

# Implementing a BYM Model in Stan to Fit Boston Housing Price Data

Marcel Gietzmann-Sanders

STAT641 - Bayesian Statistics  
University of Alaska Fairbanks

## 1 Data on Boston Housing Prices in 1978

For this study we'll be using a dataset originally published in 1978 on housing prices in Boston, Massachusetts (David and Rubinfeld (1978)). This data was made accessible through `spData` package in R (Bivand (2022)). Specifically this dataset contains median house prices and potential covariates for 506 tracts in the Greater Boston Area. Fig. 1 shows the median price per tract of land where it should be understood that this data is censored and median values over \$50,000 are capped. A full list of covariates can be found in the documentation for the `boston` dataset in `spData`. What we wish to understand is which of these covariates are indeed related to housing prices after being added to a spatial bayesian model.

Starting with some exploratory data analysis we identified four covariates of particular interest. Crime rates per capita, average number of rooms per dwelling, weighted distance to employment centers, and the nitric oxide concentrations per town.

First, as the distribution of values is right skewed (and following in the footsteps of (Moraga (2023))) we took as our target variable the logarithm of the median house value per tract instead of the median value itself. Fig. 2 shows the distribution of our target.

Fig. 3 shows the relationship between per capita crime rate and our target with a clear negatively correlated relationship between then two and a rather wide spread of values in and around 0.

Fig. 4 shows how our target varies with the average number of rooms per dwelling in each tract. Here we can see a strong positive relationship between the two which makes a great deal of sense.

In Fig. 5 we see a similar positive relationship between the logarithm of the average weighted distance to employment centers and our target.

However this relationship is not as clear or strong as the one between our target and the number of rooms per dwelling.

Finally Fig. 6 shows us the relationship between nitric oxide concentrations in parts per million and our target with a somewhat spurious negative relationship.

## 2 Defining the Model

We will be using a variant of the Besag-York-Mollié (BYM) model (Moraga (2023))(Mitzi Morris (2019)). In this model we assume that we have a series of  $Y_i$  which is a vector of our target variable (log median housing value in our case) for each tract  $i$ . Furthermore we assume our  $Y_i$  can be modeled as a normal distribution:

$$Y_i \sim Normal(\mu_i, \sigma^2)$$

where it is the  $\mu_i$  that will be affected by our covariates. Specifically we will have:

$$\mu_i = \beta_0 + \vec{\beta}x_i + \sigma_r \left( \sqrt{\rho}\phi_i + \sqrt{1-\rho}\theta_i \right)$$

where  $\beta_0$  is our intercept,  $\vec{\beta}$  are the coefficients for our models effects from each covariate,  $x_i$  are our covariates corresponding to tract  $i$ ,  $\phi_i$  and  $\theta_i$  are spatial and random effects respectively, and  $\sigma_r$  and  $\rho$  allow us to control the effect of the random variables as well as the degree to which our model has spatial and/or unstructured noise (if  $\rho = 1$  we have only spatial structure whereas if  $\rho = 0$  it is totally unstructured).

For our more straightforward priors we will have:

$$\beta_0 \sim Normal(0, 1)$$

$$\beta_i \sim Normal(0, 1)$$

$$\theta_i \sim Normal(0, 1)$$

$$\sigma_r \sim Uniform(0, 1)$$

$$\rho = \frac{e^r}{1 + e^r}, r \sim Normal(0, 1)$$

In initial fitting of the model we also had the prior:

$$\sigma \sim Normal(0, 1)$$

but consistently found  $\sigma \approx 0$  and therefore simply set  $\sigma = 0.01$  in order to not over complicate the sampling.

Our spatial terms  $\phi_i$  are a little more complicated.

Each spatial interaction term  $\phi_i$  is modeled as conditional on the other terms as:

$$\phi_i | \phi_j \sim N \left( \sum_j w_{ij} \phi_j, \sigma^2 \right), i \neq j$$

the name of which is a conditional autoregressive model (CAR). A key result that Besag proved (Julian (1974)) is that the joint distribution  $\phi$  ends up being multivariate normal random variable centered at 0

$$\vec{\phi} \sim N(0, Q^{-1})$$

where  $Q = D(I - \alpha A)$  where  $D$  is a diagonal "neighbors" matrix (each element on the diagonal is the number of neighbors unit  $i$  has),  $A$  is an adjacency matrix where if  $i, j$  are neighbors then the  $i, j$  element is 1, and  $\alpha$  lets us control spatial dependence. This results in a log probability density of  $\vec{\phi}$  which is proportional to:

$$\frac{n}{2} \log(\det Q) - \frac{1}{2} \vec{\phi}^T Q \vec{\phi}$$

Given  $\det Q$  is a constant and MCMC samplers compute the log probability up to a proportionality constant (Mitzi Morris (2019)) the first term drops out of the computation thereby reducing the computational intensity of this evaluation.

In our case, as we will be following the stan implementation from the paper (Mitzi Morris (2019)) we will be setting  $\alpha = 1$  and thereby get an intrinsic conditional autoregressive model (ICAR) which reduces  $Q$  to  $D - A$ . With an ICAR model each  $\phi_i$  is distributed with a mean equal to the average of its neighbors. If we additionally assume that  $\vec{\phi}$  is centered at zero with common variance 1, then the joint probability of  $\vec{\phi}$  becomes (Mitzi Morris (2019)):

$$p(\vec{\phi}) \propto \exp \left( -\frac{1}{2} \sum_{i \sim j} (\phi_i - \phi_j)^2 \right)$$

where  $i \sim j$  indicates that  $i$  and  $j$  are neighbors. This then is our prior for the  $\phi_i$  - an ICAR model centered at 0 with common variance 1.

### 3 Results

Parameter	95% CI	Mean	SD	SE	$\hat{R}$
$\beta_0$	2.130, 2.397	2.264	0.198	0.004	1.001
$\beta_{CRIM}$	-0.009, -0.007	-0.008	0.001	$\approx 0$	1.000
$\beta_{RM}$	0.237, 0.255	0.246	0.014	$\approx 0$	1.000
$\beta_{\log(DIS)}$	-0.171, -0.054	-0.112	0.085	0.004	1.008
$\beta_{NOX}$	-1.234, -0.972	-1.105	0.197	0.003	1.002
$\sigma_r$	0.341, 0.363	0.352	0.016	$\approx 0$	1.001
$\rho$	0.941, 0.966	0.952	0.019	$\approx 0$	1.001

Given the very small value on  $\beta_{CRIM}$  we also trained a model without that feature present.

Parameter	95% CI	Mean	SD	SE	$\hat{R}$
$\beta_0$	2.041, 2.312	2.177	0.203	0.004	1.001
$\beta_{RM}$	0.237, 0.257	0.247	0.015	$\approx 0$	1.000
$\beta_{\log(DIS)}$	-0.119, 0.003	-0.061	0.085	0.004	1.009
$\beta_{NOX}$	-1.260, -0.983	-1.122	0.205	0.004	1.003
$\sigma_r$	0.361, 0.381	0.371	0.015	$\approx 0$	1.002
$\rho$	0.952, 0.973	0.961	0.016	0.001	1.005

## 4 Code

### 4.1 Implementing the Model

Here we have the `stan` model itself:

```
functions {
  real icar_normal_lpdf(vector phi, int N, int[] node1, int[] node2) {
    return -0.5 * dot_self(phi[node1] - phi[node2])
      + normal_lpdf(sum(phi) | 0, 0.001 * N);
  }
}
data {
  int<lower=0> N; // number of tracts
  int<lower=0> N_edges; // number of unique edges
  int<lower=1, upper=N> node1[N_edges]; // start of edge
  int<lower=1, upper=N> node2[N_edges]; // end of edge
  int<lower=1> K; // number of covariates
  matrix[N, K] x; // design matrix
  real y[N]; // target
}
parameters {
  real beta0;
  vector[K] betas;

  real logit_rho;

  vector[N] phi;
  vector[N] theta;

  real<lower=0> sigma_r;
}
transformed parameters {
  real<lower=0, upper=1> rho = inv_logit(logit_rho);
  vector[N] convolved_re = sqrt(rho) * phi
    + sqrt(1 - rho) * theta;
}
model {
  y ~ normal(beta0 + x * betas + convolved_re * sigma_r, 0.01);
  target += icar_normal_lpdf(phi | N, node1, node2);
  beta0 ~ normal(0, 1);
  betas ~ normal(0, 1);
  logit_rho ~ normal(0, 1);
  theta ~ normal(0, 1);
  sigma_r ~ uniform(0, 1);
}
```

}

Note the term `normal_lpdf(sum(phi)|0,0.001*N)` used to center  $\vec{\phi}$  at zero (Mitzi Morris (2019)).

## 4.2 Fitting the Model

First we need to setup all of our data:

```
library(sf)
library(spData)
library(mapview)
map <- st_read(
  system.file("shapes/boston_tracts.shp", package = "spData"),
  quiet = TRUE
)

library(spdep)
nb <- poly2nb(map)
N = length(map$MEDV)
node1 = c()
node2 = c()
for (i in 1:N) {
  for (j in nb[[i]]) {
    if (j > i) {
      node1 = c(node1, i)
      node2 = c(node2, j)
    }
  }
}
N_edges = length(node1)

map$log_median_value <- log(map$MEDV)
y = map$log_median_value
x = cbind(map$CRIM, map$RM, log(map$DIS), map$NOX)
K = dim(x)[2]
```

## 5 Bibliography

### References

- Bivand, Roger, J. N.-R. L. (2022). spdata: Datasets for spatial analysis.
- David, H. and Rubinfeld, D. L. (1978). Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management*.
- Julian, B. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society*.
- Mitzi Morris, Katherine Wheeler-Martin, D. S. S. J. M. A. G. C. D. (2019). Bayesian hierarchical spatial models: Implementing the besag york mollié model in stan. *Spat Spatiotemporal Epidemiol*.
- Moraga, P. (2023). *Spatial Statistics for Data Science: Theory and Practice with R*. American Fisheries Society.



## 6 Figures

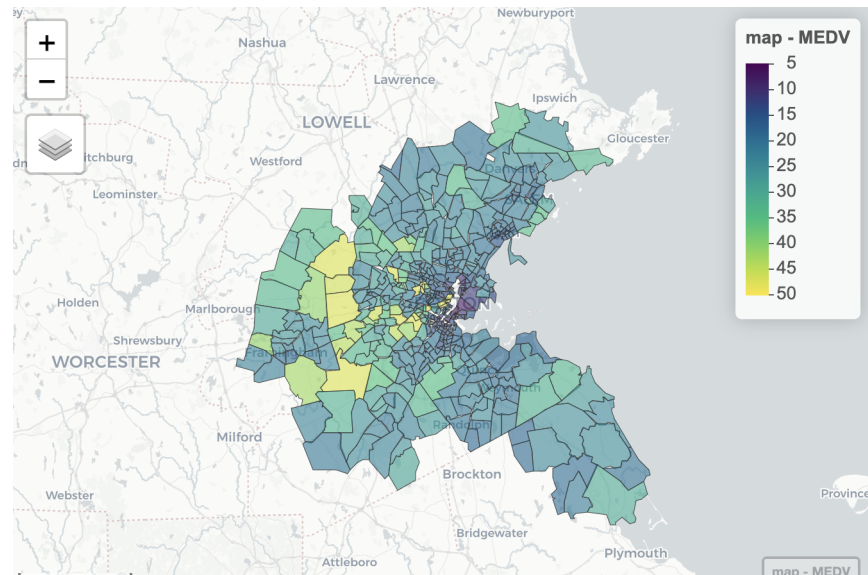


Figure 1: Median Housing Price

The median house value (in \$1000USD) by census tract in the Greater Boston Area.

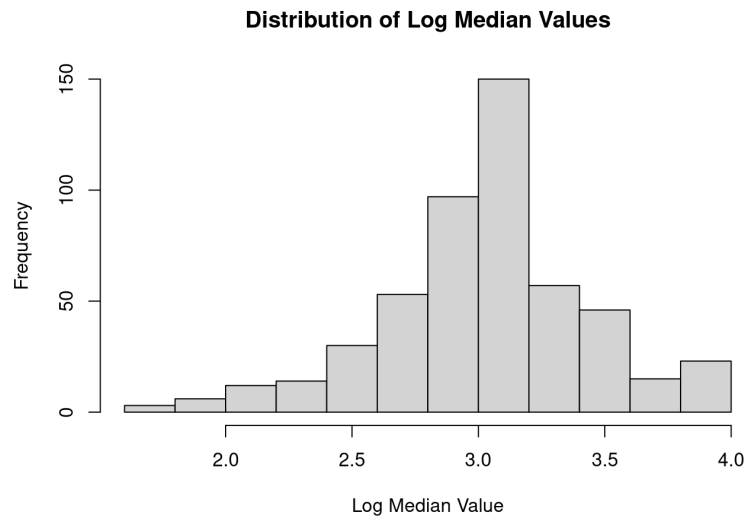


Figure 2: Log Median Housing Price Distribution

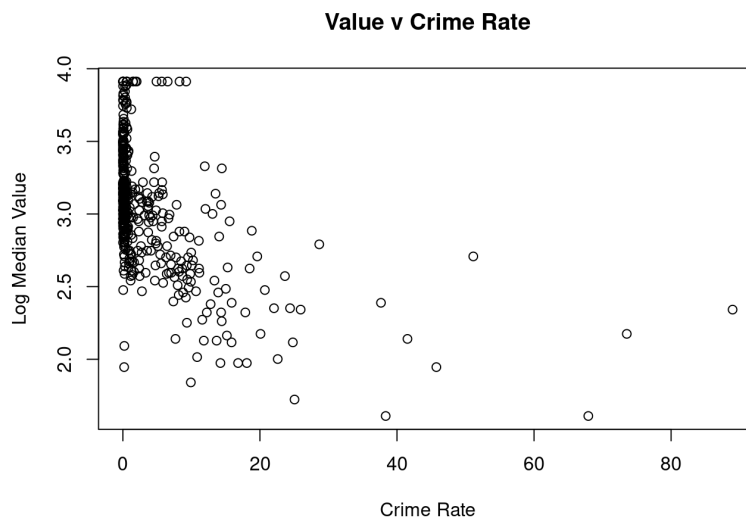


Figure 3: Log Median Price vs Crime Rate

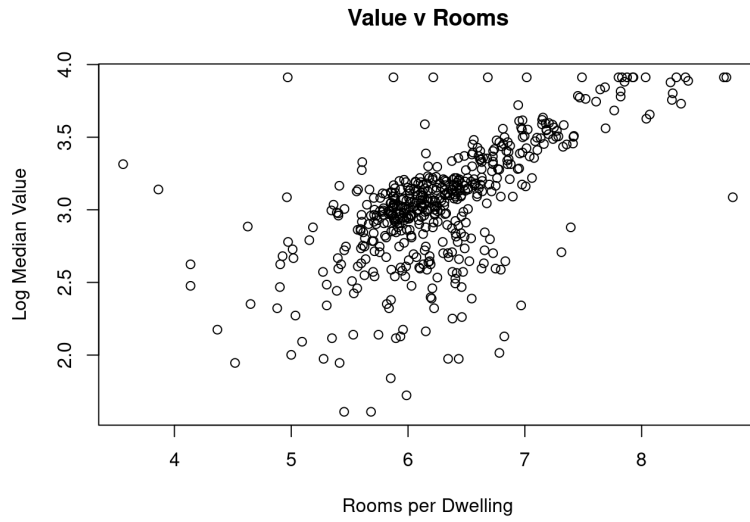


Figure 4: Log Median Price vs Rooms per Dwelling



Figure 5: Log Median Price vs Log Weighted Distance to Employment Centers

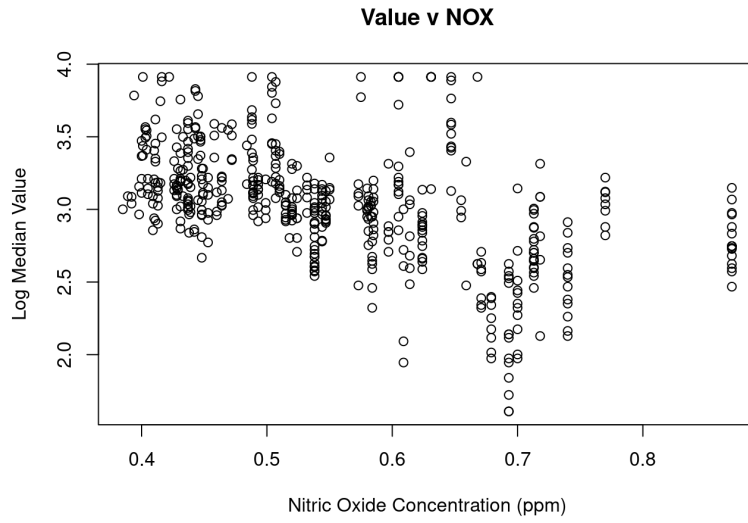


Figure 6: Log Median Price vs NOX

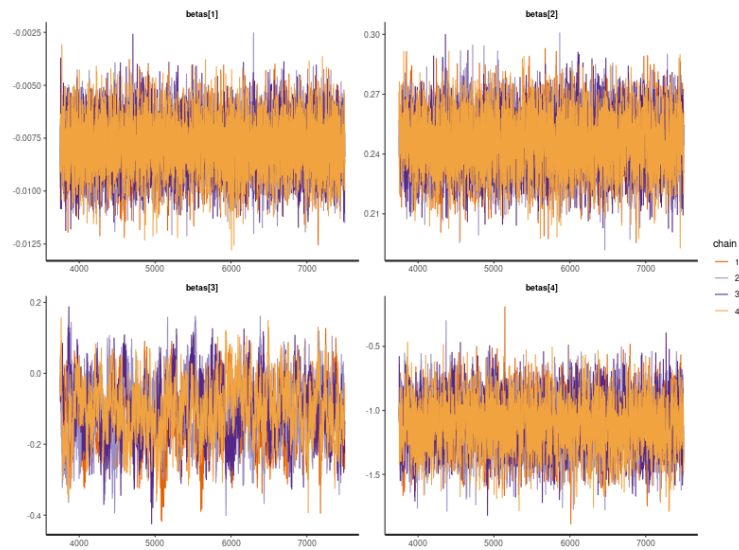


Figure 7:  $\vec{\beta}$  Traceplots for 4 Feature Model

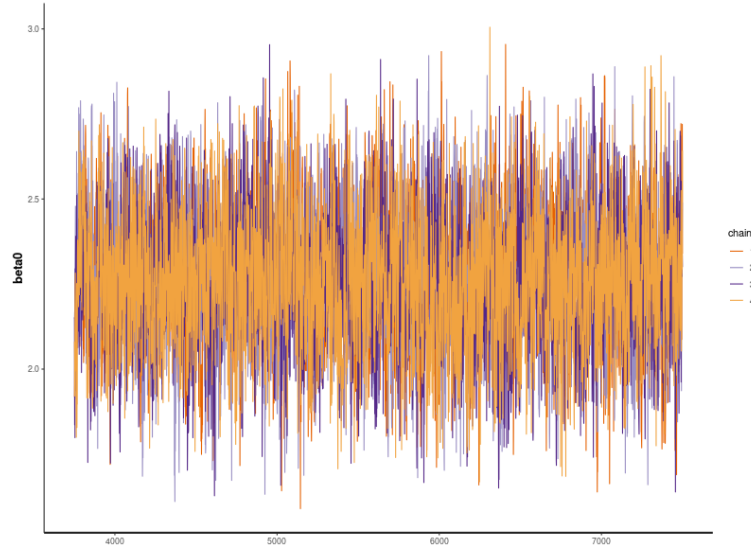


Figure 8:  $\beta_0$  Traceplot for 4 Feature Model

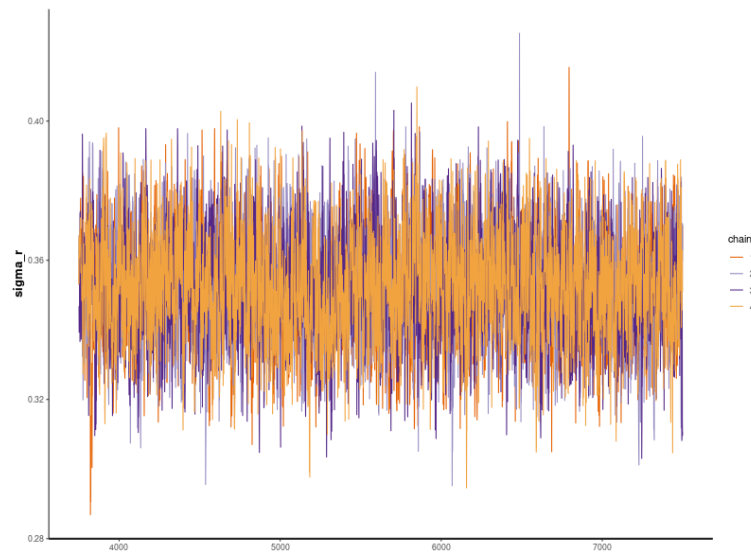


Figure 9:  $\sigma_\tau$  Traceplot for 4 Feature Model

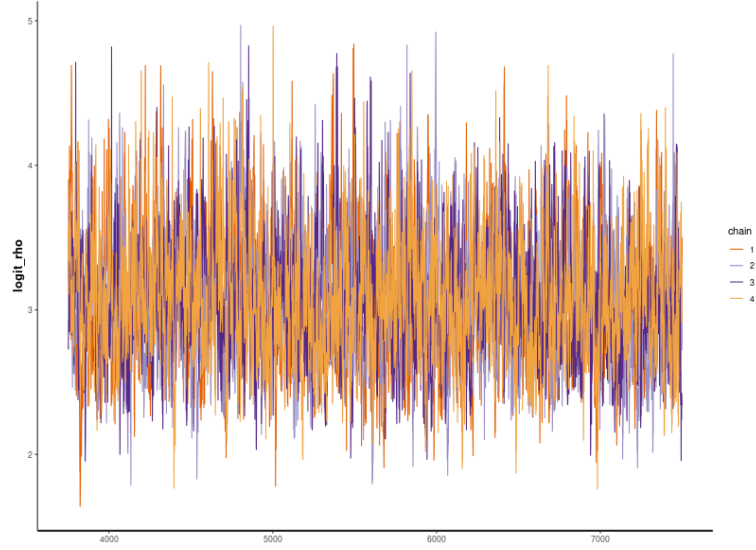


Figure 10:  $r$  Traceplot for 4 Feature Model

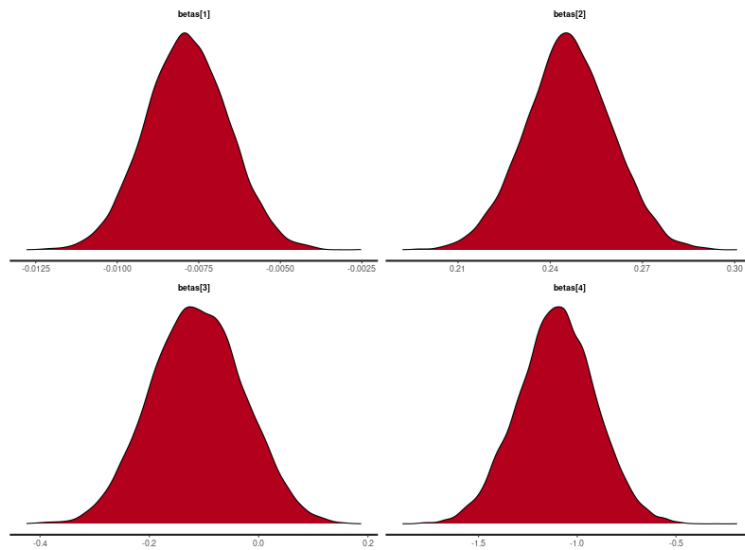


Figure 11:  $\vec{\beta}$  Densities for 4 Feature Model

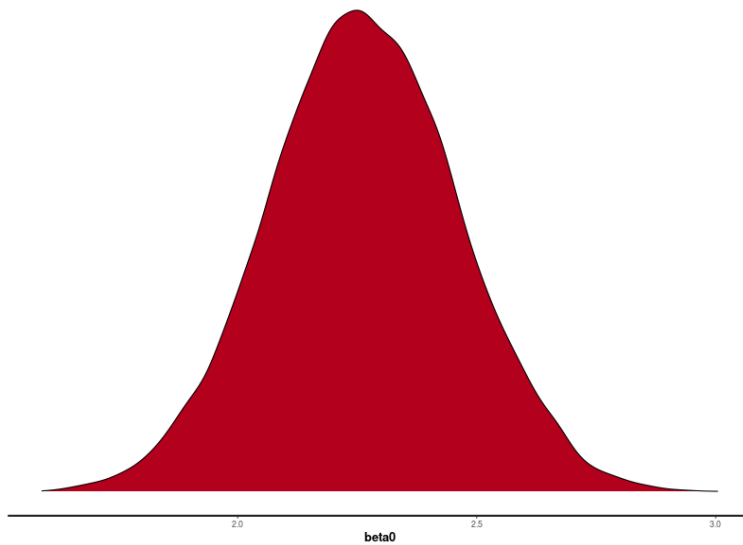


Figure 12:  $\beta_0$  Density for 4 Feature Model

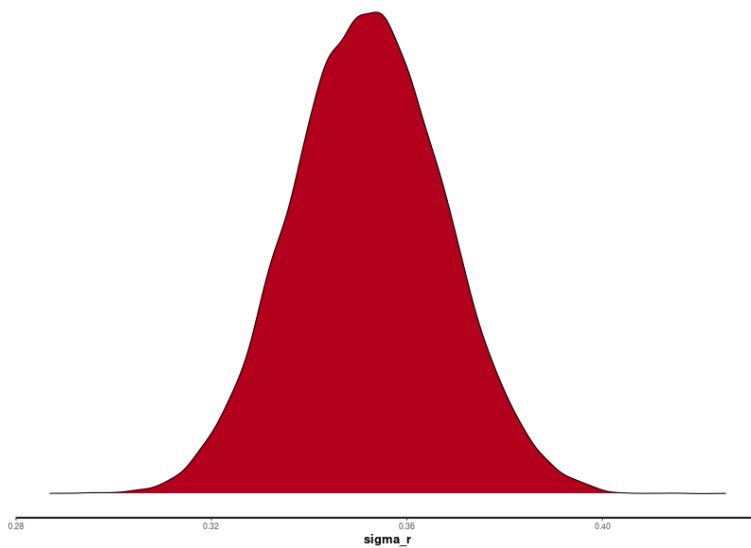


Figure 13:  $\sigma_r$  Density for 4 Feature Model

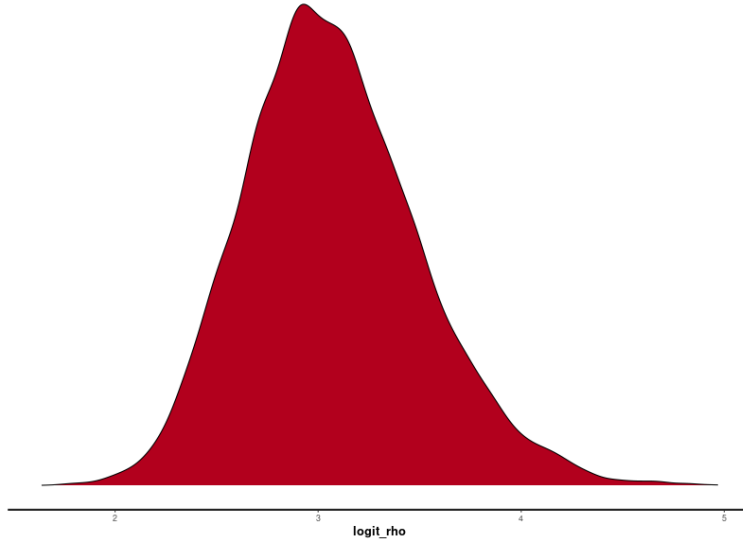


Figure 14:  $r$  Density for 4 Feature Model

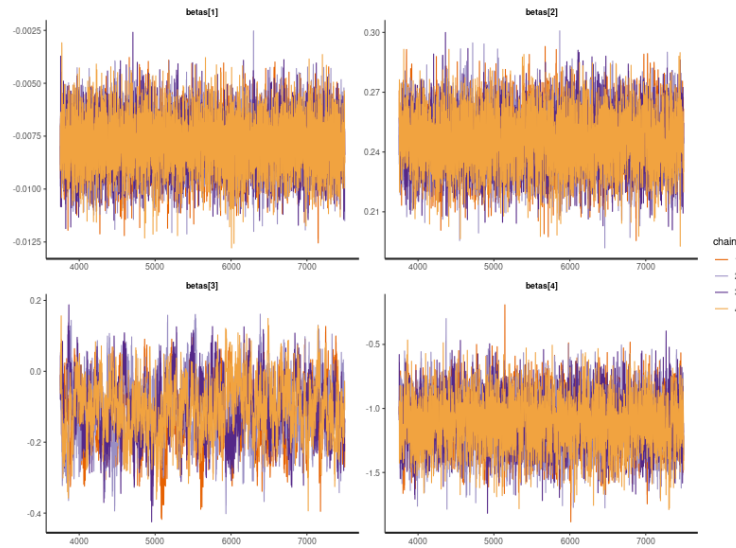


Figure 15:  $\vec{\beta}$  Traceplots for 3 Feature Model



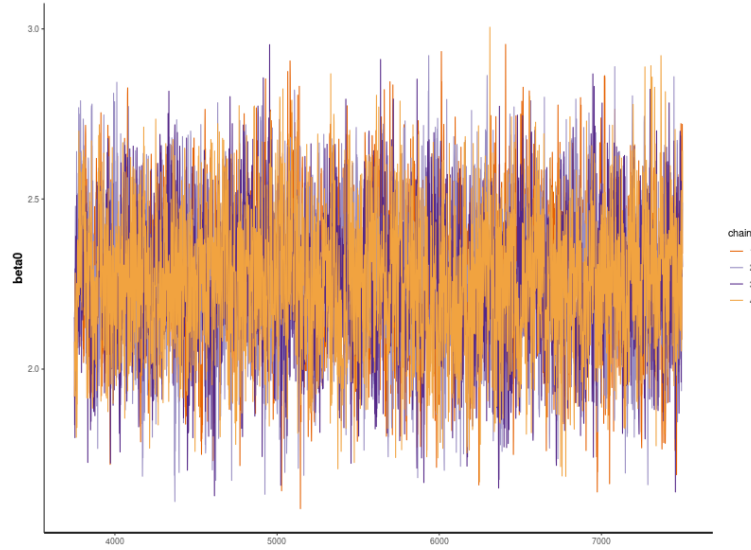


Figure 16:  $\beta_0$  Traceplot for 3 Feature Model

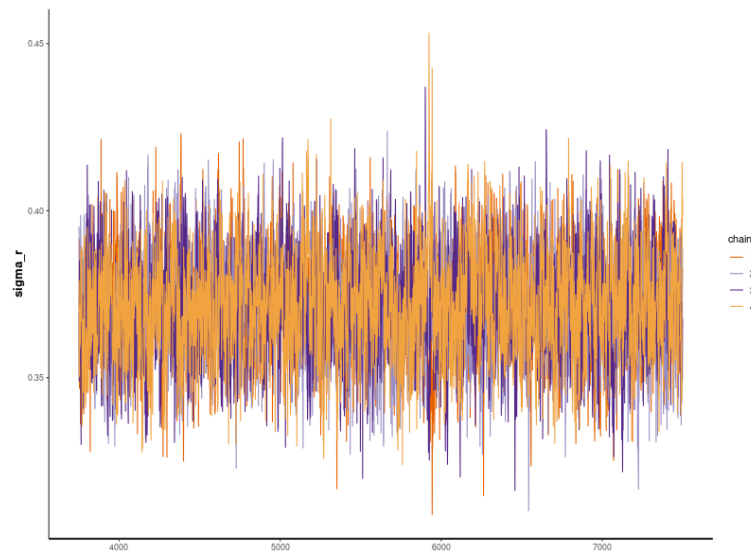


Figure 17:  $\sigma_r$  Traceplot for 3 Feature Model

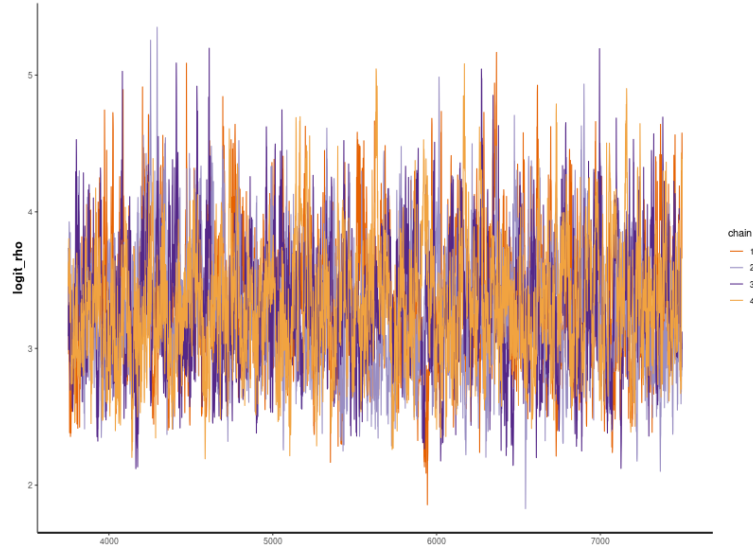


Figure 18:  $r$  Traceplot for 3 Feature Model

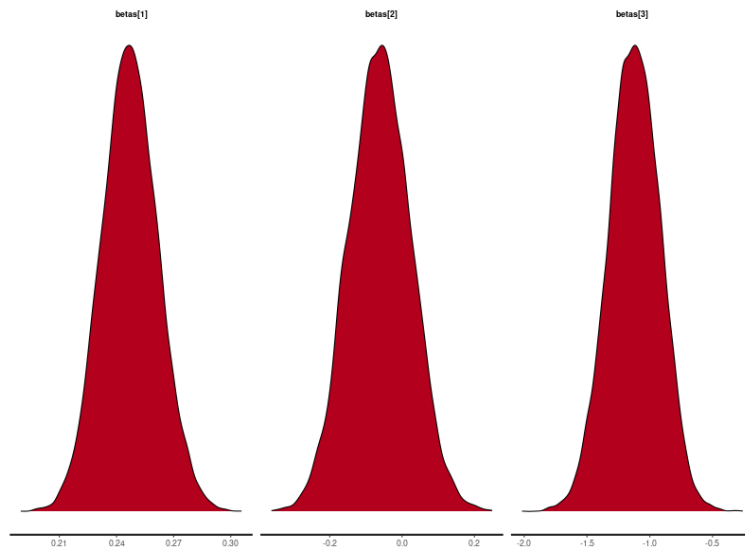


Figure 19:  $\vec{\beta}$  Densities for 3 Feature Model

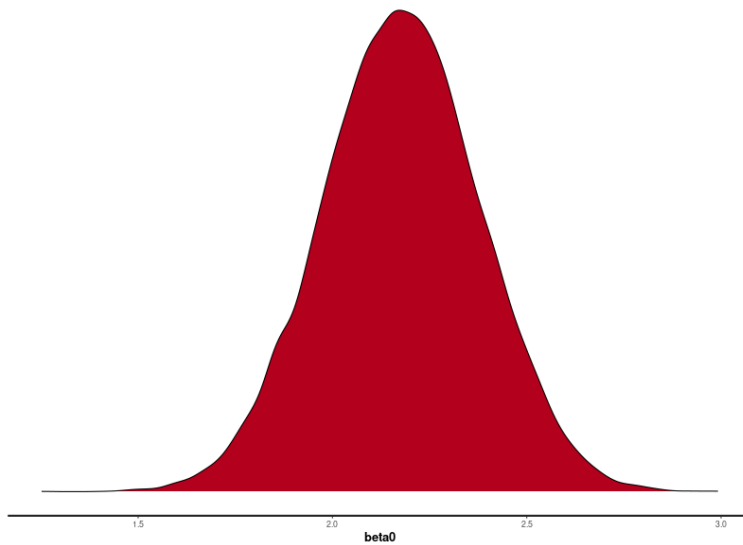


Figure 20:  $\beta_0$  Density for 3 Feature Model

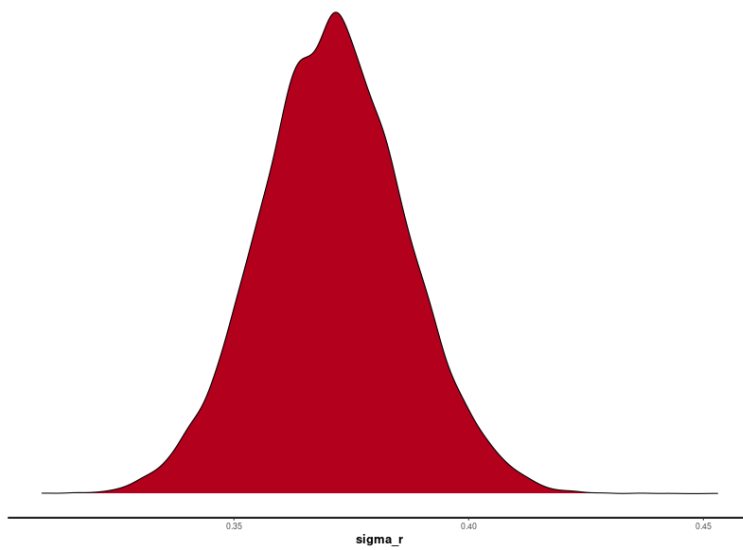


Figure 21:  $\sigma_r$  Density for 3 Feature Model

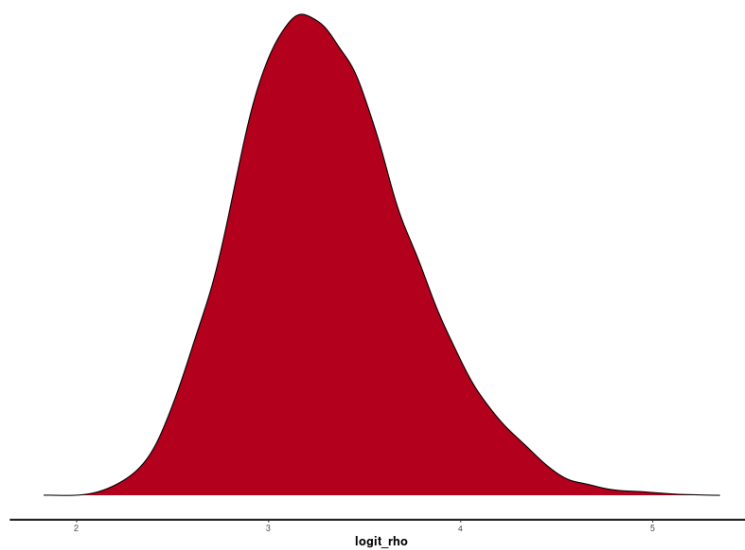


Figure 22:  $r$  Density for 3 Feature Model