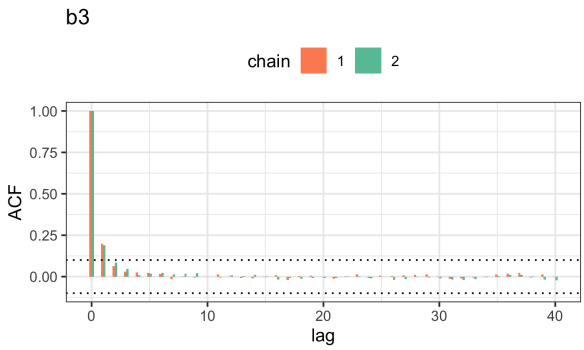
1. European Social Survey. (2022). European Social Survey Data Portal. ESS: https://ess-search.nsd.no/en/all/query/
2. Gu, X., Hoijtink, H. J. A., Mulder, J., & Van Lissa, C. J. (2019). bain: Bayes factors for Informative Hypotheses. R package version 0.2. 1. R package version 0.2. 0.

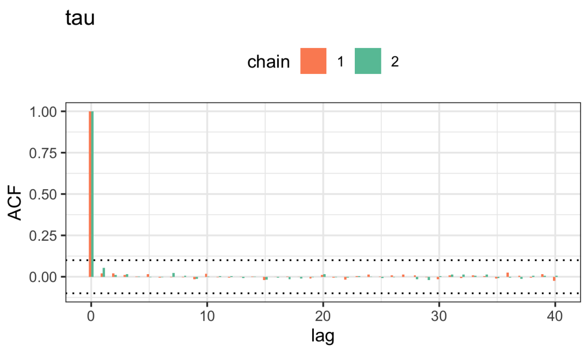
Appendix A - Auto Correlation Plots





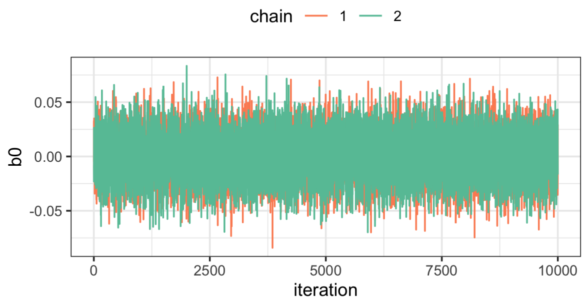




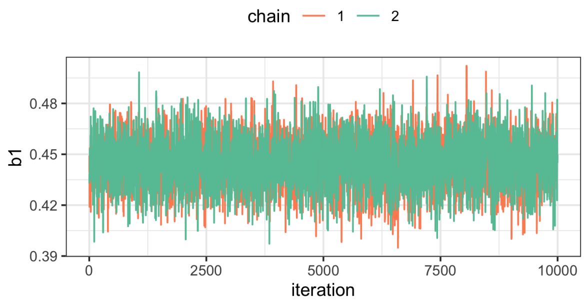


Appendix B - Trace Plots

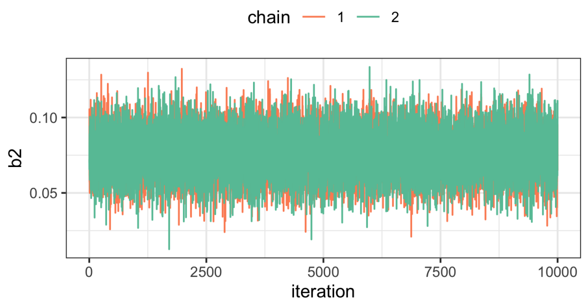
b0



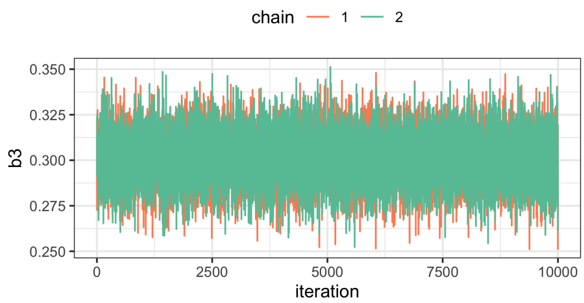
b1



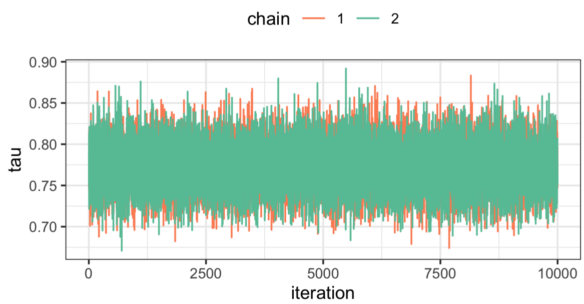
b2



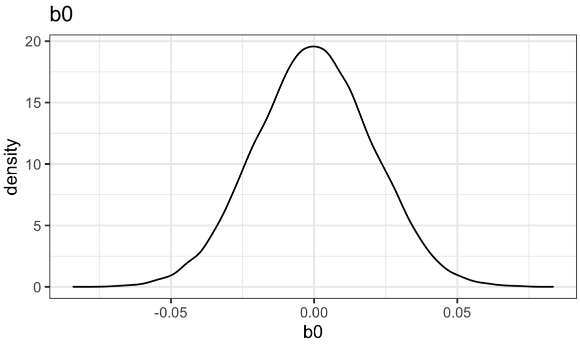
b3

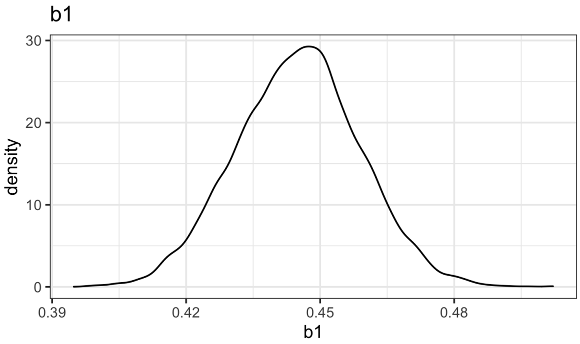


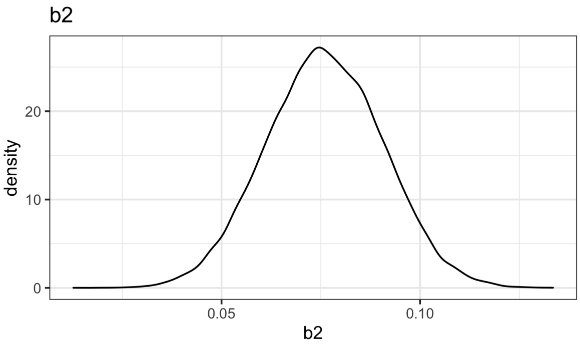
tau

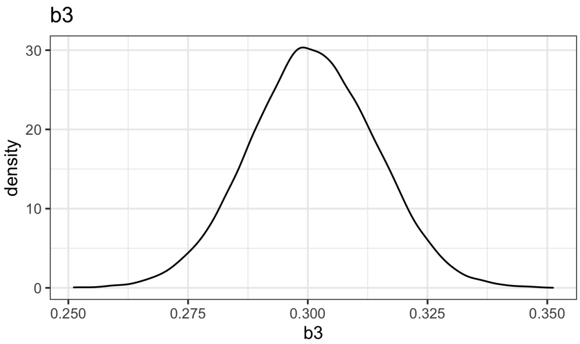


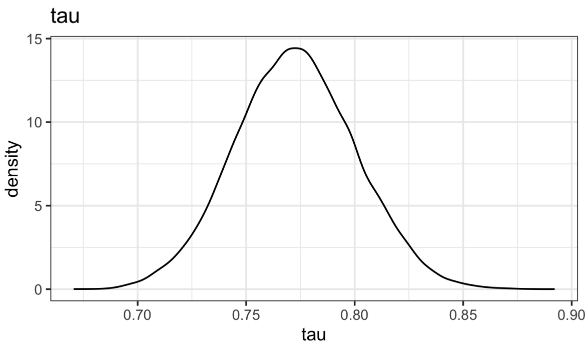
Appendix C - Conditional Posterior Distributions











Appendix D - Residuals Plot