Bayesian learning and Monte Carlo Simulations: fnal project

Inês Jorge da Silva e Ferreira and André Guilherme dos Santos Neto

2022-09-06

Problem Statement

Forest fires are a major environmental issue, creating economical and ecological damage while endangering animal and human lives. Therefore, fast detection is a key element for controlling such phenomenon. In this project we modelled the burned area of the forest as a function of the variables x-axis coordinate (from 1 to 9), y-axis coordinate (from 1 to 9), Month of the year (January to December) day Day of the week (Monday to Sunday), FFMC code, DMC code, DC code, ,ISI index ,Outside temperature (in $\circ$C) ,Outside relative humidity (in %) wind ,Outside wind speed (in km/h), Outside rain (in mm/m2) to assist in prediction of the future eventuality.

Data Description

The Forest Fire Index (FWI) is the Canadian system for rating fire danger and it includes six components: Fine Fuel Moisture Code (FFMC), Duff Moisture Code (DMC), Drought Code (DC), Initial Spread Index (ISI), Buildup Index (BUI) and FWI. The first three are related to fuel codes, the FFMC denotes the moisture content surface litter and influences ignition and fire spread while DMC and DC represent the moisture content of shallow and deep organic layers which affect fire intensity. The ISI is a score that correlates with fire velocity spread while BUI represents the amount of available fuel. The FWI index is an indicator of fire intensity. Although different scales are used for each of the FWI elements, high values suggest more severe burning conditions. Also, the fuel moisture codes require a memory (time lag) of past weather conditions i.e., 16 hours for FFMC, 12 days for DMC and 52 days for DC. This data contains 13 variables and 517 observations as described above.

For the purpose of reading correctly and data analysis we highlight the 13 data variables as follows

1. **X** (int) - x-axis spatial coordinate within the Montesinho park map: 1 to 9
2. **Y** (int) - y-axis spatial coordinate within the Montesinho park map: 2 to 9
3. **month** (chr) - month of the year: "jan" to "dec"
4. **day** (chr) - day of the week: "mon" to "sun"
5. **FFMC** (num) - Fine Fuel Moisture Code index from the FWI (Fire Weather Index) system: 18.7 to 96.20
6. **DMC** (num) - Duff Moisture Code index from the FWI system: 1.1 to 291.3
7. **DC** (num) - Drought Code index from the FWI system: 7.9 to 860.6
8. **ISI** (num) - Initial Spread Index from the FWI system: 0.0 to 56.10
9. **temp** (num) - temperature in Celsius degrees: 2.2 to 33.30

10. **RH** (int) - Relative Humidity in %: 15.0 to 100
11. **wind** (num) - wind speed in km/h: 0.40 to 9.40
12. **rain** (num) - outside rain in mm/m2 : 0.0 to 6.4
13. **area** (num) - the burned area of the forest (in ha): 0.00 to 1090.84

Model Specification

The model is represented as

$Y_i|\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta 9, \sigma \overset{\text{ind}}{\sim} Normal(\beta_0 + \beta_1 x_{i,FFMC} + \beta_2 x_{i,DMC} + \beta_3 x_{i,DC} + \beta_4 x_{i,ISI} + \beta_5 x_{i,temp} + \beta_6 x_{i,RH} + \beta_7 x_{i,windspeed} + \beta_8 x_{i,rain}, \sigma)$
where $\beta_i' s$ are the coefficients of the variables

Likelihood

The likelihood function of the model is given by;

$y_i|\mu_i, \sigma \overset{\text{ind}}{\sim} Normal(\mu_i, \sigma), i = 1, \ldots, n,$ μ_i is the response which linear to the coefficient parameter. Where given the data point, the equation is represented as

$N(y_i|Xw, \sigma^2 I_n) = (2\pi\sigma^2)^{-n/2} exp(-\frac{1}{2\sigma^2}(y - Xw)^T(y - Xw))$
where $w = (X^T X)^{-1} X^T y$

Priors

Since we have limited information on the regression parameters we assume weakly informative priors this . Assuming independence, the prior density for the parameters $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_8, \sigma$

where $\beta 0 \sim Normal(mu_0, sigma_0)$, $\beta 1 \sim Normal(mu_1, sigma_1)$, $\beta 2 \sim Normal(mu_2, sigma_2)$, $\beta 3 \sim Normal(mu_3, sigma_3)$, $\beta 4 \sim Normal(mu_4, sigma_4)$, $\beta 5 \sim Normal(mu_5, sigma_5)$, $\beta 6 \sim Normal(mu_6, sigma_6)$, $\beta 7 \sim Normal(mu_7, sigma_7)$, $\beta 8 \sim Normal(mu_8, sigma_8)$, $\beta 9 \sim Normal(mu_9, sigma_9)$, and the precision parameter $\phi = 1/\sigma^2$, the inverse of the variance σ^2, is $Gamma(a, b)$.

Because we have little information about the location of the regression parameters $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8$ and β_9, we assigns the respective prior means to be 0 and the prior standard deviations as the standard deviations of the data varibales. And the values of the hyperparameters a and b are assumed as $a = b = 0.001$ for Gamma prior the precision is given by $\phi = 1/\sigma^2$.

```
## Compiling rjags model...
## Calling the simulation using the rjags method...
## Note: the model did not require adaptation
## Burning in the model for 2000 iterations...
## Running the model for 5000 iterations...
## Simulation complete
## Calculating summary statistics...
## Calculating the Gelman-Rubin statistic for 11 variables....
## Finished running the simulation
```

```
##
## JAGS model summary statistics from 10000 samples (chains = 2; adapt+burnin = 3000):
##
##       Lower95    Median Upper95     Mean     SD Mode    MCerr MC%ofSD SSeff
## beta0   -18.6    -0.285    18.7   -0.302   9.56   --    0.471     4.9   412
## beta1  -0.523   -0.0388   0.506  -0.0342  0.263   --   0.0301    11.4    77
## beta2 -0.0191     0.107   0.242    0.107 0.0671   --  0.00288     4.3   541
## beta3 -0.0514   -0.0162  0.0217  -0.0162 0.0186   -- 0.000999     5.4   345
## beta4   -1.89      -0.5   0.803   -0.513  0.688   --   0.0228     3.3   907
## beta5  -0.722     0.848    2.35    0.862  0.802   --   0.0685     8.5   137
## beta6  -0.594    -0.186   0.217   -0.186  0.213   --    0.015     7.1   201
## beta7   -1.46     0.919    3.23    0.923    1.2   --   0.0324     2.7  1371
## beta8   -0.59 -0.000324   0.555 -0.00268  0.294   --  0.00294       1 10000
## beta9  -0.575      0.91    2.35      0.9  0.754   --   0.0291     3.9   673
## sigma    59.8      63.5    67.6     63.6   1.99   --   0.0212     1.1  8857
##
##           AC.10 psrf
## beta0     0.434 1.04
## beta1     0.858 1.04
## beta2     0.347    1
## beta3     0.504    1
## beta4     0.179    1
## beta5     0.759 1.01
## beta6     0.558 1.01
## beta7    0.0716    1
## beta8 -0.000207    1
## beta9     0.267    1
## sigma -0.000526    1
##
## Total time taken: 10.4 seconds
```

Posterior Analysis

When the model includes all the variables the following posterior results were generated, From the table we observe that no variable was predictor which was significant since the parameters confidence intervals covers zeros, this indicates that not all the variables are helpful in predicting the burn area.

Table 1: Table for parameter confidence intervals, median and mean

	Lower95	**Upper95**	**Median**	**Mean**	**SD**
beta0	-18.1	20.4	0.665	0.524	9.92
beta1	-0.584	0.415	-0.0727	-0.0729	0.257
beta2	-0.0237	0.247	0.112	0.113	0.0698
beta3	-0.0518	0.019	-0.0176	-0.0175	0.0184
beta4	-1.9	0.83	-0.512	-0.509	0.697
beta5	-0.533	2.56	0.913	0.932	0.766
beta6	-0.553	0.244	-0.168	-0.165	0.203
beta7	-1.49	3.33	0.945	0.94	1.23
beta8	-0.582	0.569	0.00202	-0.00103	0.296
beta9	-0.553	2.41	0.968	0.963	0.759
sigma	59.8	67.7	63.6	63.6	2.03

```
## Compiling rjags model...
## Calling the simulation using the rjags method...
## Note: the model did not require adaptation
## Burning in the model for 2000 iterations...
## Running the model for 5000 iterations...
## Simulation complete
## Calculating summary statistics...
## Calculating the Gelman-Rubin statistic for 6 variables....
## Finished running the simulation
```

```
##
## JAGS model summary statistics from 10000 samples (chains = 2; adapt+burnin = 3000):
##
##       Lower95    Median Upper95      Mean    SD Mode   MCerr MC%ofSD SSeff
## beta0   -17.3    -0.674    16.8    -0.624  8.64   --   0.358     4.1   583
## beta5   0.188     0.903    1.65       0.9 0.366   --  0.0129     3.5   804
## beta6  -0.416    -0.135   0.138    -0.137 0.143   -- 0.00449     3.1  1008
## beta7   -1.56     0.641    2.78     0.641  1.11   --  0.0271     2.4  1667
## beta8   -0.59 -0.000293   0.568 -0.00268 0.296   -- 0.00296       1 10000
## sigma    59.6      63.5    67.3      63.5  1.98   --    0.02       1  9759
##
##          AC.10 psrf
## beta0    0.284    1
## beta5    0.187    1
## beta6    0.102    1
## beta7 -0.00844    1
## beta8  0.00457    1
## sigma   0.0061    1
##
## Total time taken: 4.5 seconds
```

When the model was fitted with the independent variables temperature, relative humidity, wind speed and rain the following posterior results were found, From the table we observe that temperature variable is the only significant predictor since the parameter confidence intervals (0.199,0.916) do not cover zero, this indicates that temperature is helpful in predicting the burn area.

Table 2: Table for parameter confidence intervals, median and mean

	Lower95	**Upper95**	**Median**	**Mean**	**SD**
beta0	-19.4	16.2	-1.01	-1.06	9.02
beta5	0.199	1.63	0.916	0.918	0.366
beta6	-0.405	0.154	-0.136	-0.136	0.143
beta7	-1.53	2.9	0.642	0.657	1.14
beta8	-0.581	0.592	-0.0106	-0.00474	0.297
sigma	59.7	67.4	63.5	63.5	1.97

After finding out that temperature is one of the main factor in predicting the burn area, we now fit the new model with the variable. The following table shows the summary posterior results of the selected model with significant variable (temperature).

```
## Compiling rjags model...
## Calling the simulation using the rjags method...
```

```
## Note: the model did not require adaptation
## Burning in the model for 2000 iterations...
## Running the model for 5000 iterations...
## Simulation complete
## Calculating summary statistics...
## Calculating the Gelman-Rubin statistic for 3 variables....
## Finished running the simulation
```

```
##
## JAGS model summary statistics from 10000 samples (chains = 2; adapt+burnin = 3000):
##
##       Lower95 Median Upper95 Mean   SD Mode  MCerr MC%ofSD SSeff  AC.10 psrf
## beta0   -1.36   9.18    19.8 9.12 5.43   --  0.144     2.7  1415  0.061    1
## beta5    -6.8   2.96    12.9 3.02  5.1   --  0.135     2.7  1422 0.0617    1
## sigma      60   63.7    67.7 63.8 1.96   -- 0.0196       1 10000  0.013    1
##
## Total time taken: 1.3 seconds
```

Table 3: Table for parameter confidence intervals, median and mean

	Lower95	**Upper95**	**Median**	**Mean**	**SD**
beta0	-1.49	19.8	8.92	8.91	5.44
beta5	-6.42	13.4	3.11	3.2	5.11
sigma	59.8	67.6	63.7	63.8	2.01

```
## Obtaining DIC samples from 5000 iterations...
```

```
## Mean deviance:  5760
## penalty 8.586
## Penalized deviance: 5768
```

```
## Obtaining DIC samples from 5000 iterations...
```

```
## Mean deviance:  5759
## penalty 3.735
## Penalized deviance: 5763
```

```
## Obtaining DIC samples from 5000 iterations...
```

```
## Mean deviance:  5763
## penalty 1.97
## Penalized deviance: 5765
```

Moodel Comparison

From the table we observe that the model with independent variables temperature, relative humidity, wind speed and rain is the most plausible model to predict the burn area, the model has DIC 5763 compared with the model with all the varibale included DIC 5768 and the model with only temperature as independent varible DIC 5765.

Table 4: Model Comparison

Model	**DIC**
Model1	5768
Model2	5763
Model3	5765

Plots of the posterior distributions

The posterior distribution plots are presented in the appendix

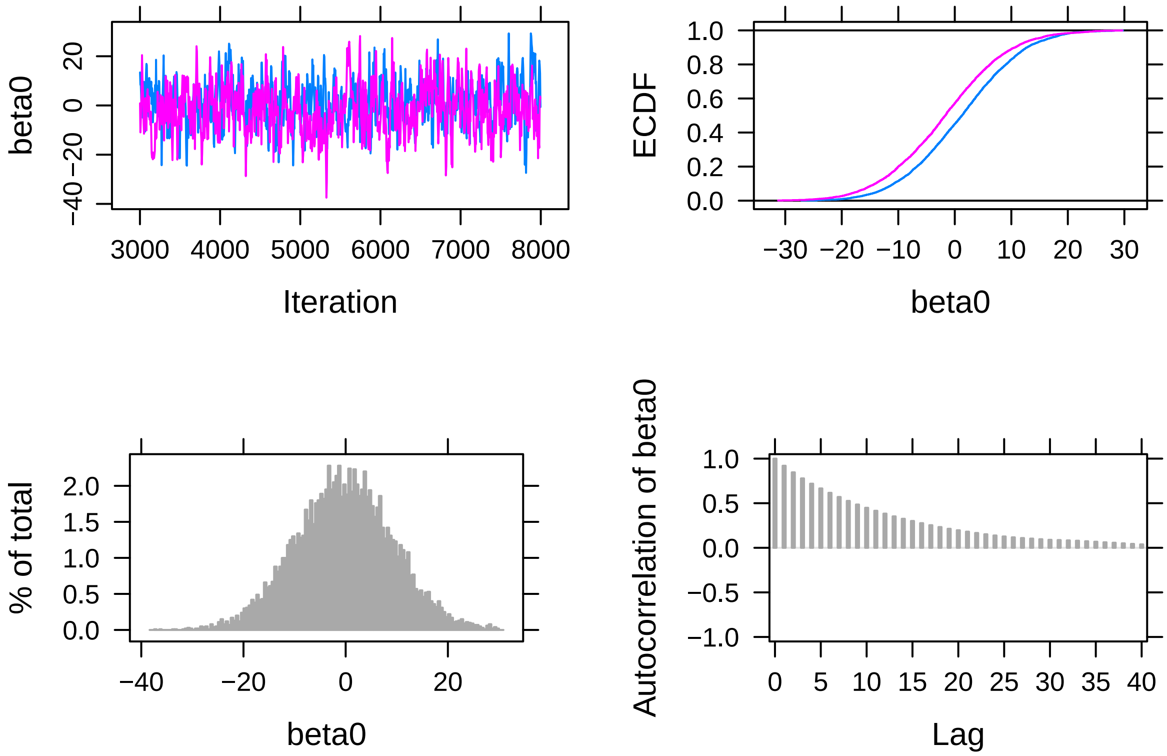
Conclusion

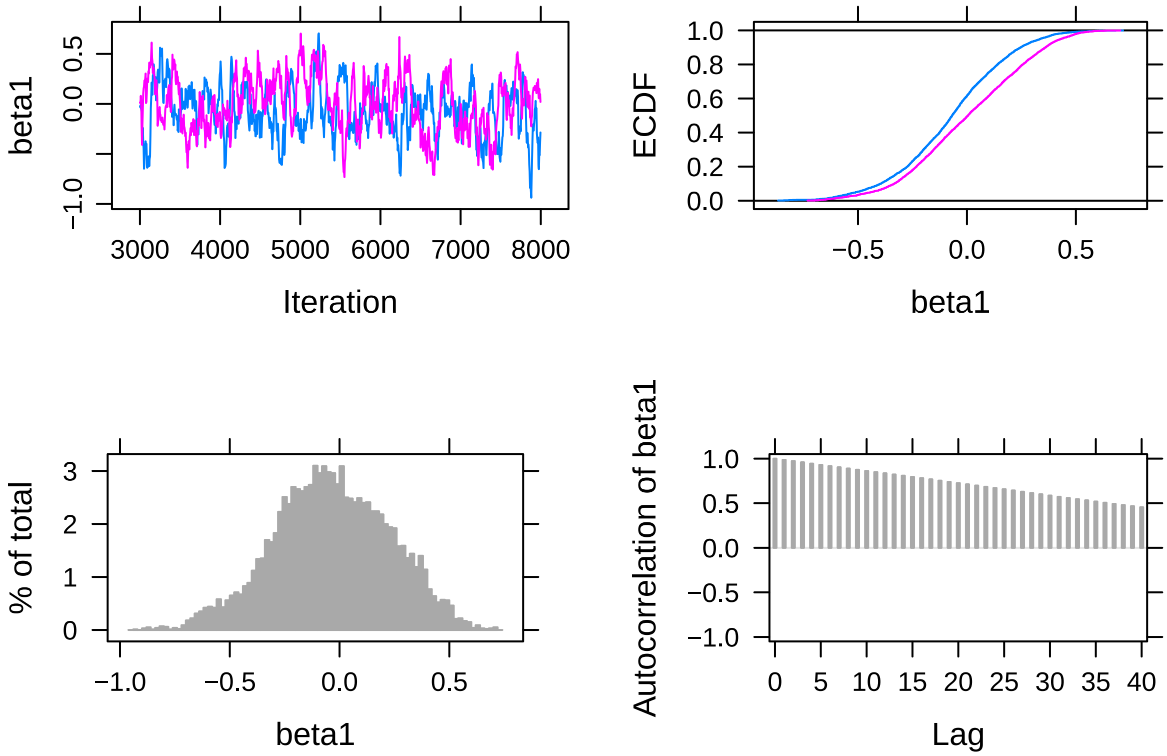
From the analysis we observe that temperature in Celsius degrees: 2.2 to 33.30 was the significant variable in predicting the area the burned area of the forest (in ha) 0.00 to 1090.84, we also observe that a model with independent variables temperature, relative humidity, wind speed and rain is the most plausible model to predict the burned area.

Appendix

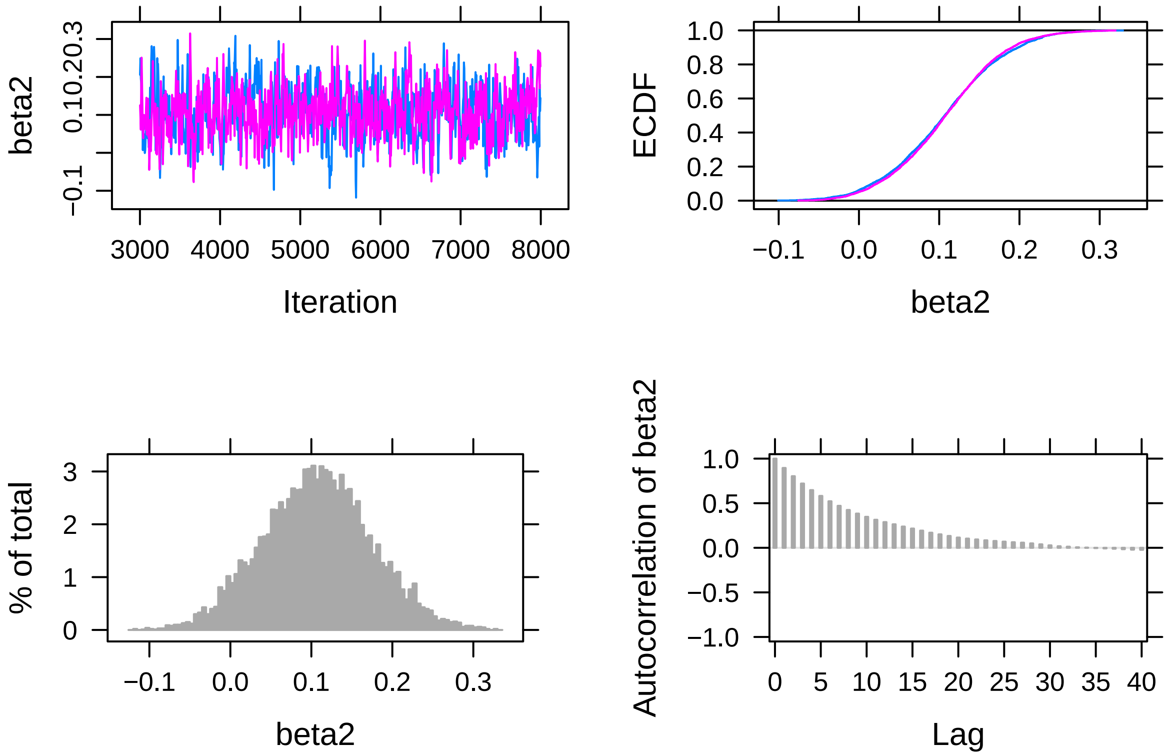
The following plots shows the trace plots and autocorrelation of the parameters

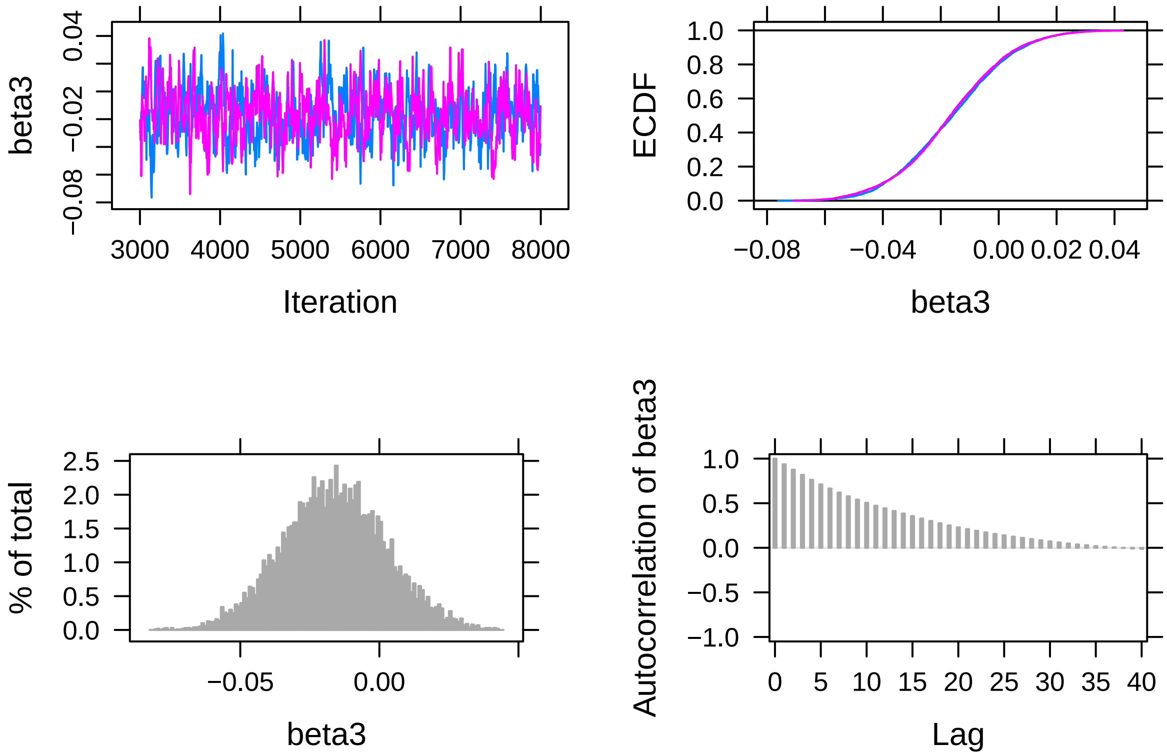
```
## Generating plots...
```

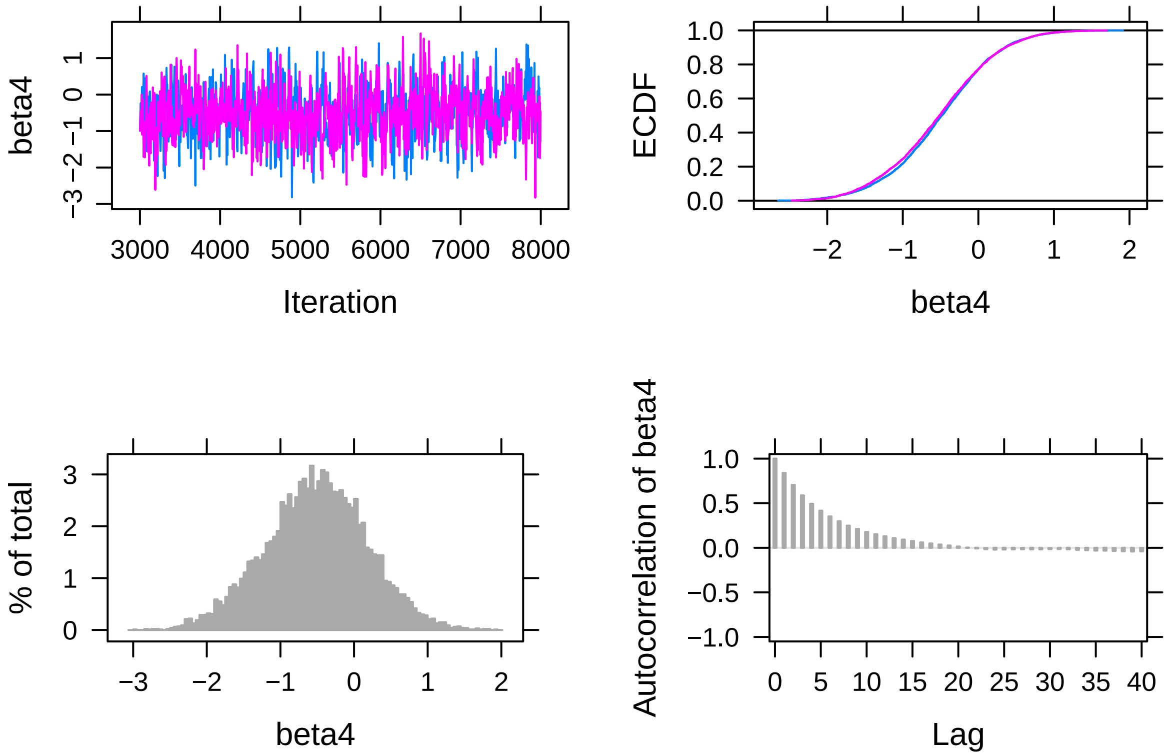

beta0
20
0
−20
−40
3000
4000
5000
6000
7000
8000
Iteration
ECDF
1.0
0.8
0.6
0.4
0.2
0.0
−30
−20
−10
0
10
20
30
beta0
% of total
2.0
1.5
1.0
0.5
0.0
−40
−20
0
20
beta0
Autocorrelation of beta0
1.0
0.5
0.0
−0.5
−1.0
0
5
10
15
20
25
30
35
40
Lag

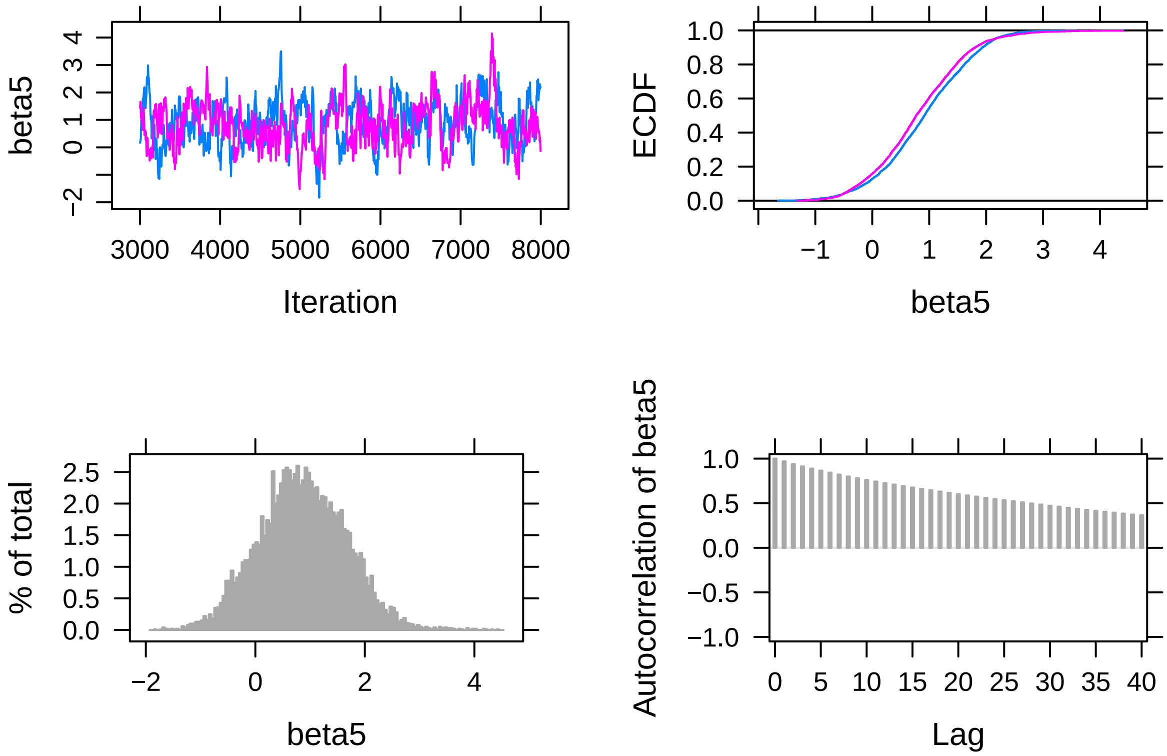


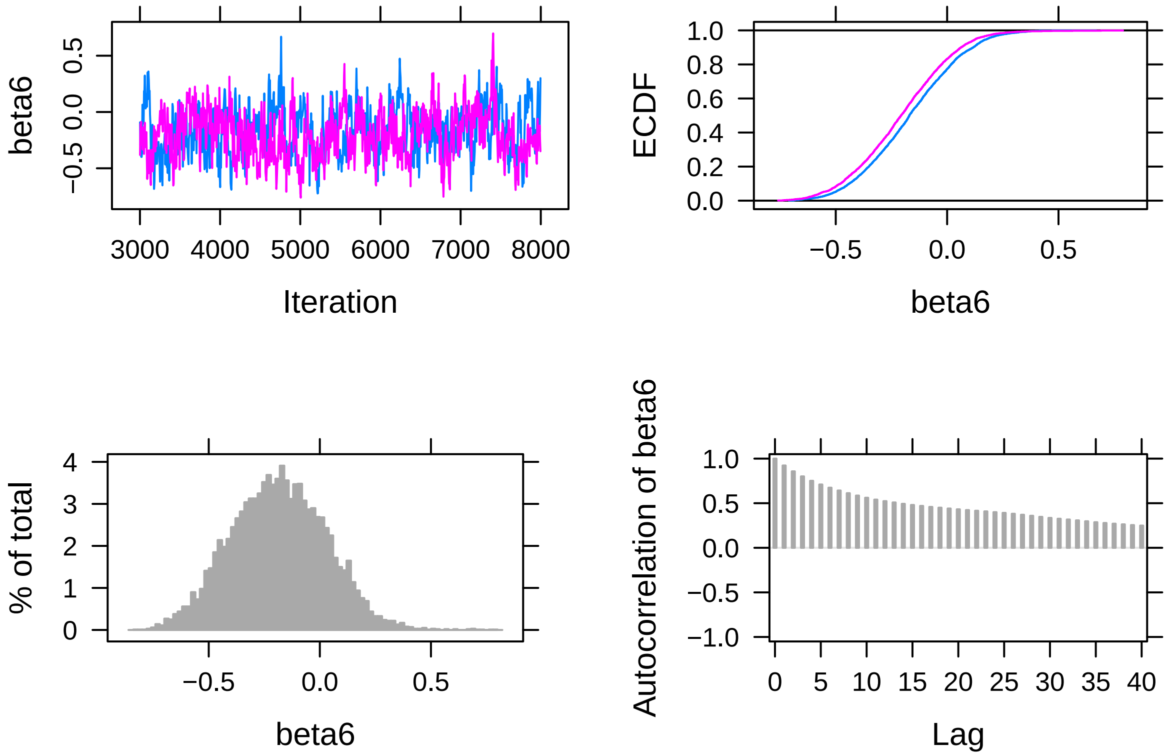
beta1
Iteration
3000 4000 5000 6000 7000 8000
−1.0 0.0 0.5
ECDF
1.0 0.8 0.6 0.4 0.2 0.0
−0.5 0.0 0.5
beta1
% of total
3 2 1 0
−1.0 −0.5 0.0 0.5
beta1
Autocorrelation of beta1
1.0 0.5 0.0 −0.5 −1.0
0 5 10 15 20 25 30 35 40
Lag

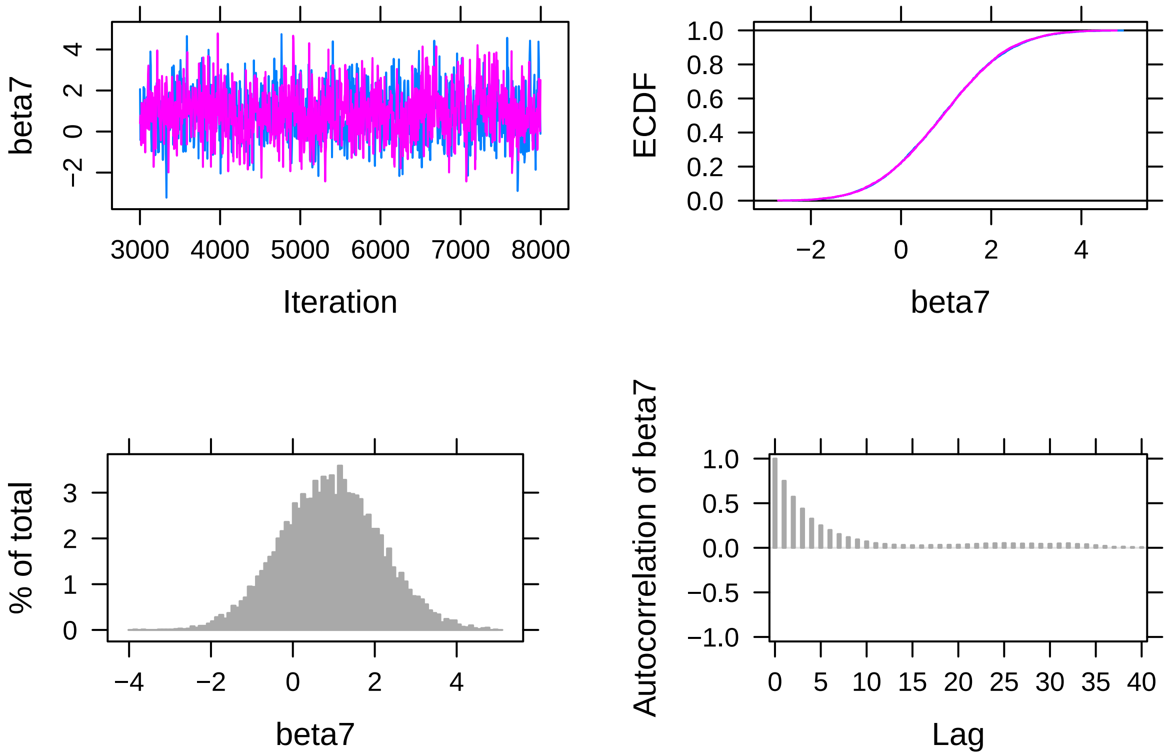

beta2
Iteration
ECDF
beta2
% of total
beta2
Autocorrelation of beta2
Lag

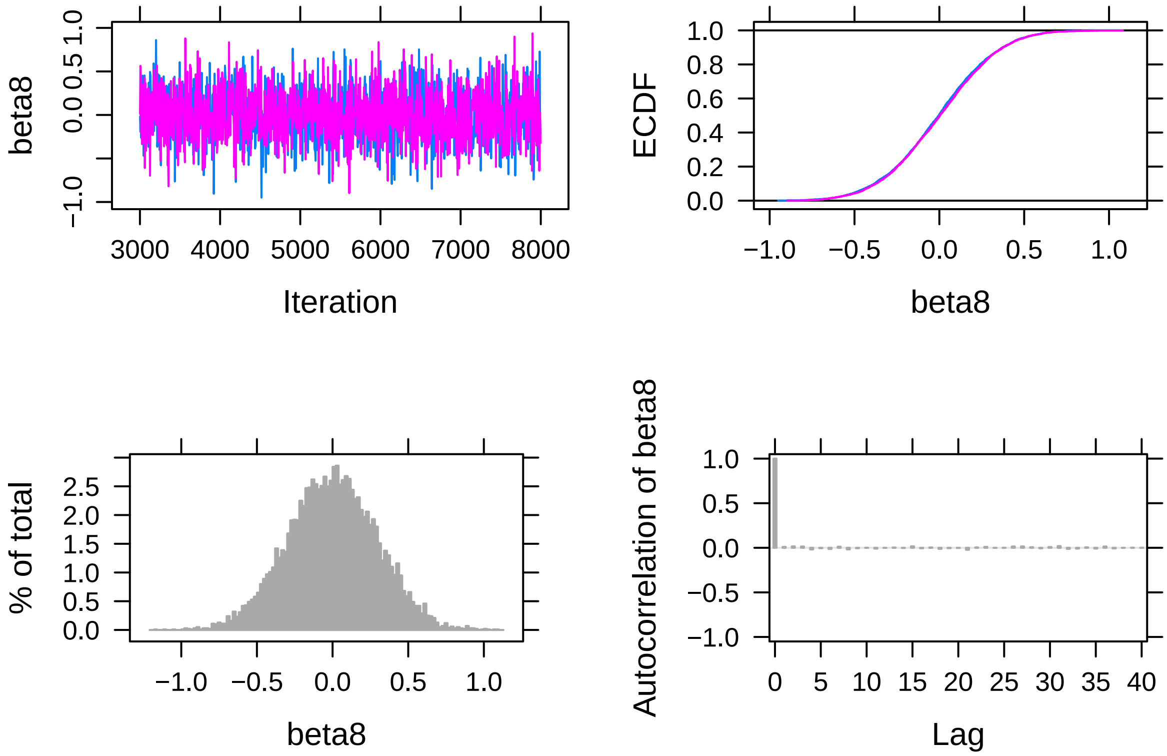

beta3
Iteration
3000
4000
5000
6000
7000
8000
0.04
−0.02
−0.08
ECDF
1.0
0.8
0.6
0.4
0.2
0.0
−0.08
−0.04
0.00
0.02
0.04
beta3
% of total
2.5
2.0
1.5
1.0
0.5
0.0
−0.05
0.00
beta3
Autocorrelation of beta3
1.0
0.5
0.0
−0.5
−1.0
0
5
10
15
20
25
30
35
40
Lag

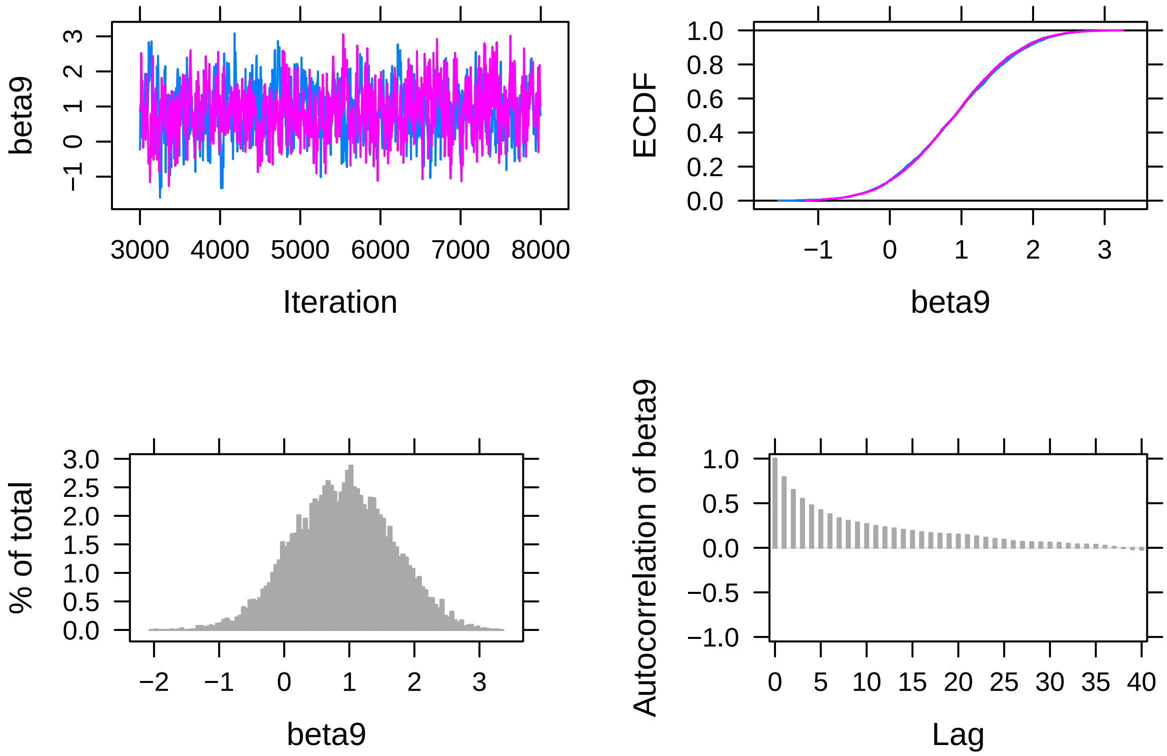

beta4
Iteration
3000 4000 5000 6000 7000 8000
ECDF
beta4
% of total
beta4
Autocorrelation of beta4
Lag

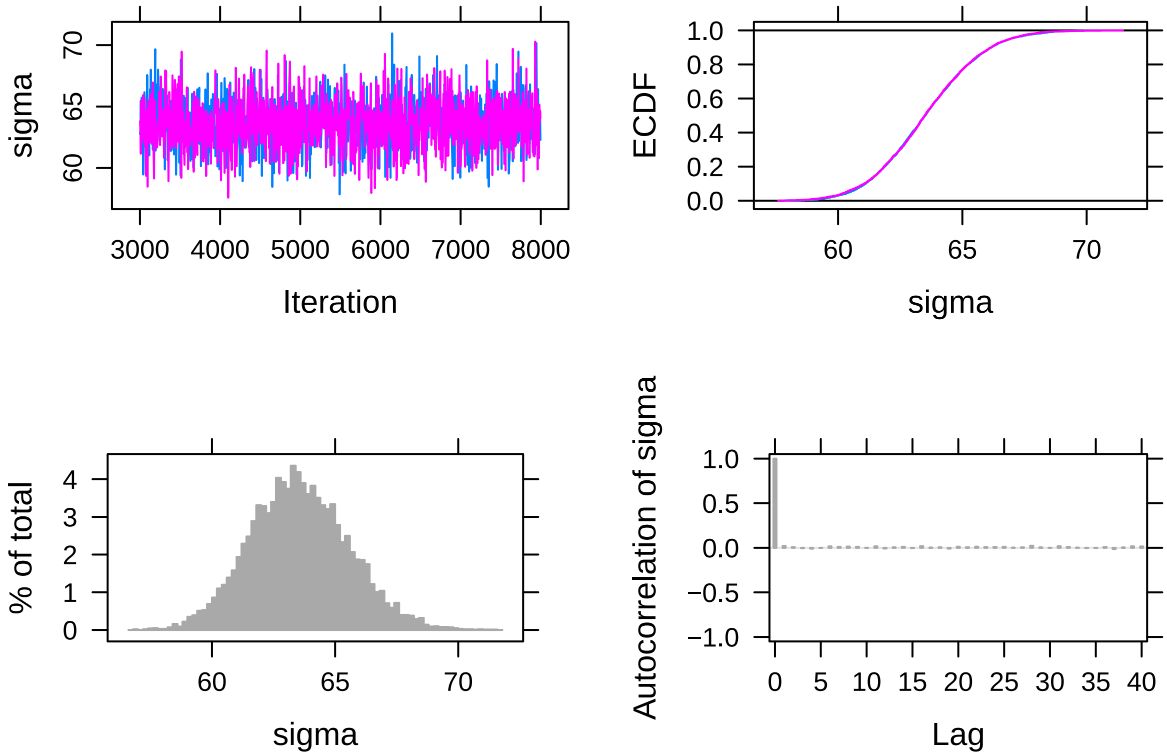

beta5
Iteration
3000
4000
5000
6000
7000
8000
−2
0
1
2
3
4
ECDF
0.0
0.2
0.4
0.6
0.8
1.0
−1
0
1
2
3
4
beta5
% of total
0.0
0.5
1.0
1.5
2.0
2.5
−2
0
2
4
beta5
Autocorrelation of beta5
−1.0
−0.5
0.0
0.5
1.0
0
5
10
15
20
25
30
35
40
Lag

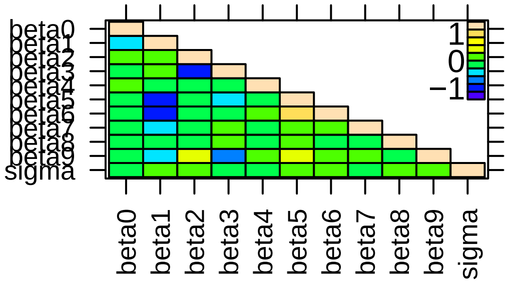

beta6
−0.5
0.0
0.5
3000
4000
5000
6000
7000
8000
Iteration
ECDF
0.0
0.2
0.4
0.6
0.8
1.0
−0.5
0.0
0.5
beta6
% of total
0
1
2
3
4
−0.5
0.0
0.5
beta6
Autocorrelation of beta6
−1.0
−0.5
0.0
0.5
1.0
0
5
10
15
20
25
30
35
40
Lag


beta7
Iteration
3000 4000 5000 6000 7000 8000
−2 0 2 4
ECDF
0.0 0.2 0.4 0.6 0.8 1.0
beta7
−2 0 2 4
% of total
0 1 2 3
beta7
−4 −2 0 2 4
Autocorrelation of beta7
−1.0 −0.5 0.0 0.5 1.0
Lag
0 5 10 15 20 25 30 35 40


beta8
Iteration
3000
4000
5000
6000
7000
8000
−1.0
0.0
0.5
1.0
ECDF
beta8
0.0
0.2
0.4
0.6
0.8
1.0
−1.0
−0.5
0.0
0.5
1.0
% of total
beta8
0.0
0.5
1.0
1.5
2.0
2.5
Autocorrelation of beta8
Lag
0
5
10
15
20
25
30
35
40


beta9
Iteration
ECDF
beta9
% of total
beta9
Autocorrelation of beta9
Lag


sigma
Iteration
ECDF
sigma
% of total
sigma
Autocorrelation of sigma
Lag


beta0
beta1
beta2
beta3
beta4
beta5
beta6
beta7
beta8
beta9
sigma
1
0
−1