Chloe Segale

Problem Set 2

1. Error models, such as the classical linear regression model, define a likelihood based on some relevant principles such as (1) The sign of the error doesn't matter, only the magnitude does, and (2) The prior should be exchangeable, since it doesn't matter what order we see the observations in. If the data as a list of pairs: $\{\{y_1, x_1\}, ..., \{y_n, x_n\}\}$, let the deterministic prediction of whatever model we are considering be written as $y[x;\theta]$. Then the errors are $e = \{e_1 = y_1 - y[x_1;\theta], e_2 = y_2 - y[x_2;\theta], ..., e_n = y_n - y[x_n;\theta]\}$ Given the class of models represented by the function $y[\cdot]$ the statistical problem is to recover a posterior probability $p[\theta|\{y,x\}]$. One common measure of the size of the errors a model makes is the average size of the squared errors, $\frac{1}{n}\sum_{i=1}^{n} e_i^2$, which measures the square of the average mistake the model made in predicting the data. Since we are interested in the size of the average mistake, we might base the likelihood for the model on the square root of the average squared scaled error, rmse $= \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2}$ which is called the *root mean squared error.* The most common likelihood based on the $rmse$ includes a penalty to the model for the number of chances the model had to make a mistake and is written:

$$p[\{y_1, ..., y_n\}|y[x;\theta]] \propto \text{rmsq}^{-n} = \left(\frac{1}{n}\sum_{i=1}^{n} e_i^2\right)^{-n/2} \quad (1.1)$$

   a) If the model class we are interested in is the set of linear relations $y[x;\beta_0,\beta_2] = \beta_0 + \beta_1 x$, what is the root mean squared error likelihood?

$$e = \{e_1 = y_1 - y[x_1;\beta_0,\beta_2], \dots, e_n = y_n - y[x_n;\beta_0,\beta_2]\}$$

$$e = \{e_1 = y_1 - y(\beta_o + \beta_2 x_1), e_2 = y_2 - y(\beta_o + \beta_2 x_2), \dots, e_n = y_n - y(\beta_o + \beta_2 x_n)\}$$

$$rmse = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - (\beta_o + \beta_2 x_i))^2}$$

$$p[\{y_1, \dots, y_n\}|y[x;\beta_o,\beta_2]] \propto \left(\frac{1}{n}\sum_{i=1}^{n} (y_i - (\beta_o + \beta_2 x_i))^2\right)^{-\frac{n}{2}}$$

   b) With uniform priors over β_0 and β_1 what are the maximum posterior probability estimates of β_0 and β_1?

uniform priors so posterior proportional to likelihood

$$\beta o \ @ \ \frac{\partial}{\partial \beta_o}\left(\frac{1}{n}\sum_{i=1}^{n}(y_i - (\beta_o + \beta_2 x_i))^2\right)^{-\frac{n}{2}} = 0$$

$$\beta_1 \ @ \ \frac{\partial}{\partial \beta_1}\left(\frac{1}{n}\sum_{i=1}^{n}(y_i - (\beta_o + \beta_2 x_i))^2\right)^{-\frac{n}{2}} = 0$$

2. Consider the following (3×3) transition matrix

0.0	0.1	0.5
p_1	0.0	p_3
0.6	p_2	0

a) What values of p_1, p_2 and p_3 make this matrix a proper left stochastic transition matrix?

b) Use the Chapman-Kolmogorov equation to find the probability mass function of the system after 10 time steps using the initial conditions $\pi[0] = (0, .3, .7)$.

c) Find the stationary (ergodic) distribution using the Perron-Frobenius theorem.

d) Is a Markov process defined by this transition matrix irreducible, that is, do all states communicate with each other?

e) Is a Markov process defined by this transition matrix aperiodic?

a) Proper left stochastic transition matrix if each columns sums to one:

$$p_1 + 0.6 = 1 \implies p_1 = 0.4$$

$$0.5 + p_3 = 1 \implies p_3 = 0.5$$

$$p_2 + 0.1 = 1 \implies p_2 = 0.9$$

b) Probability mass function of system after 10 time steps, initial condition: $\pi[0] = \begin{pmatrix} 0.0 \\ 0.3 \\ 0.7 \end{pmatrix}$

Chapman-Kolmogorov equation: $\pi[t] = P^t \pi[0]$

$$\pi[t] = \begin{pmatrix} 0.0 & 0.1 & 0.5 \\ 0.4 & 0.0 & 0.5 \\ 0.6 & 0.9 & 0 \end{pmatrix}^{10} \begin{pmatrix} 0.0 \\ 0.3 \\ 0.7 \end{pmatrix} = \begin{pmatrix} 0.25 & 0.26 & 0.24 \\ 0.32 & 0.32 & 0.31 \\ 0.43 & 0.41 & 0.45 \end{pmatrix} \begin{pmatrix} 0.0 \\ 0.3 \\ 0.7 \end{pmatrix}$$

$$\pi[10] = \begin{pmatrix} 0.25 \\ 0.31 \\ 0.44 \end{pmatrix}$$

c) Perron-Frobenius Theorem: if P is irreducible, aperiodic stochastic matrix, then P is non-negative and largest eigenvalue of P is real and corresponds to a non-negative eigenvector.

Matrix is stochastic $\Longrightarrow$ $largest\ eigenvalue,\ \lambda = 1$, find corresponding eigenvector:

$$P - \lambda I = \begin{pmatrix} -1 & 0.1 & 0.5 \\ 0.4 & -1 & 0.5 \\ 0.6 & 0.9 & -1 \end{pmatrix}$$

$$\begin{pmatrix} -1 & 0.1 & 0.5 \\ 0.4 & -1 & 0.5 \\ 0.6 & 0.9 & -1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} .57 \\ .73 \\ 1 \end{pmatrix}$$

$$\pi^* = \begin{pmatrix} .57 \\ .73 \\ 1 \end{pmatrix}$$

d) Yes, the Markov process is irreducible as all states in the transition matrix are able communicate with eachother- it is possible to get from any state in the transition matrix to another.

e) Yes, the the Markov process is aperiodic as the transition matrix is stochastic and thus the possible path lengths of a transition has a common factor of one.

2 R EXERCISES

1. Use the web to find the NBER business cycle reference dates that classify quarters as recession or recovery periods. Count how many months are recession months and how many months are recovery months, and use this information to set up a Markov chain model of the business cycle. That is, use this frequency data to develop a (2×2) empirical Markov transition kernel of the form:

	Expansion	Contraction
Expansion	p_{11}	p_{12}
Contraction	p_{21}	p_{22}

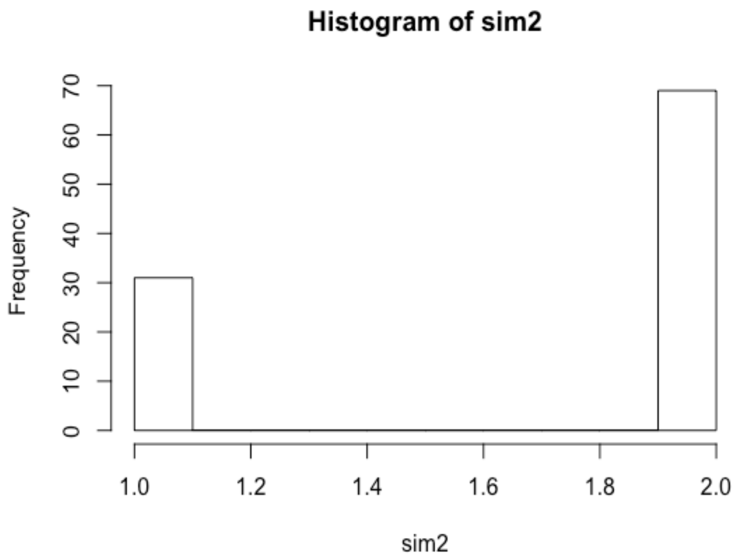
Simulate this model, and compare the results to the NBER data (Hint: there are 32 transitions from expansion to contraction and you will need to normalize the month counts by the total number of transitions). Use the Markov Chain simulation function

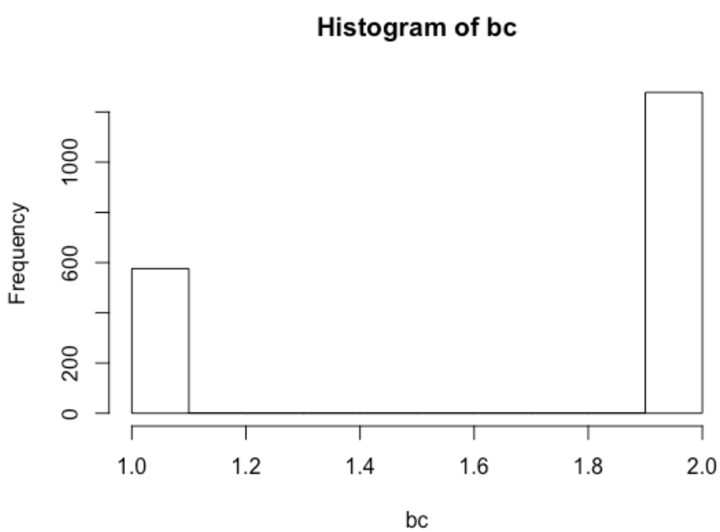
...

to compare the histogram of simulated stochastic process to the histogram of actual expansions and contractions. Does the empirical Markov transition model seem to produce the same type of data as the economy? Note, once you create the variables "expansion" and "contraction" you can transform them into states with:

```
data<-c(rbind(expansion,contraction))
dd<-data.frame("length"=data,"state"=rep(c(1,2),33))
bc<-rep(dd$state, dd$length)
```

Yes:


Histogram of sim2
Frequency
0
10
20
30
40
50
60
70
1.0
1.2
1.4
1.6
1.8
2.0
sim2


Histogram of bc
Frequency
0
200
600
1000
1.0
1.2
1.4
1.6
1.8
2.0
bc

2. Download the Extended Penn World Tables from blackboard.

   a) Subset the data to include (1) only observations between 1983 and 2007 (2) only observations of quality "A" or "B", (3) only the variables "Country", "Id", "Year", "x", and "c", and (4) drop all missing values using the compete.cases function.

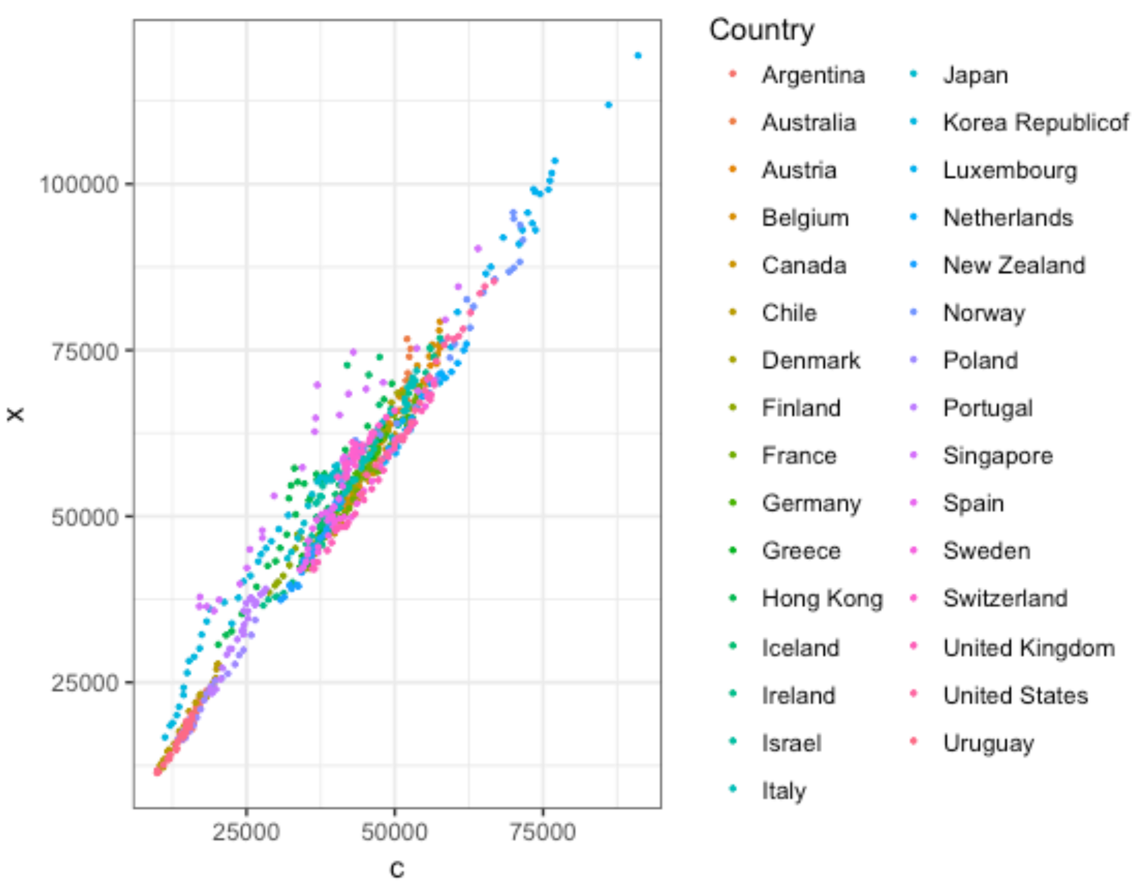
   b) Plot a scatter plot of consumption per capital (c) against income per capita (x).

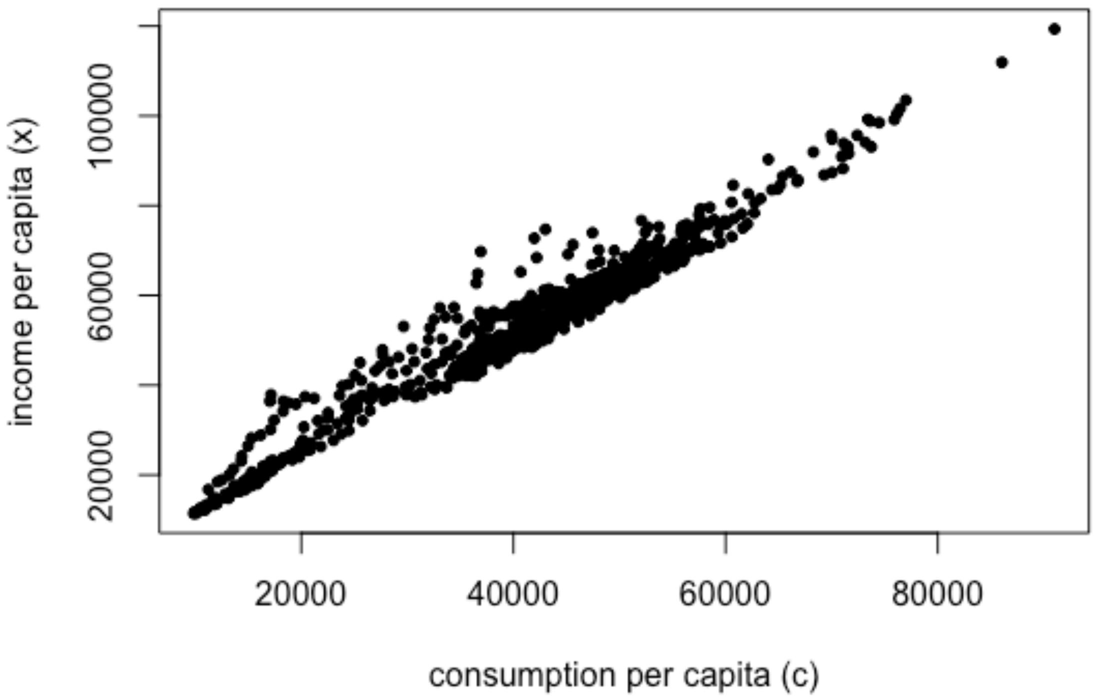
   c) Write a model in Stan or JAGS to estimate the consumption function $c = c_0 + c_1 x$.

   d) Diagnose you model with traceplots of the MCMC runs, simulated posterior densities of the parameters, autocorrelation, partial correlation, Geweke statistic, and Rhat.

   e) Use the mean of the posterior for the marginal propensity to consume (c_1) to calculate the multiplier $\frac{1}{1-c_1}$.

a) In r code
b) (experimenting w gg plot)



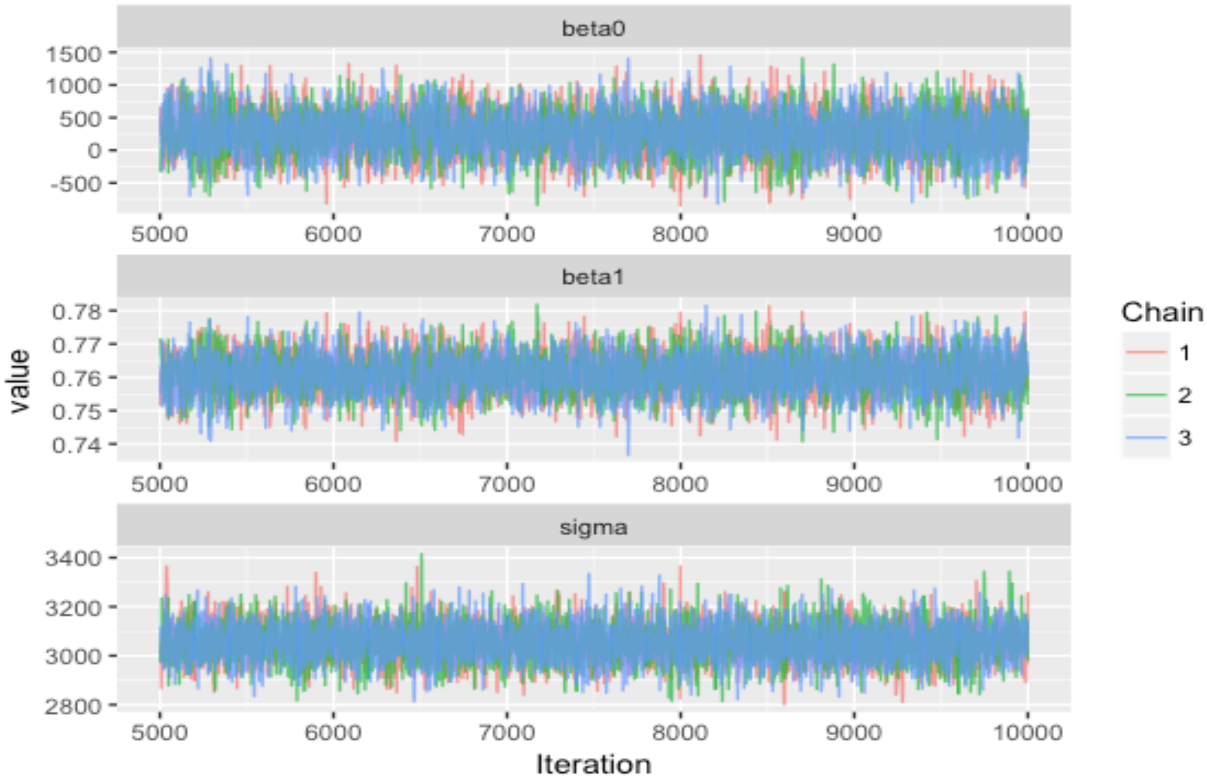


c) in R code

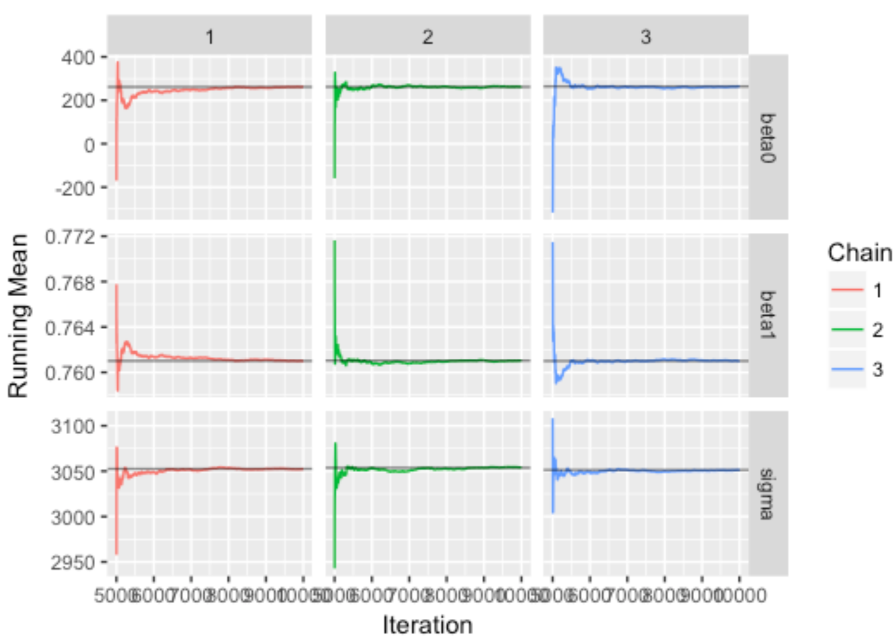
```
3 chains, each with iter=10000; warmup=5000; thin=2;
post-warmup draws per chain=2500, total post-warmup draws=7500.

         mean se_mean     sd     2.5%      25%      50%      75%    97.5% n_eff Rhat
beta0  260.56    5.08 342.33  -409.05    28.19   258.42   492.96   940.60  4534    1
beta1    0.76    0.00   0.01     0.75     0.76     0.76     0.77     0.77  4474    1
sigma 3053.67    1.00  78.58  2904.11  3000.16  3051.93  3105.44  3211.56  6180    1
lp__ -6597.04    0.02   1.27 -6600.39 -6597.61 -6596.71 -6596.12 -6595.60  4058    1
```

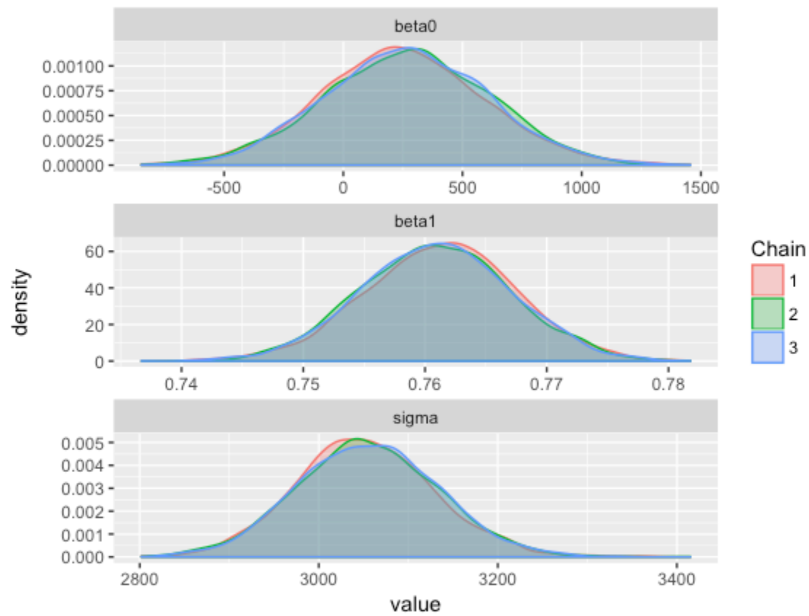
d) Traceplots:



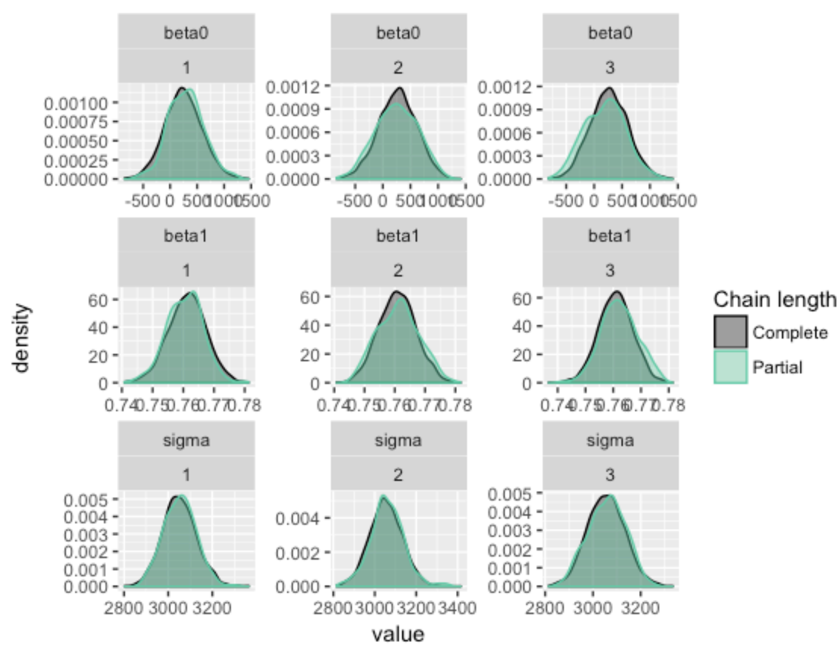
running mean(below) showing that chains are quickly reaching convergence



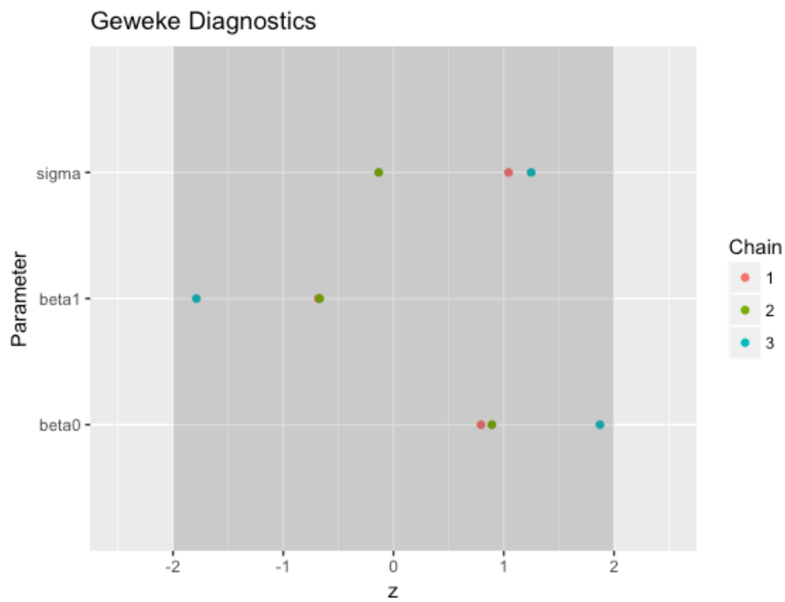
simulated posterior densities of parameters (below): showing that posterior distributions have converged in a similar space.



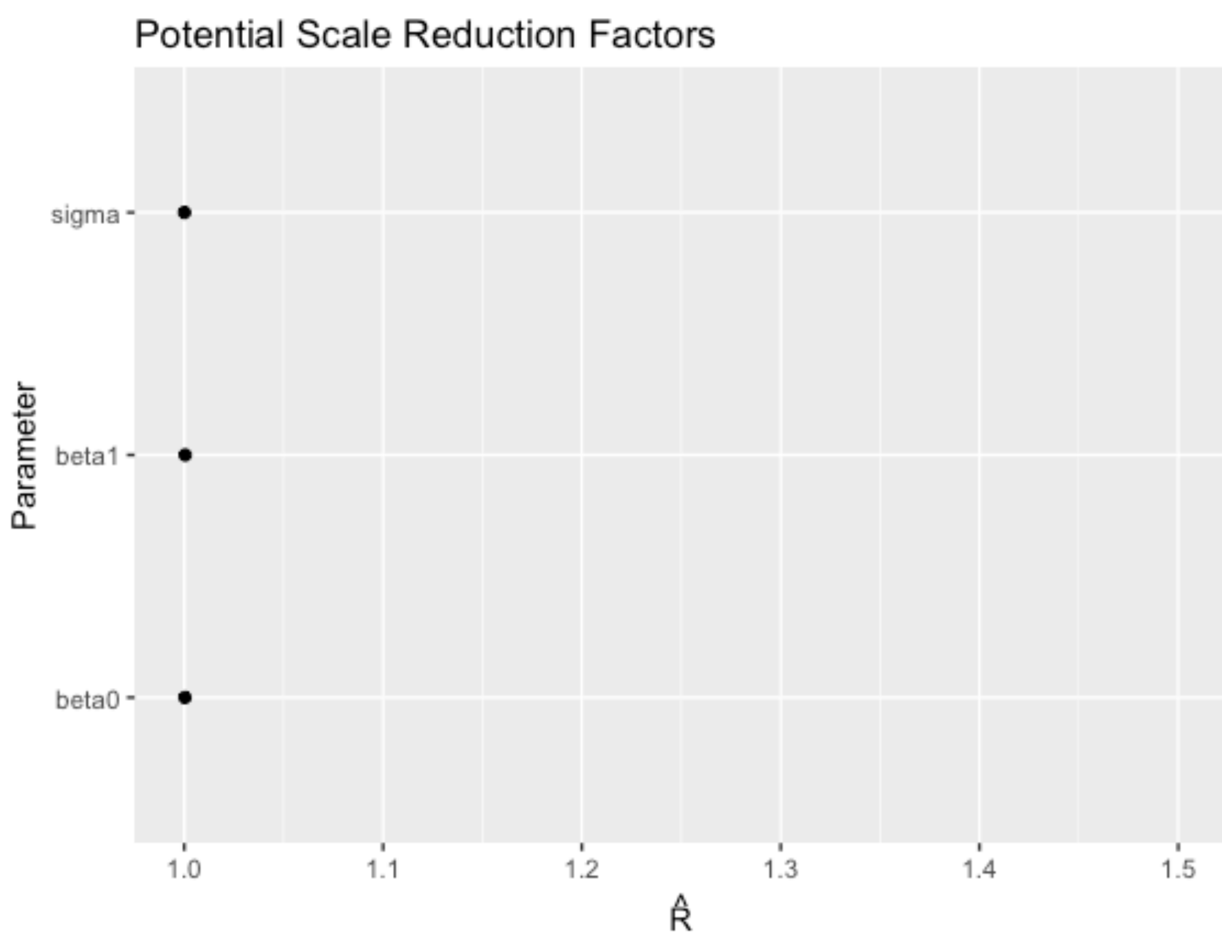
partial correlation: compare last part of the chain with the whole chain. Overlapping densities look similar.



geweke:all values contained within (-2,2).



rhat looks good, all at 1.



e)

```
3 chains, each with iter=10000; warmup=5000; thin=2;
post-warmup draws per chain=2500, total post-warmup draws=7500.

          mean se_mean     sd     2.5%      25%      50%      75%    97.5% n_eff Rhat
beta0   260.56    5.08 342.33  -409.05    28.19   258.42   492.96   940.60  4534    1
beta1     0.76    0.00   0.01     0.75     0.76     0.76     0.77     0.77  4474    1
sigma  3053.67    1.00  78.58  2904.11  3000.16  3051.93  3105.44  3211.56  6180    1
lp__  -6597.04    0.02   1.27 -6600.39 -6597.61 -6596.71 -6596.12 -6595.60  4058    1
```

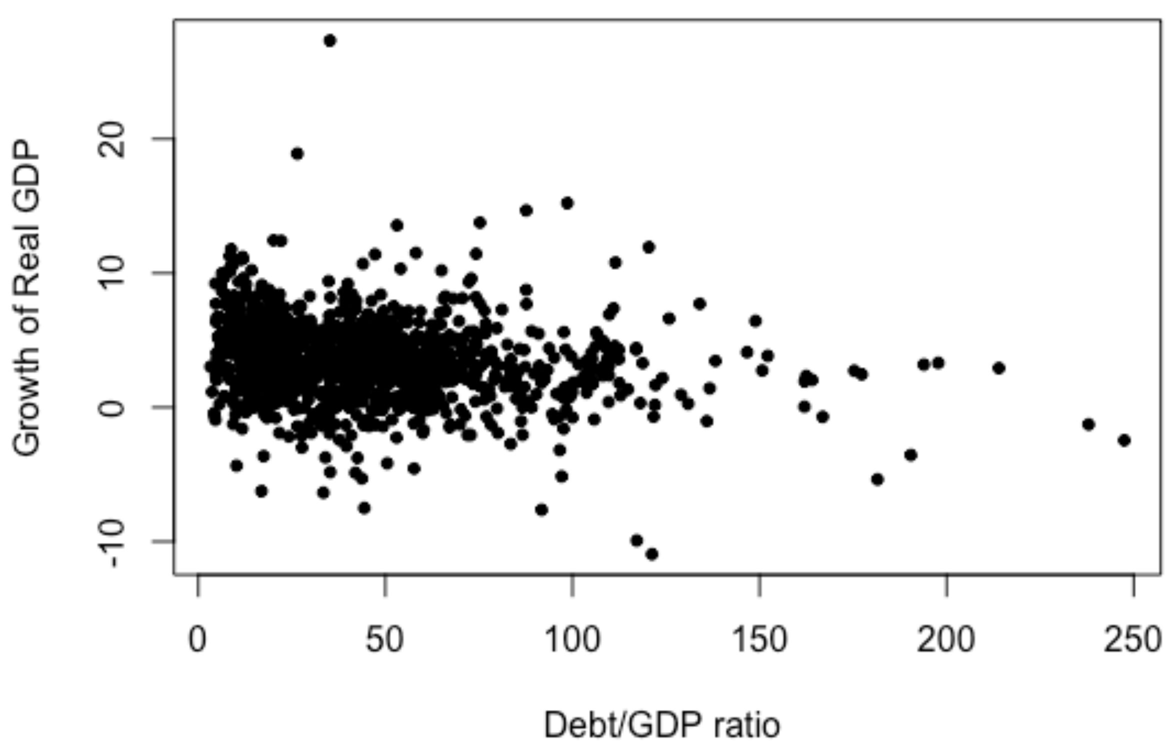
$$\beta_1 = 0.76$$

$$multiplier = \frac{1}{1 - 0.76} = 4.16$$

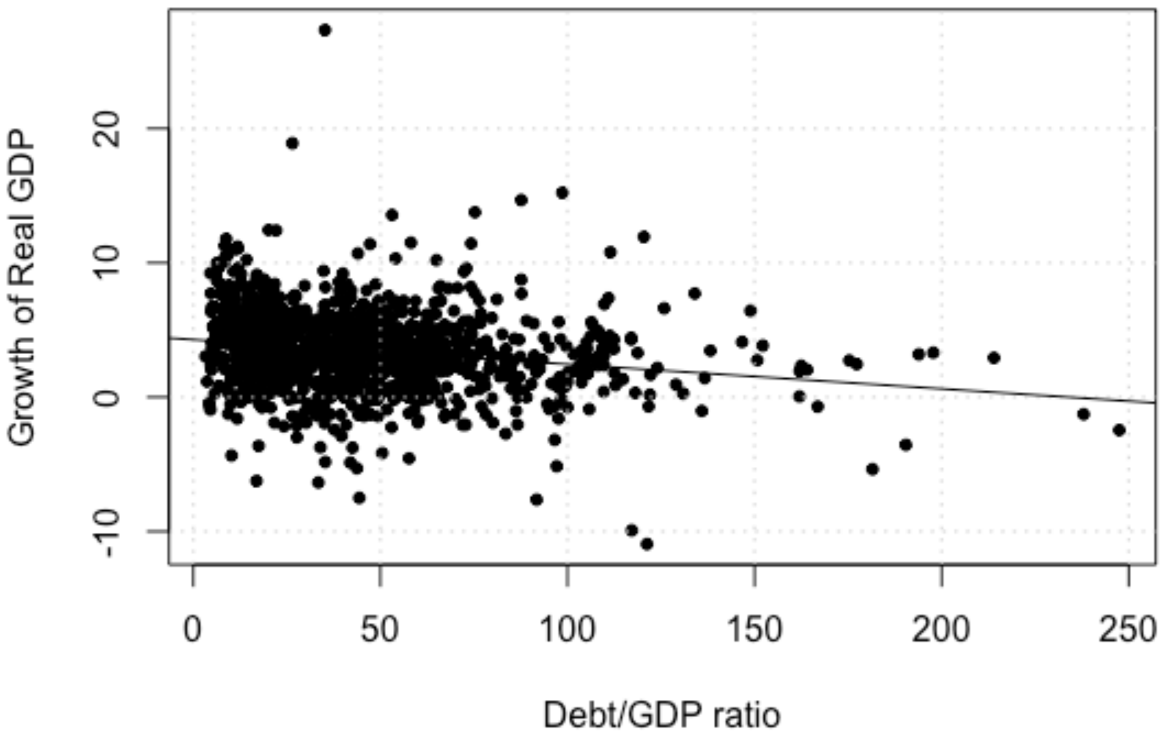
3. Download Reinhart and Rogoff's data from blackboard.

    a) Read in the data and plot the debt/GDP ratio ($debtgdp$) against the growth of real GDP ($dRGDP$). Is it evident that there is negative relation between debt/gdp ratio and real growth? Does it appear that there is a critical threshold at debt/gdp ratio around 90% at which real growth turns sharply down?
    b) Write a linear model in Stan or JAGS to estimate the elasticity of real GDP growth to the debt/GDP ratio.
    c) Plot the posteriors densities for the model parameters with HPD credibility intervals and report the posterior summary statistics.
    d) Diagnose your model with traceplots of the MCMC runs, simulated posterior densities of the parameters, autocorrelation, partial correlation, Geweke statistic, and Rhat.
    e) Write a program that simulates fitted regression lines from the posterior distributions of your model parameters and plot 100 simulated regression lines over the scatterplot of data.

a)



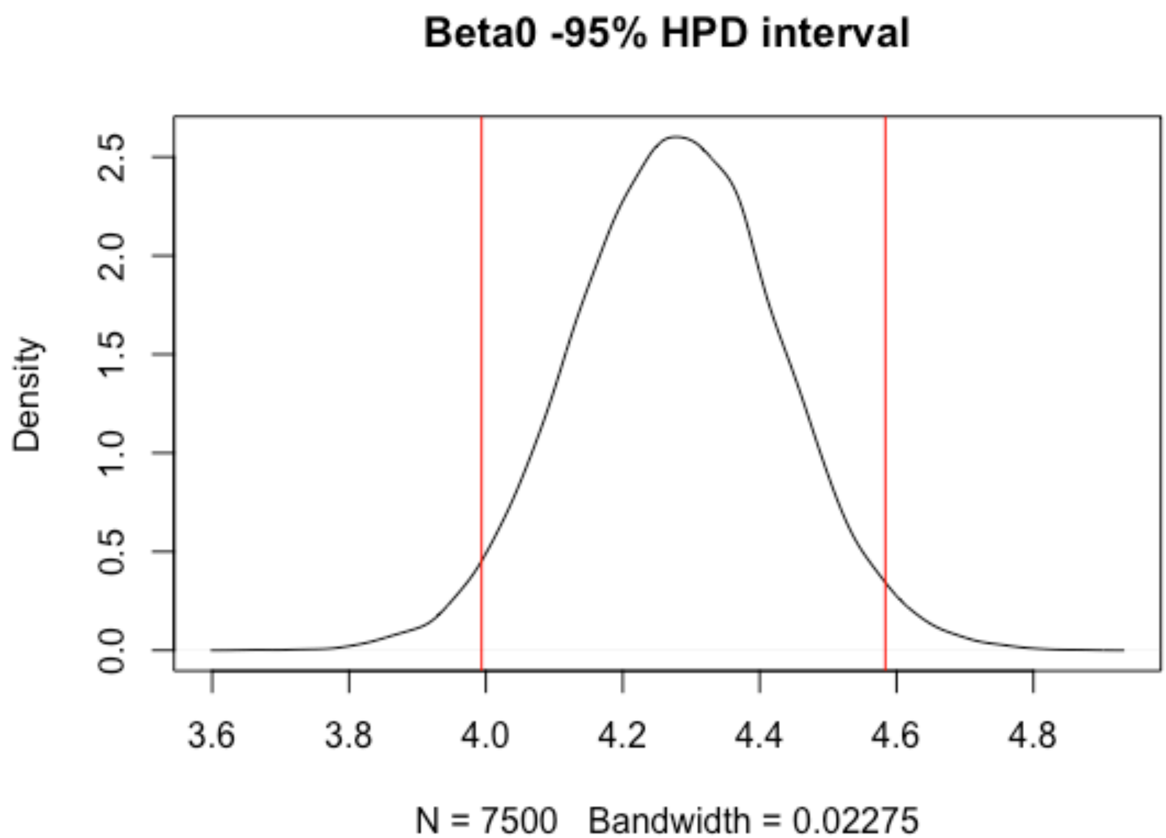
with trend line:

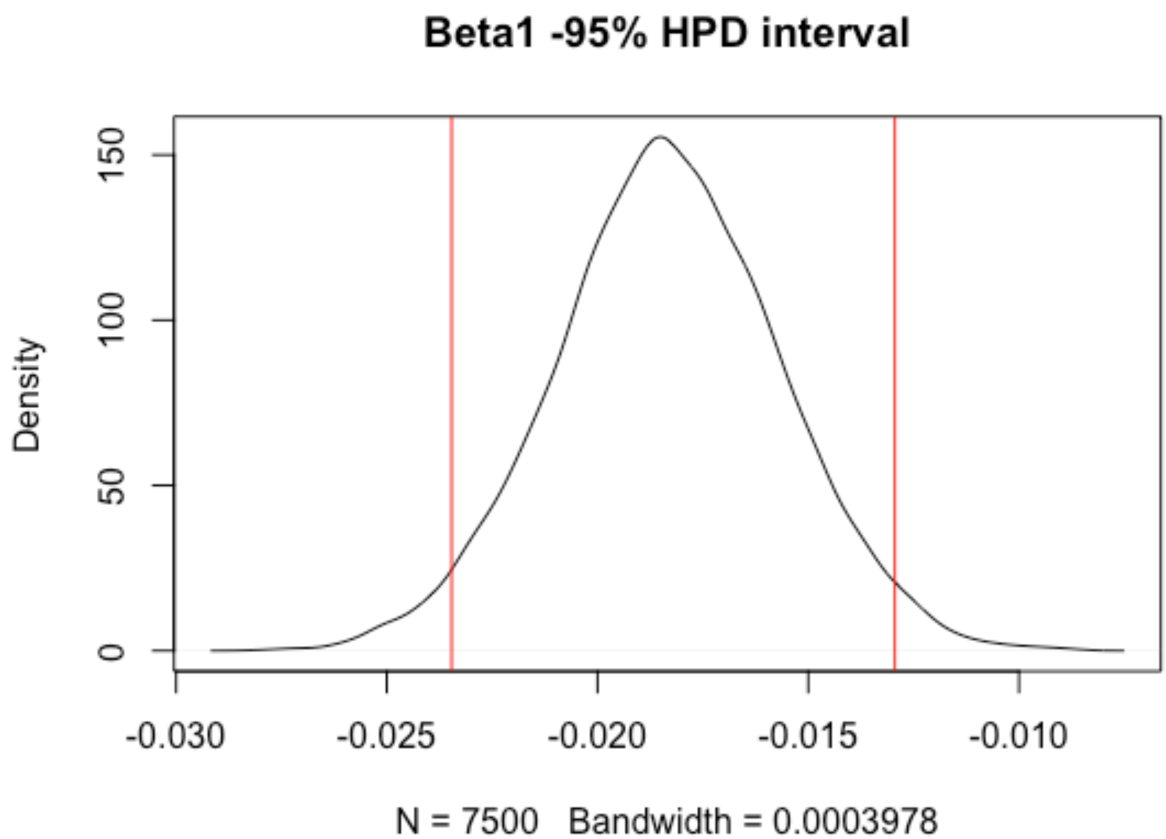


no? the only drop off at 95% seems to be a lack of data beyond that threshold- not enough evidence to support claim.

b) in r

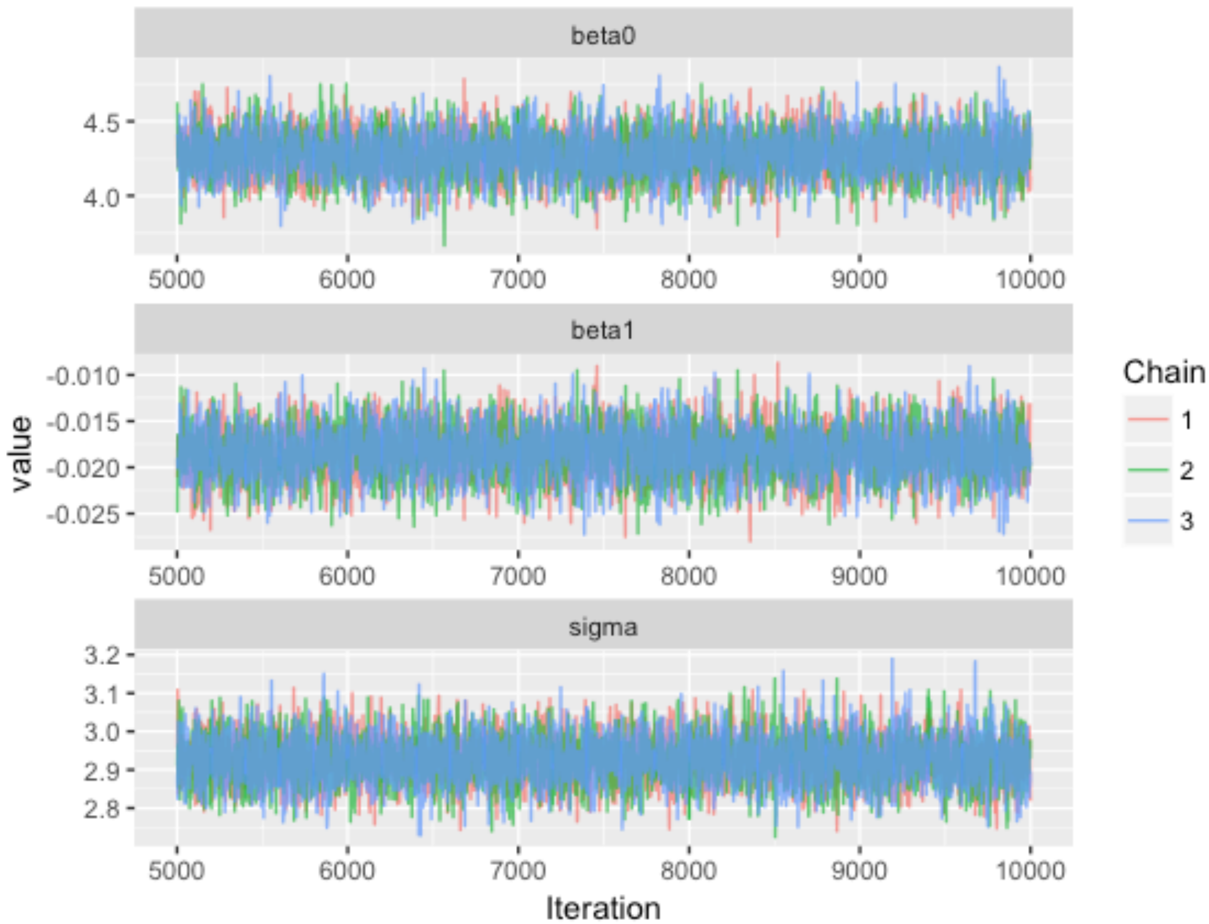
c)


Beta0 -95% HPD interval
Density
0.0
0.5
1.0
1.5
2.0
2.5
3.6
3.8
4.0
4.2
4.4
4.6
4.8
N = 7500 Bandwidth = 0.02275

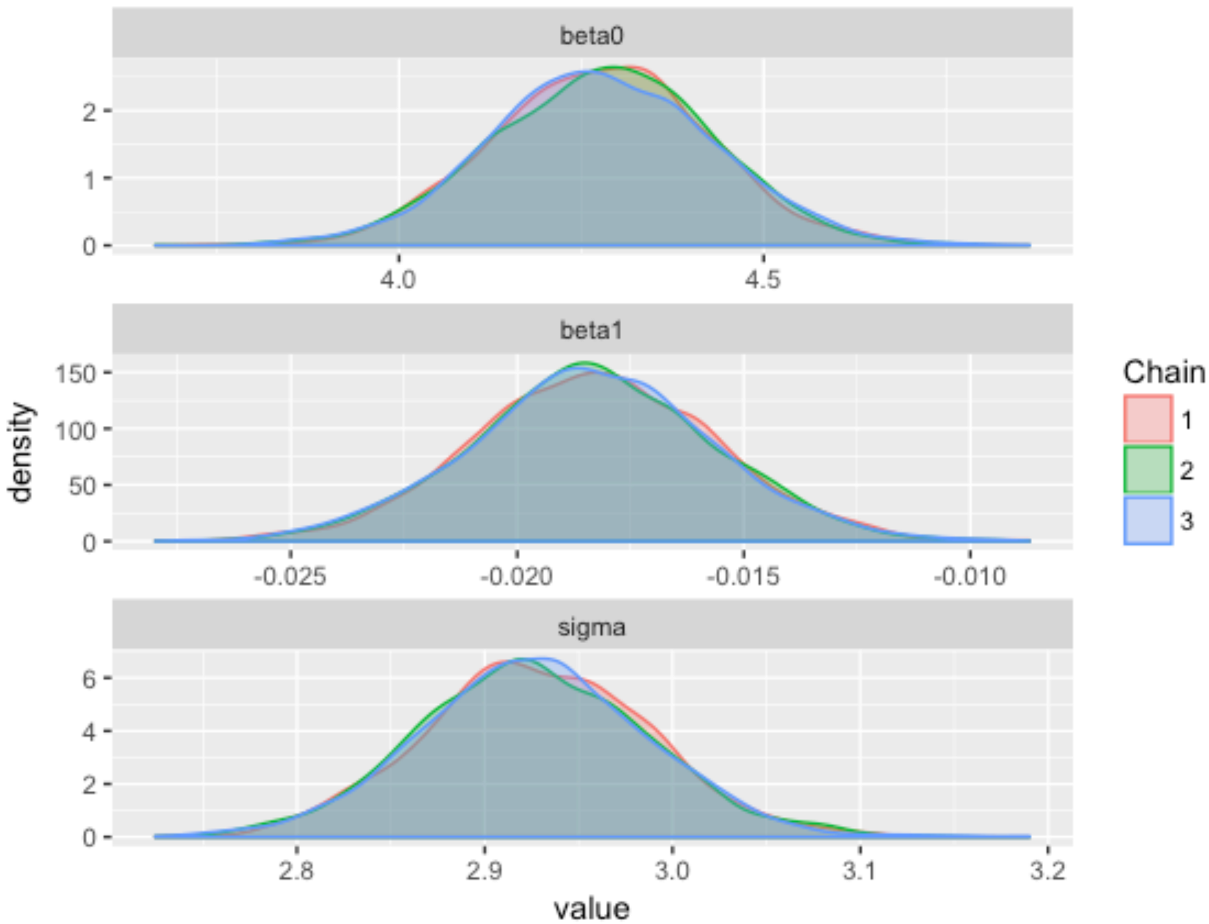


	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
beta0	4.28	0.00	0.15	3.98	4.18	4.28	4.38	4.58	5556	1
beta1	-0.02	0.00	0.00	-0.02	-0.02	-0.02	-0.02	-0.01	5868	1
sigma	2.93	0.00	0.06	2.81	2.89	2.92	2.97	3.05	6815	1
lp__	-1840.70	0.02	1.27	-1843.95	-1841.25	-1840.35	-1839.79	-1839.29	5022	1

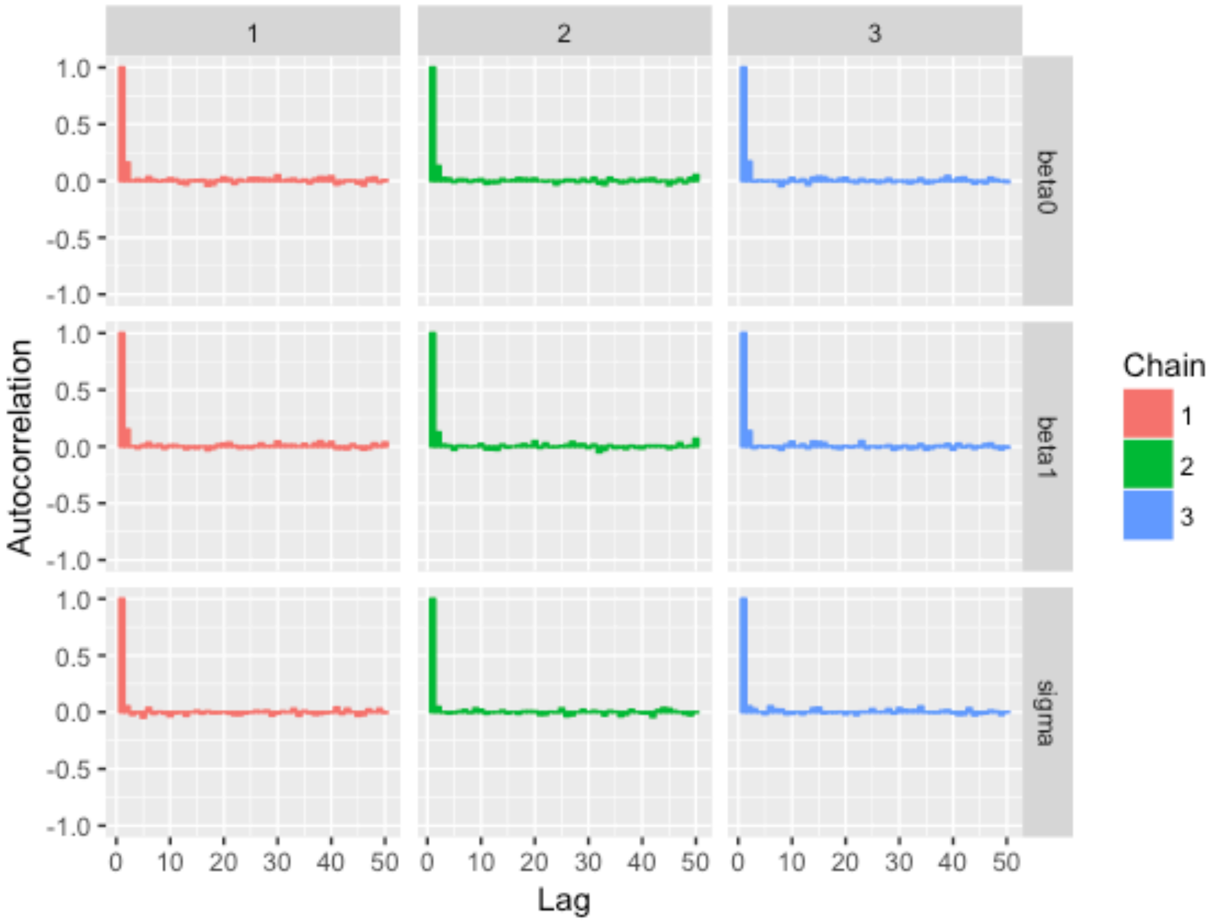
e) Traceplot:



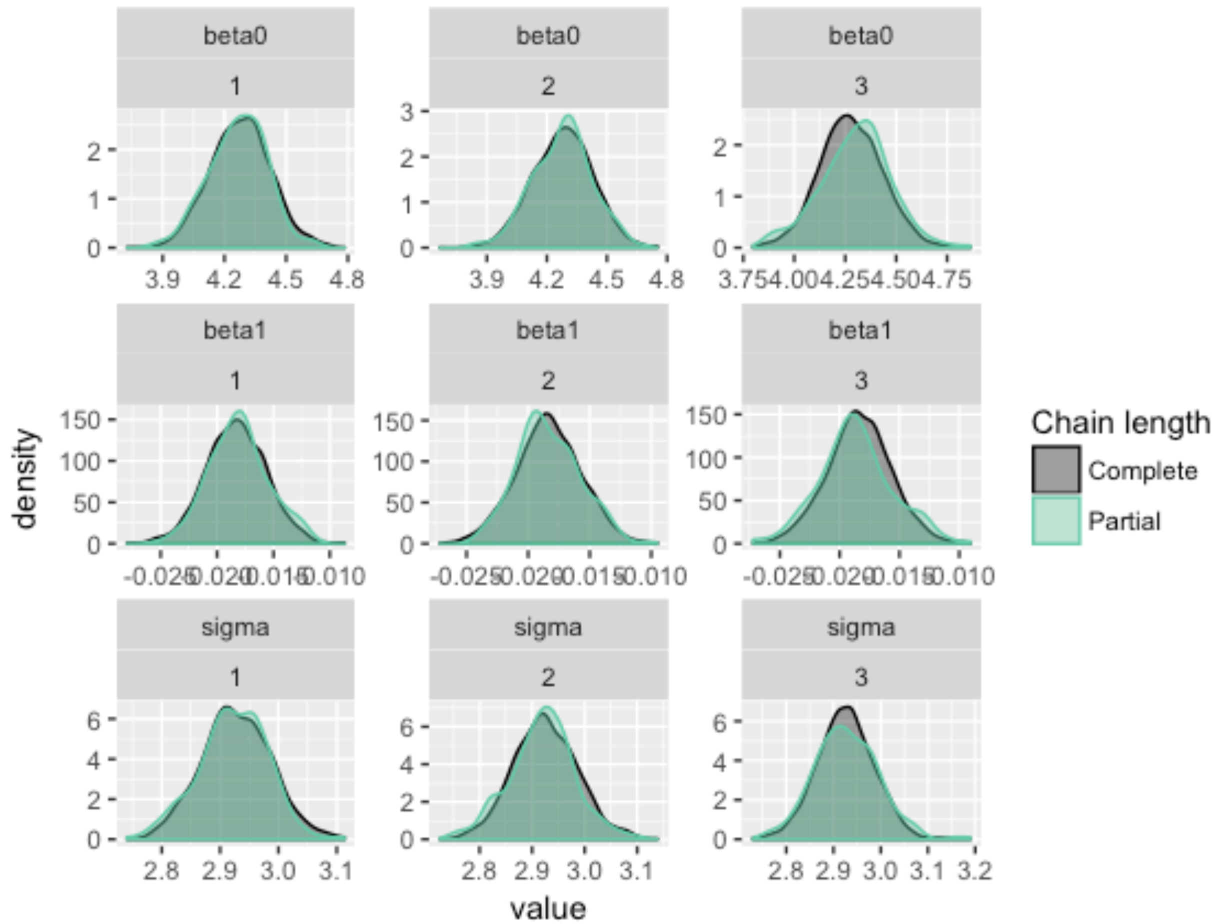
simulated posteriors:



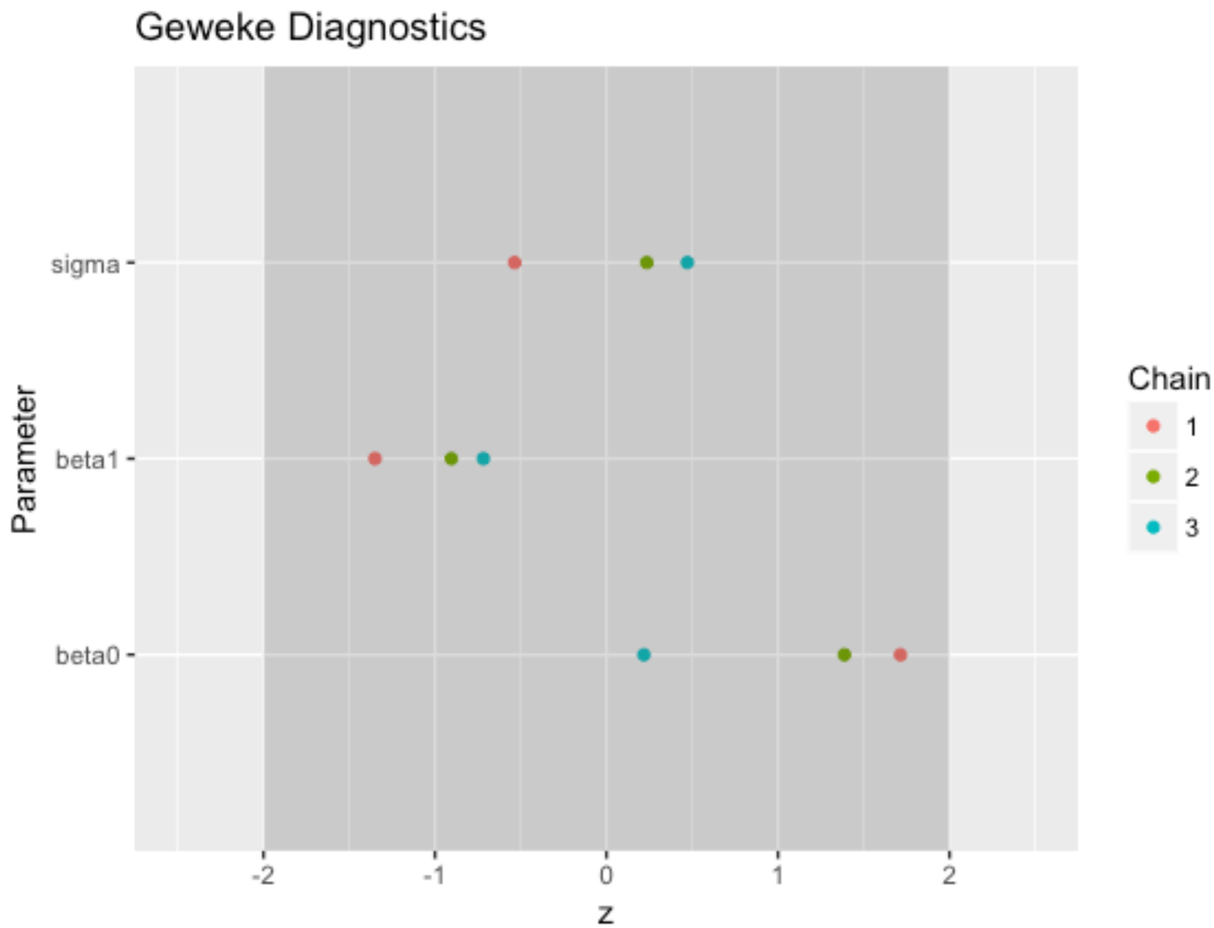
autocorrelation

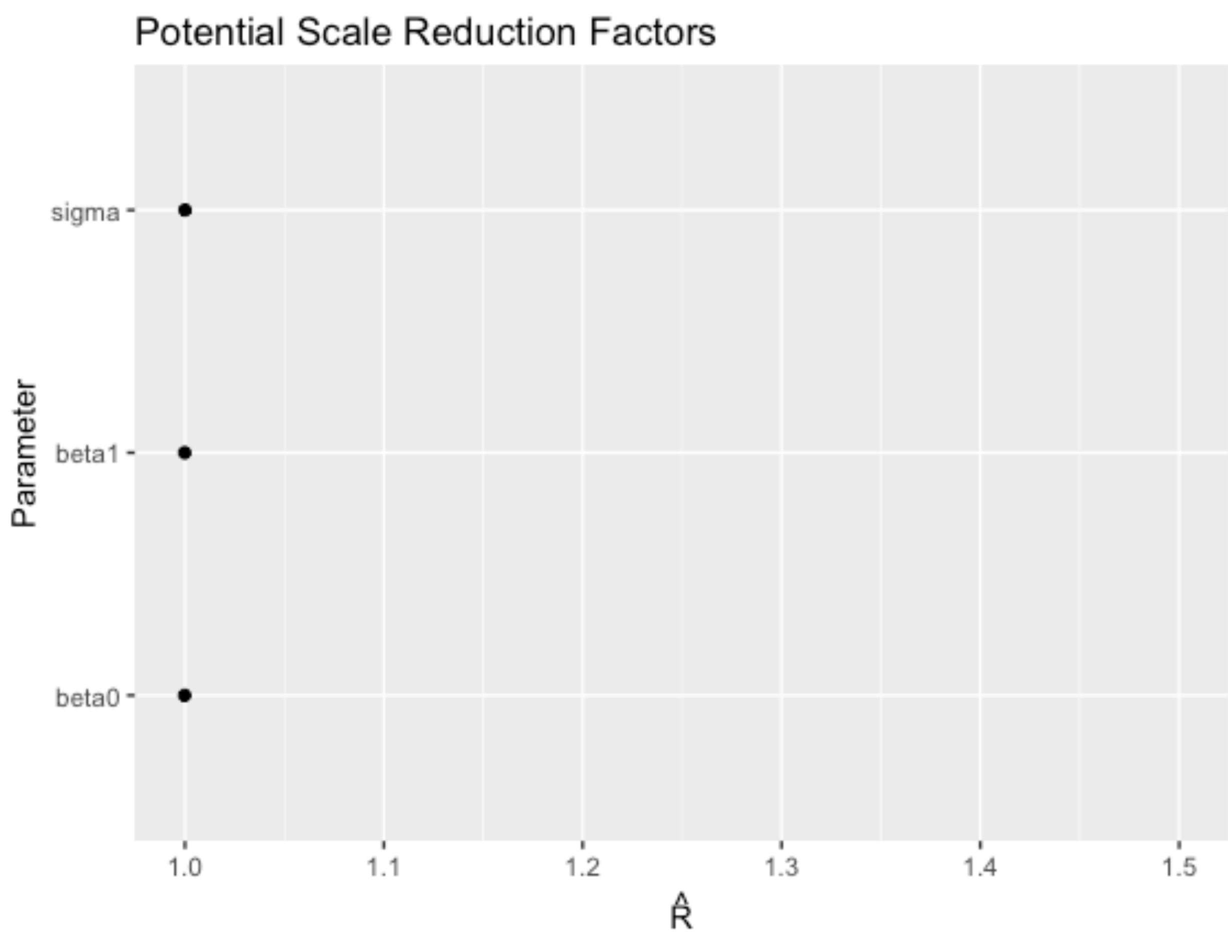


partial correlation



gweke statistic:


Geweke Diagnostics
sigma
beta1
beta0
Parameter
-2
-1
0
1
2
z
Chain
1
2
3


Potential Scale Reduction Factors
sigma
beta1
beta0
Parameter
1.0
1.1
1.2
1.3
1.4
1.5
R̂