

# Bayesian Statistics Assignment

Sofía Villalobos Aliste (6060714)

10-01-2022

## Research question

The aim of this project is to examine if Political Self-Positioning in the Netherlands can be predicted from variables concerning Institutional Trust: Trust on country's parliament, Trust in legal system, and Trust in Police. This would be addressed modelling the following regression:

$$Y = B_0 + B_1 * X1 + B_2 * X2 + B_3 * X3 + \varepsilon$$

## Data

The data for this research had been obtained from the European Social Survey 9th round, collected in 2018 in the Netherlands.

The dependent variable that will be used as political self-positioning is "Placement on left right scale" and consists in 10 categories going from 0 (Left) to 10 (Right). Independent variables Trust in country's parliament (prl), Trust in the legal system (lgl), and Trust in the police (plc), has are categorical variables that goes from 0 (No trust at all) to 10 (Complete trust). All these variables will be here treated as continuous.

Two models of multiple linear regression are proposed here that differ in the type of previous information used: for the first model we included specified priors by taking in account the round 8 of the European Social Survey conducted in the Netherlands in 2016. The second model was constructed with the same variables but uninformative priors instead.

The total cases for this survey used in the analysis is 1524 after using listwise deletion.

## Estimation and convergence

The estimation process was made through Gibbs Sampling, for  $B_0, B_2, B_3$ , and  $\sigma^2$ , and with Metropolis Hasting sampling for the estimation of  $B_1$ . First, we define the initial values for each of the two chains that we will be using to sample, and the number of iterations (10000 for each).

$B_0$  conditional posterior was obtained by derivating first the initial values to obtain conditional posterior mean and variance, and then sampling a value from conditional posterior of  $B_0$ , and we keep this value for the next step of sampling  $B_1$ .  $B_2$  and  $B_3$  are sampled in the same way.

$B_1$  was obtained through a Metropolis Hasting step, were we start taking the previously obtained value for  $B_0$ , to derive the conditional mean and variance, and with these values we sample from a normal proposal distribution a value for  $B_1$ , that could be accepted or rejected. according if it the value we sample from a uniform distribution is larger or smaller in comparison to the acceptance ratio that is calculated computing the target densities that can be computed from the conditional mean and variance. This is done as a

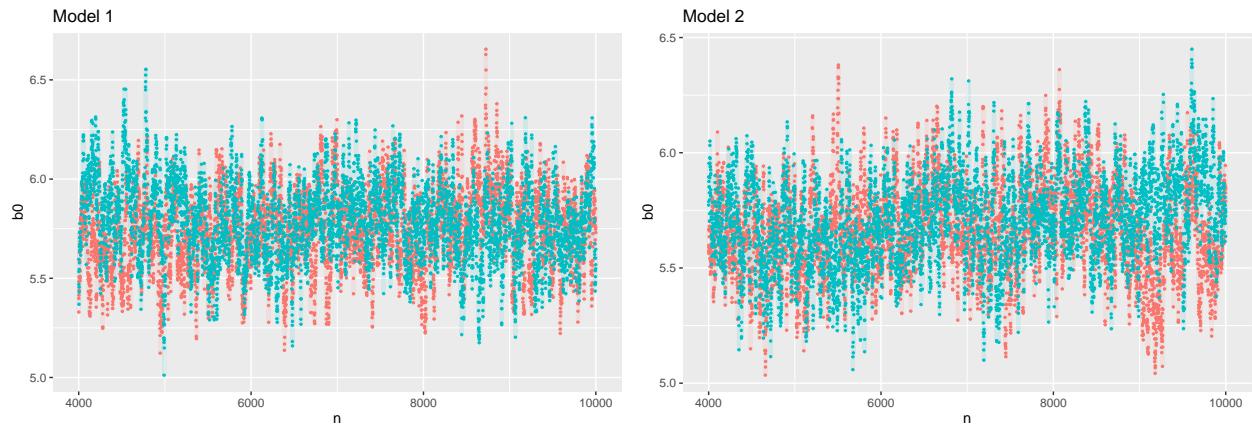
correction in case of obtaining too improbable values. Then, if the proposal value of  $B_1$  is rejected for being larger than the ratio, the last value of  $B_1$  is retained until there is a proposal value of  $B_1$  that is equal or smaller to this ratio.

Finally, we sample the posterior distribution for the residual variances. Using the sampled values of  $B_0, B_1, B_2$  and  $B_3$  to compute the condition posterior of  $\sigma^2$  by sampling posterior shape parameter alfa and posterior scale parameter beta from where we can compute the sum of squared residuals.

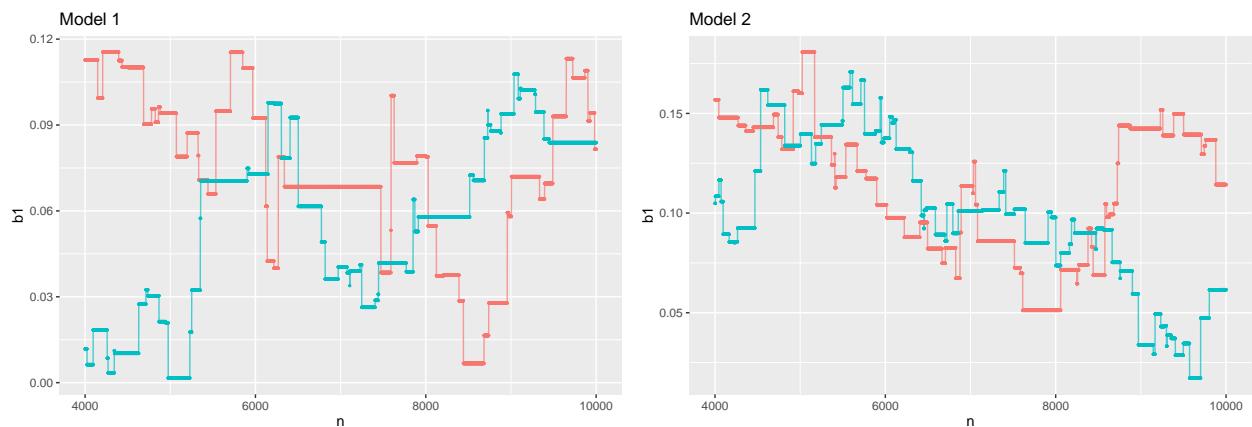
That is how the sampler works and was used to estimate the parameters for Model 1 (with informative priors) and Model 2 (with uninformative priors).

To evaluate convergence of this iterations, we first check at trace plots considering the total number of iterations and deleted the first ones on where we can see that the chains were far to overlap. 4000 first iterations were deleted for each chain on each model, resulting in the following plots where convergence can be observed as a first check:

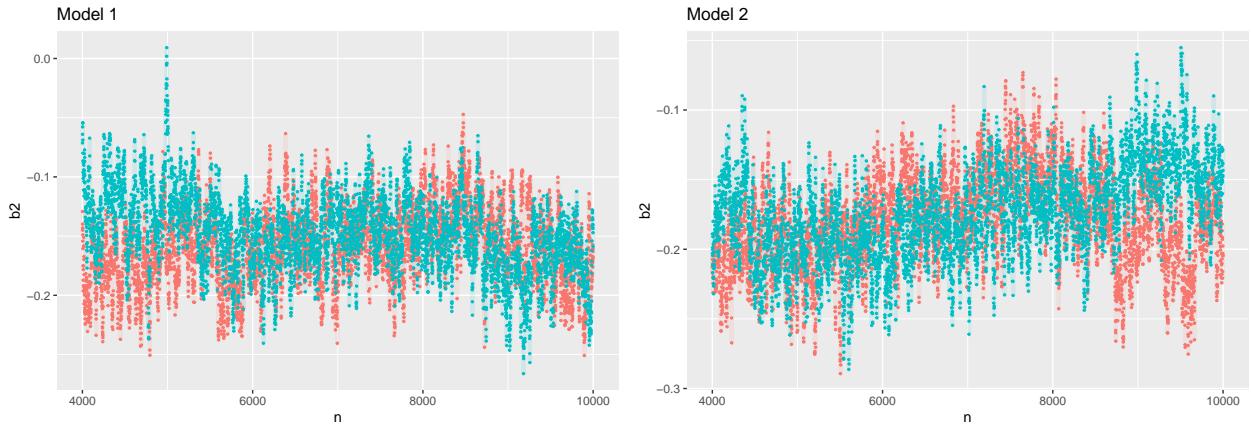
### Trace plots for $B_0$



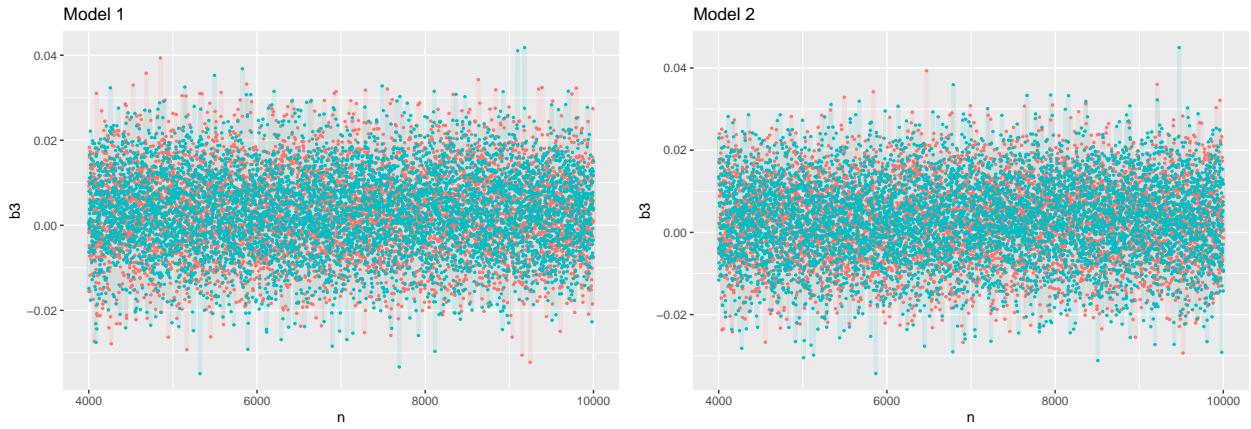
### Trace plots for $B_1$



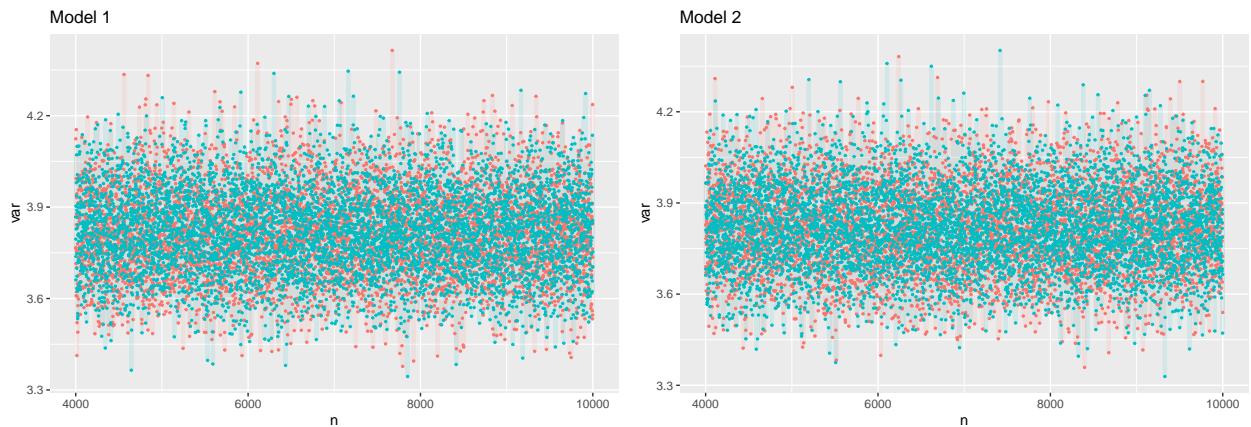
### Trace plots for $B_2$



### Trace plots for $B_3$



### Trace plots for $\sigma^2$



We checked density plots as an approximation of the marginal posterior distribution, and in this case the ones for  $B_0, B_2, B_3$  and variance look normal, for both models. Also, for both models, since  $B_1$  was computed with Metropolis Hasting it was expected to look far away for normal.

Finally, we checked Monte Carlo Error and they all stay  $<.05\%$ . So, we can conclude that the models had converged, and we can continue with further analysis.

## Posterior predictive check

To check the assumption of normality of residuals we first establish the null hypotheses as our proposed linear regression model:

$$y_i = B_0 + B_1 * X1_i + B_2 * X2_i + B_3 * X3_i + e_i$$

with

$$e \sim \mathcal{N}(\mu, \sigma^2)$$

Second, we sample the posterior distribution of the model parameter, this is for each sampled parameters  $\theta = [B_0, B_1, B_2, B_3, \sigma^2]$  computed in every 12000 iterations of our Gibbs sampler, and then we modelled  $Y$  as a regression with fixed effect of  $X1, X2$  and  $X3$ . So, we end up a matrix of 12000 columns and 1524 rows, where every box reflects and simulated outcome per subject of the observed data using a different set of estimates for every column.

Now we compute the residuals for the observed data: This is, we subtract to the observed  $Y$  of every subject the estimate predicted with the 12000 sets of  $\theta$ . To compute the residuals for the simulated data, we subtract the same estimated with the 12000  $\theta$  to the modelled  $Y$  that we previously obtain by simulating the outcomes maintaining  $X1, X2$ , and  $X3$  fixed. Then, we compute the skewness in the absolute value  $= |3 * (\text{Mean} - \text{Median}) / \text{Standard Deviation}|$ , to avoid problems after when calculating the bayesian p-value like for example having a skewness coefficient of  $-.5$  being considered as less skewed than a coefficient of  $1$ . Finally, we count how many times the skewness on the simulated data is larger than the skewness in the observed data, and we storage the results in a matrix of 12000 columns with 1 row where every time we will put a  $0$  if the coefficient of the observed data is larger than the simulated, and a  $1$  if it is the other way around. This procedure is computed in the exact same way for both models.

For model 1, there were 11526 occasions where skewness was larger for simulated data, leading to a bayesian p-value of  $0.96$ . For model two, there were 7714 occasions where skewness was larger for simulated data, leading to a p-value of  $0.64$ . This shows that for both models, assumption of normality of residuals is met.

## Results and interpretation

Table 1: Results for informative priors

	Mean	SD	Q025	Q975
b0	5.7647598	0.1983241	5.3760910	6.1551057
b1	0.0654934	0.0301140	0.0034442	0.1153862
b2	-0.1537901	0.0326255	-0.2190708	-0.0901948
b3	0.0030348	0.0096830	-0.0159706	0.0221803
var	3.8168007	0.1391219	3.5535557	4.0984122

Table 2: Results for uninformative priors

	Mean	SD	Q025	Q975
b0	5.6825778	0.1970284	5.2992095	6.0778240
b1	0.1059245	0.0358671	0.0339017	0.1618901

	Mean	SD	Q025	Q975
b2	-0.1775598	0.0333221	-0.2415244	-0.1136630
b3	0.0027655	0.0097097	-0.0162951	0.0214100
var	3.8179029	0.1375935	3.5570703	4.0931386

As we can see in both tables, on average all estimates for the intercept of  $B_0$  are positive,

For  $B_0$  average results are positive, with a mean of 5.764 and CI [5.3-6.1] in first model and mean of 5.683 and CI [5.3-6-1] in second model. For  $B_1$  average results are positive, with a mean of 0.065 and CI [0.0-0.11] in first model and mean of 0.1 and CI [0.03-0.16] in second model. For  $B_2$  average results are negative, with a mean of -0.15 and CI [-0.21 - -0.09] in first model and mean of -0.177 and CI [-0.24- -0.113] in second model. For  $B_3$  average results are positive, with a mean of 0.0 and CI [-0.01- 0.02] in first model and mean of 0.002 and CI [-0.01-0.02] in second model.

For both models, coefficients  $B_0, B_2$  and  $B_3$  could be interpret as evidence that they are different from 0 and have a positive association with the outcome, since their credible intervals do not include 0. In the case of  $B_3$  there is not enough clear evidence that the coefficient is different from 0 and its direction its not clear since their credible intervals include 0.

## Model selection using the DIC

To compare these models, we refer to the Deviance Information Criterion.

AIC for model with informative priors is 6365.9, whereas for model with uninformative priors is 6366.3. This difference of 0.4 does not allows to discriminate for AIC which one is the better model. Since both might be equivalent, we will stay with the one with prior information for next analysis.

## Model selection using bayes factor

After concluding that model that uses prior information was the better model when selecting using DIC, 10 different hypotheses were tested concerning each of the predictor being larger or smaller than 0:

Table 3: Testing of hypotheses with bayesian factor

Hypotheses	BF.u	BF.c	PMPa	PMPb
H0 = B1 = 0, B2 = 0, B3 = 0	0.005	0.005	0.001	0.001
H1 = B1 > 0, B2 > 0, B3 > 0	0.000	0.000	0.000	0.000
H2 = B1 > 0, B2 > 0, B3 < 0	0.000	0.000	0.000	0.000
H3 = B1 > 0, B2 < 0, B3 > 0	5.457	103.47	0.934	0.797
H4 = B1 > 0, B2 < 0, B3 < 0	0.070	0.060	0.012	0.010
H5 = B1 < 0, B2 < 0, B3 < 0	0.002	0.002	0.000	0.000
H6 = B1 < 0, B2 < 0, B3 > 0	0.243	0.218	0.042	0.035
H7 = B1 < 0, B2 > 0, B3 > 0	0.000	0.000	0.000	0.000
H8 = B1 > 0, B2 < 0, B3 < 0	0.066	0.056	0.011	0.010
H9 = B1 < 0, B2 > 0, B3 < 0	0.000	0.000	0.000	0.000

These hypotheses were tested using the package Bain in R, trying different sets to ensure the stability of the results. From it we can observe that according to the column BF.u that indicates the Bayes factor as a ratio of fit and complexity, the most strong hypotheses to be supported among the one tested is the third one.

That indicates that to predict out outcome variable  $Y$ - Political Self-positioning, the coefficient  $B_1$ - Trust on country's parliament would have a positive association with  $Y$ , Trust in legal system would have a negative one, and Trust in Police would have a positive one. This is, the more a subject trust in parliament and police, would be place more to the right in the scale of political self-positioning. And works in the opposite way for trust in the legal system, for each point of trust on it that a person has, it would be placed more to the left.

## **Bayesian and frequentist approach**

Finally, we estimate a linear regression in the frequentist approach, and we can see that the results barely differ from the ones obtained with a bayesian approach. Coefficients for  $B_0 = 5.41177$ ,  $B_1 = 0.06598$ ,  $B_2 = -0.19384$ ,  $B_3 = 0.08784$  and Residual standard error = 1.95. All coefficients significant to  $p < .05$ . Because of this is difficult to observe any benefit that could have to adopt this approach, since we also could observe that using informative and uninformative priors did not make any difference.

## **Source**

ESS Round 8: European Social Survey Round 8 Data (2016). Data file edition 2.2. NSD - Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC. doi:10.21338/NSD-ESS8-2016.