1 Introduction

This report will focus on Bayesian knowledge, based on prior and sample information to get information with posterior. Next section methodology includes two parts, part A correspond to estimate value for parameter mu and variance for posterior distribution. Part B correspond to predict SalePrice variable base on multiple linear regression model. Both two parts will be judged by diagnostic figure, estimated value distribution and prediction value distribution.

2 Methodology

2.1 Part A

2.1.1 JAGS model diagram

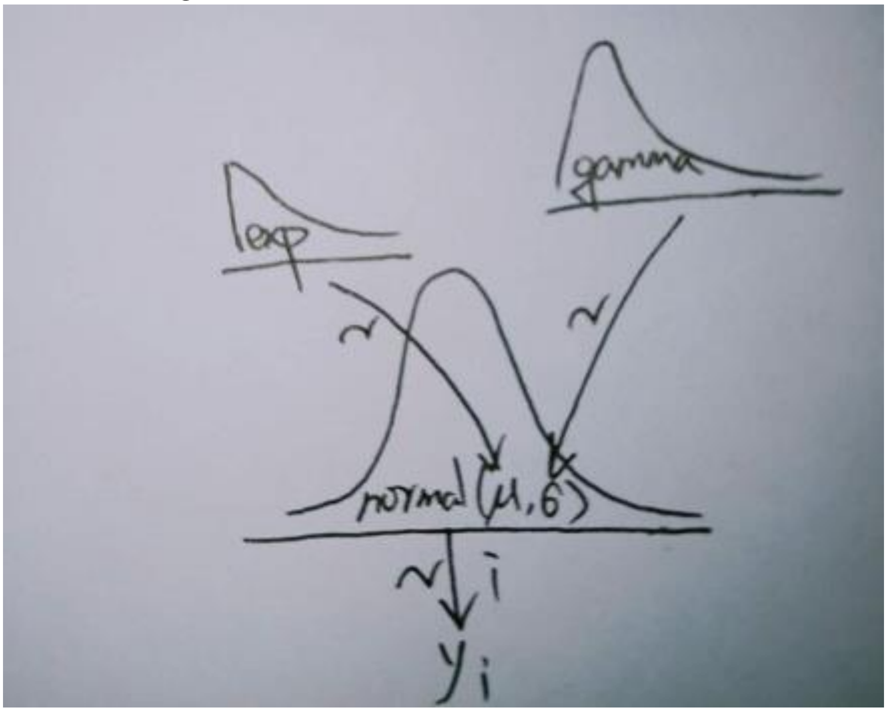


Figure1: JAGS model diagram

2.1.2 Model definition for jags

Considering domain for mu and sigma, they all between 0 to infinite, so, exponential and gamma distribution are suitable.

```
modelString = "
                model {
                  for (i in 1:NTotal){
                    y[i] ~ dnorm(mu, 1/sigma^2)
                    # condition is normal for posterior parameter
                  }
                  # noninformative for prior(mu, sigma)
                  mu ~ dexp(1.0E-8)
                  sigma ~ dgamma(0.1, 1.0E+8)
                  variance <- sigma * sigma
                }
              "
```

2.1.3 Diagnostic

For this part, based on condition that SalePrice is distributed as Normal (μ, σ2), and both are unknown. Say, just put y[i] as normal distribution with unknown parameters mu and sigma. Because normal distribution should define with mean and variance, just put mu and sigma square to represent them. However, the normal distribution for Jags does not same as our general distribution, it represents by precision, so, just set 1/sigma^2 for second parameter.

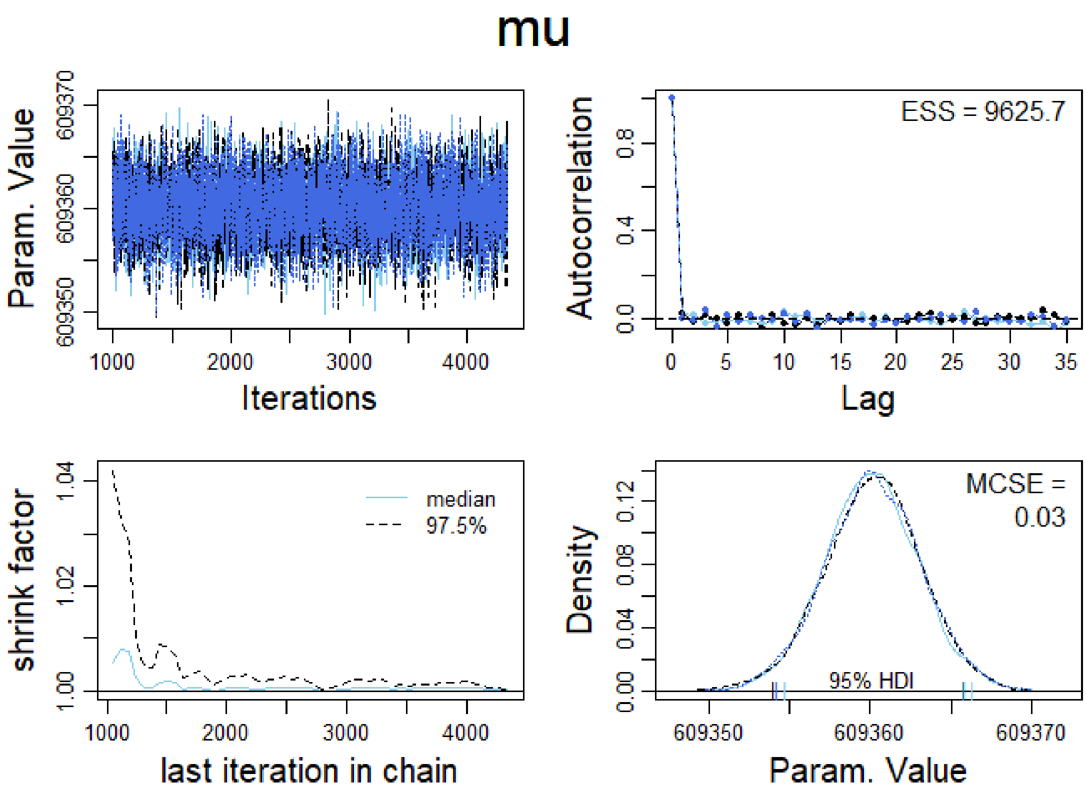
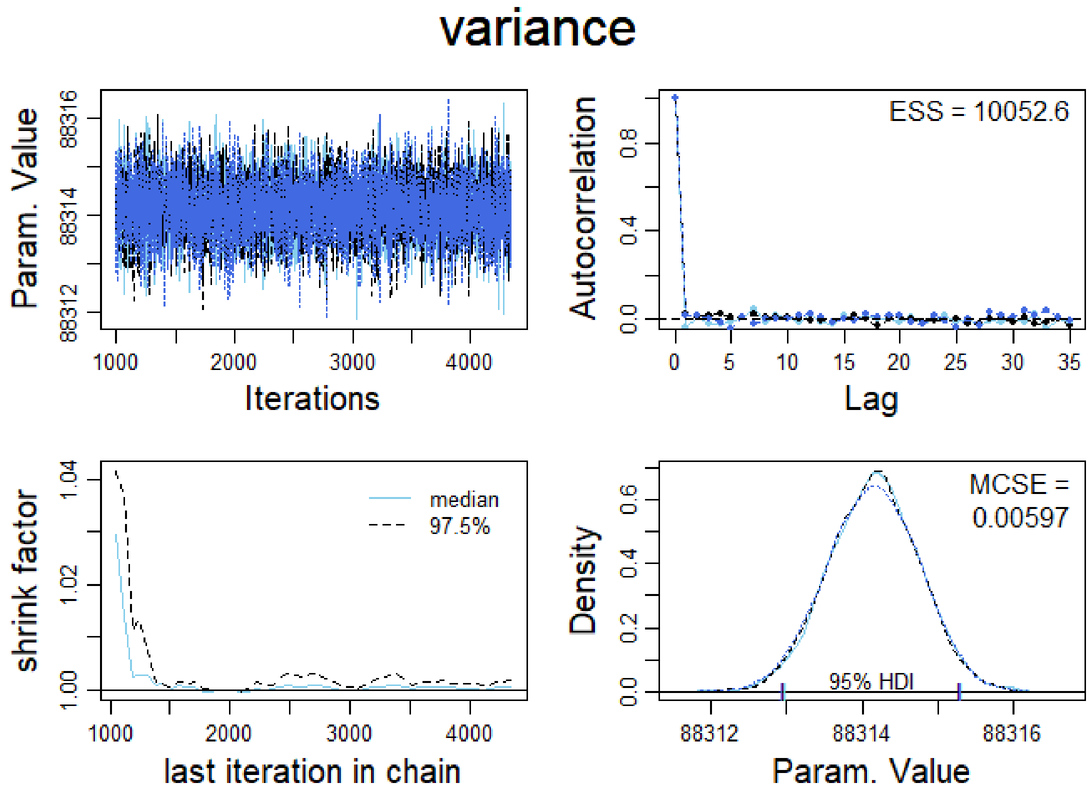


Figure2: Diagnostic of parameter mu

From figure2, it seems that shrink factor going down to a lower value that below 1.1 quickly, parameter value with iteration around mean value, which means that this one could well represent posterior. ESS value is high enough with auto-correlation value nearly equal to zero, although we would like to get a value more than 10000 generally. The density for mu with a low MCSE value that is acceptable.



Fugure3: Diagnostic of parameter variance

For figure3, say, the shrink factor is going to below 1.1 quickly, iteration part just around mean of value, ESS is high with a low auto-correlation nearly to zero, and density of variance with a low MCSE value.

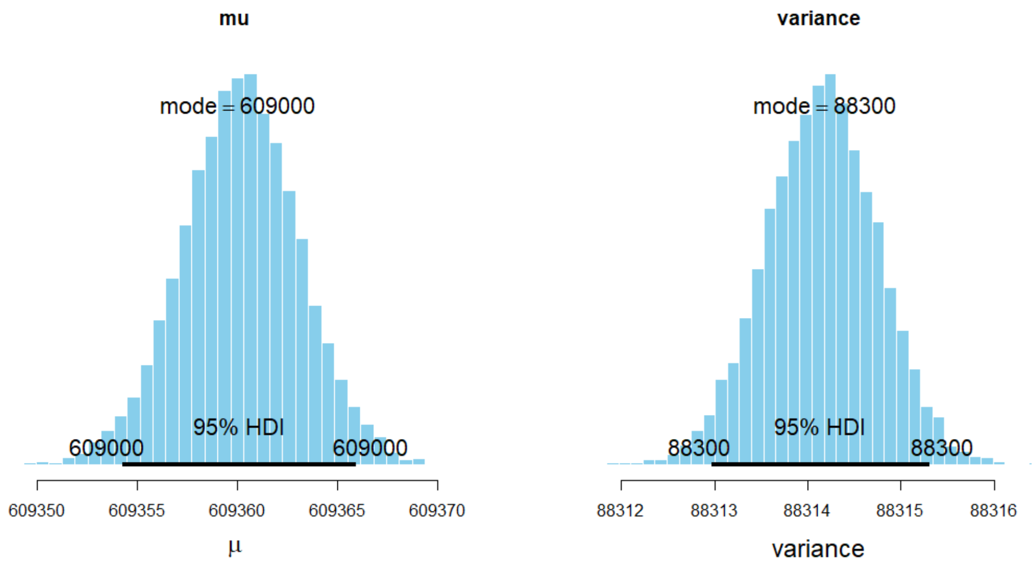


Figure4: Distribution for mu and variance

Summary for mu

	ESS	mean	median	mode	hdiMass	hdiLow	hdiHigh
mu	10002	609360.3	609360.3	609360.9	0.95	609354.6	609366.1

Summary for variance

	ESS	mean	median	mode	hdiMass	hdiLow	hdiHigh
variance	10002	88314.15	88314.15	88314.15	0.95	88313	88315.27

Form figure4, the mode of SalePrice for sample posterior shows as 609000, and the mode for variance shows as 88300. However, the summary for mu and variance show that exact mode for mu is 609360.9, the range from 609354.6 to 609366.1 with 95% high density interval. For variance is 88314.15, the range for 95% high density interval is 88313 to 88315.27. The reason for this plot is that the scale is huge for this dataset, it just keeps first three digits here. In fact, if we would like to see detail information, the summary shown clearly. Cause condition for this Bayesian distribution is that prior is non-informative, it means that likelihood domain prior distribution. So, estimated mu close to sample SalePrice value, and estimated variance smaller than sample variance that make sense, because we also include prior information for this posterior.

2.2 Part B

This part works with multiple linear regression, cause SalePrice is not just relate to itself, it relates to other factors as well. According to condition given by task, 5 features (Area, Bedrooms, Bathrooms, CarParks and PropertyType) as independent variables for dependent variable SalePrice. And the degree of belief for each variable is different. Say, very strong expert knowledge for Area and PropertyType, strong expert knowledge for CarParks, and weak knowledge for Bedrooms, and there is no expert knowledge for Bathroom variable. According to degree of belief, reasonable variance value should be set to prior information. Say, a high degree of belief corresponds to a low variance value, and vice versa. Moreover, for Bathrooms, cause no knowledge about this, considering high variance for it. The information for these variances should base on the variance value of SalePrice, otherwise, we could not get good result.

2.2.1 JAGS model diagram

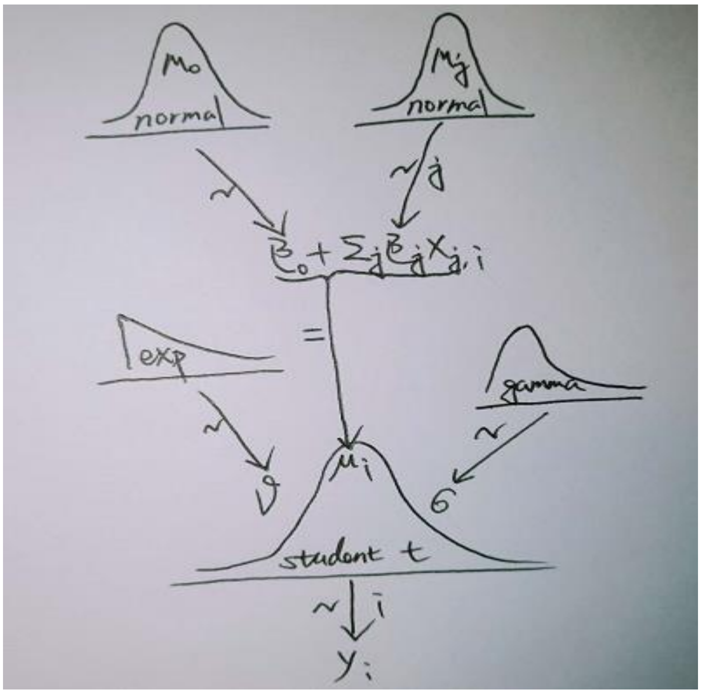


Figure5: JAGS multiple linear regression model diagram

2.2.2 Model definition for jags

In order to reduce collinearlity between parameters, considering standardization for them. After got result of parameters, just convert all of standardization random variables into original scale. Cause domain of parameters for dependent variables are between negative infinite to infinite, just using normal distribution for them. For y[i] come from posterior, using student t distribution, for degree of freedom which is non-informative, the domain is 0 to infinite, considering exponential distribution, and sigma is non-informative, the domain is 0 to infinite, considering gamma distribution is fine.

```
# THE MODEL.
modelString = "
# Standardize the data:
data {
ym <- mean(y)
ysd <- sd(y)
for ( i in 1:Ntotal ) {
zy[i] <- ( y[i] - ym ) / ysd
}
for ( j in 1:Nx ) {
xm[j]  <- mean(x[,j])
xsd[j] <-   sd(x[,j])
for ( i in 1:Ntotal ) {
zx[i,j] <- ( x[i,j] - xm[j] ) / xsd[j]
}
}

# Specify the priors for original beta parameters
# Prior locations to reflect the expert information
mu0 <- ym
# Set to overall mean a prior based on the interpretation of constant term in regression
mu[1] <- 90 # Area
mu[2] <- 100000 # Bedroom
mu[3] <-  0 # Bathroom, no expert knowledge
mu[4] <- 120000 # CarParks
mu[5] <- -150000 # PropertyType
```

```
# Prior variances to reflect the expert information
# variance base on sample variance
Var0 <- 1.0E+12 * (ysd)^2 # inference to sample variance
Var[1] <- 0.1 * (ysd)^2 # Area, very strong
Var[2] <- 1.0E+4 * (ysd)^2 # Bedroom, weak
Var[3] <- 1.0E+5 * (ysd)^2 # Bathroom, no expert knowledge
Var[4] <- 0.5 * (ysd)^2 # Carparks, strong
Var[5] <- 0.1 * (ysd)^2 # PropertyType, very strong

# Compute corresponding prior means and variances for the standardised parameters
muZ[1:Nx] <-  mu[1:Nx] * xsd[1:Nx] / ysd

muZ0 <- (mu0 + sum( mu[1:Nx] * xm[1:Nx] / xsd[1:Nx] )*ysd - ym) / ysd

# Compute corresponding prior variances and variances for the standardised parameters
VarZ[1:Nx] <- Var[1:Nx] * ( xsd[1:Nx]/ ysd )^2
VarZ0 <- Var0 / (ysd^2)
```

```
# Specify the model for standardized data:
model {
for ( i in 1:Ntotal ) {
zy[i] ~ dt( zbeta0 + sum( zbeta[1:Nx] * zx[i,1:Nx] ) , 1/zsigma^2 , nu )
}

# Priors vague on standardized scale:
# set all of prior parameters for mu[i] as normal distribution
# cause domain is (-infinite, +infinite)
zbeta0 ~ dnorm( muZ0 , 1/VarZ0 )
for ( j in 1:Nx ) {
zbeta[j] ~ dnorm( muZ[j] , 1/VarZ[j] )
}
zsigma ~ dgamma(0.01, 0.01) #zsigma is noninformative, similar as uniform
nu ~ dexp(1/30.0) # nu is noninformative

# Transform to original scale:
beta[1:Nx] <- ( zbeta[1:Nx] / xsd[1:Nx] )*ysd
beta0 <- zbeta0*ysd  + ym - sum( zbeta[1:Nx] * xm[1:Nx] / xsd[1:Nx] )*ysd
sigma <- zsigma*ysd

# Compute predictions at every step of the MCMC
pred <- beta0 + beta[1] * xPred[1] + beta[2] * xPred[2] + beta[3] * xPred[3] +
        beta[4] * xPred[4] + beta[5] * xPred[5]
}
" # close quote for modelString
```

2.2.3 Diagnostic

According to figure6 to figure50 show below, they all show the same type of diagnostics, and all of them look good. Say, for shrink factor, they all going to zero quickly, and parameter value for each one with iteration around their own mean levels, it means that all of them are well representative. The ESS values are all high enough with low auto-correlation values nearly to zero, it means that there is no correlation between those observation that come from posterior, although some of them like beta[2] for second instance looks a little bit small, but all of other plots are good for this parameter, this ESS is acceptable. And for MCSE, it seems that all of them are not too high, and they all acceptable. In fact, for zbetas, the MCSE are nearly equal to zero, cause by large variance, when transfer standardized values to original values, the MCSE will increase as well. And we also could see that some distribution for MCSE, few parts for 3 chains are not overlap well like beta[5] for first instance, for this one, for this issue, we could increase iteration to solve this problem. Overall, all of them seems reasonable.

2.2.3.1 Diagnostic for instance c (600, 2, 2, 1, 1)

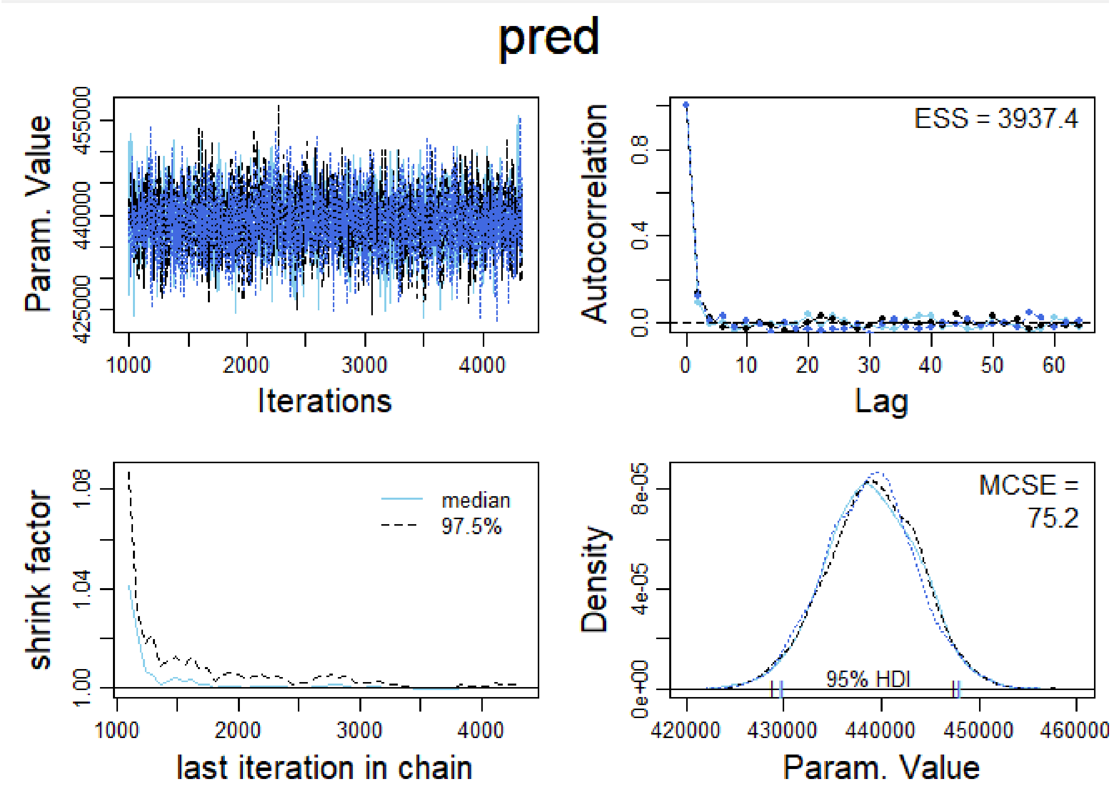


Figure6: Diagnostic of parameter prediction

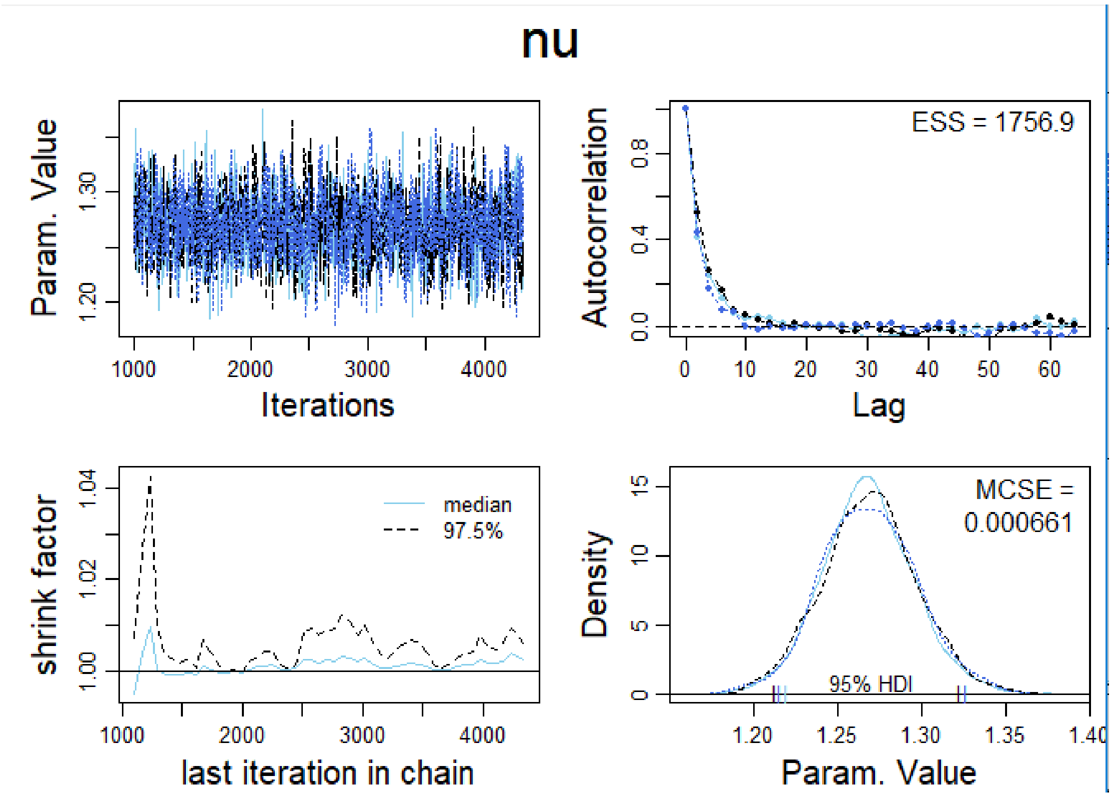


Figure7: Diagnostic of parameter nu

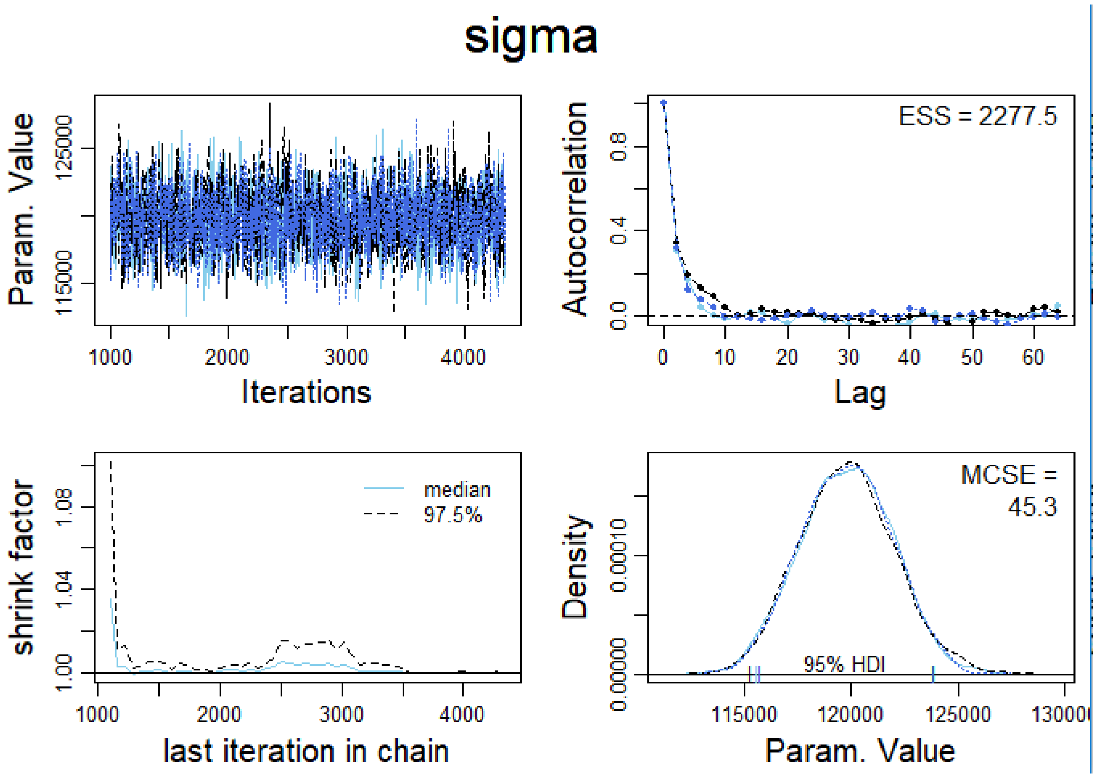


Figure8: Diagnostic of parameter sigma

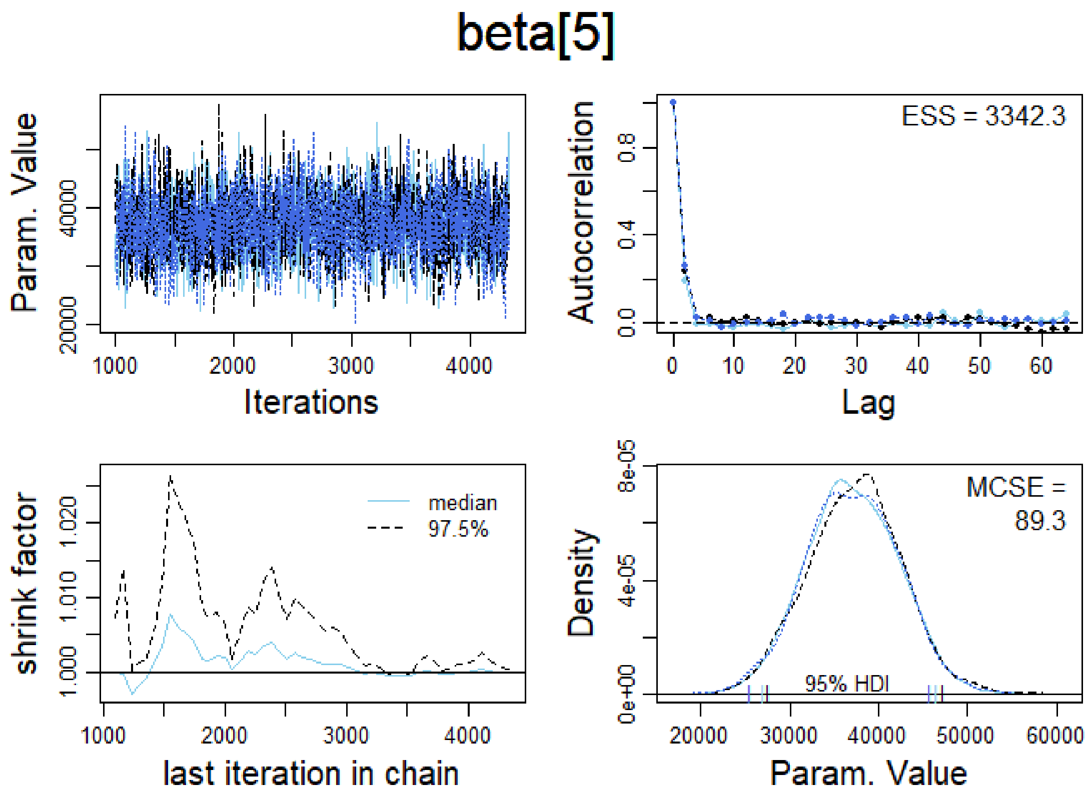


Figure9: Diagnostic of parameter beta[5]

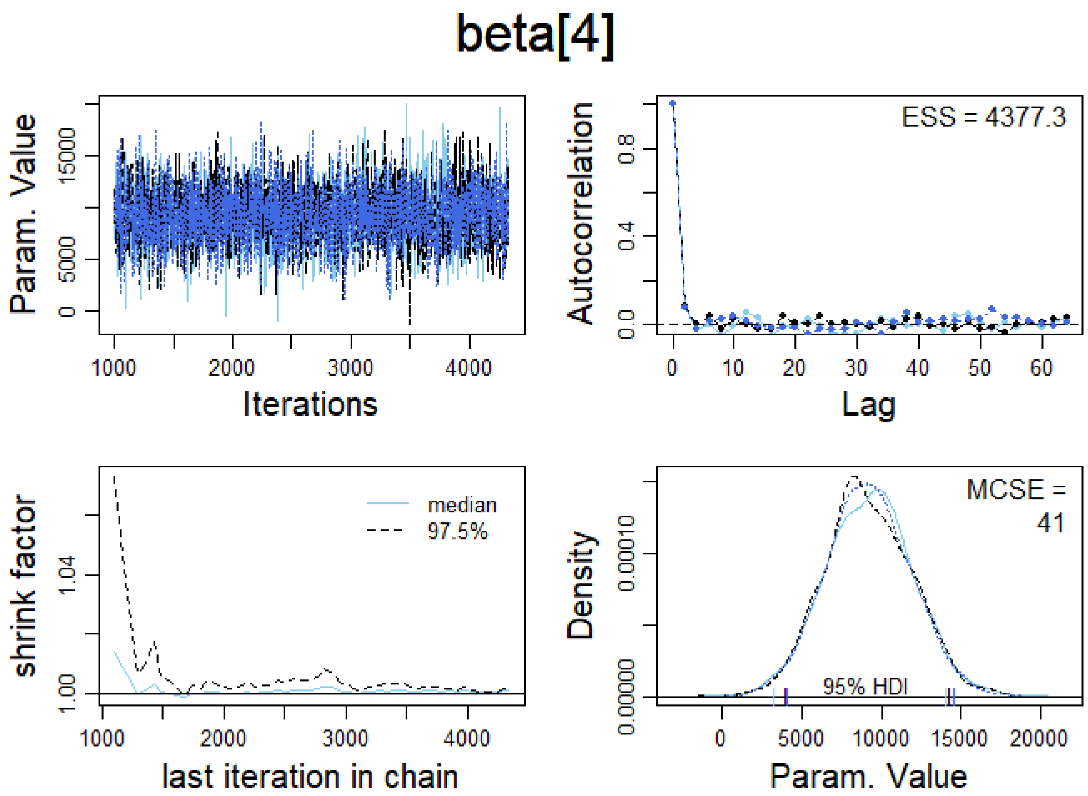


Figure10: Diagnostic of parameter beta[4]

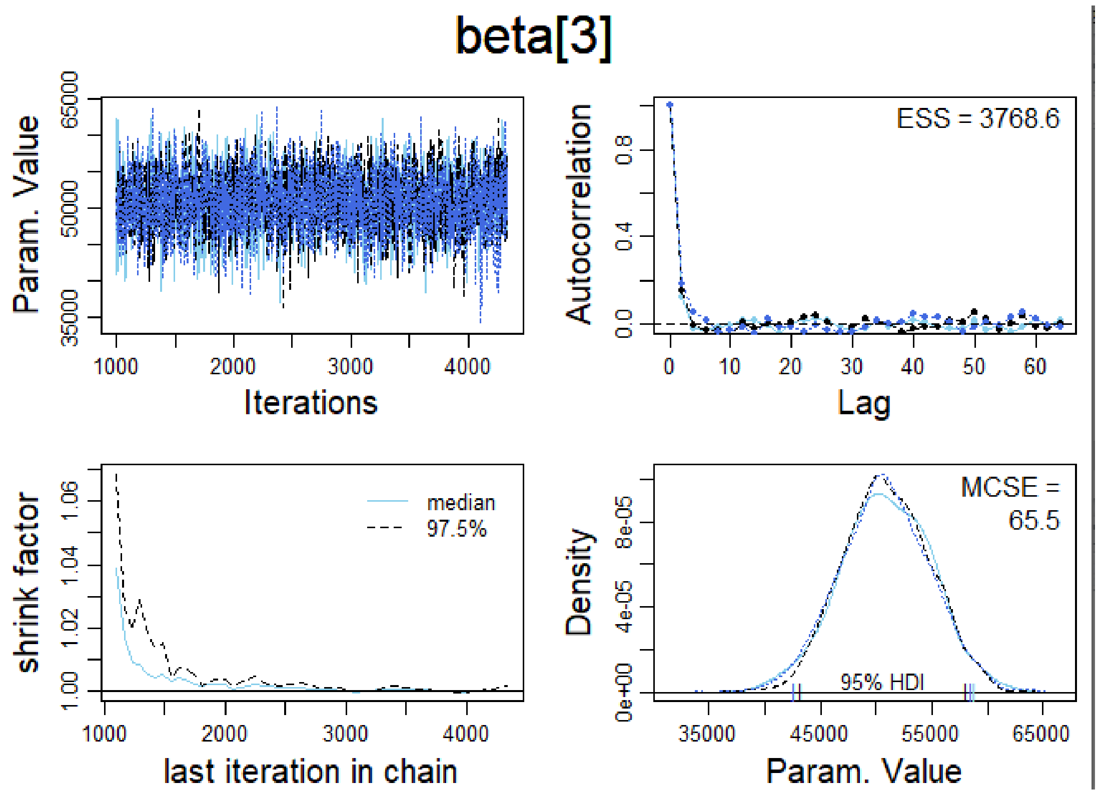


Figure11: Diagnostic of parameter beta[3]

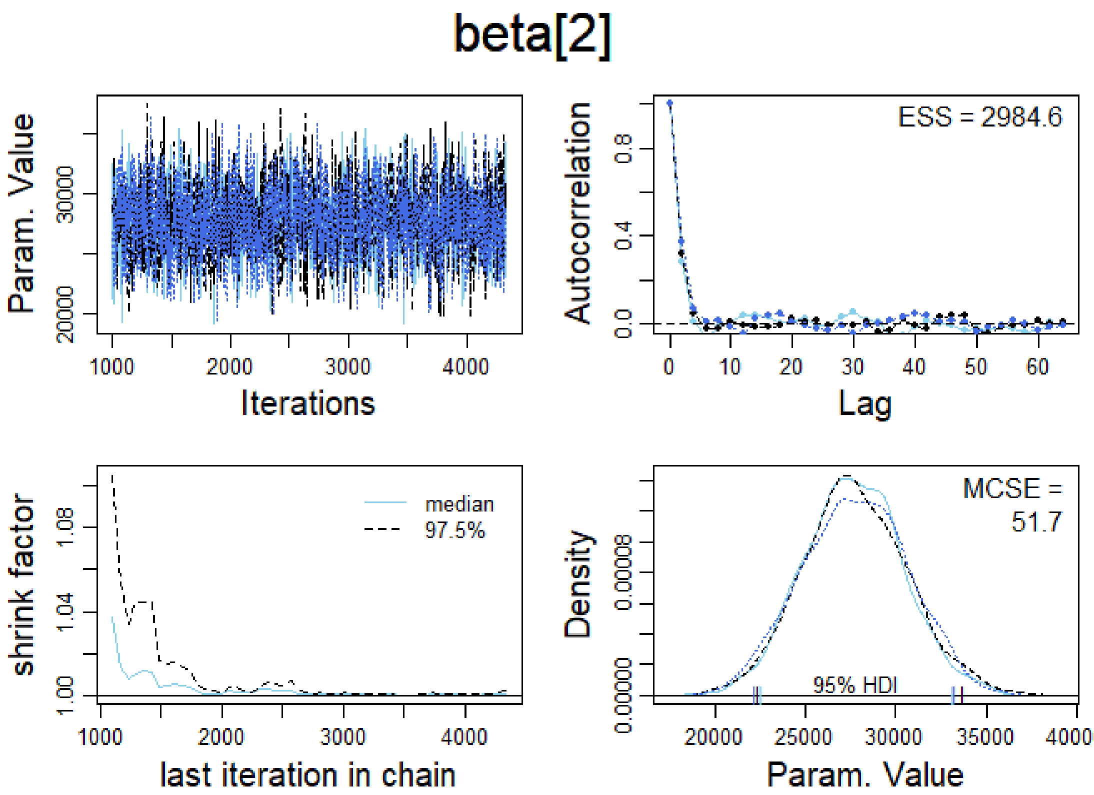


Figure12: Diagnostic of parameter beta[2]

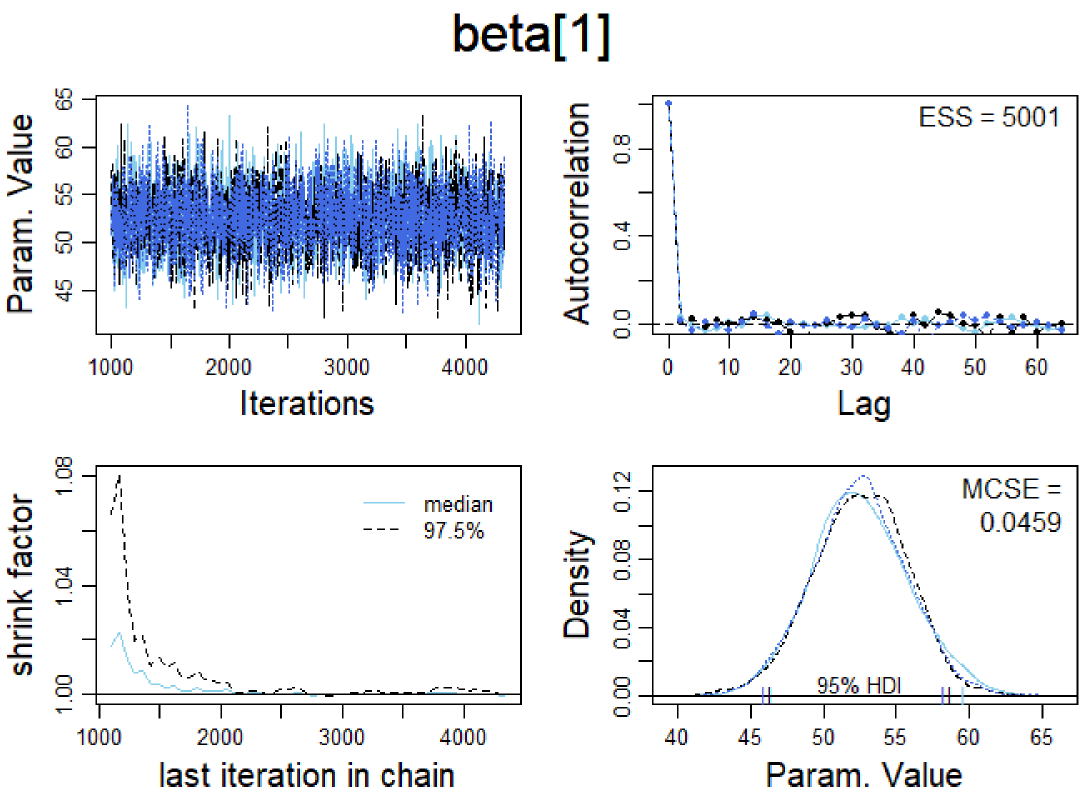


Figure13: Diagnostic of parameter beta[1]

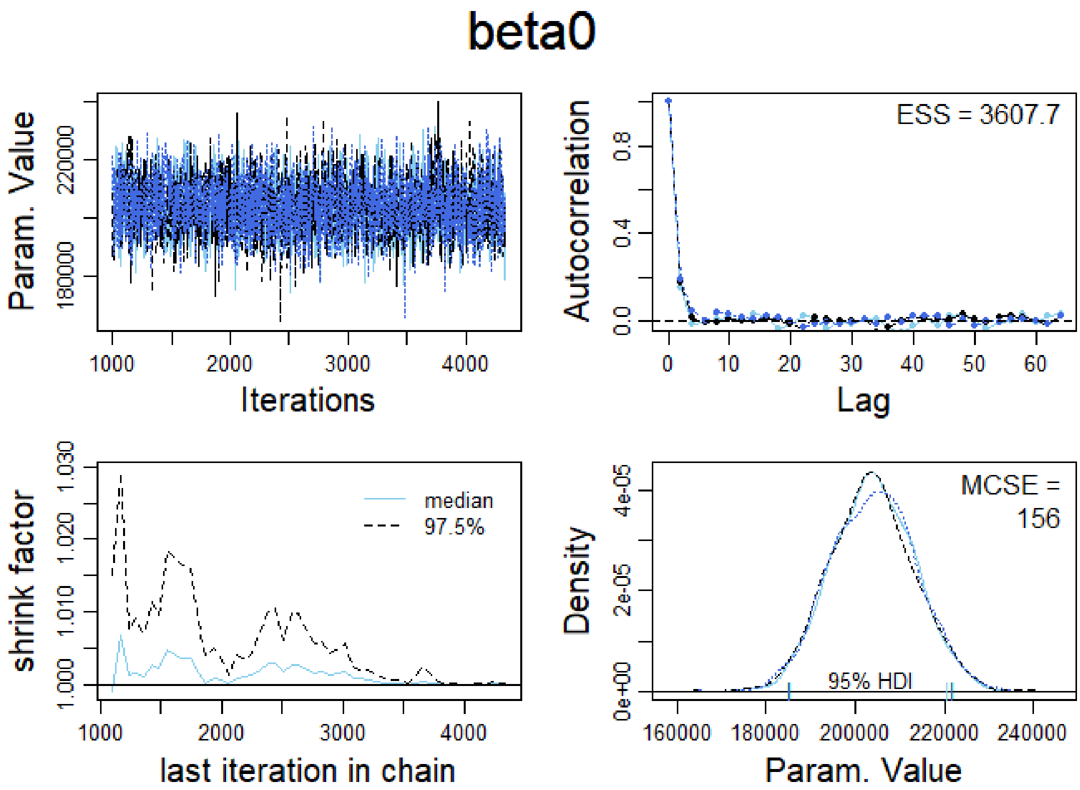


Figure14: Diagnostic of parameter beta0

2.2.3.2 Diagnostic for instance c (800, 3, 1, 2, 0)

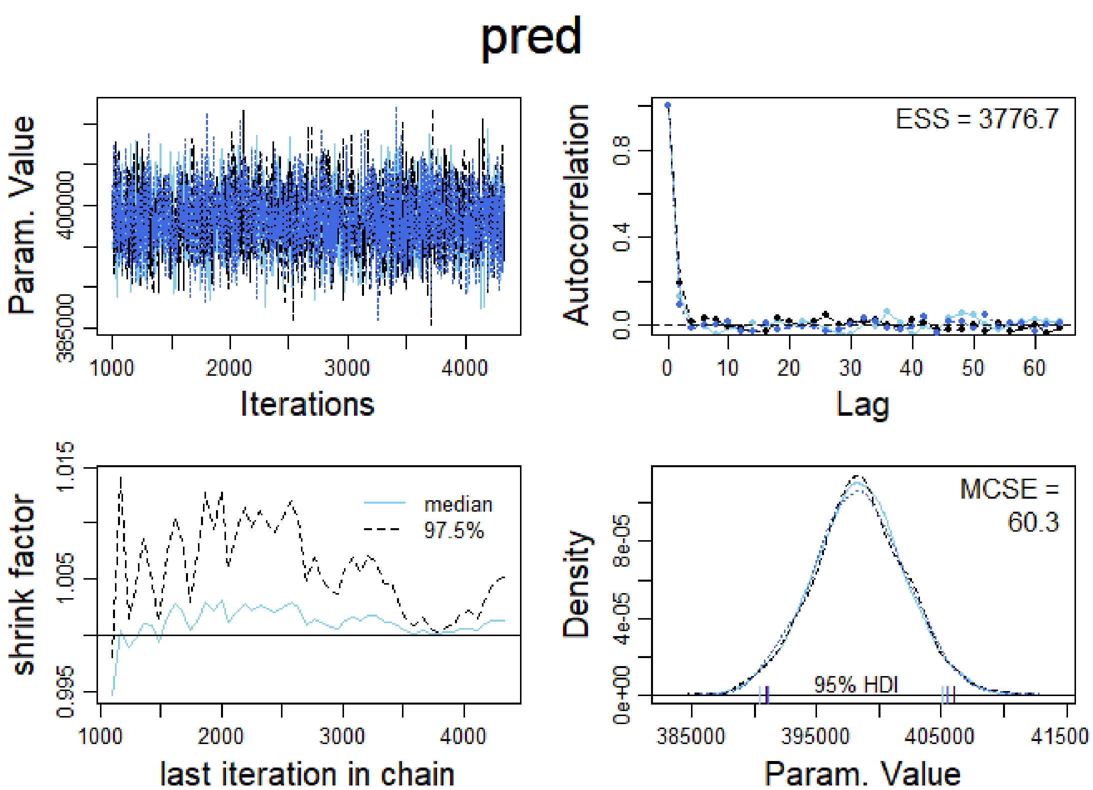


Figure15: Diagnostic of parameter prediction

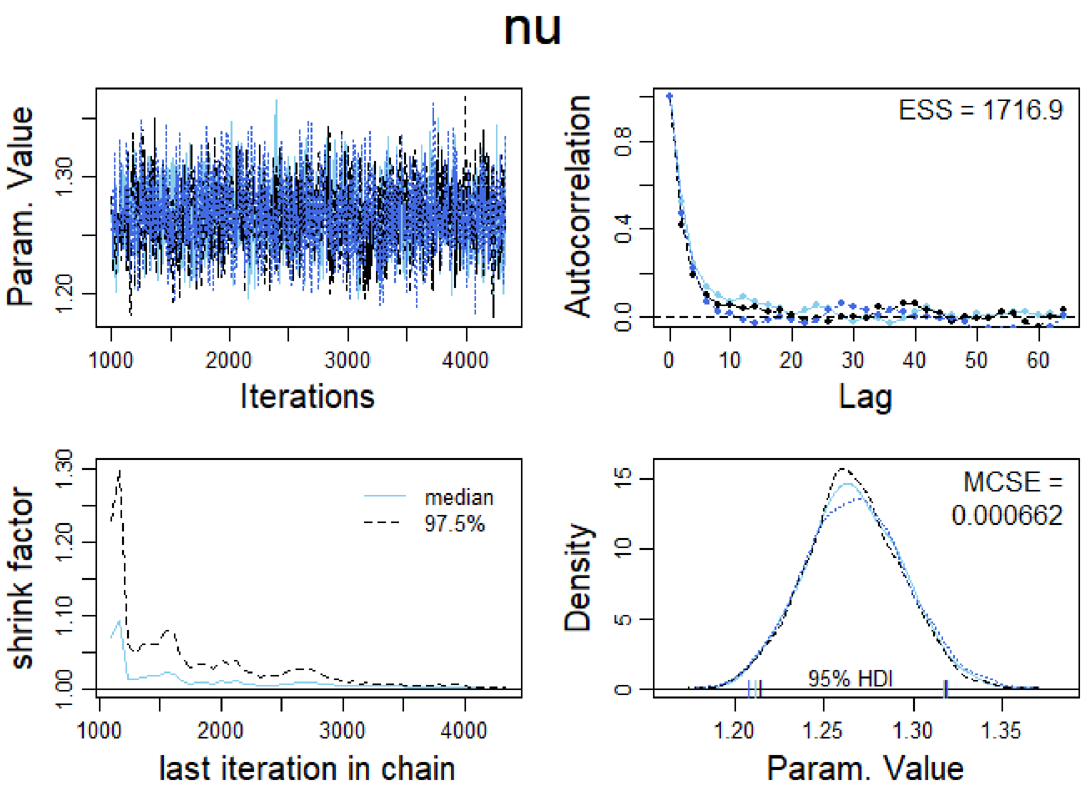


Figure16: Diagnostic of parameter nu

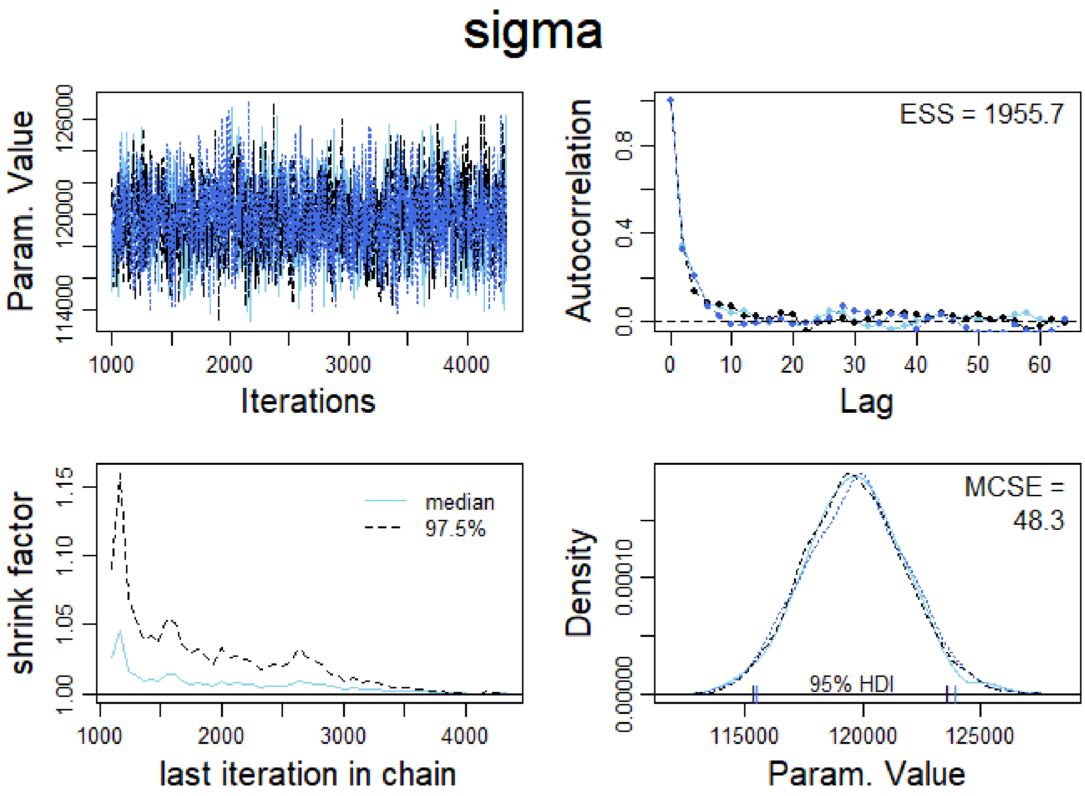


Figure17: Diagnostic of parameter sigma

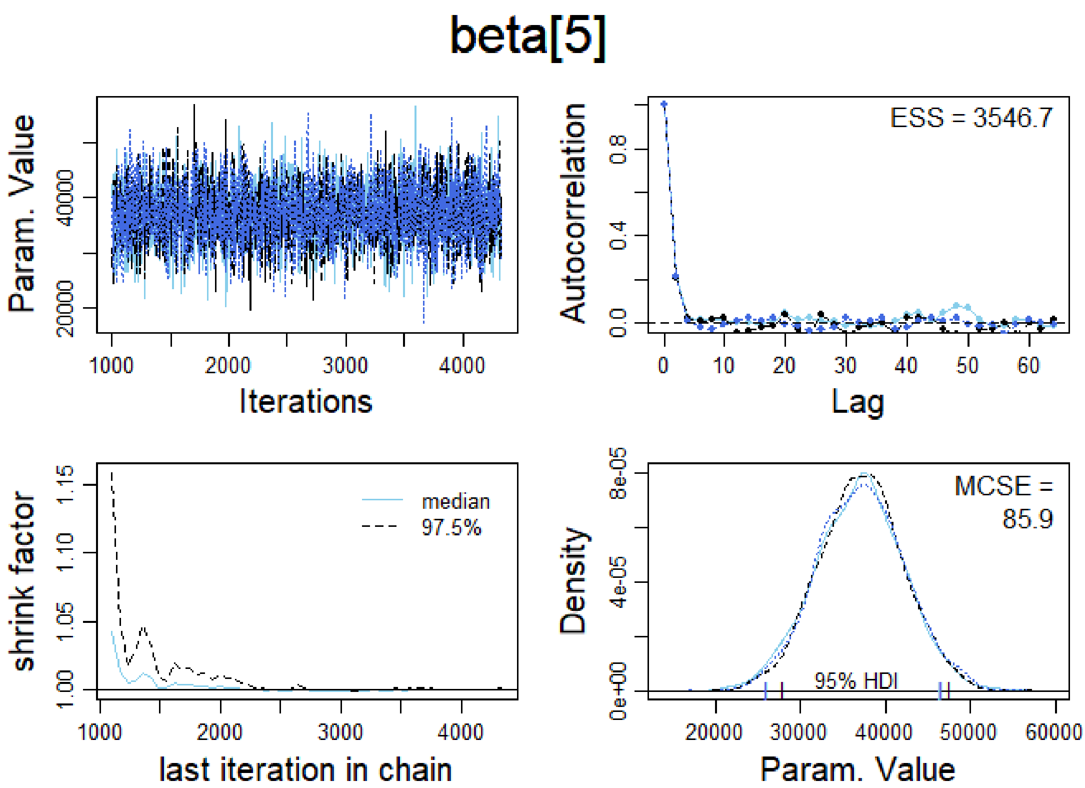


Figure18: Diagnostic of parameter beta[5]

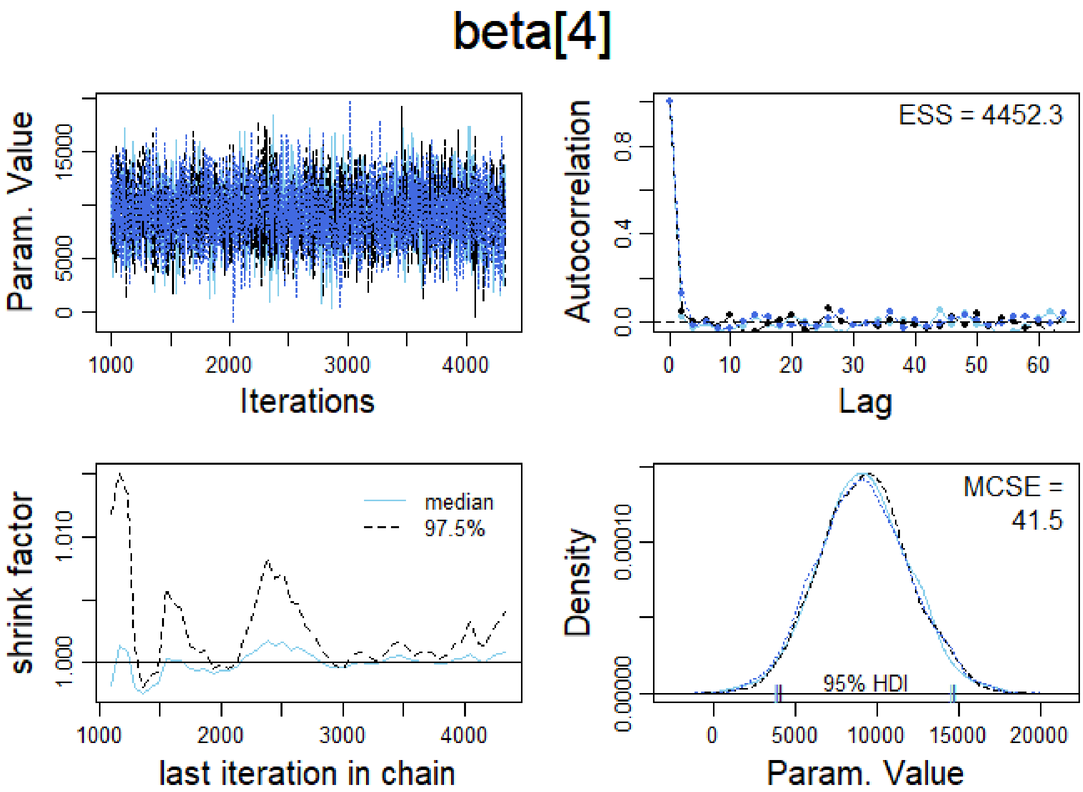


Figure19: Diagnostic of parameter beta[4]

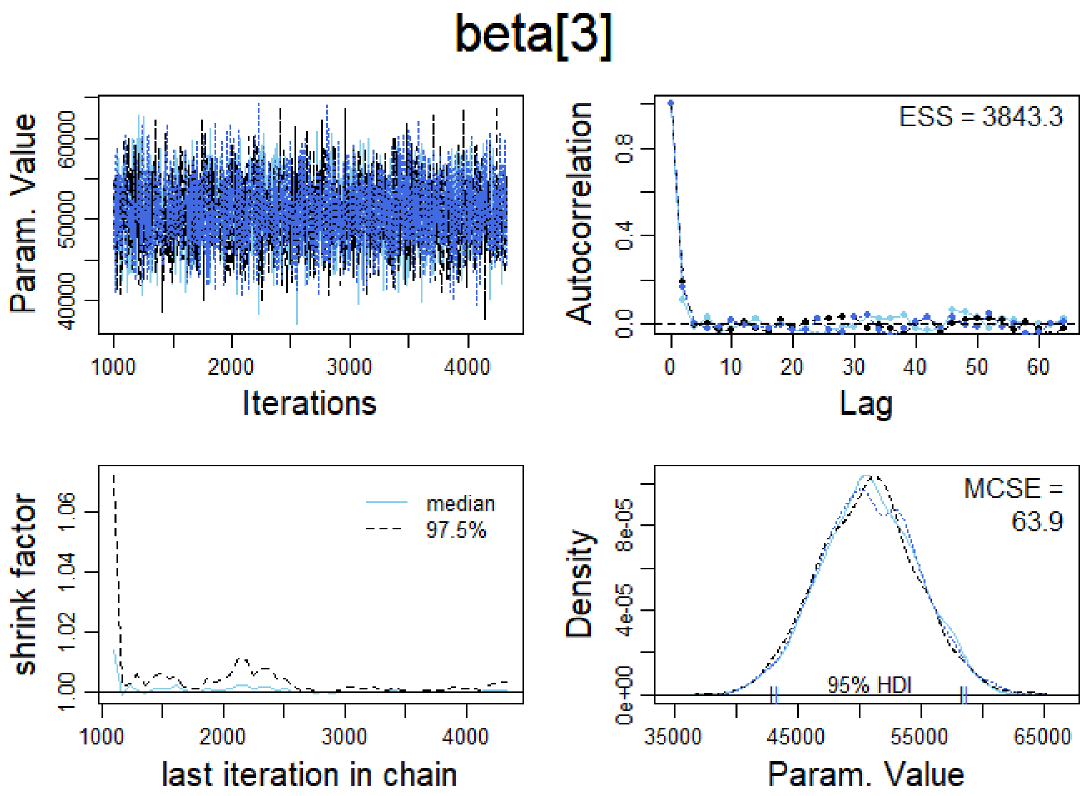


Figure20: Diagnostic of parameter beta[3]

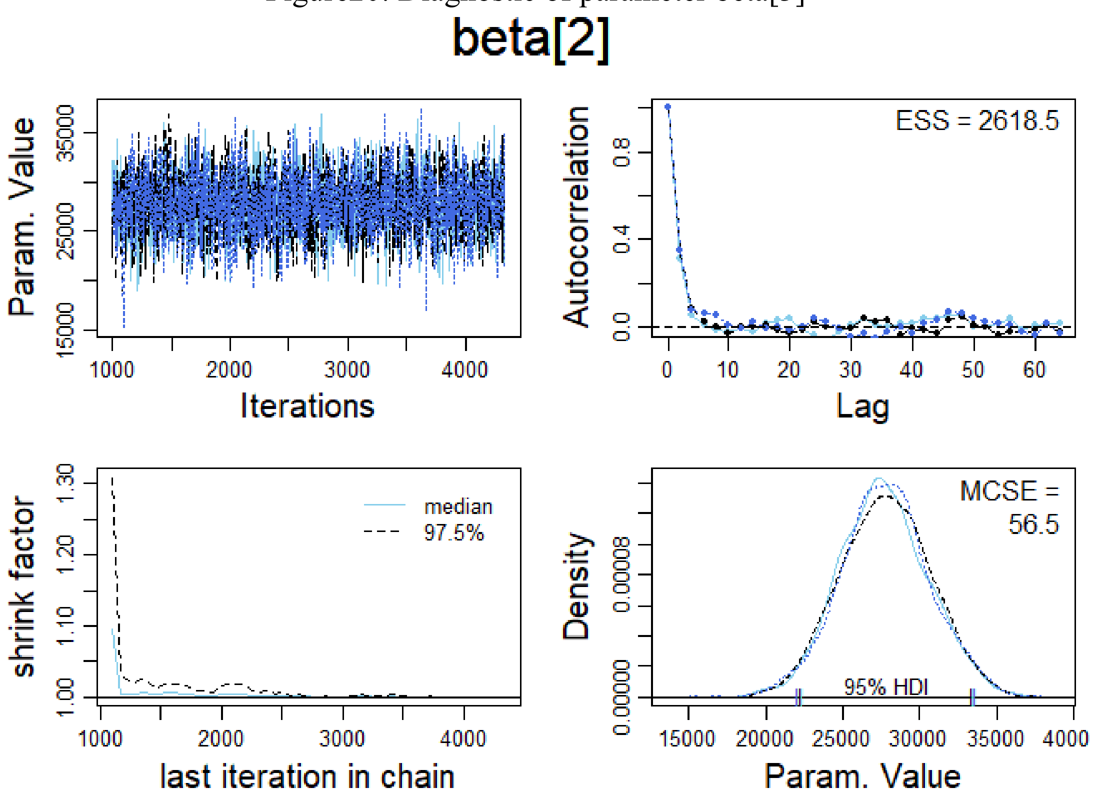


Figure21: Diagnostic of parameter beta[2]

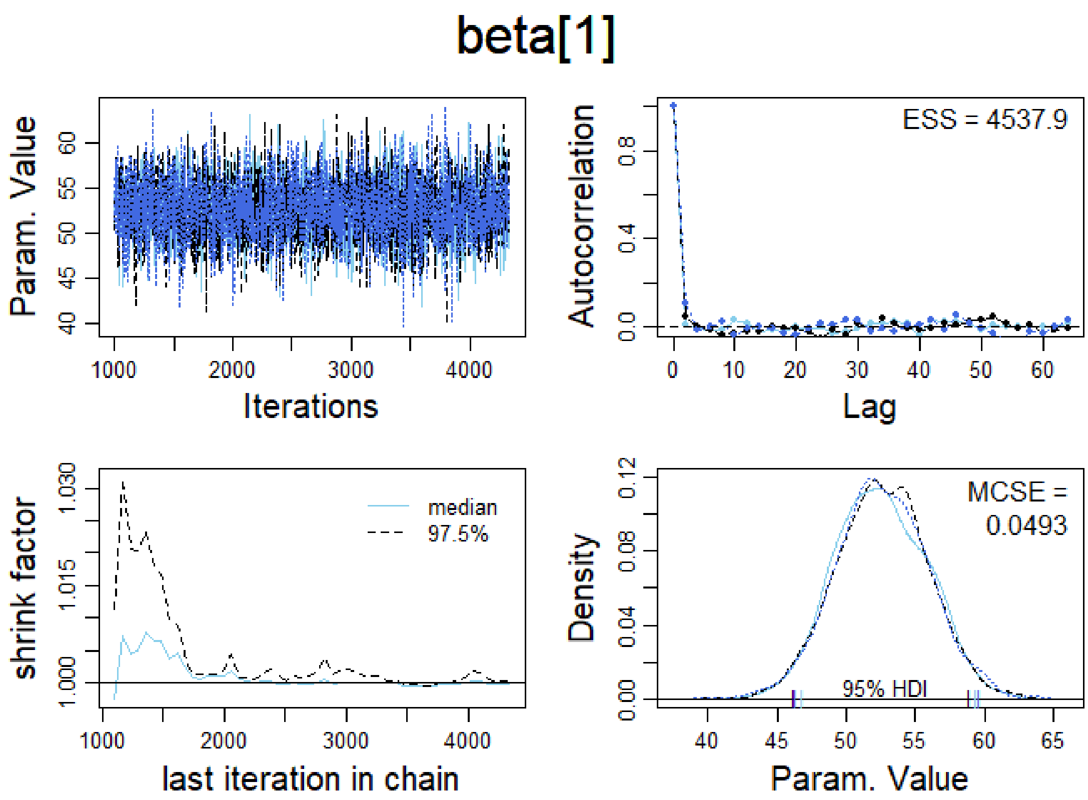


Figure22: Diagnostic of parameter beta[1]

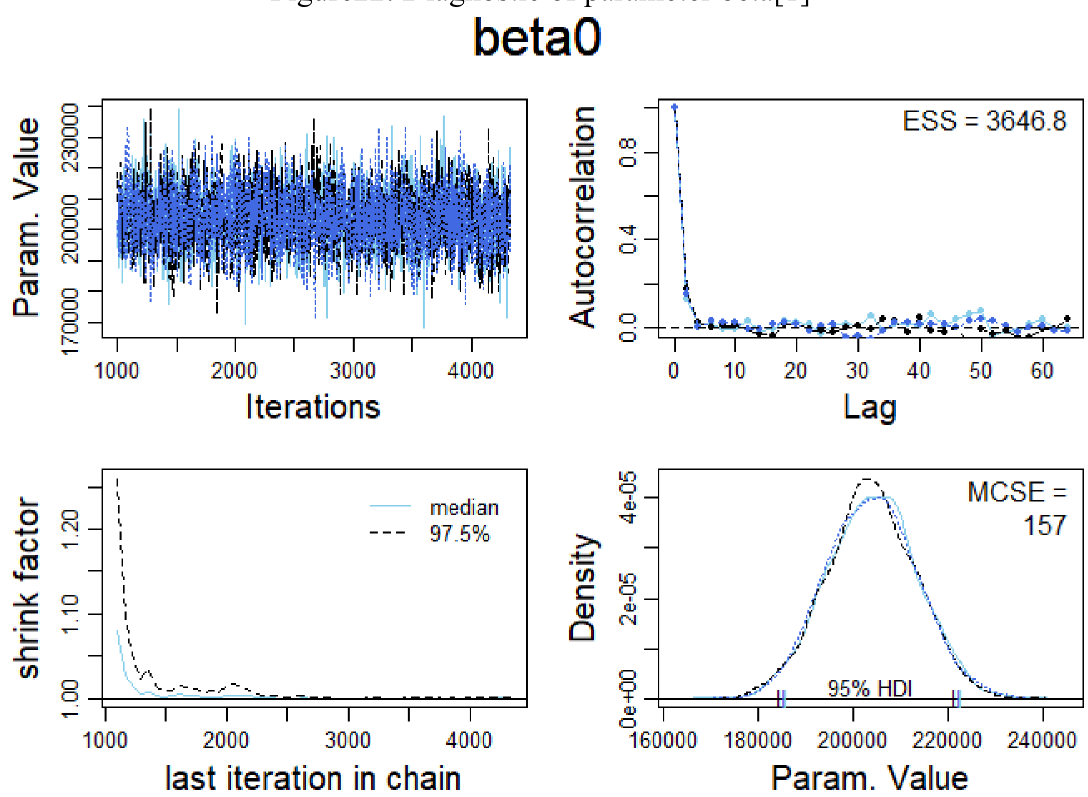


Figure23: Diagnostic of parameter beta0

2.2.3.3 Prediction for instance c (1500, 2, 1, 1, 0)

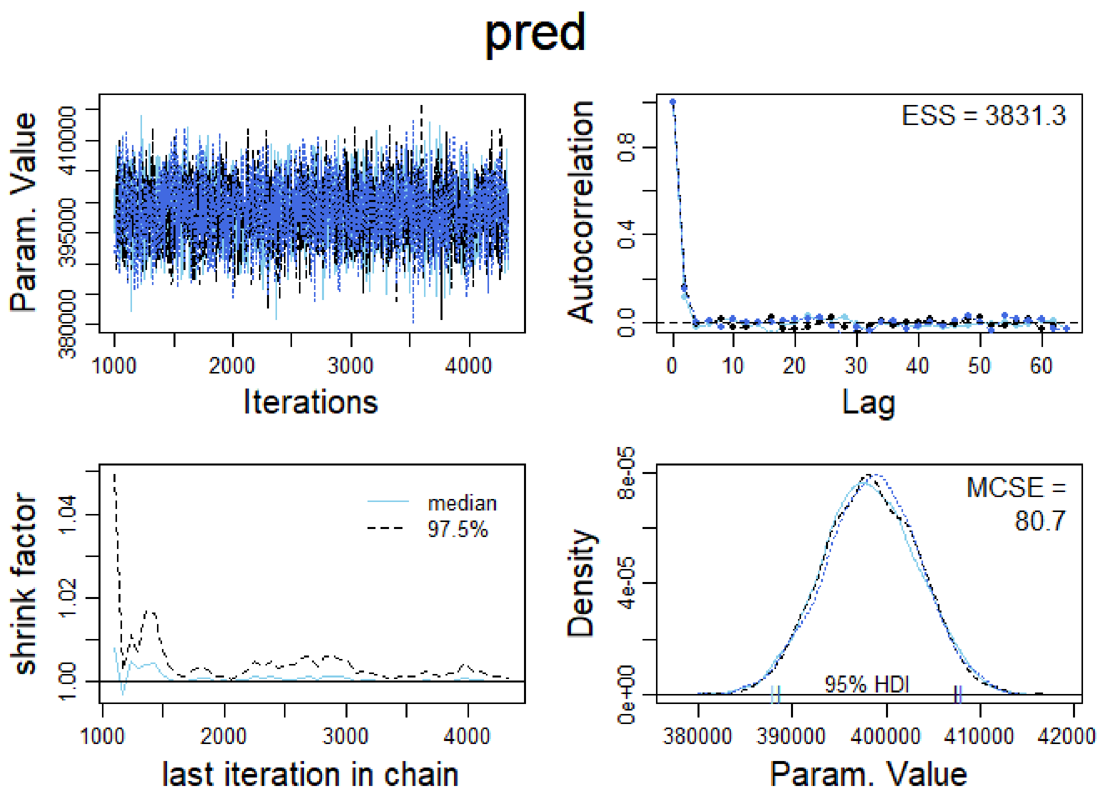


Figure24: Diagnostic of parameter prediction

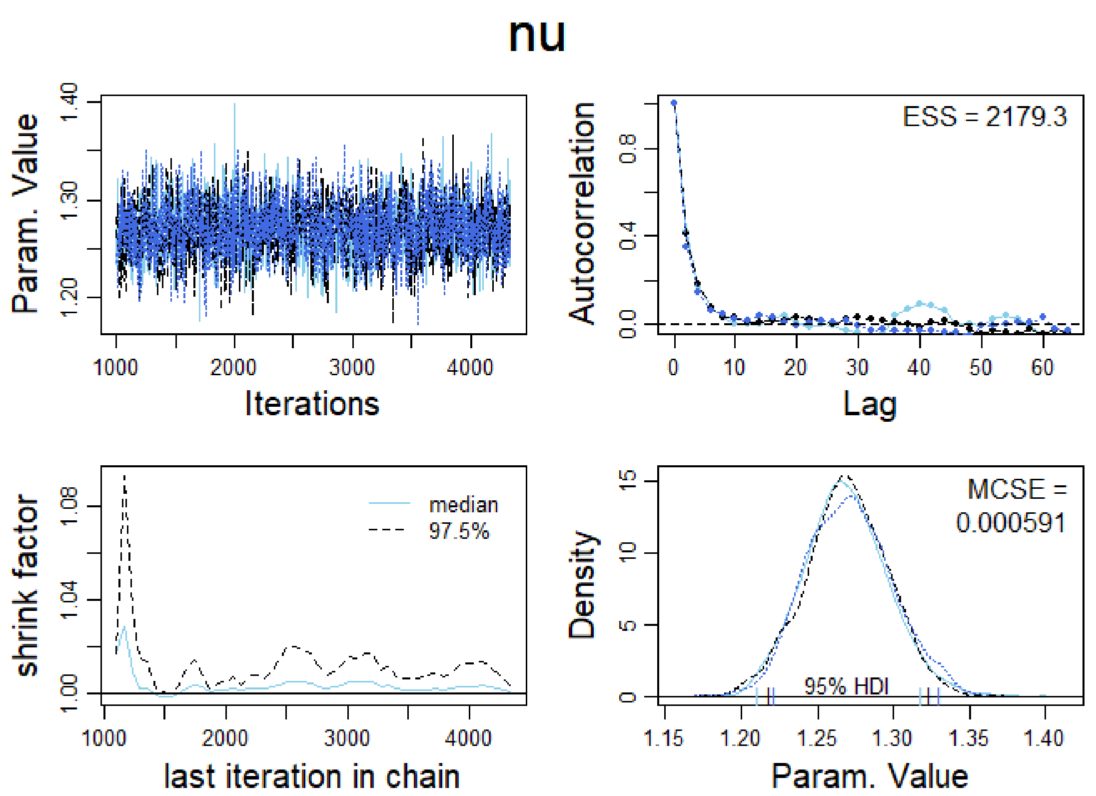


Figure25: Diagnostic of parameter nu

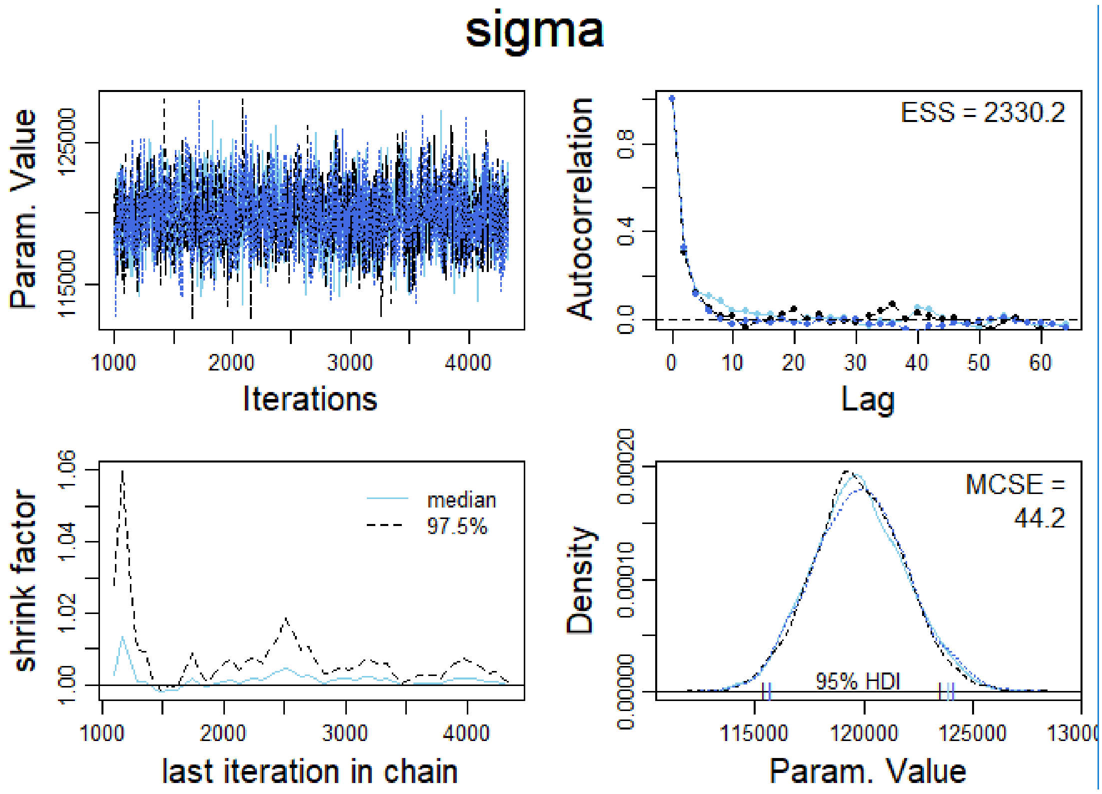


Figure26: Diagnostic of parameter sigma

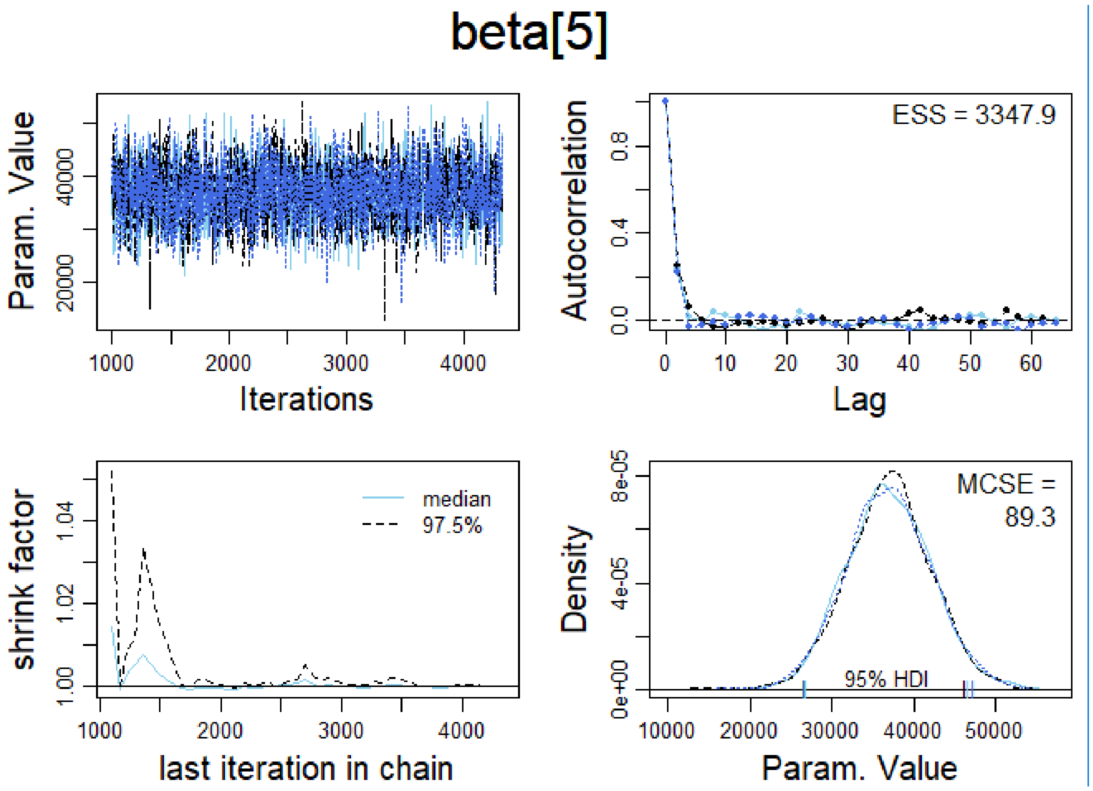


Figure27: Diagnostic of parameter beta[5]

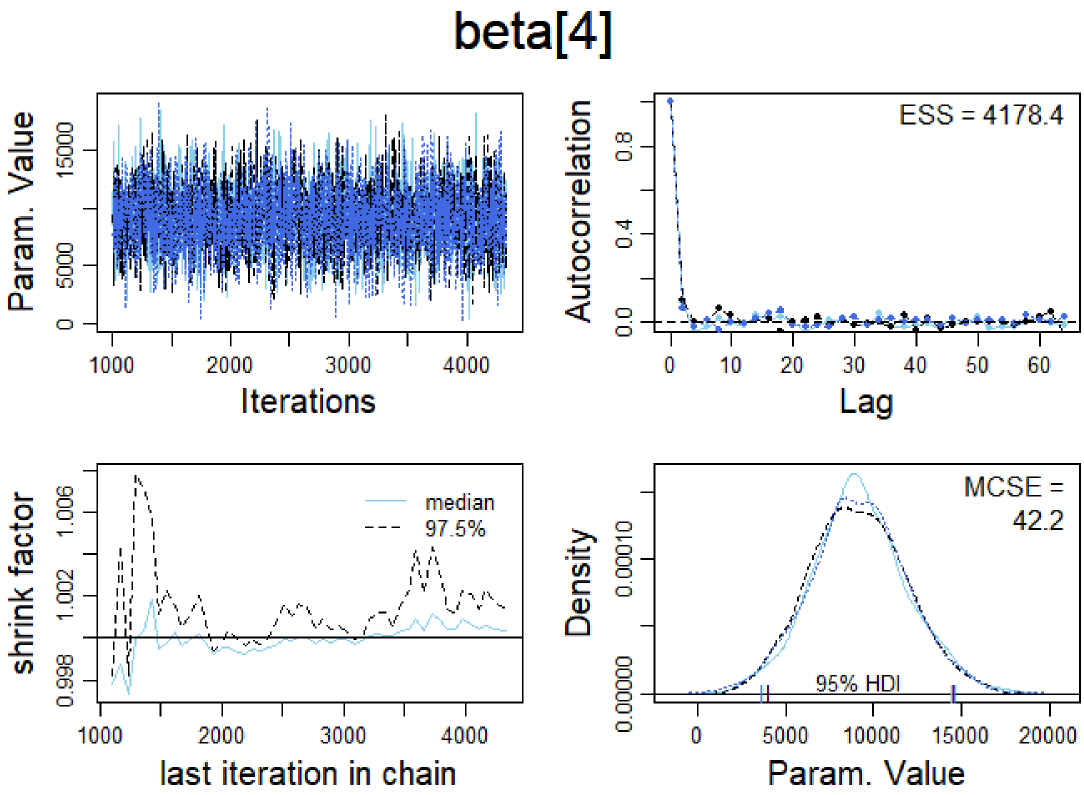


Figure28: Diagnostic of parameter beta[4]

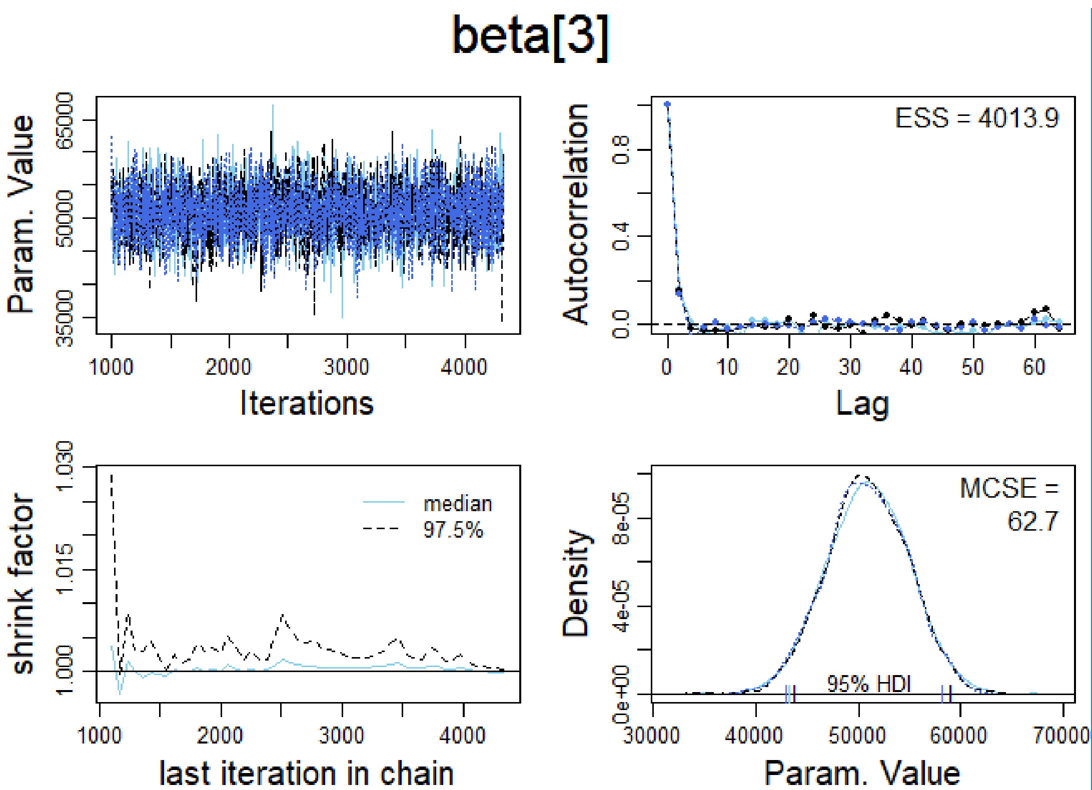


Figure29: Diagnostic of parameter beta[3]

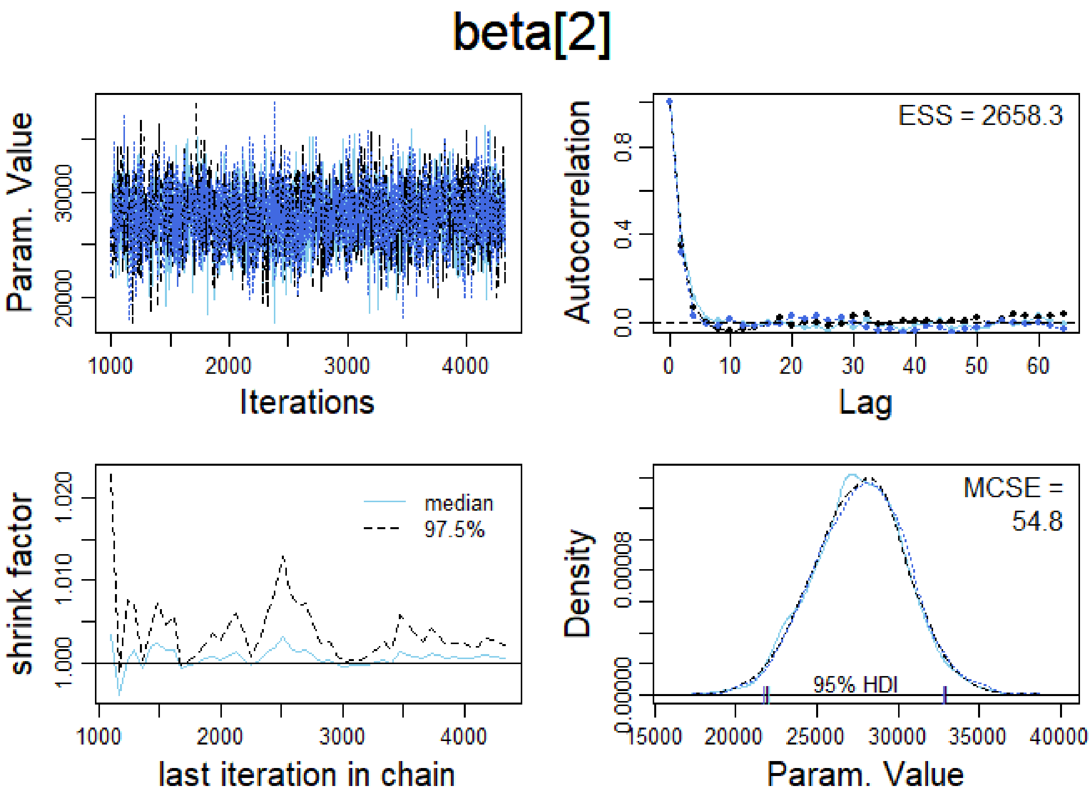


Figure30: Diagnostic of parameter beta[2]

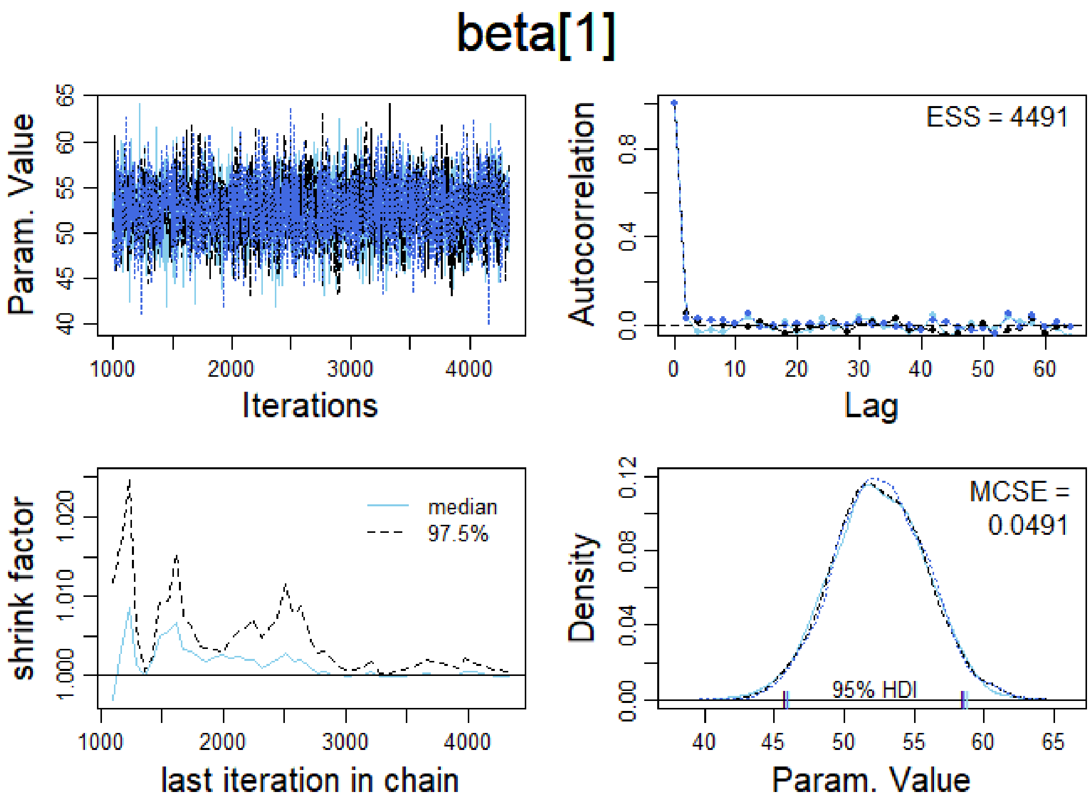


Figure31: Diagnostic of parameter beta[1]

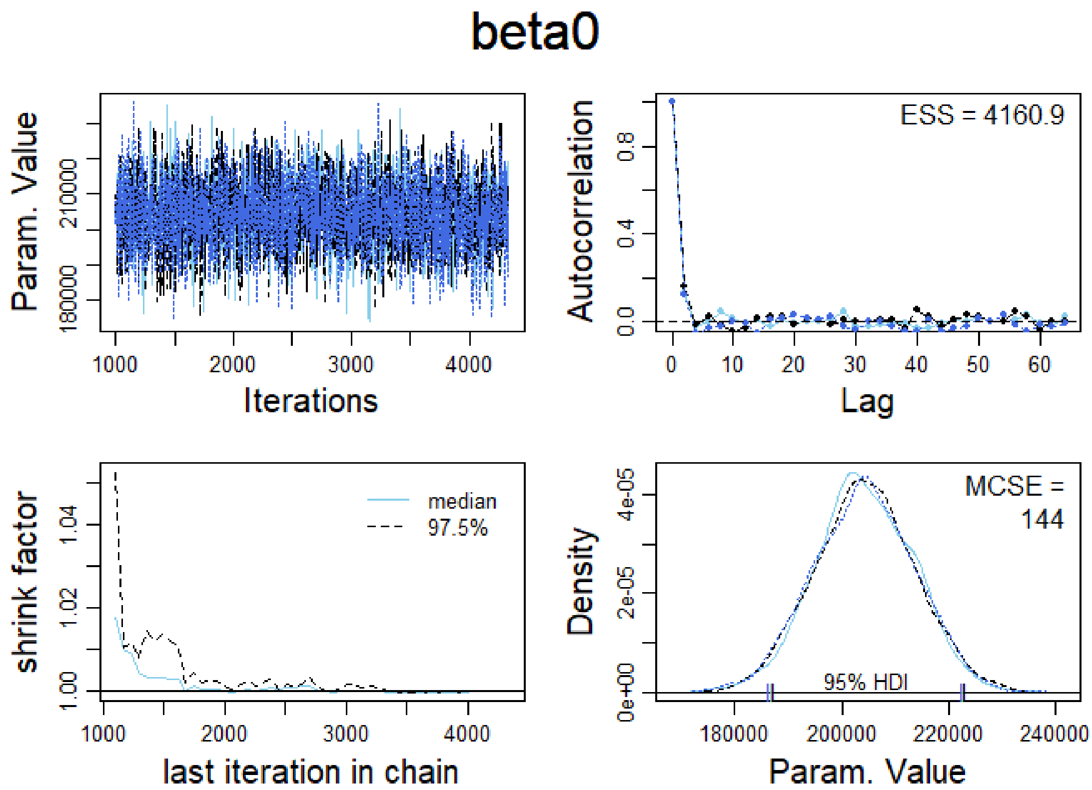


Figure32: Diagnostic of parameter beta0

2.2.3.4 Diagnostic for instance c (2500, 5, 4, 4, 0)

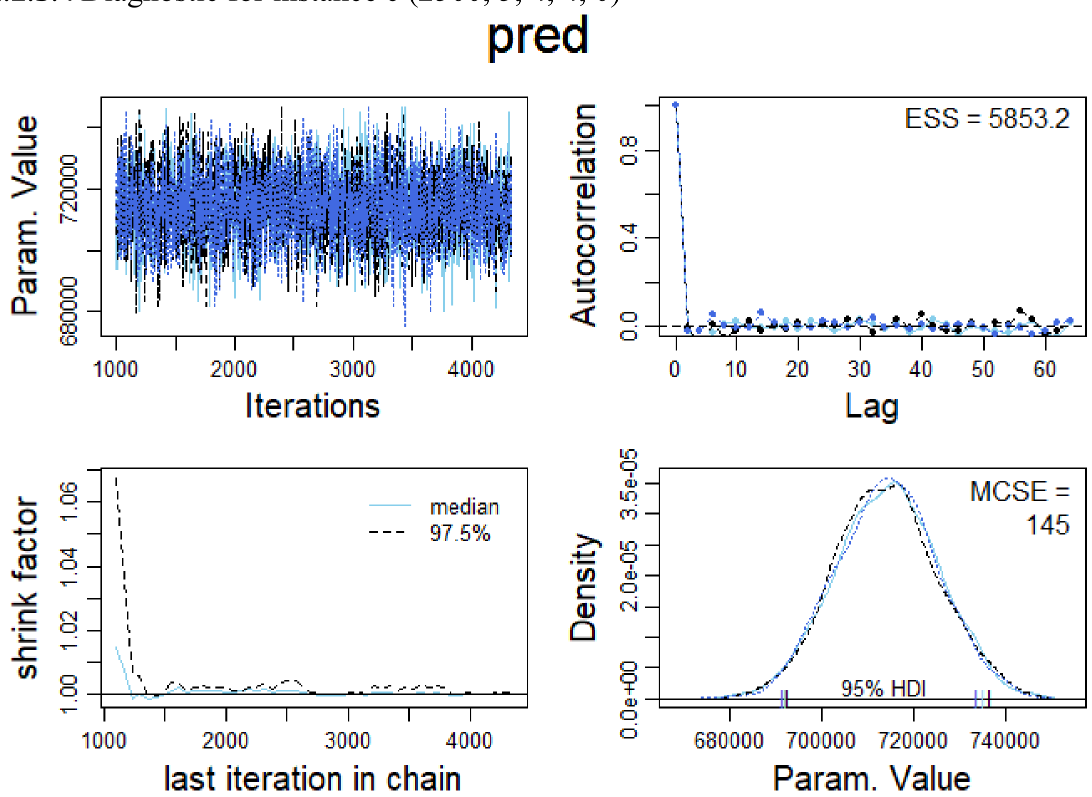


Figure33: Diagnostic of parameter prediction

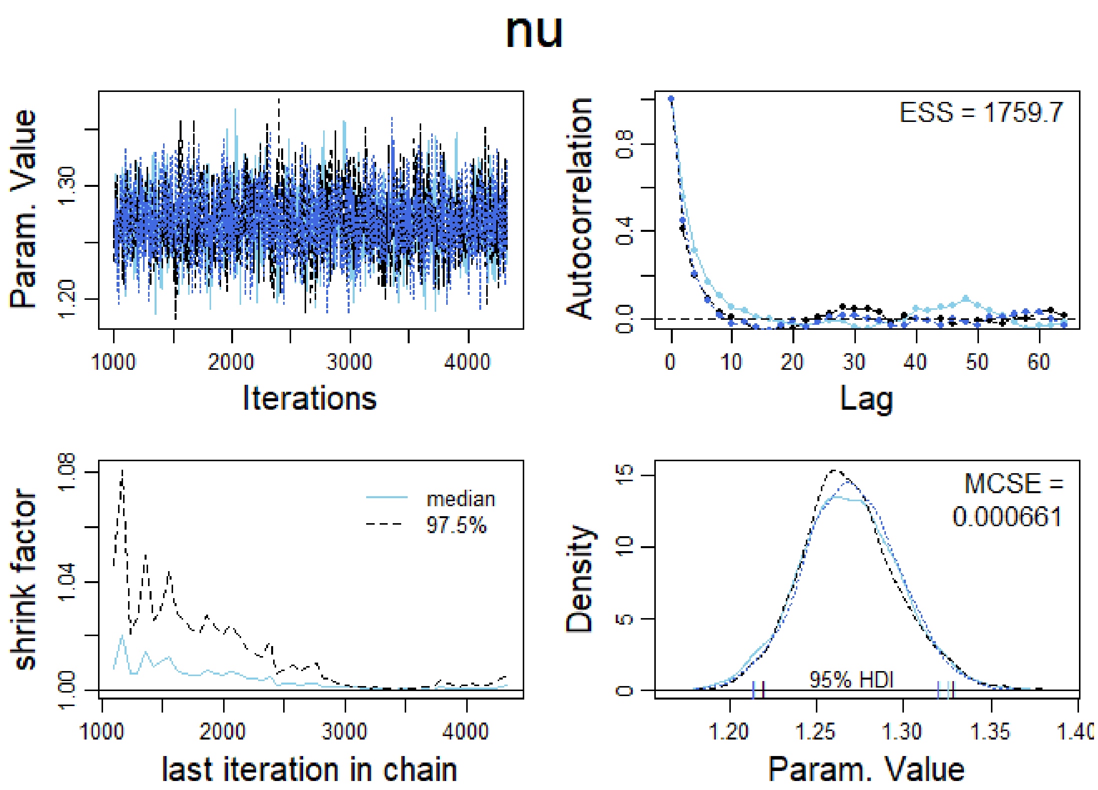


Figure34: Diagnostic of parameter nu

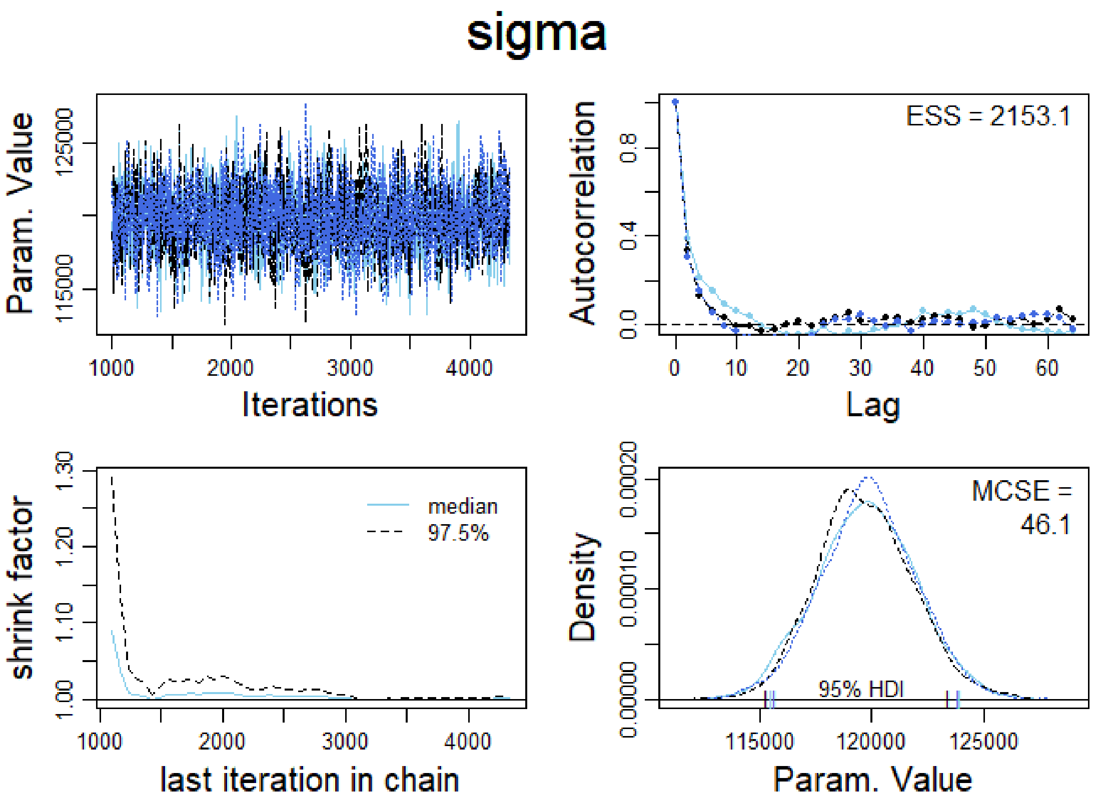


Figure35: Diagnostic of parameter sigma

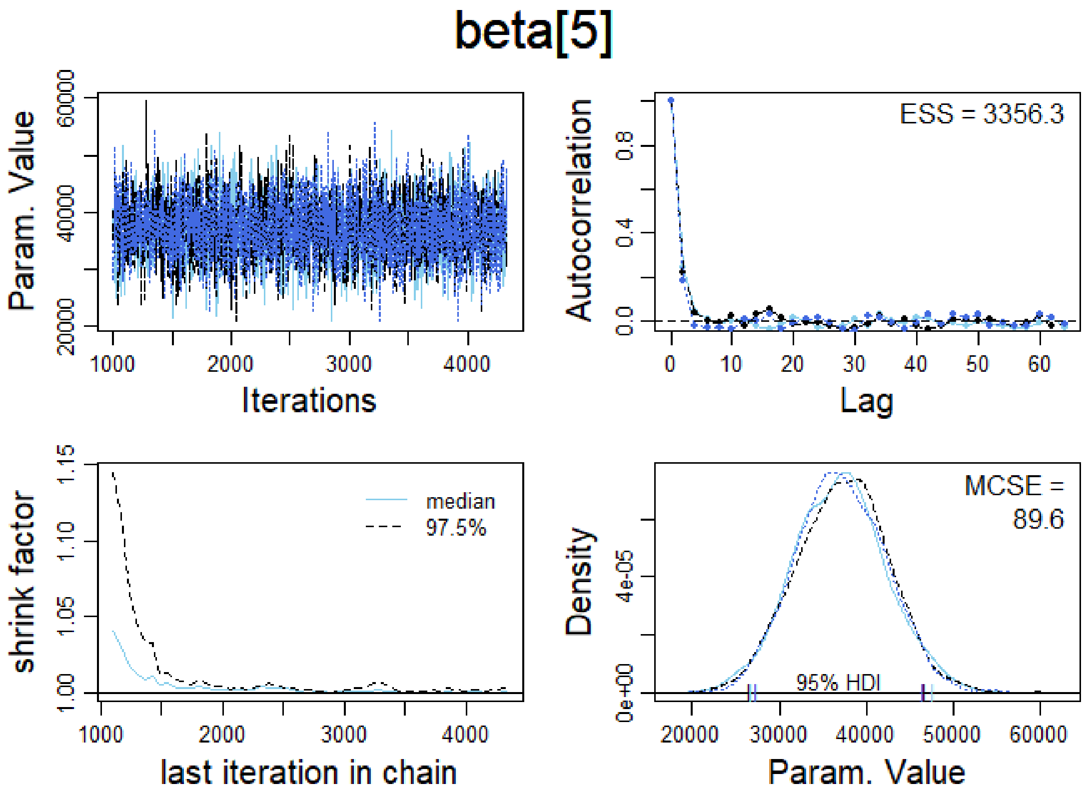


Figure36: Diagnostic of parameter beta[5]

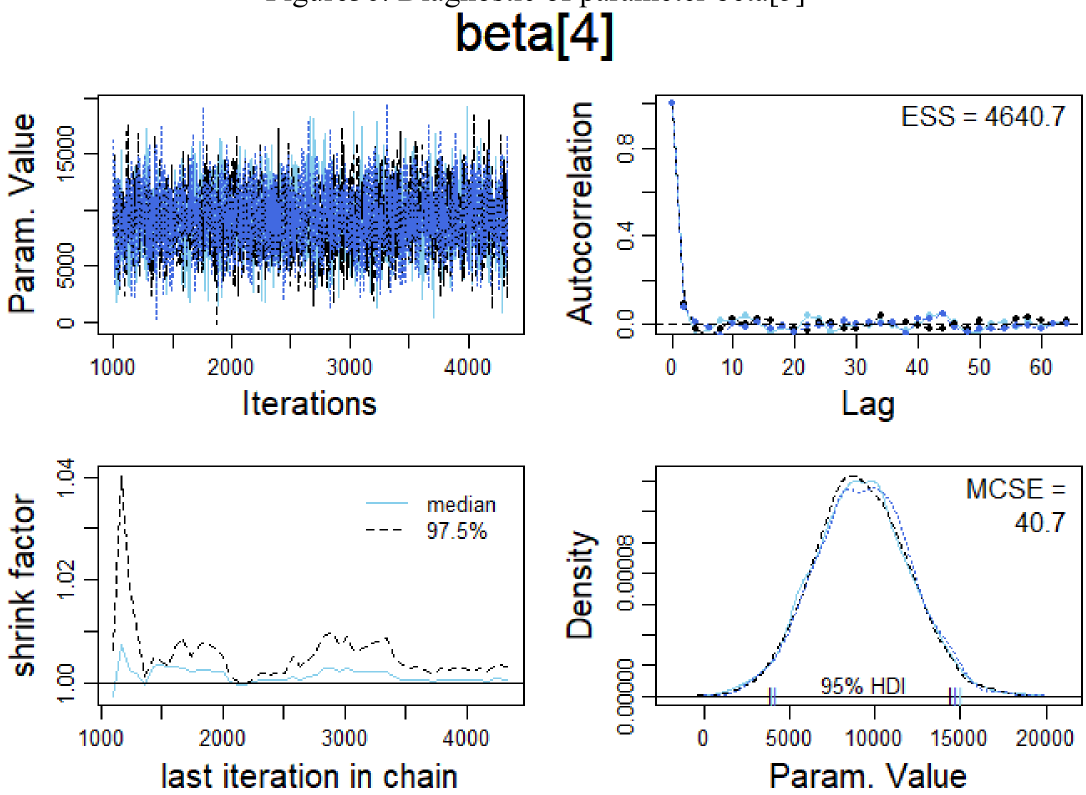


Figure37: Diagnostic of parameter beta[4]

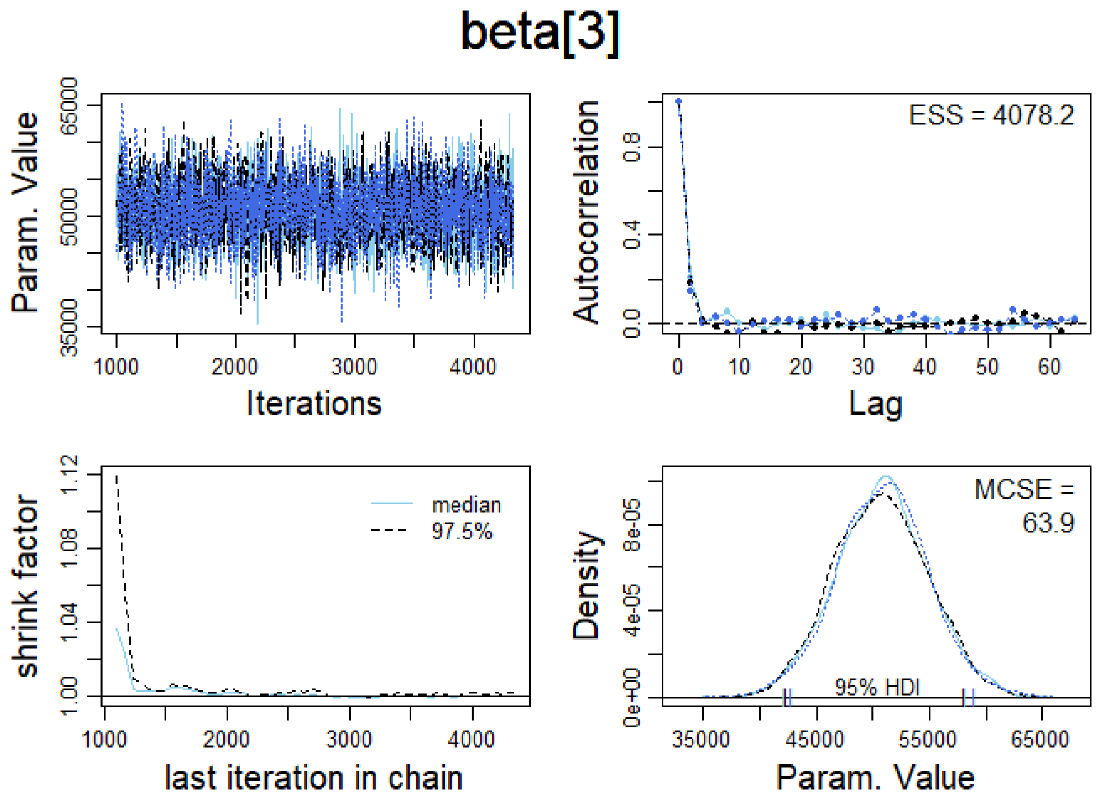


Figure38: Diagnostic of parameter beta[3]

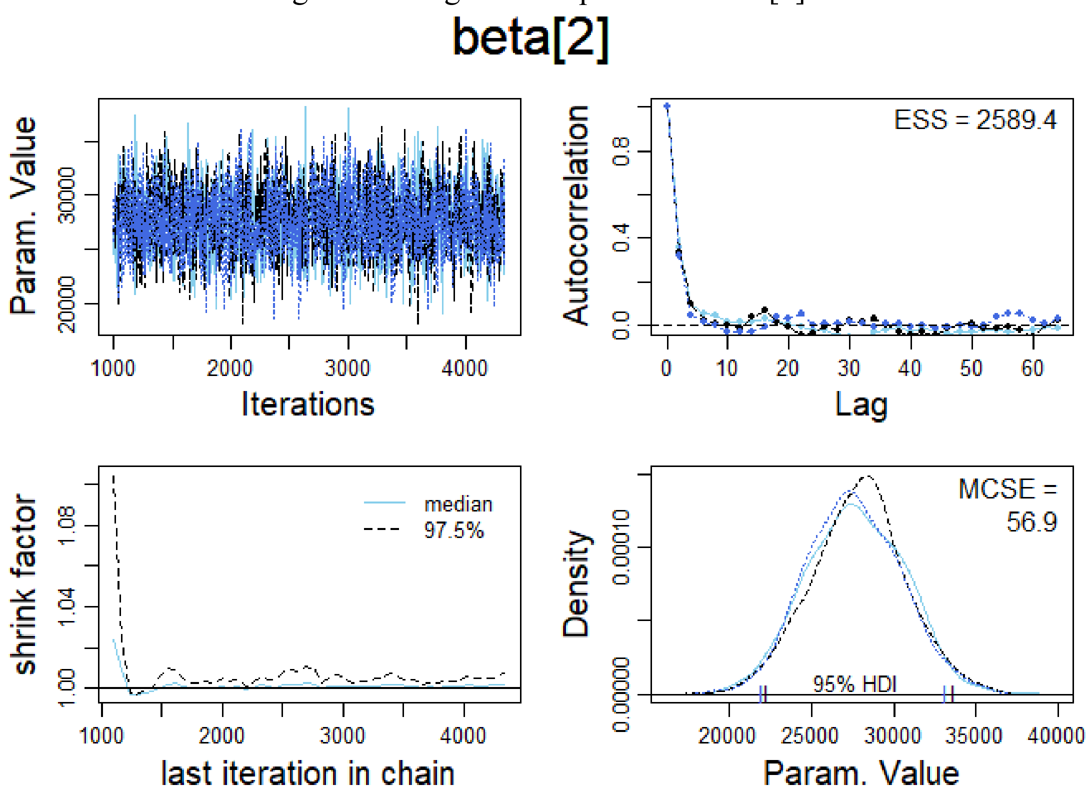


Figure39: Diagnostic of parameter beta[2]

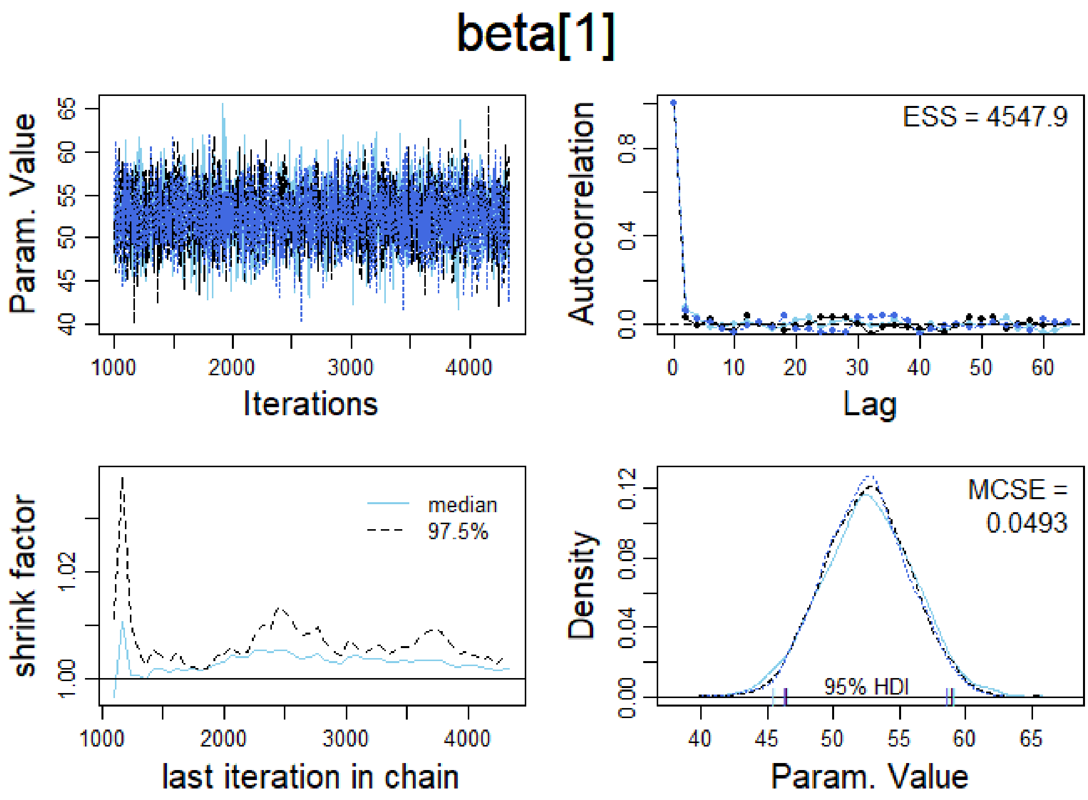


Figure40: Diagnostic of parameter beta[1]

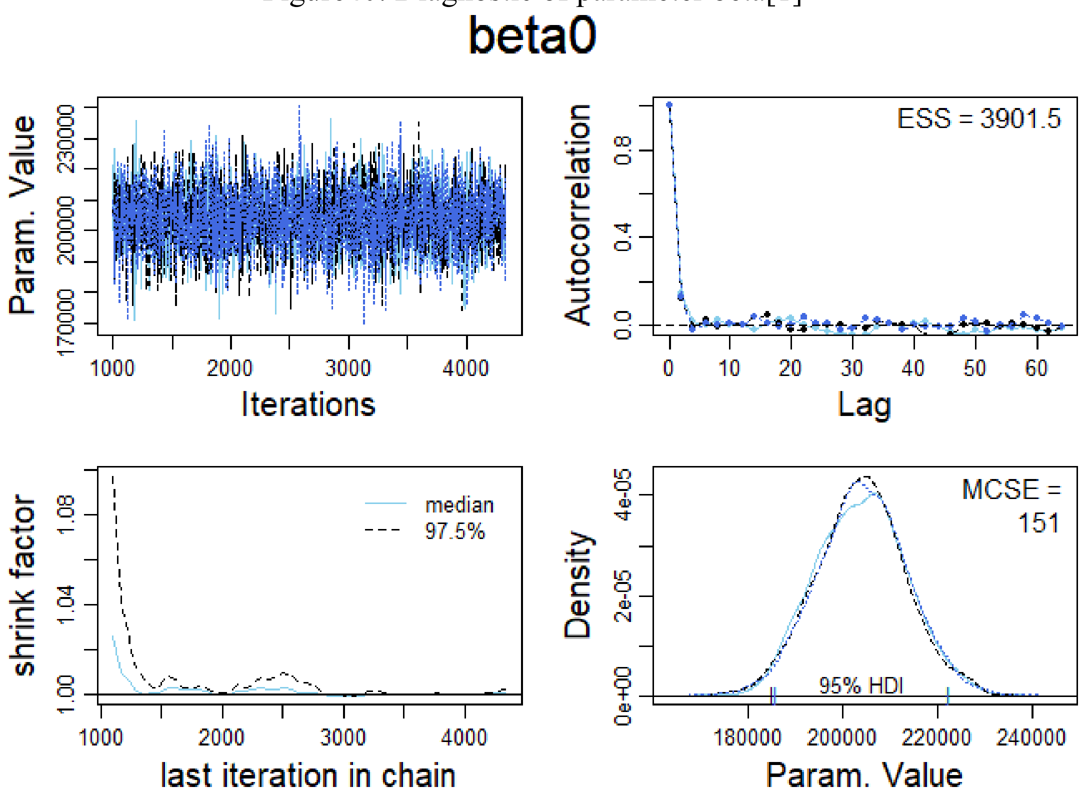


Figure41: Diagnostic of parameter beta0

2.2.3.5 Diagnostic for instance c (250, 3, 2, 1, 1)

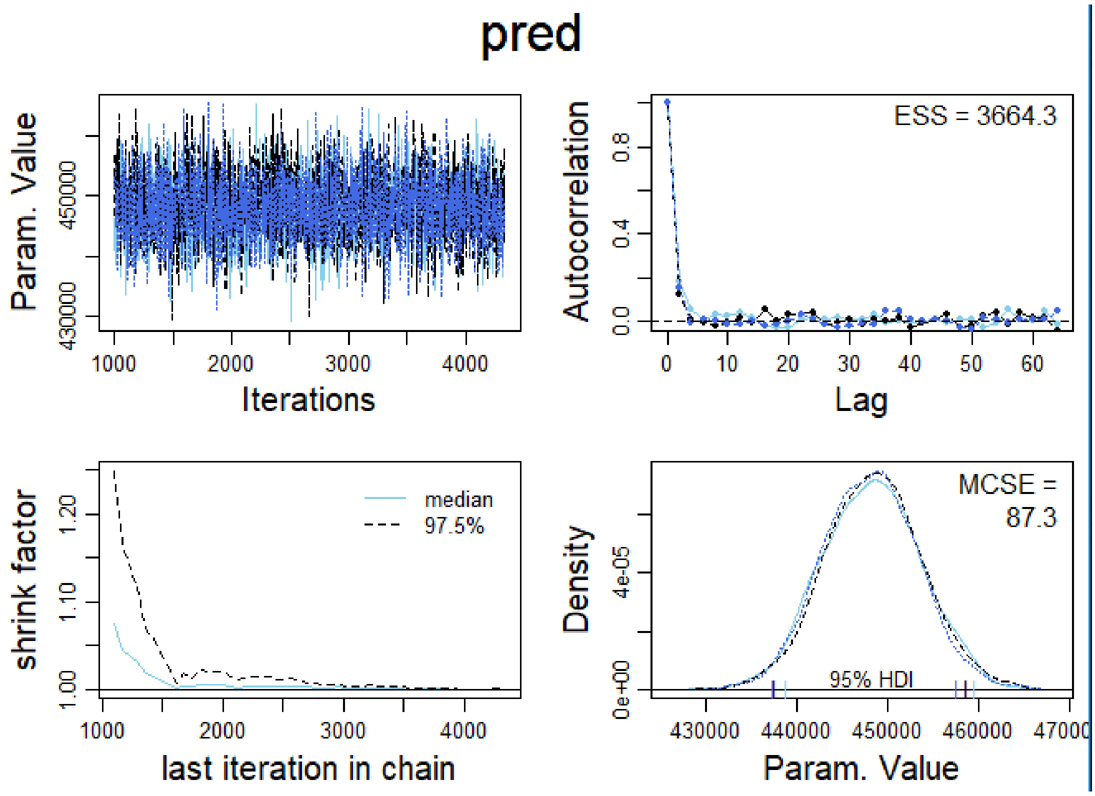


Figure42: Diagnostic of parameter prediction

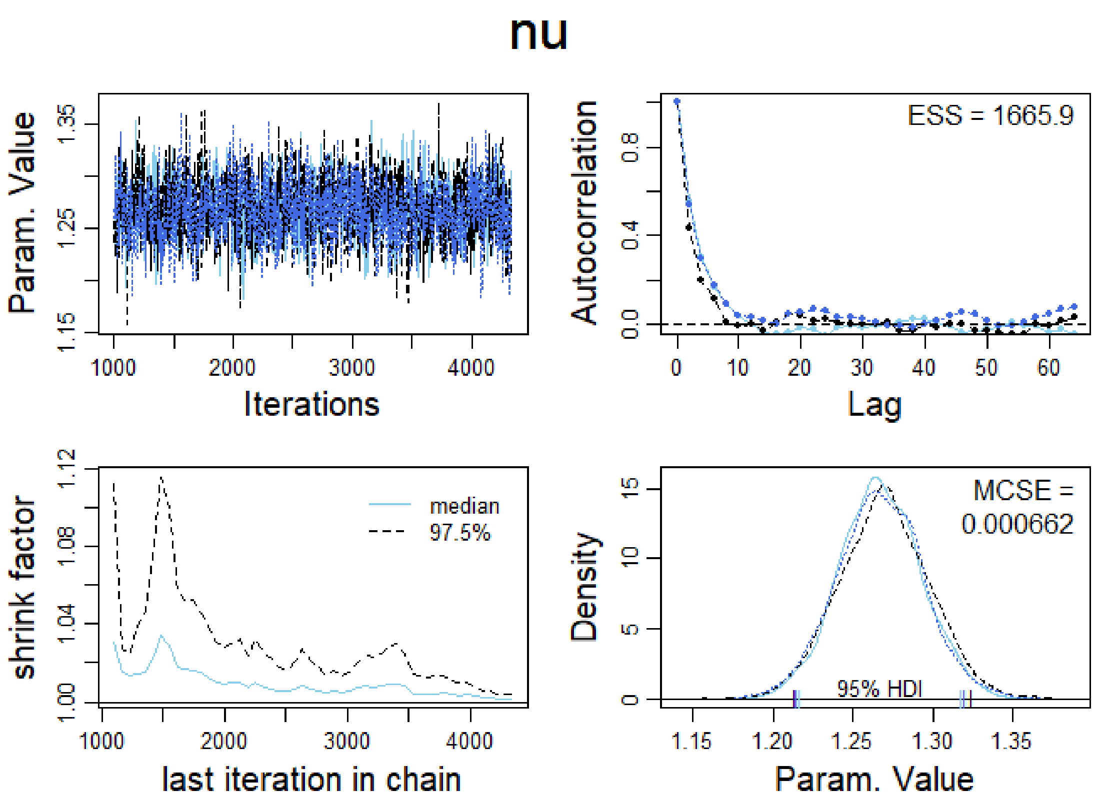


Figure43: Diagnostic of parameter nu

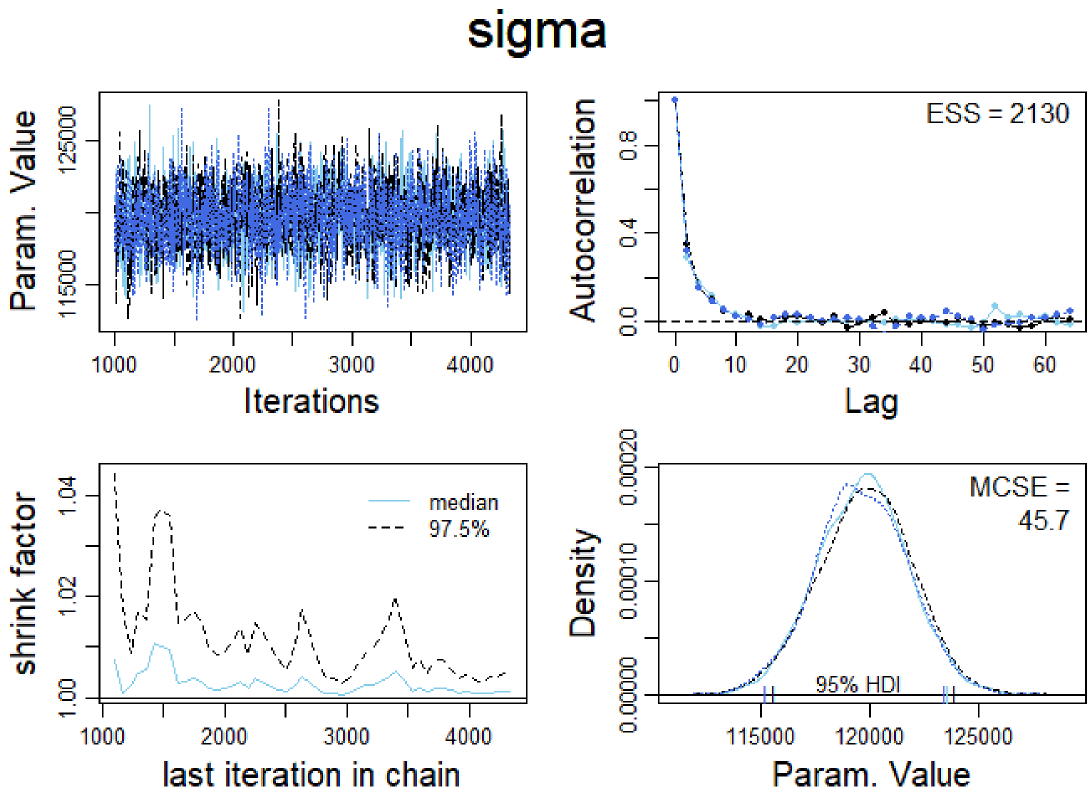


Figure44: Diagnostic of parameter sigma

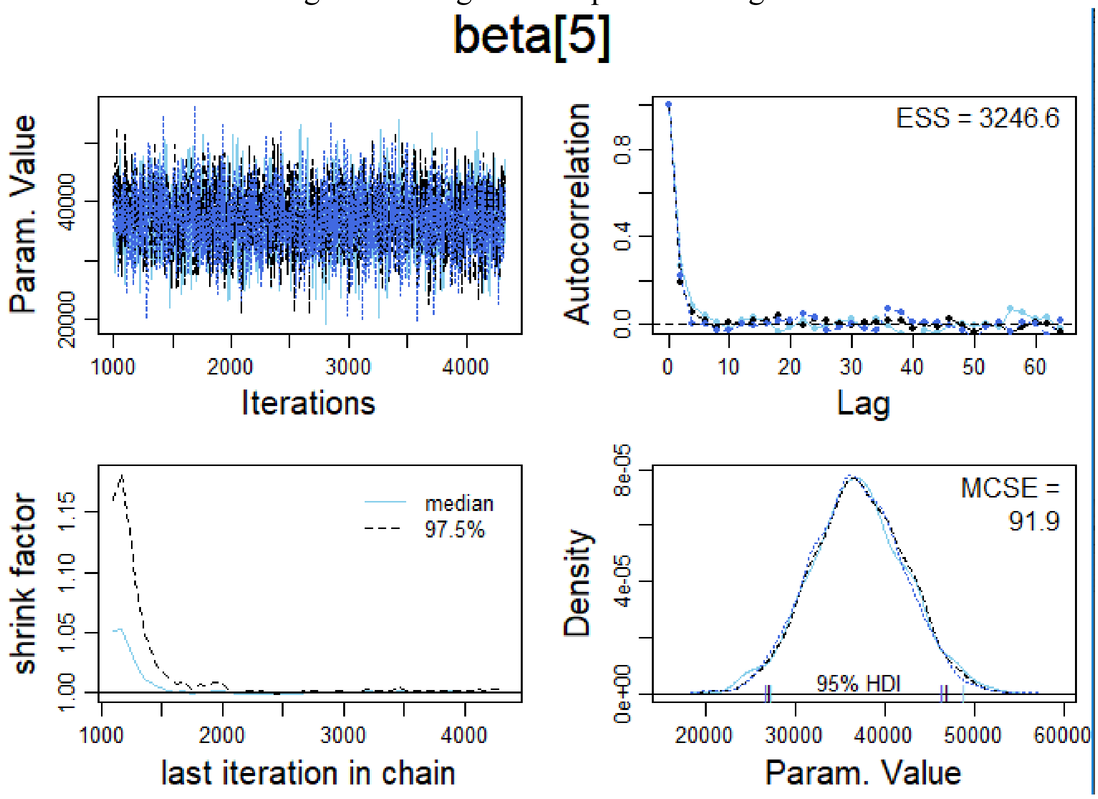


Figure45: Diagnostic of parameter beta[5]

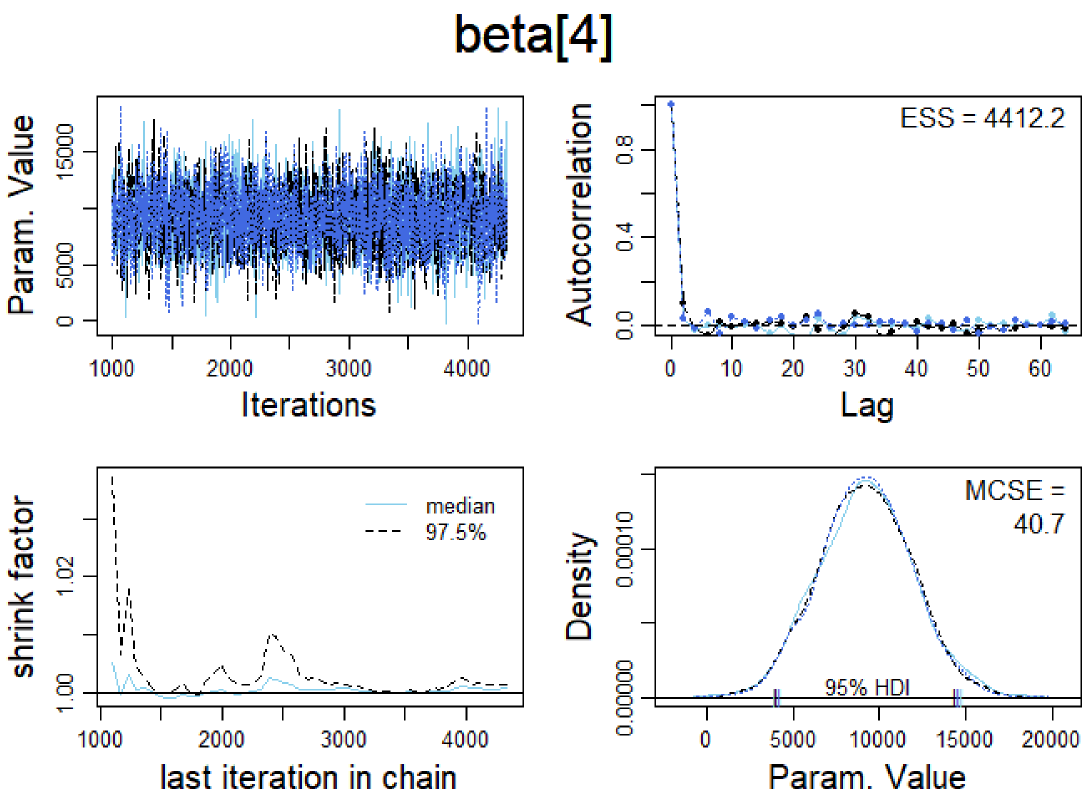


Figure46: Diagnostic of parameter beta[4]

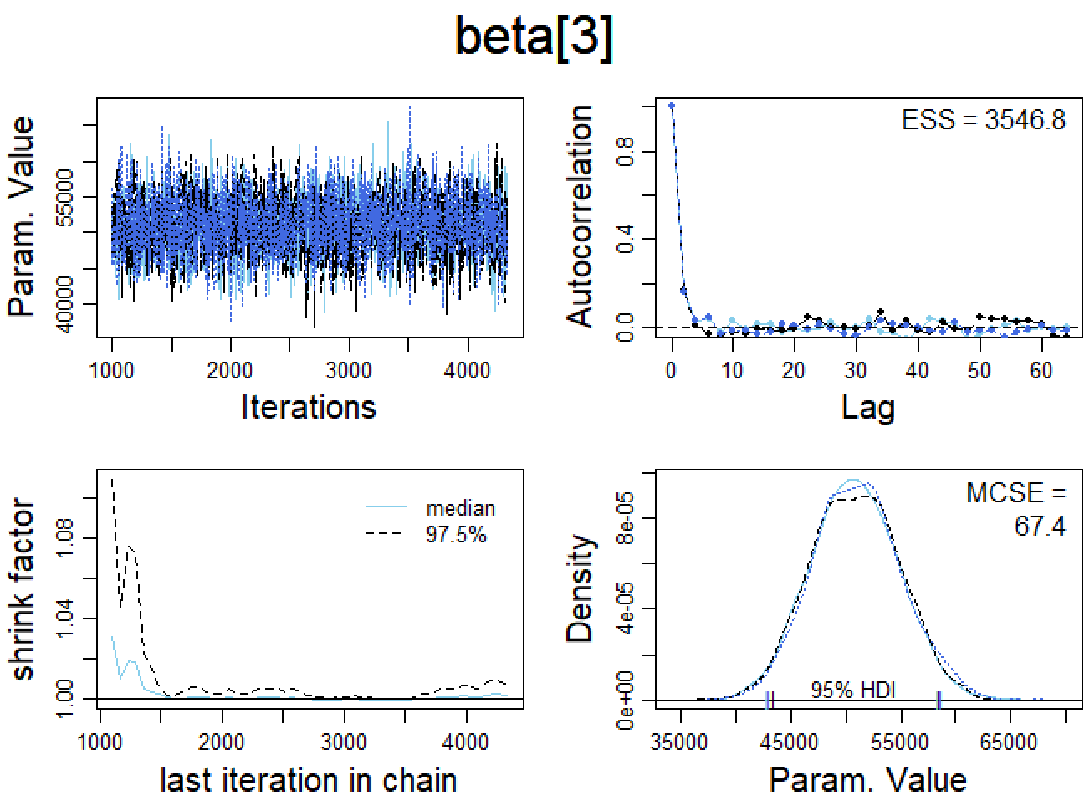


Figure47: Diagnostic of parameter beta[3]

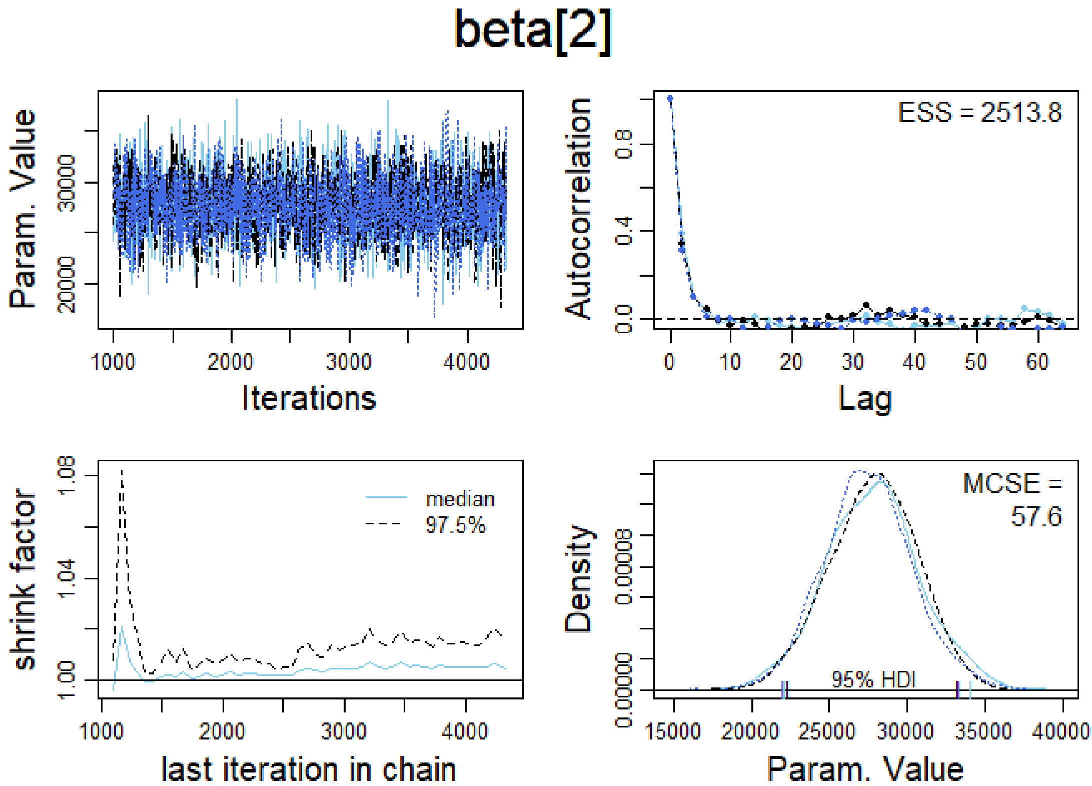


Figure48: Diagnostic of parameter beta[2]

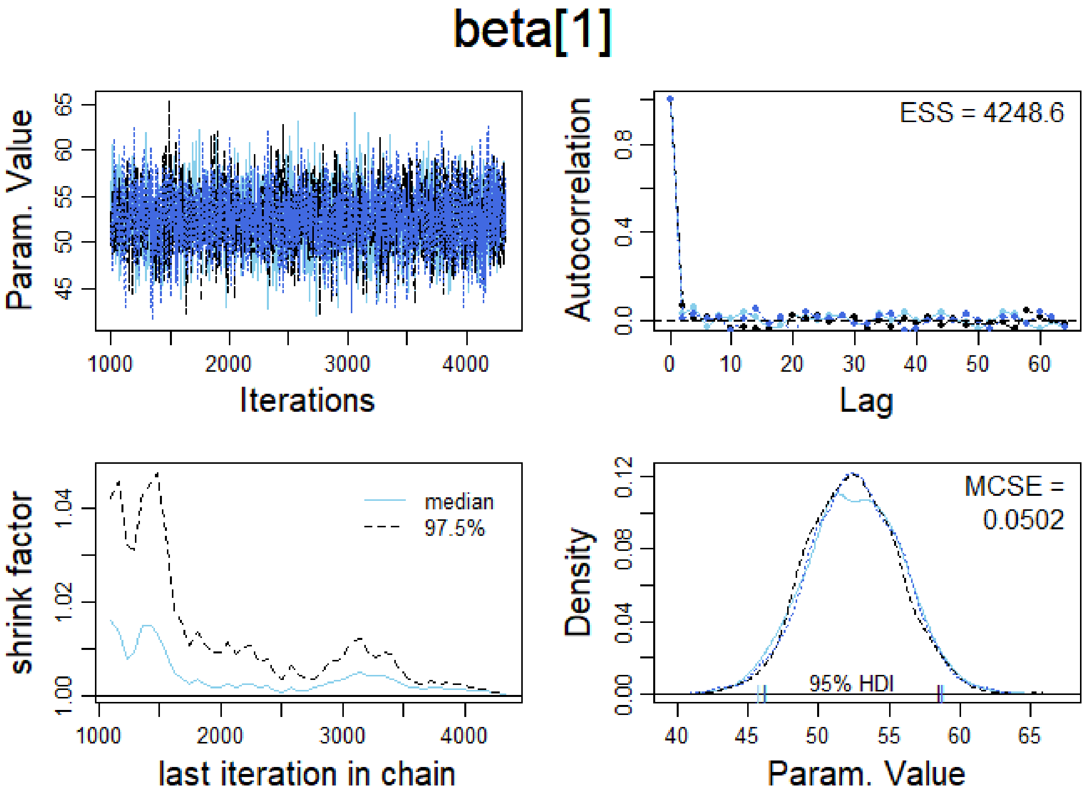


Figure49: Diagnostic of parameter beta[1]

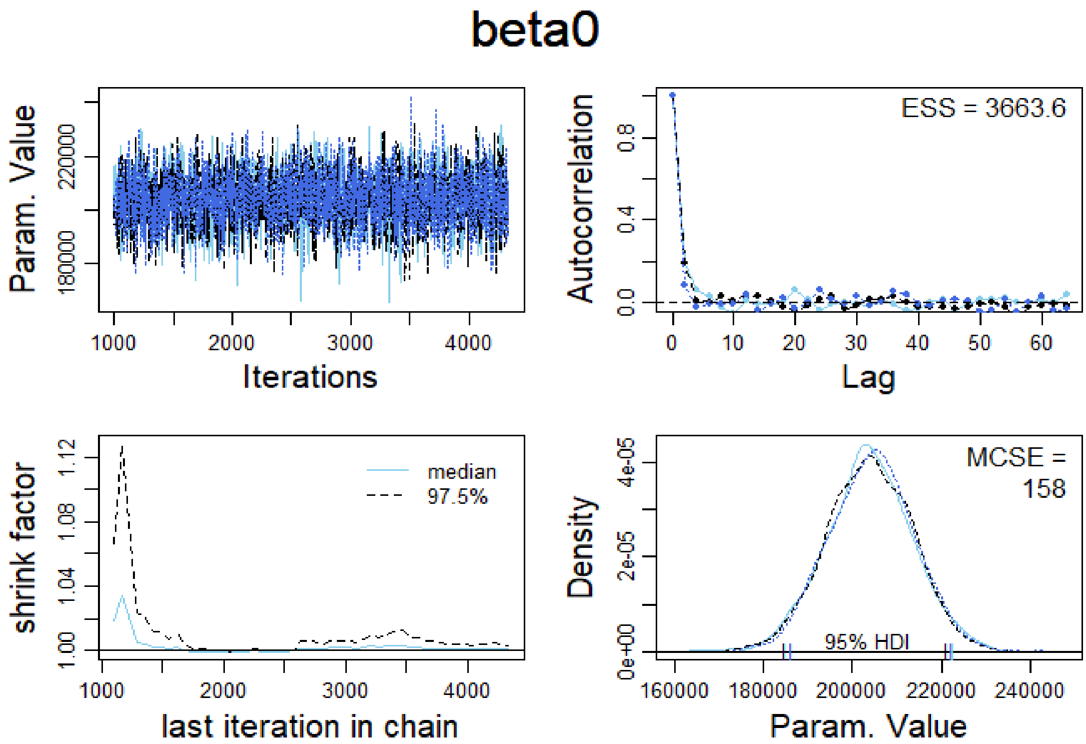


Figure50: Diagnostic of parameter beta0

2.2.4 Prediction distribution

This part will show the predicted value with SalePrice for each observation. Through these prediction values, we could get a basic idea for the SalePrice. For this part, we still using prediction distribution for each parameter base on each instance, and then to see whether the prediction is reasonable. If some value is much high or low that compare with sample and prior information, we should better go back to adjust value for parameters with prior information in our model block.

2.2.4.1 Prediction for instance c (600, 2, 2, 1, 1)

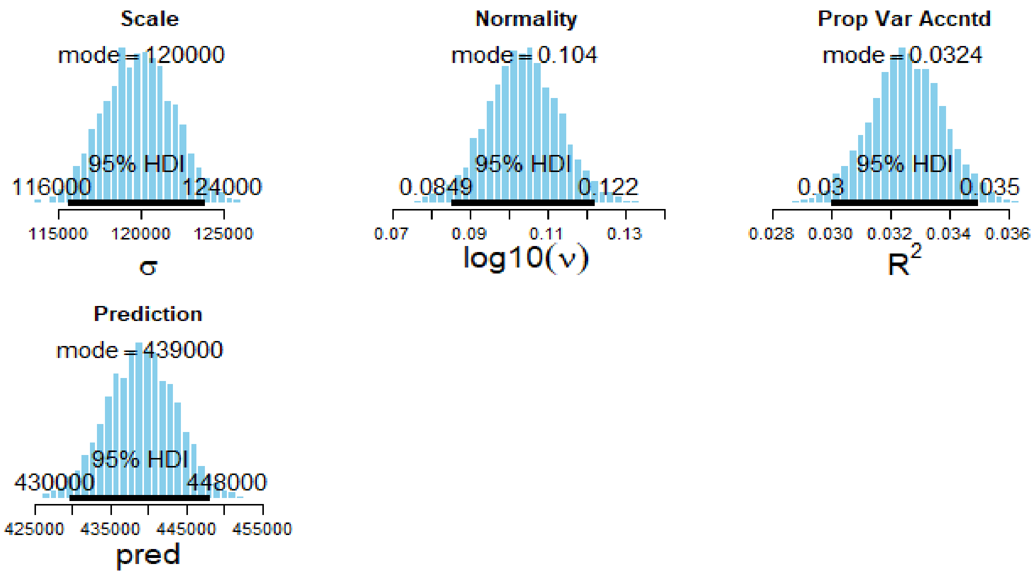


Figure51: Distribution for prediction price

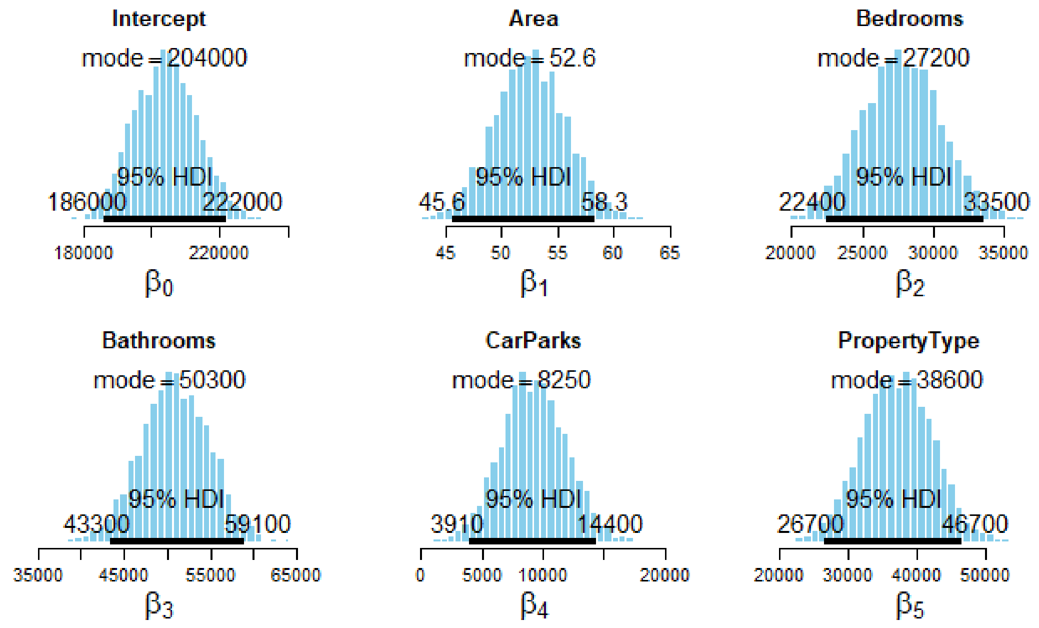


Figure52: Distribution for each parameter

Base on figure, the mode of prediction value for price is 439000, and the range from 430000 to 448000 with 95% high density interval.

The formula this model is:

y1 = 204000 + 52.6*beta[1] + 27200*beta[2] + 50300*beta[3] + 8250*beta[4] + 38600*beta[5]

2.2.4.2 Prediction for instance c (800, 3, 1, 2, 0)

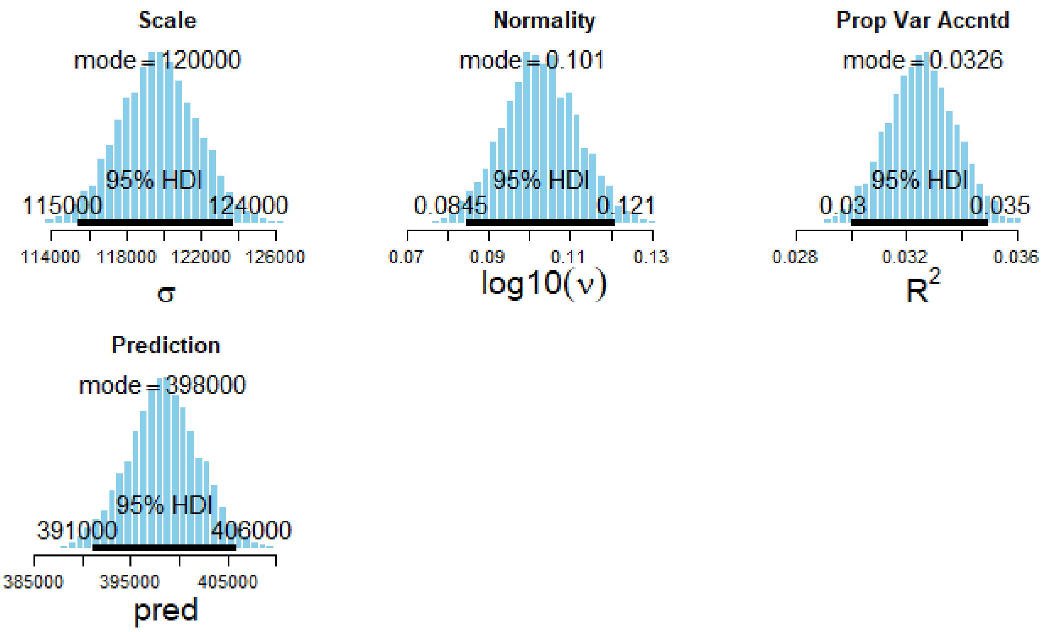


Figure53: Distribution for prediction price

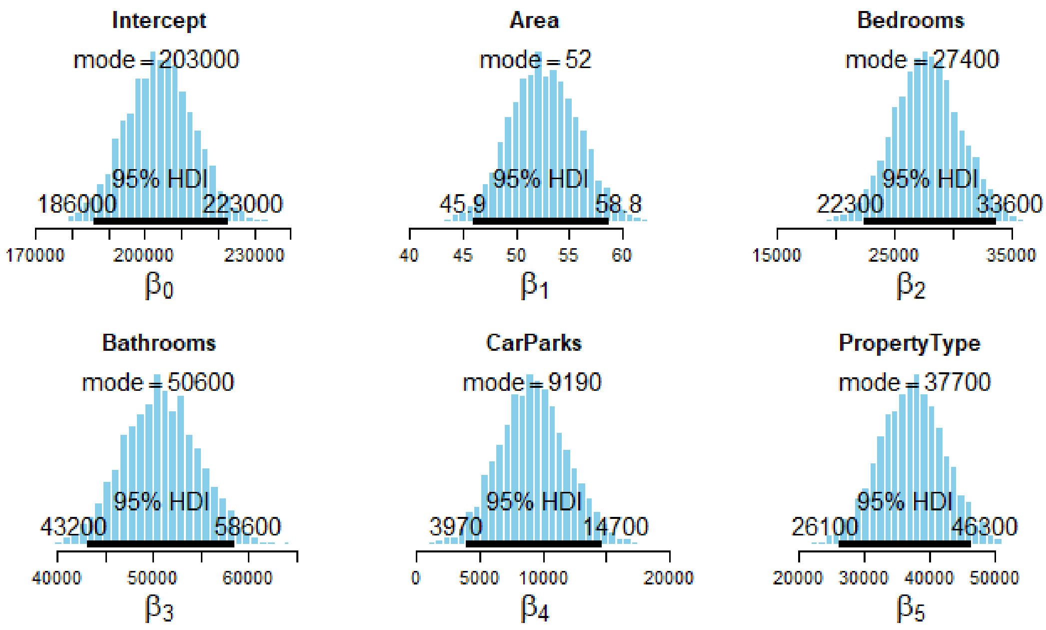


Figure54: Distribution for each parameter

Base on figure, the prediction value for price is 398000, and the range from 391000 to 406000 with 95% high density interval. This one seems reasonable.

The formula this model is:

y2 = 203000 + 52*beta[1] + 27400*beta[2] + 50600*beta[3] + 9190*beta[4] + 37700*beta[5]

2.2.4.3 Prediction for instance c (1500, 2, 1, 1, 0)

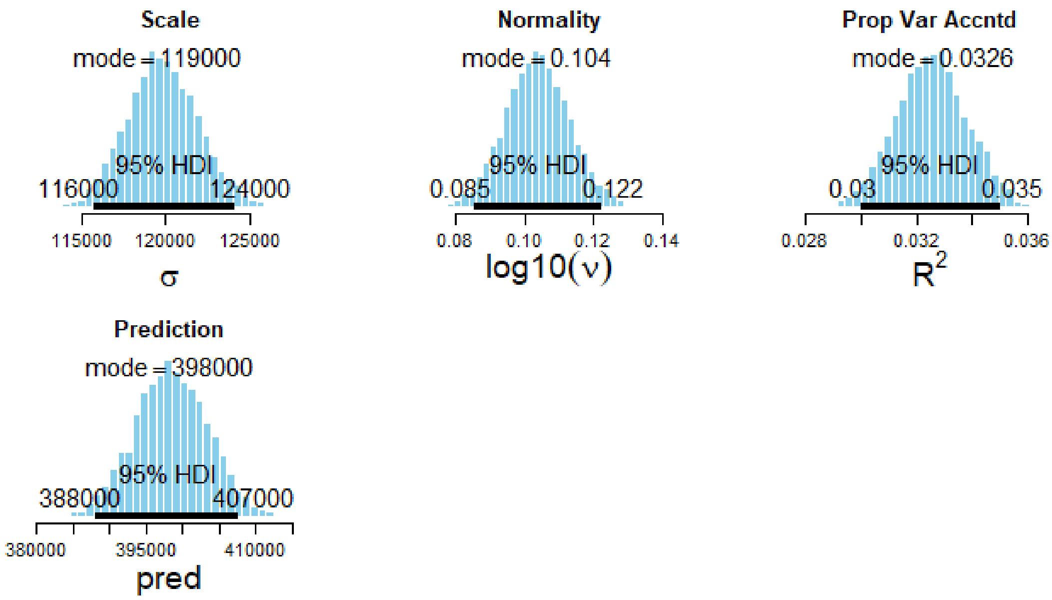


Figure55: Distribution for prediction price

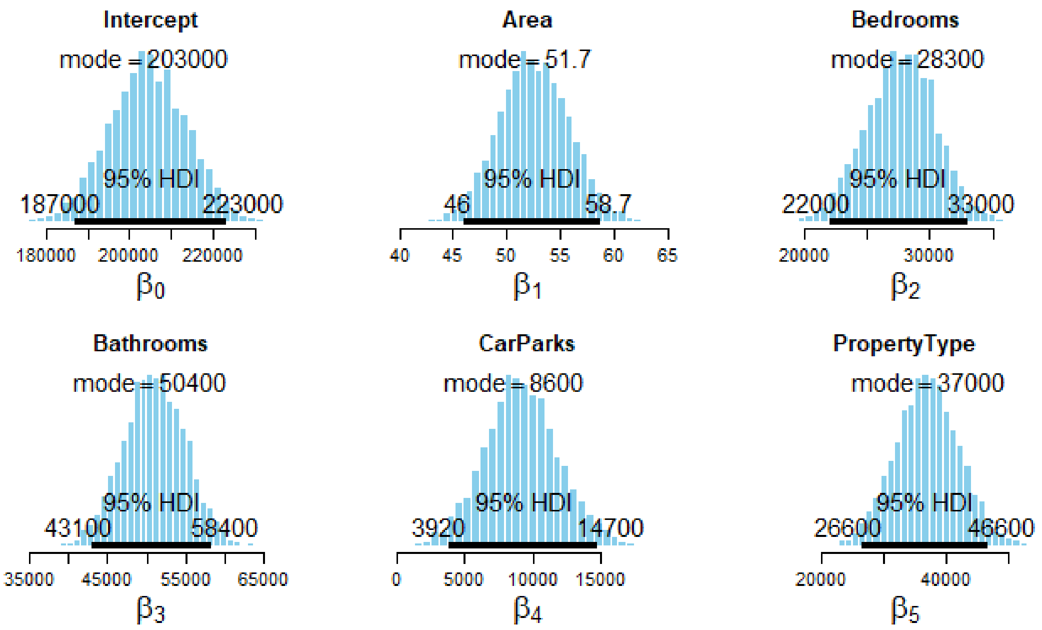


Figure56: Distribution for each parameter

Base on figure, the prediction value for price is 398000, and the range from 388000 to 407000 with 95% high density interval. This one seems reasonable.

The formula this model is:

y3 = 203000 + 51.7*beta[1] + 28300*beta[2] + 50400*beta[3] + 8600*beta[4] + 37000*beta[5]

2.2.4.4 Prediction for instance c (2500, 5, 4, 4, 0)

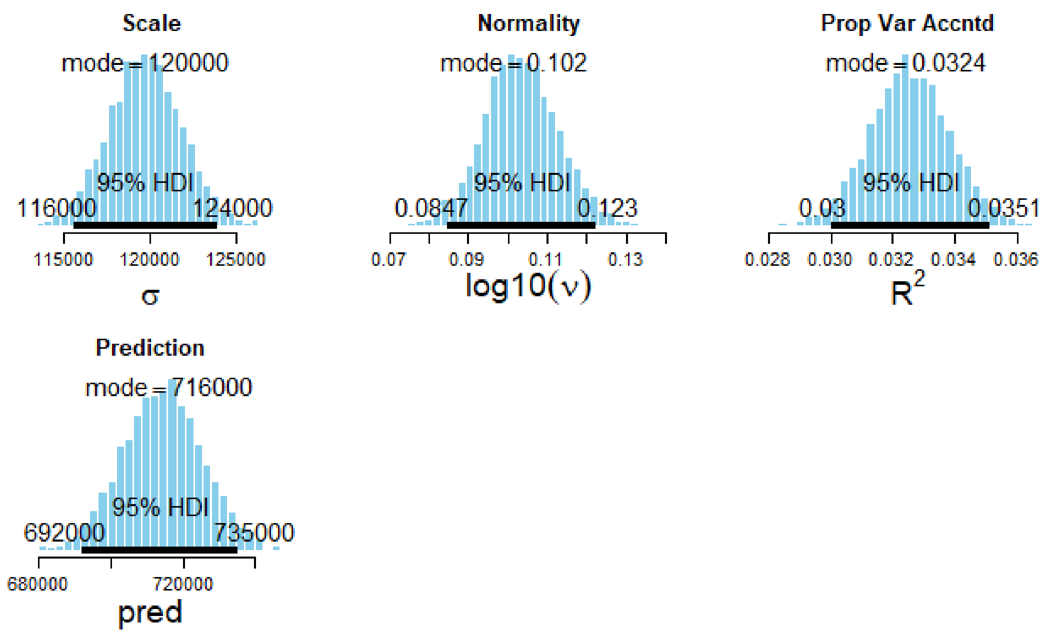


Figure57: Distribution for prediction price

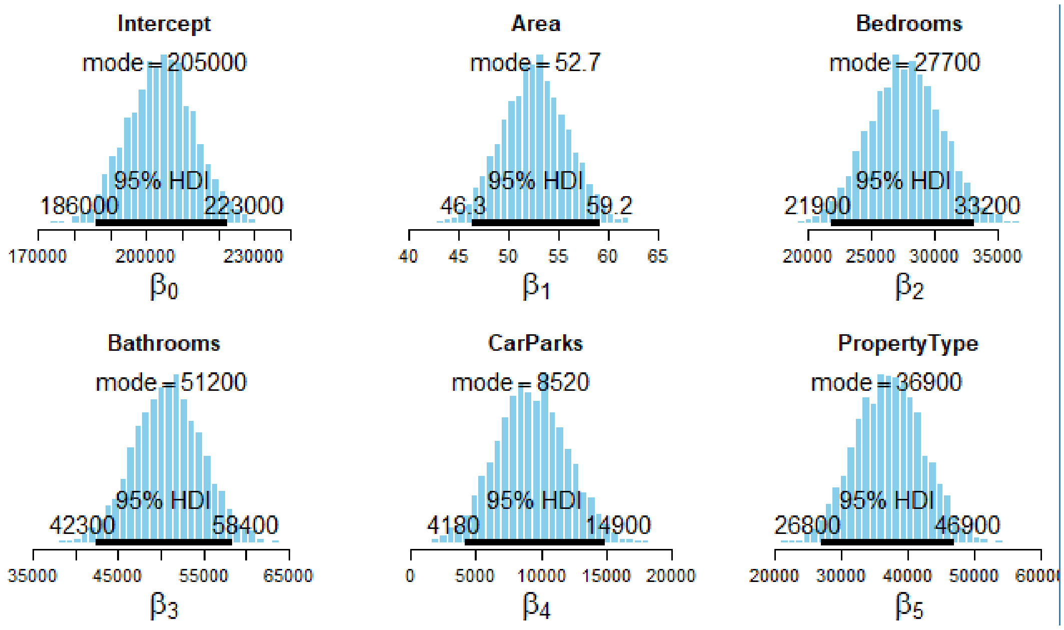


Figure58: Distribution for each parameter

Base on figure, the prediction value for price is 716000, and the range from 692000 to 735000 with 95% high density interval. This one seems reasonable.

The formula this model is:

y4 = 205000 + 52.7*beta[1] + 27700*beta[2] + 51200*beta[3] + 8520*beta[4] + 36900*beta[5]

2.2.4.5 Prediction for instance c (250, 3, 2, 1, 1)

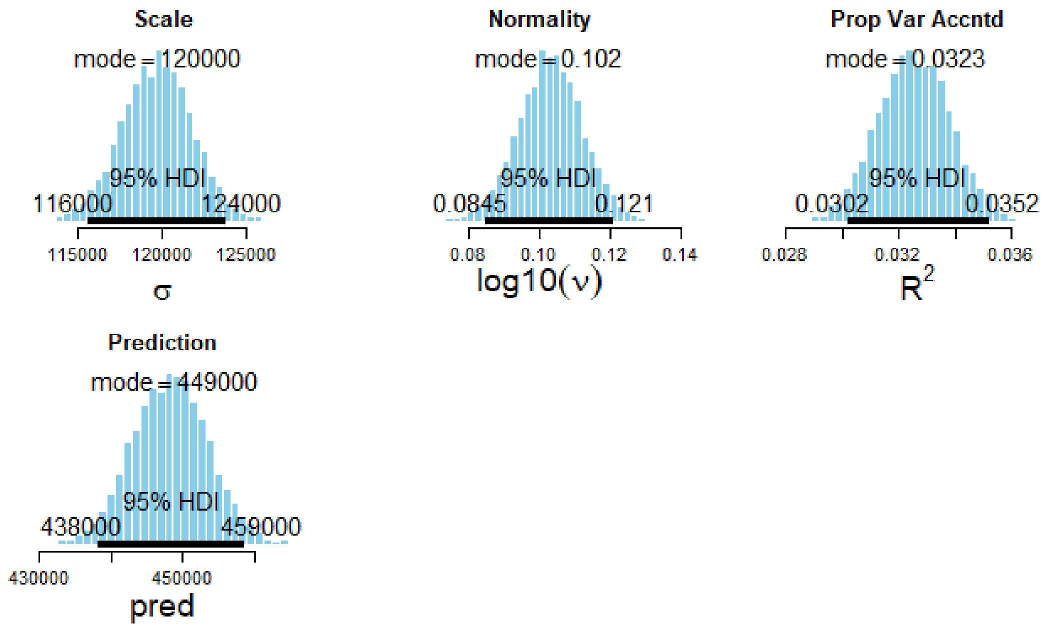


Figure59: Distribution for prediction price

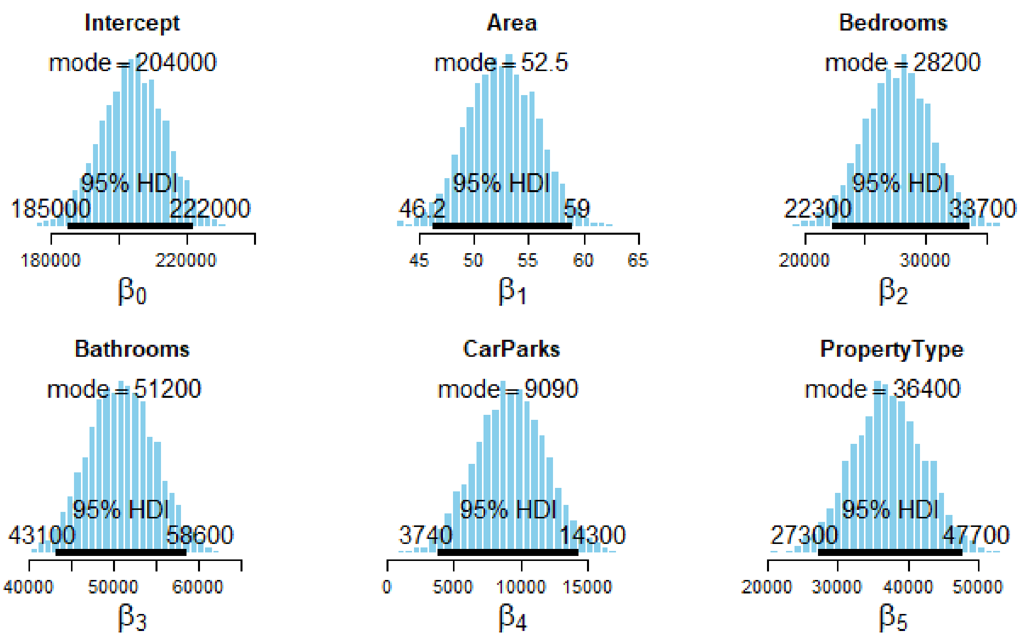


Figure60: Distribution for each parameter

Base on figure, the prediction value for price is 449000, and the range from 438000 to 459000 with 95% high density interval. This one seems reasonable.

y5 = 204000 + 52.5*beta[1] + 28200*beta[2] + 51200*beta[3] + 9090*beta[4] + 36400*beta[5]

2.2.5 Comparison

For each instance, the R square is small, just around 0.032, it seems that this value is not good enough. However, all predicted prices seem like reasonable. In fact, estimated parameters are similar for each instance that could be seen below.

y1 = 204000 + 52.6*beta[1] + 27200*beta[2] + 50300*beta[3] + 8250*beta[4] + 38600*beta[5]

y2 = 203000 + 52*beta[1] + 27400*beta[2] + 50600*beta[3] + 9190*beta[4] + 37700*beta[5]

y3 = 203000 + 51.7*beta[1] + 28300*beta[2] + 50400*beta[3] + 8600*beta[4] + 37000*beta[5]

y4 = 205000 + 52.7*beta[1] + 27700*beta[2] + 51200*beta[3] + 8520*beta[4] + 36900*beta[5]

y5 = 204000 + 52.5*beta[1] + 28200*beta[2] + 51200*beta[3] + 9090*beta[4] + 36400*beta[5]

Table1: Prediction sale price for instance

Property No	Area	Bedrooms	Bathrooms	CarParks	PropertyType	Prediction(mode)
1	600	2	2	1	Unit	439000
2	800	3	1	2	House	398000
3	1500	2	1	1	House	398000

4	2500	5	4	4	House	716000
5	250	3	2	1	Unit	449000
Note: Property Type is represented by number (Unit: 1, House: 0)						

3 Conclusion

Through this report, we got estimated value for model parameters with non-informative condition for part A. How likelihood domain prior distribution base on condition could be seen clearly. Furthermore, a reasonable mu value that close to sample mean with a lower variance shown in the figures. For part B, applying multiple linear regression to define model in jags, and then use the model to predict SalePrice for each instance. Through this part, how the multiple linear regression works with models, and the formula for prediction variable could be seen clearly. Overall, all of prediction values for variable SalePrice are reasonable, and they all base on knowledge with both prior and likelihood information, could be more informative than other methodologies.