Prepared by group 3

Bayesian Mini Project

MATTHEW ANTHONY TJAHJADI - 2702404576
JAMES ANDERSEN - 2702350402

Introduction

This project analyzes a dataset from Walmart, containing 6,435 observations across 8 variables that record weekly sales data for multiple stores. The purpose of the analysis is to evaluate factors that influence Weekly Sales.

The dataset includes:

- Store: Identifier for the store (1 to 45).
- Date: Week-ending date of the observation.
- Weekly_Sales: Total sales for the store in the specified week.
- Holiday_Flag: Binary indicator (0 or 1), where 1 indicates a special holiday week.
- Temperature: Average temperature for that week (in Fahrenheit).
- Fuel_Price: Price of fuel per gallon in the region.
- CPI: Consumer Price Index, indicating changes in prices.
- Unemployment: Unemployment rate in the region

Models used

Linear Regression with Uninformative priors

```
model_string <- textConnection("model {
   # Likelihood
   for (i in 1:n) {
      y[i] ~ dnorm(mu[i], tau)
      mu[i] <- alpha + inprod(x[i,], beta[])
      like[i] <- logdensity.norm(y[i], mu[i], tau)
   }

   # Priors
   for (j in 1:p) {
      beta[j] ~ dnorm(0, 0.001)
   }
   alpha ~ dnorm(0, 0.001)
   tau ~ dgamma(0.1, 0.1)
}")
```

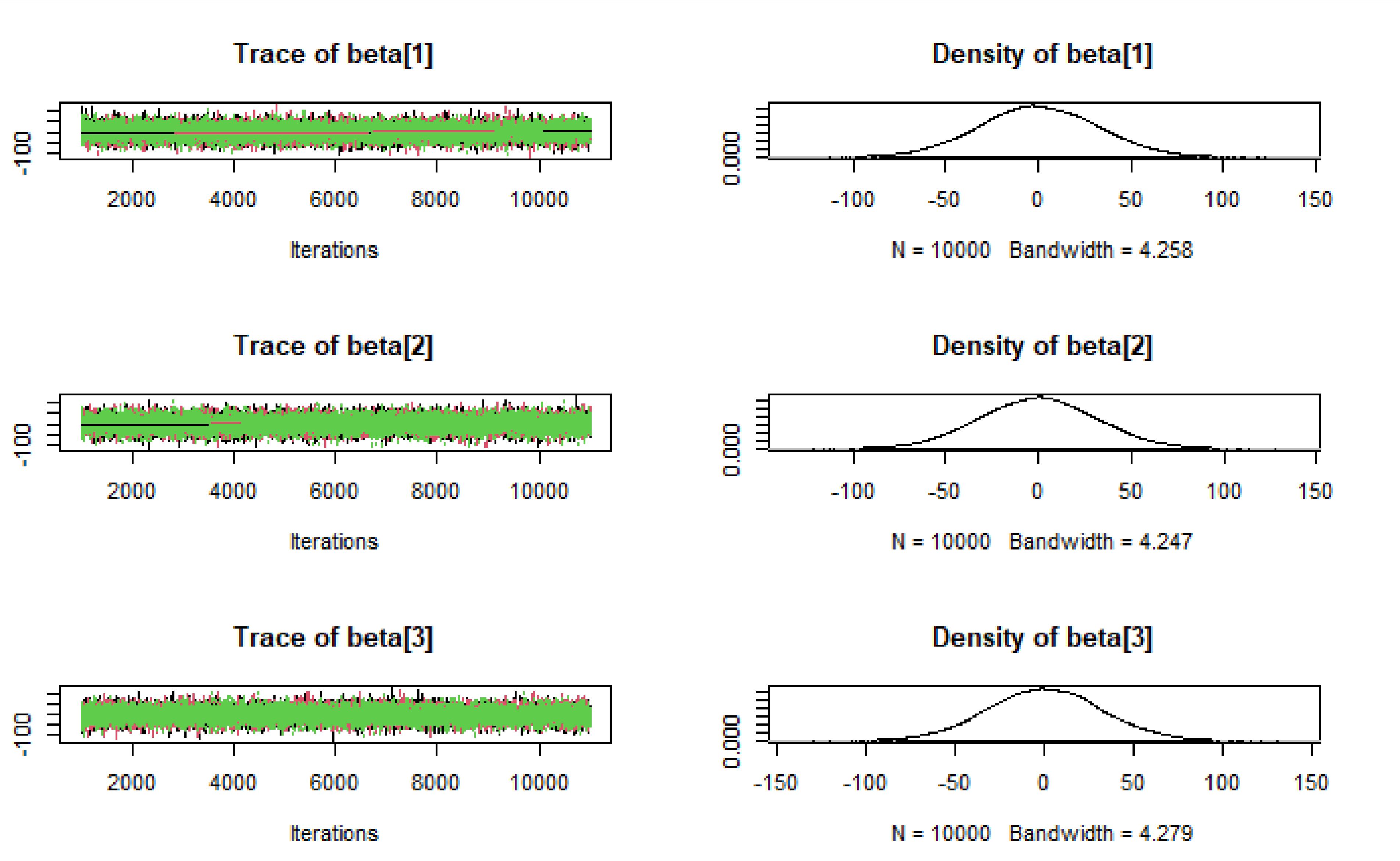
Logistic Regression with Bernoulli likelihod

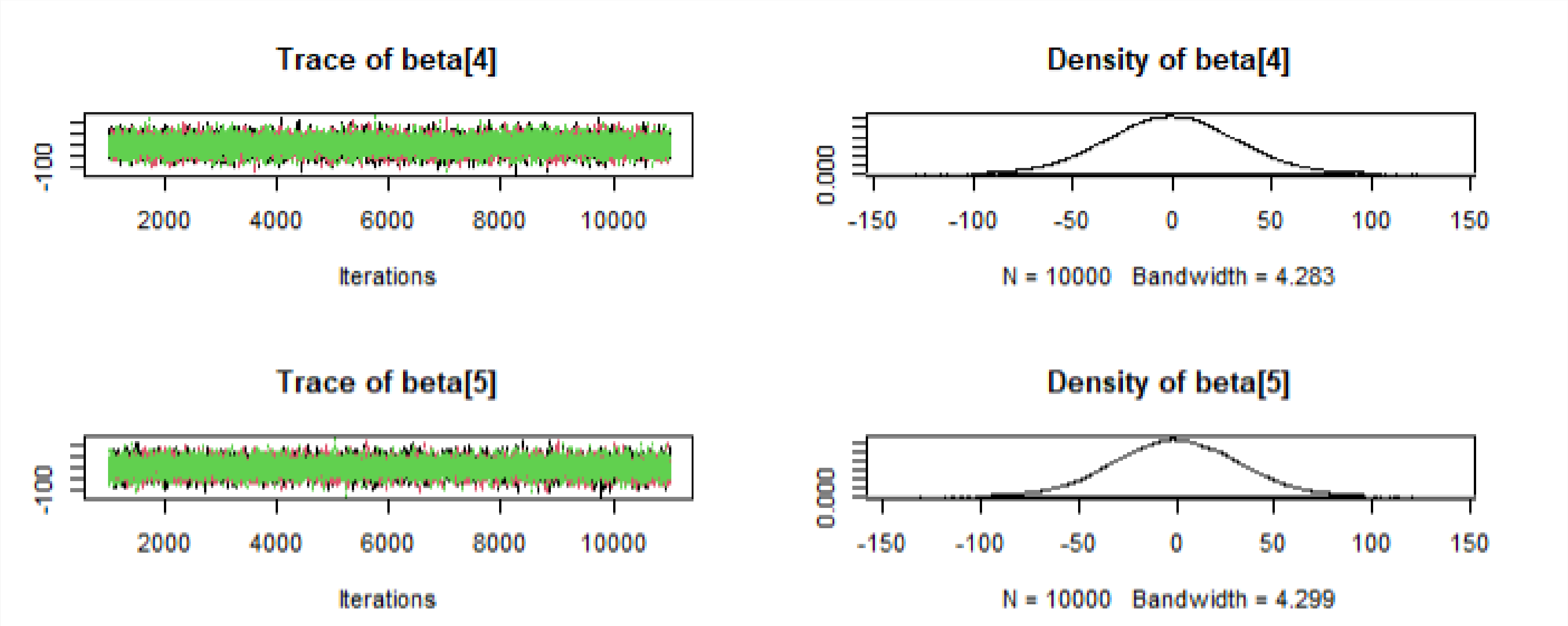
```
model_string <- textConnection("model{
  # Likelihood
  for(i in 1:n){
    y[i] ~ dbern(pi[i])
    logit(pi[i]) <- alpha + inprod(x[i,], beta[])
    like[i] <- logdensity.bern(y[i], pi[i])
  }

  # Priors
  alpha ~ dnorm(0, 0.01)
  for(j in 1:p){
    beta[j] ~ dnorm(0, 0.01)
  }
}")
```

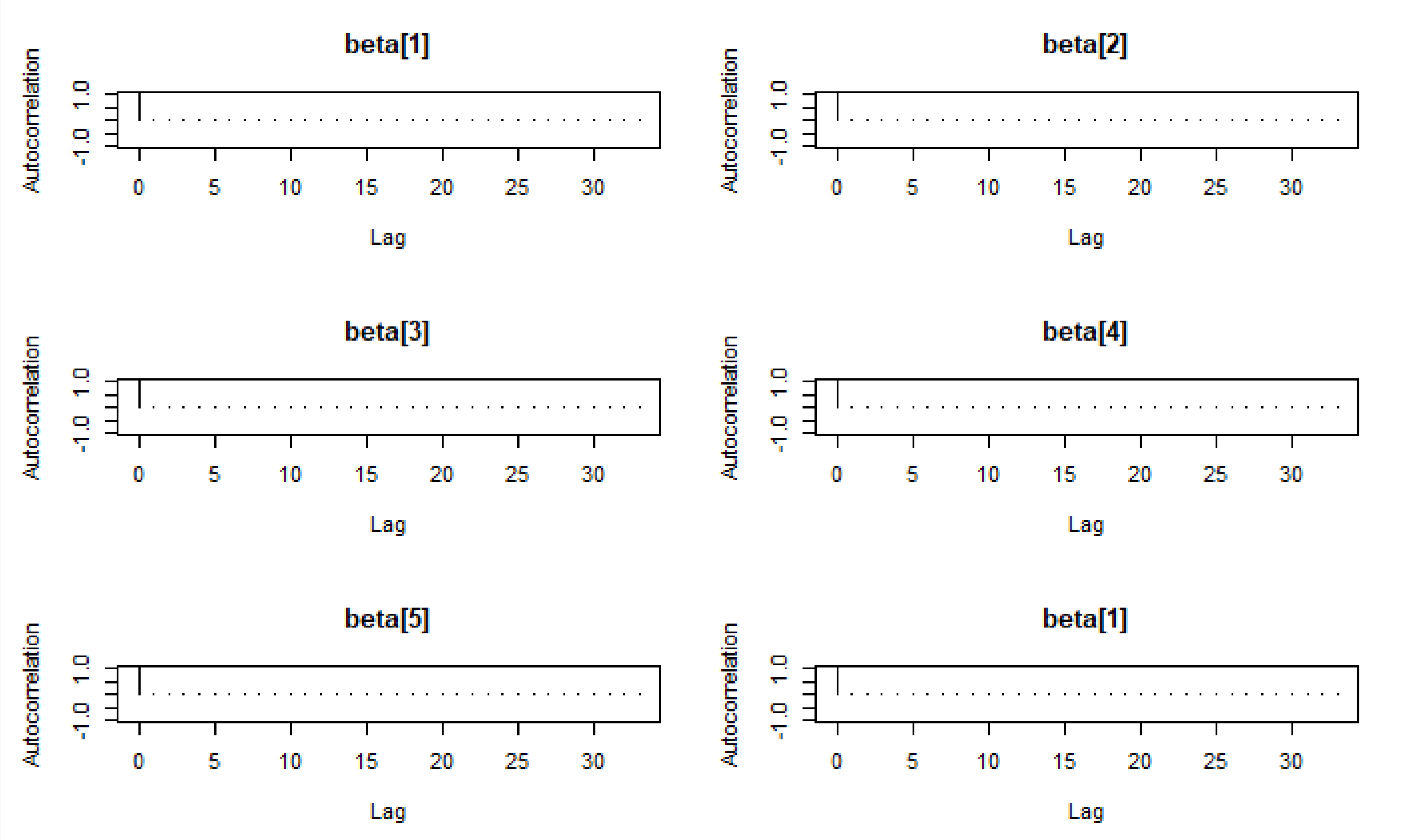
MCMC Settings

```
data <- list(y=y,x=x,n=n,p=p)
params <- c("alpha","beta","like")
burn <- 1000
n.iter <- 10000
n.chains <- 3
```

Convergency of Model 1







Convergency of Model 1

Autocorr

```
, , alpha

            alpha      beta[1]     beta[2]      beta[3]      beta[4]    beta[5]
Lag 1 0.0003419879 -0.0004675971 0.006864755 -0.003229136 -0.005137743 0.02106169

, , beta[1]

          alpha    beta[1]    beta[2]     beta[3]    beta[4]      beta[5]
Lag 1 -0.0178868 0.02050194 0.00967554 -0.01014493 0.01174759 -0.003398033

, , beta[2]

           alpha     beta[1]     beta[2]      beta[3]      beta[4]      beta[5]
Lag 1 0.001133542 0.008975517 0.003551694 0.0009107969 -0.005226773 -0.007214513

, , beta[3]

           alpha     beta[1]     beta[2]     beta[3]     beta[4]    beta[5]
Lag 1 0.000219442 0.004104976 0.002321614 0.001932869 -0.01281911 0.01161244

, , beta[4]

            alpha   beta[1]     beta[2]     beta[3]      beta[4]      beta[5]
Lag 1 -0.003648812 0.0186348 -0.01344082 0.009636201 -0.007016942 -0.006807325

, , beta[5]

          alpha    beta[1]       beta[2]     beta[3]     beta[4]     beta[5]
Lag 1 0.01299603 0.00043296 -0.0005497778 0.001993984 0.009100789 -0.01783862
```

Gelman

```
        Point est. Upper C.I.
alpha            1          1
beta[1]          1          1
beta[2]          1          1
beta[3]          1          1
beta[4]          1          1
beta[5]          1          1
```

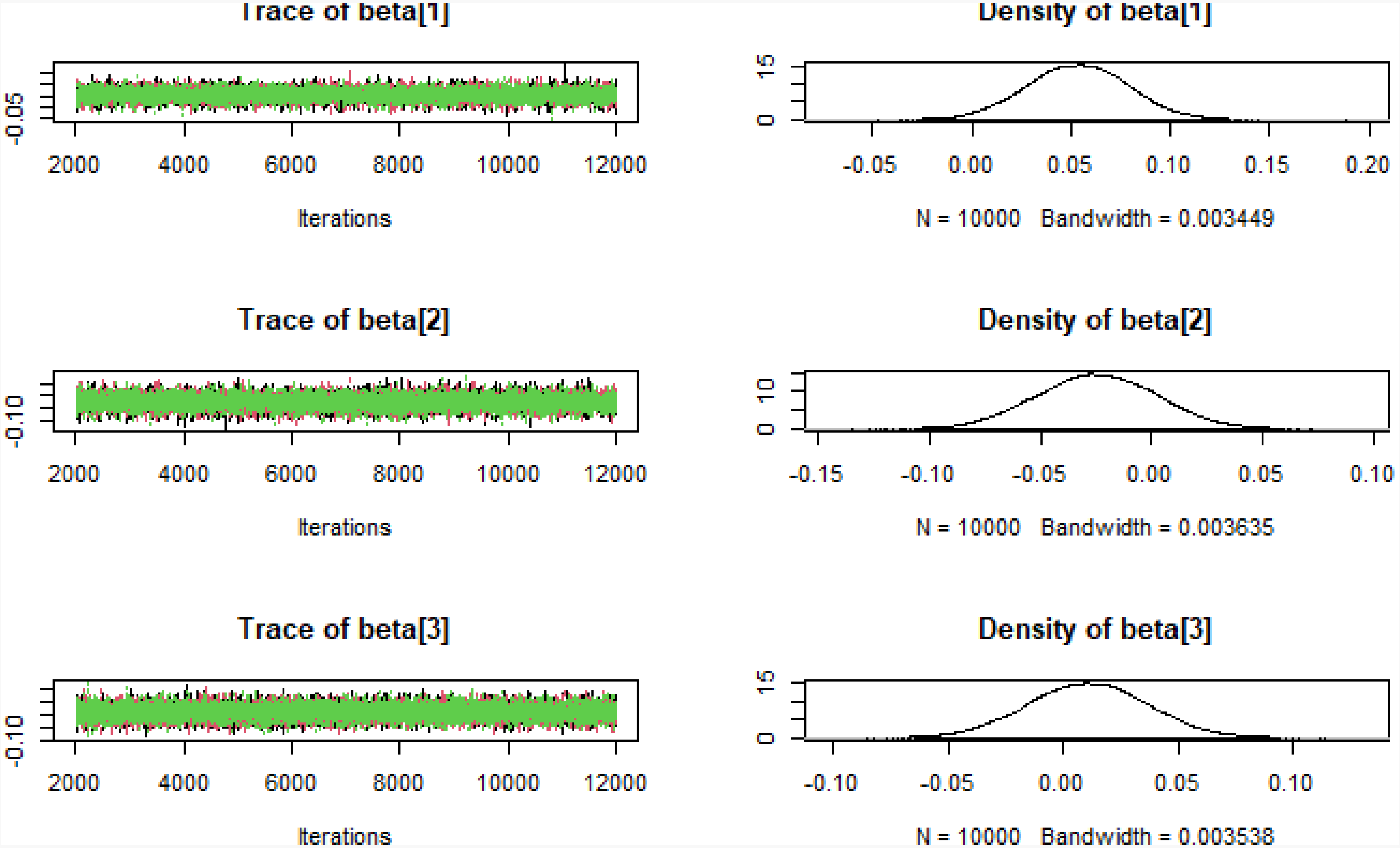
Geweke

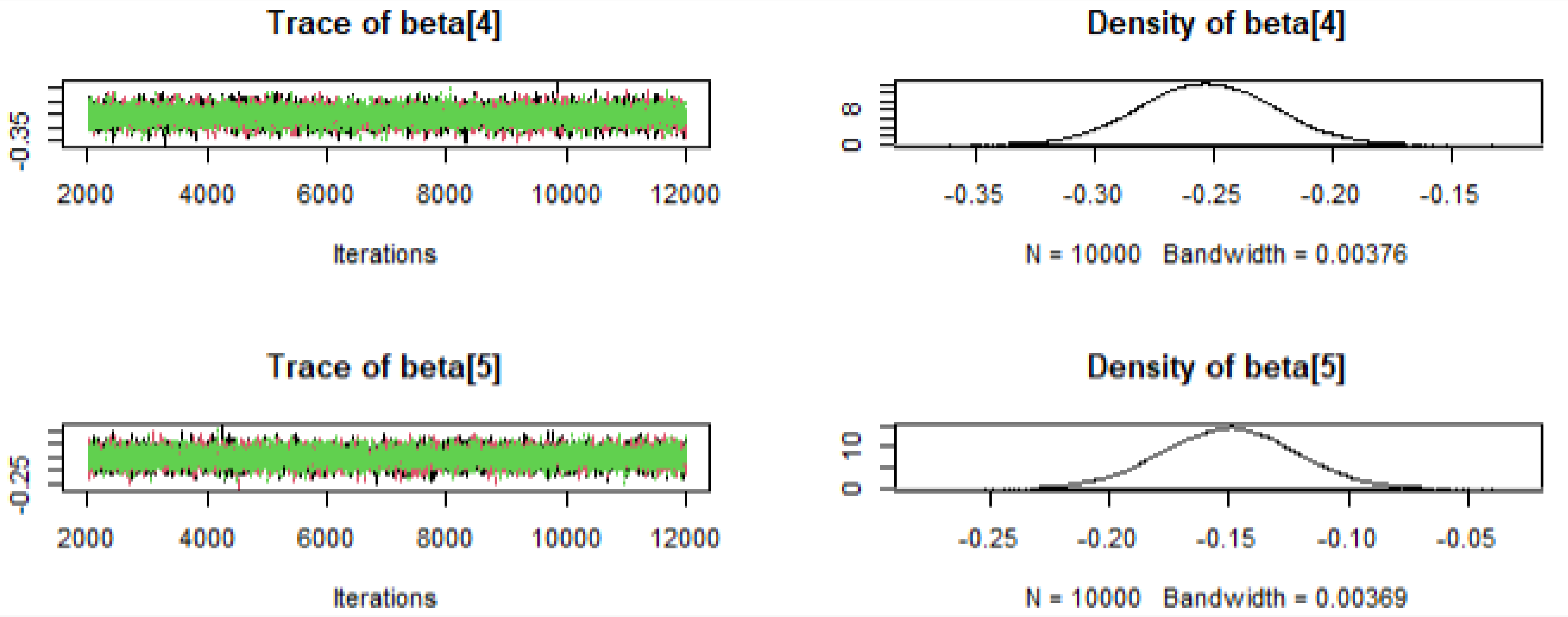
```
  alpha beta[1] beta[2] beta[3] beta[4] beta[5]
-0.5568  0.5438  1.0132  1.2498  1.4775  1.2603
```

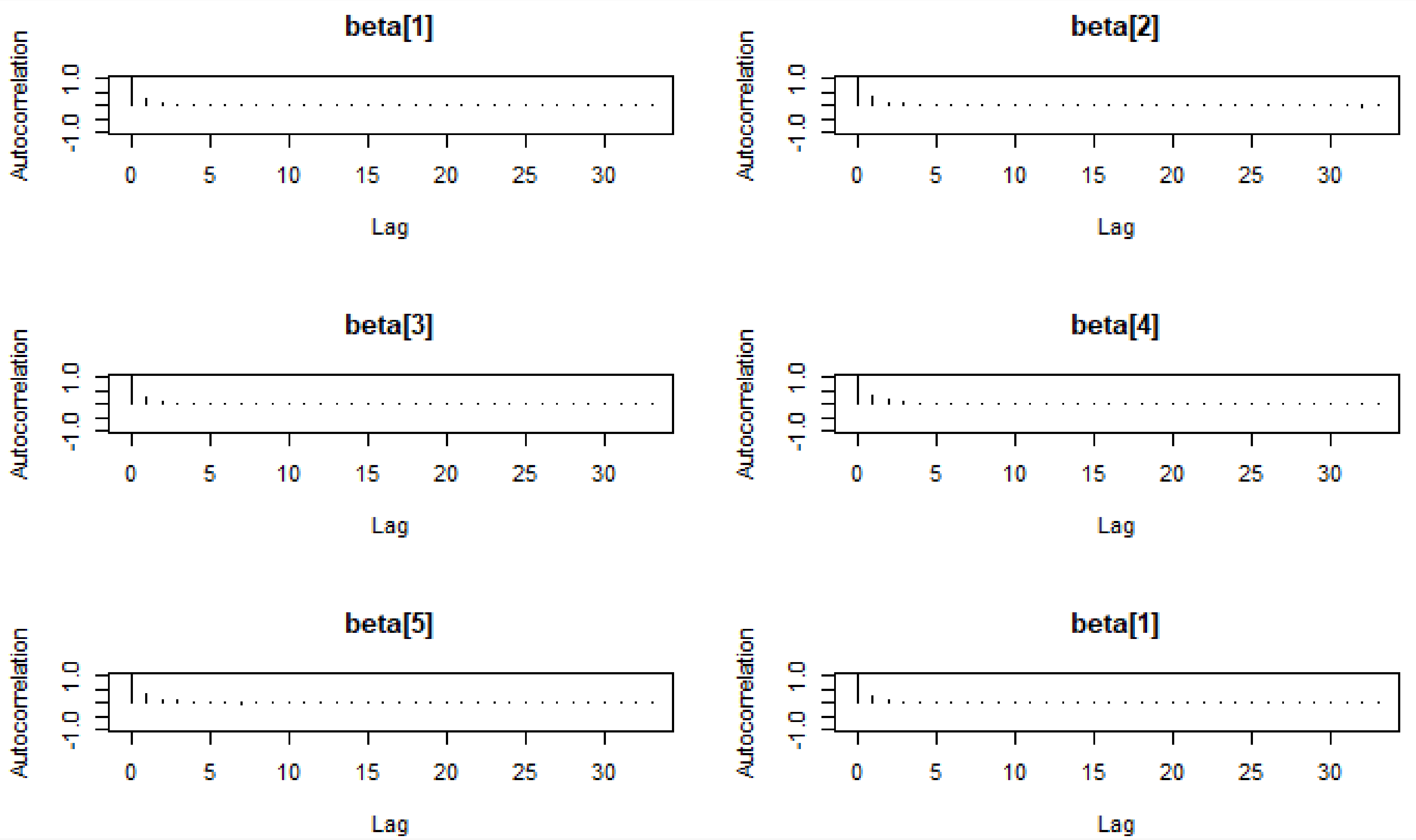
ESS

```
   alpha  beta[1]  beta[2]  beta[3]  beta[4]  beta[5]
30000.00 29597.24 30000.00 30273.72 31095.25 30896.54
```

Convergency of Model 2







Convergency of Model 2

Autocorr

```
, , alpha

          alpha     beta[1]     beta[2]    beta[3]     beta[4]     beta[5]
Lag 1 0.2211348 0.004793978 0.008081352 0.00148246 -0.02839639 -0.02217336

, , beta[1]

             alpha   beta[1]   beta[2]    beta[3]     beta[4]     beta[5]
Lag 1 -0.005734285 0.2601626 0.1682365 0.02596745 -0.04511806 -0.03047953

, , beta[2]

           alpha    beta[1]   beta[2]    beta[3]    beta[4]    beta[5]
Lag 1 0.01471342 0.08223682 0.2930216 -0.2052587 -0.2543852 -0.1718834

, , beta[3]

            alpha     beta[1]   beta[2]   beta[3]   beta[4]   beta[5]
Lag 1 -0.02572107 0.007582148 -0.118226 0.2881494 0.2380904 0.1211735

, , beta[4]

           alpha    beta[1]    beta[2]   beta[3]   beta[4]   beta[5]
Lag 1 -0.0205865 -0.0217674 -0.1417478 0.1234242 0.3620204 0.3461183

, , beta[5]

            alpha      beta[1]  beta[2]    beta[3]   beta[4]   beta[5]
Lag 1 -0.01339151 -0.004580519 -0.08735 0.07866207 0.1933323 0.3322198
```

Gelman

```
        Point est. Upper C.I.
alpha            1          1
beta[1]          1          1
beta[2]          1          1
beta[3]          1          1
beta[4]          1          1
beta[5]          1          1
```

Geweke

```
   alpha  beta[1]  beta[2]  beta[3]  beta[4]  beta[5]
-0.48454  0.02163  0.40315 -0.56351  0.64132 -0.06699
```

ESS

```
   alpha  beta[1]  beta[2]  beta[3]  beta[4]  beta[5]
19395.35 17709.13 14747.82 16090.46 12810.22 13785.10
```

Model Comparison

Model 1

```
[1] "DIC Model1"
Mean deviance:  198300
penalty 1.009
Penalized deviance: 198301
```

```
$waic
[1] 918497.4

$lppd
[1] -459248.2

$p_waic
[1] 0.5247965
```

Model 2

```
[1] "DIC Model3"
Mean deviance:  8815
penalty 6.032
Penalized deviance: 8821
```

```
$waic
[1] 40571.66

$lppd
[1] -20279.85

$p_waic
[1] 5.985729
```

Model Interpretation

Based off the Information criteria and bayes factor, the second model, Logistic Regression is highly favored as the better model

```
Iterations = 2001:12000
Thinning interval = 1
Number of chains = 3
Sample size per chain = 10000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

            Mean      SD  Naive SE Time-series SE
beta[1]  0.05316 0.02557 0.0001476      0.0001924
beta[2] -0.02462 0.02696 0.0001556      0.0002220
beta[3]  0.01077 0.02624 0.0001515      0.0002068
beta[4] -0.25309 0.02788 0.0001610      0.0002465
beta[5] -0.15010 0.02736 0.0001580      0.0002331

2. Quantiles for each variable:

            2.5%       25%      50%       75%    97.5%
beta[1]  0.00289  0.035987  0.05332  0.070589  0.10330
beta[2] -0.07692 -0.042860 -0.02467 -0.006107  0.02786
beta[3] -0.04062 -0.006951  0.01078  0.028399  0.06236
beta[4] -0.30752 -0.271953 -0.25317 -0.234190 -0.19845
beta[5] -0.20418 -0.168623 -0.14996 -0.131682 -0.09724
```

We can conclude from the means and the confidence interval of each covariate, that beta4 (Consumer Price Index) and beta5 (Unemployment) negatively affects our target variable

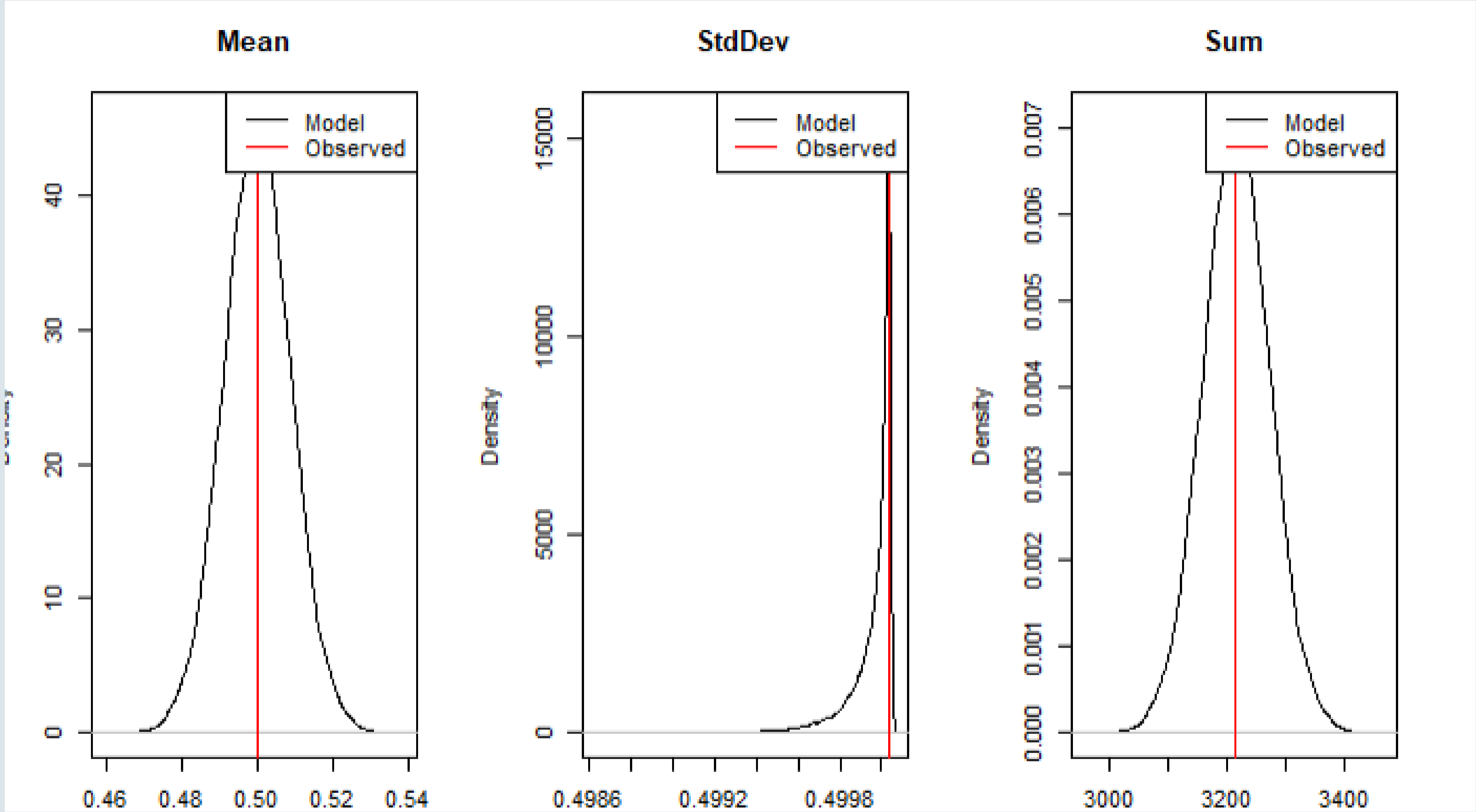
While beta1 (presence of holiday flag) positively affects our target variable

Model Interpretation

Basically, more holiday flags more sales.

Higher CPI and Unemployment less sales

Posterior Predictive Checks



For each statistic, our model is able to replicate the values of observed data well. This can be seen by the red that crosses through the centre of every distribution