STATS 205 Final Project

Yingxin Cao, Yuhshang Lai

3/18/2020

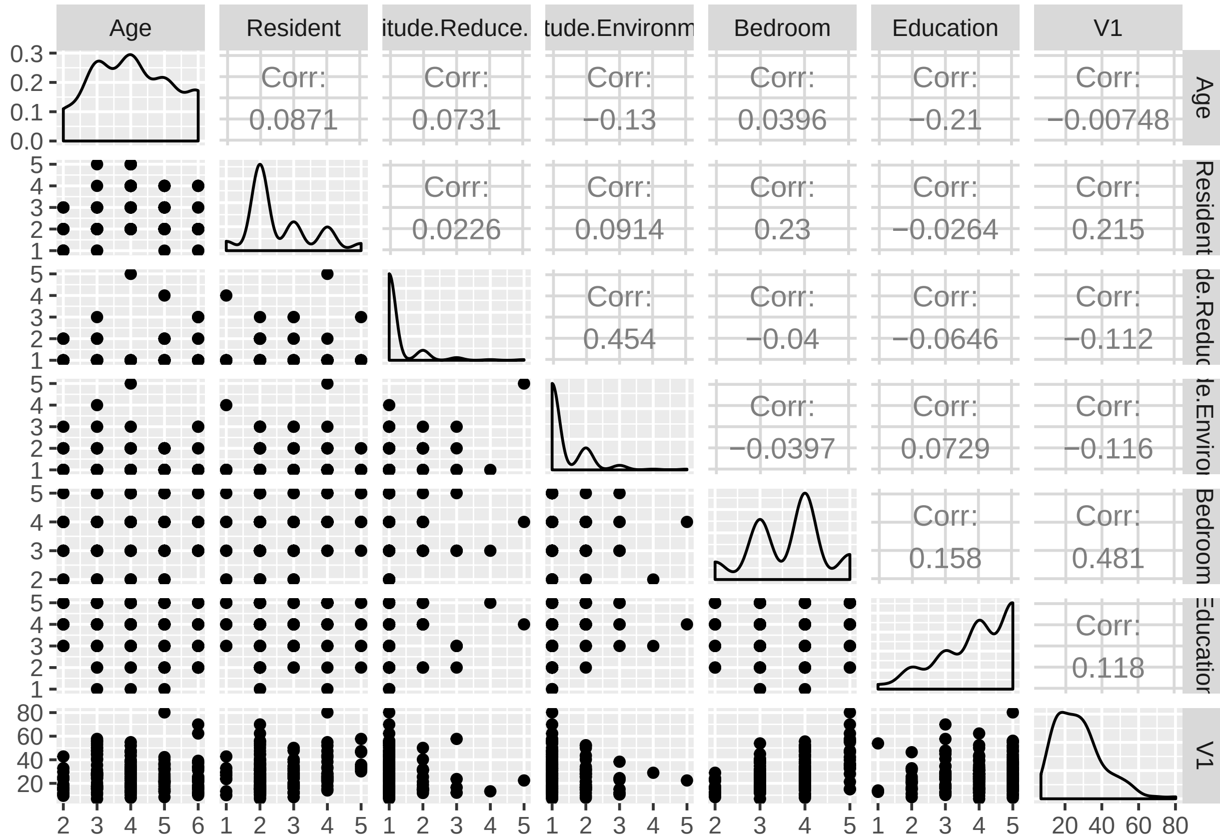
Problem 1

(1a)

We denote consumption of households on the first day as y. The parameters vector (including the intercept) we denote as β. The design matrix we denote as X with shape (n, r+1), where n represents number of samples and r represents number of covariates. Then we propose the following model for analysing the data,

$$y \sim N(\mu, \frac{1}{\tau})$$
$$\mu \sim X\beta$$
$$\beta|\tau \sim N(\beta_0, \frac{g}{\tau}(X'X)^{-1})$$
$$\tau \sim Ga(a, b)$$
$$\beta_0 = (X'X)^{-1}X'y$$
$$g = n$$

(1b)



The above figure tells us that Age of the head of household has little influence on the first day Electorcity usage; thus, we decide not to include it in our model. In contrast, Resident number and Bedroom number have have two largest correlation respect to the Electorcity usage. Consequently, we decide to include them in all of the potential model we will consider. For the features Attitude of Reduce bill, Attitude of Environment as well as Education which have moderate effect on the usage. We decide to try either with of without each of them. Thus, their are total 8 models.

(1c)

The details of model has shown in part (a). We use standard g prior with unit information prior $g = n$ since we don't have strong evidence on the corvaiance.

$$\beta|\tau \sim N(\beta_0, \frac{g}{\tau}(X'X)^{-1})$$

$$\tau \sim Ga(a, b)$$
$$\beta_0 = (X'X)^{-1}X'y$$

The diffuse prior we use for τ with $a = b = 0.001$.

Pros: g-prior is a conjugate prior. It's inofrmative. It's variance resembles the frequentists perspective of estimation. It's invariant to scale of regressors.

Cons: Though it's informative, it's not as good as BCJ priors. All the prior information are from design matrix X, which could be problematic in situations like incomplete data.

(1d)

From the paried scatter plots as well as the correlations, we consider Age as irrelavent to our study. Since Number of Residents and Number of Bedrooms show high correlation, we fix these two features, and explore all combinations of all other features, resulting in following 8 models:

Model 1: Resident + Bedroom

Model 2: Resident + Bedroom + Attitude.Reduce.Bill

Mddel 3: Resident + Bedroom + Attitude.Environment

Model 4: Resident + Bedroom + Education

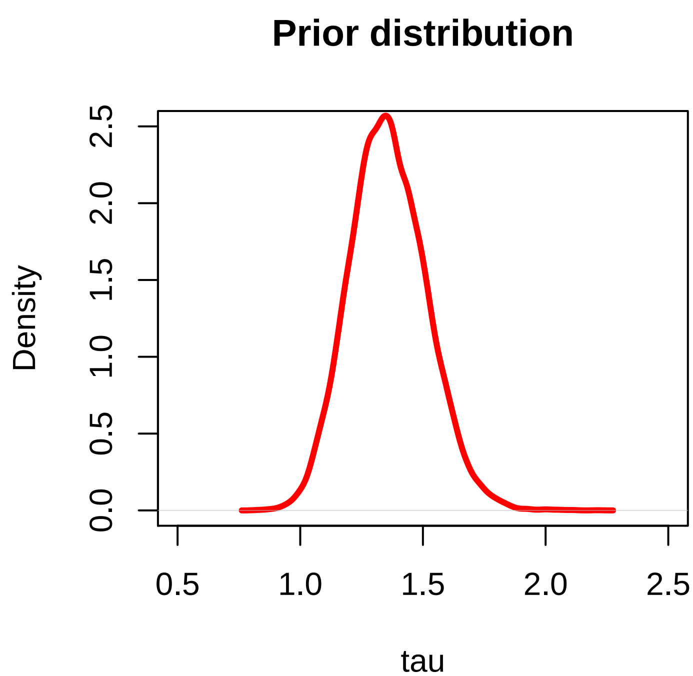
Model 5: Resident + Bedroom + Attitude.Reduce.Bill + Attitude.Environment

Model 6: Resident + Bedroom + Attitude.Reduce.Bill + + Education

Model 7: Resident + Bedroom + Attitude.Environment + Education

Model 8: Resident + Bedroom + Attitude.Reduce.Bill + Attitude.Environment + Education

(1e)



	DICs	BICs	LPMLs
model 1	393.517503835045	405.560261999977	-197.058761460155
model 2	393.481325788465	408.725172801034	-196.955230205741
model 3	393.092784477866	408.185166104452	-197.013778846707
model 4	394.847908791453	410.079947113839	-197.991404383612
model 5	394.667659282799	412.646430138086	-197.709089609129
model 6	395.147733816388	413.342621316378	-198.012493765005
model 7	394.186464830644	412.479061019568	-197.866880258642
model 8	396.036732914626	417.062688447826	-198.748203938467

Based on the information we get from data anaylsis in part (b), we decide to include Resident and Bedroom in all of the models. And for features Attitude.Reduce.Bill, Attitude.Environment and Education. We try either include or exclude each of them. And we decide to exclude Age since it has little influence on Electorcity Usage. Based on the DICs, BICs and LPMLs, we figure out that model 1,2 and 3 has similar (also better) BICs DICs and LPMLs which are the lowest three among all. Thus, we decide to choose to simpliest one Model 1 for our final choice model.

Model 1: Resident + Bedroom

Model 2: Resident + Bedroom + Attitude.Reduce.Bill

Mddel 3: Resident + Bedroom + Attitude.Environment

```
cat('Smallest DICs is at: Model',which(DICs==min(DICs)),'\n')
```

```
## Smallest DICs is at: Model 3
```

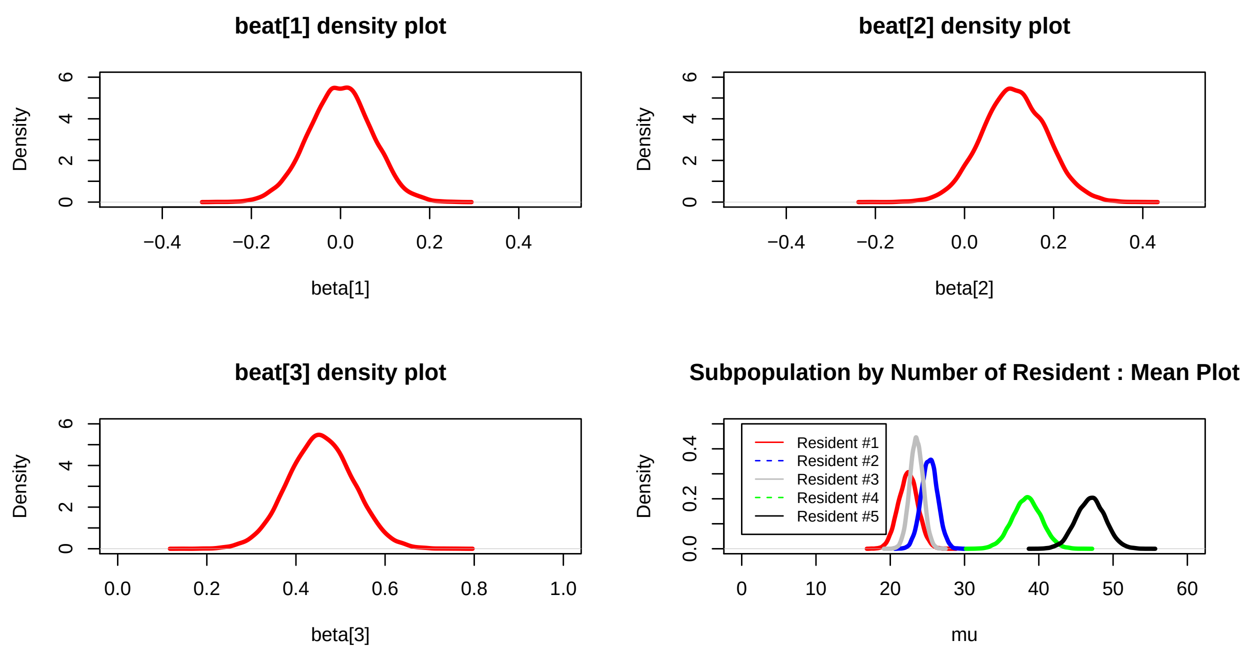
```
cat('Smallest BICs is at: Model',which(BICs==min(BICs)),'\n')
```

```
## Smallest BICs is at: Model 1
```

```
cat('Largest  LPMLs is at: Model',which(LPMLs==max(LPMLs)),'\n')
```

```
## Largest  LPMLs is at: Model 2
```

Model 1: Resident + Bedroom



The above figure is the plot of posterterior inference of for the regression parameter beta[1], beta[2], beta[3]. We also include the subpopulation of the households with different number of Resident. Five distributions are considerably reasonable, with the tendency that the more residents in the household, the more electorcity they will use. Households of three is an outlier since it may include samll child who does not use much electorcity.

	DICs	BICs	LPMLs
model 1	393.517503835045	405.560261999977	-197.058761460155
model 2	393.481325788465	408.725172801034	-196.955230205741

	DICs	BICs	LPMLs
model 3	393.092784477866	408.185166104452	-197.013778846707
model 4	394.847908791453	410.079947113839	-197.991404383612
model 5	394.667659282799	412.646430138086	-197.709089609129
model 6	395.147733816388	413.342621316378	-198.012493765005
model 7	394.186464830644	412.479061019568	-197.866880258642
model 8	396.036732914626	417.062688447826	-198.748203938467

(1f)

```
exp(LPMLs[4]-LPMLs[1])
```

```
## [1] 0.3935123
```

Based on the Week 9 Slides Bayes Factor page 61/105. We have Bayes Factor

$$BF_{41} = exp(LPML_4 - LPML_1) \approx 0.4$$

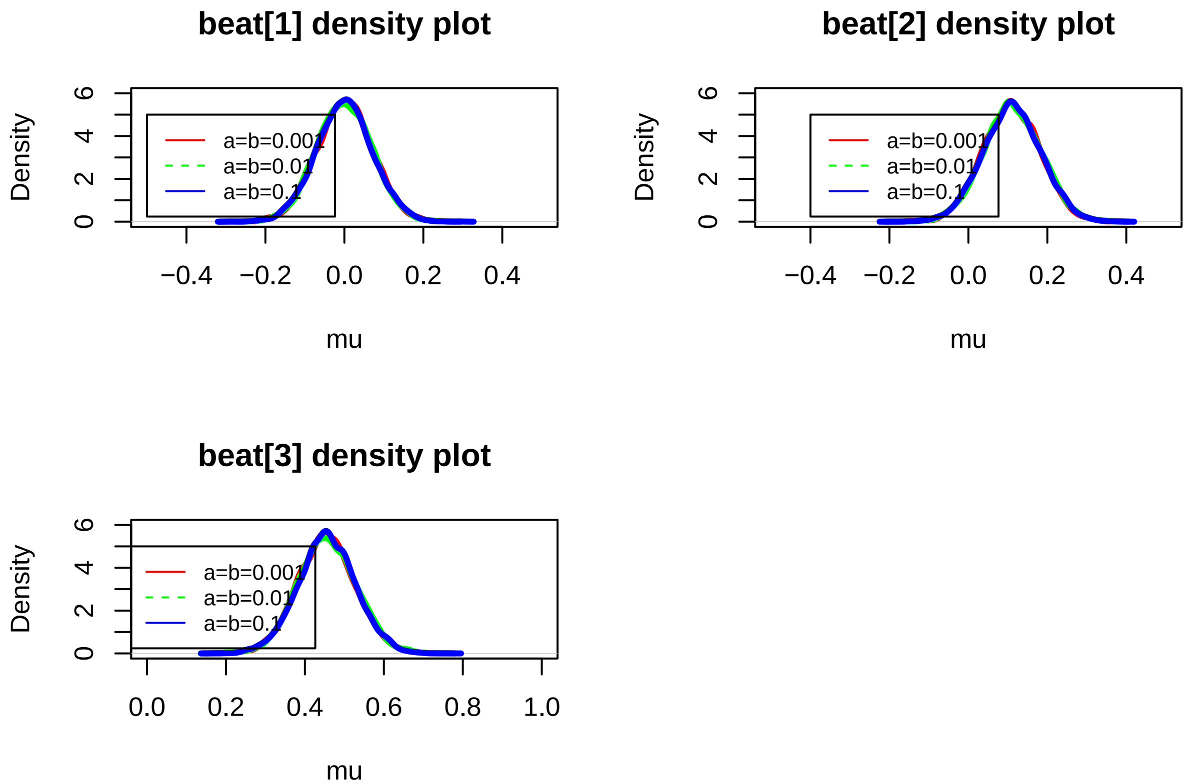
which < 1, then we conclude that level of Education is not stongly related to the energy consumption.

#(2) ## (2a)

Model	BICs	DICs	LPMLs
model 1	393.517503835045	405.560261999977	-197.058761460155
model 9	428.810170762363	447.084242348624	-215.311905960128

Compare the model 1 and the new interaction model propose, we find that model 1 both have lower BICs and DIcs while harboring lager LPML. Thus, we prefer model 1 to the new model.

##(2b)

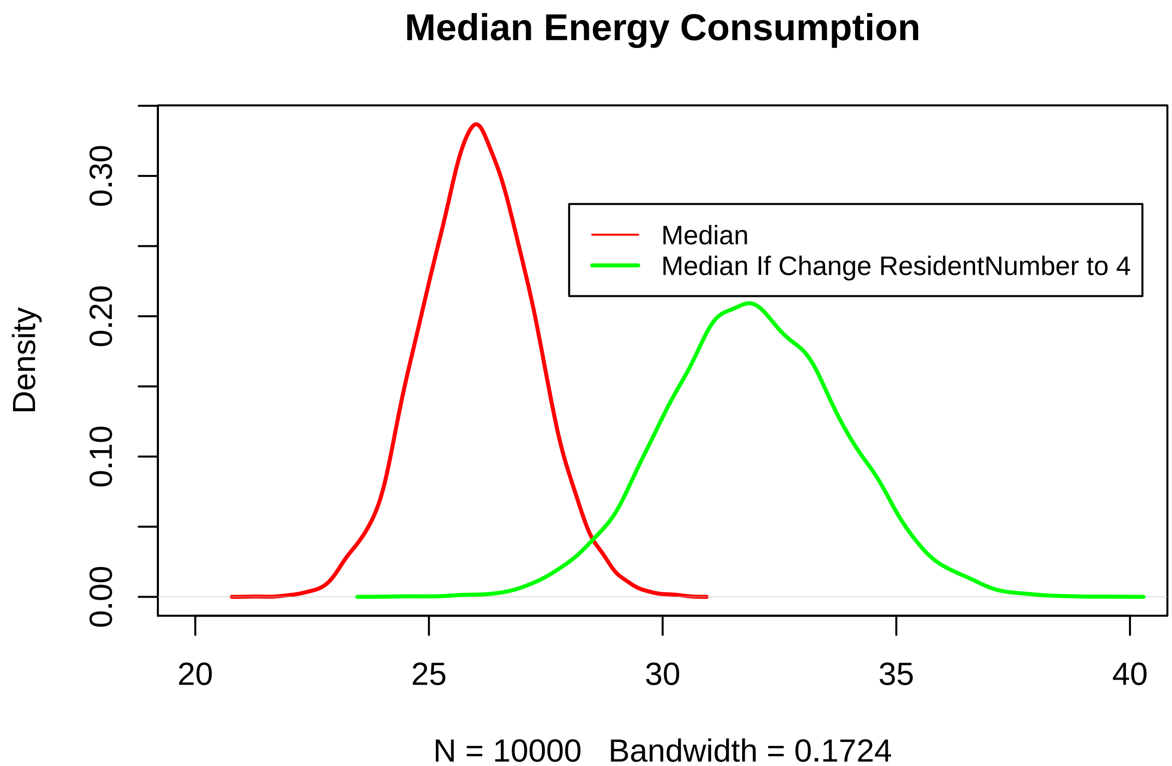


From the above figure, we can see that when we change prior tau from (a=b=0.001) to (a=b=0.01) and then to (a=b=0.1). The posterior distribution of regression parameters don't change much. Thus the best model we choose is not sensitive on priors.

(2c)

```
## [1] 26.04027
```

```
## [1] 31.90898
```




```
jags.fit.1$BUGSoutput$summary[7:8,]
```

```
##           mean       sd      2.5%      25%      50%      75%     97.5%
## meaV  26.03974 1.208862 23.63064 25.22382 26.04027 26.85373 28.39760
## meaV4 31.94533 1.903529 28.25109 30.66763 31.90898 33.22529 35.69397
```

From the above plot and table, we find that tange of meadian level households we will predict use from (23.6,28.4) with 95% credible interval with mean 26. If change Resident number to 4, we will predict (28.2,35.6) as 95 credible intervel and mean 31.9.

(3)

(a)

Here, differ than problems above, we consider y for all days (y's for different t). So here, our y is a (N, T) matrix, with rows indicates housholds (N in total) and columns indicates days (T in total).

We propose to model the consumption with the following hierarchical model. We model effect at individual level as $\beta_{i,t}$. All individual effects come from same daily population effect prior with mean β_t. And all daily population effect come from a common prior with mean β_0. We model the coefficients with standard g-prior. Details are shown below.

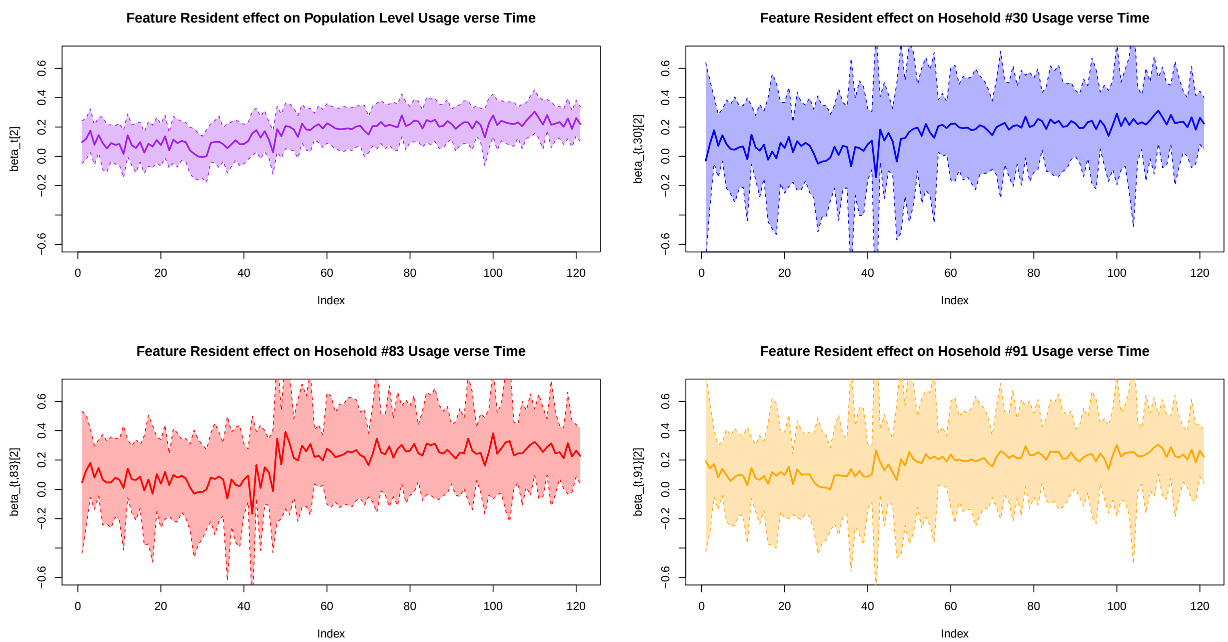
This design of model allow us to model effect of different variables at individual levels, instead of all individuals have same effect at time t.

$$y_{i|t} \sim N(\mu_{i,t} \frac{1}{\tau_{0_t}})$$

$$\mu_{i|t} \sim X_i \beta_{i,t}$$

$$\beta_{i,t} \sim N(\beta_t, \tau_{i,t})$$
$$\beta_t \sim N(\beta_0, \frac{g}{\tau_t}(X'X)^{-1})$$
$$\tau_t \sim Gamma(a, b)$$
$$\tau_{0_t} \sim Gamma(a, b)$$
$$\tau_{i,t} \sim Gamma(a, b)$$
$$\beta_0 \sim N((X'X)^{-1}X'\bar{y}, \frac{1}{\tau_0} \cdot I)$$
$$\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y[:, t]$$
$$\tau_0 \sim Ga(a_0, b_0)$$
$$g = n$$

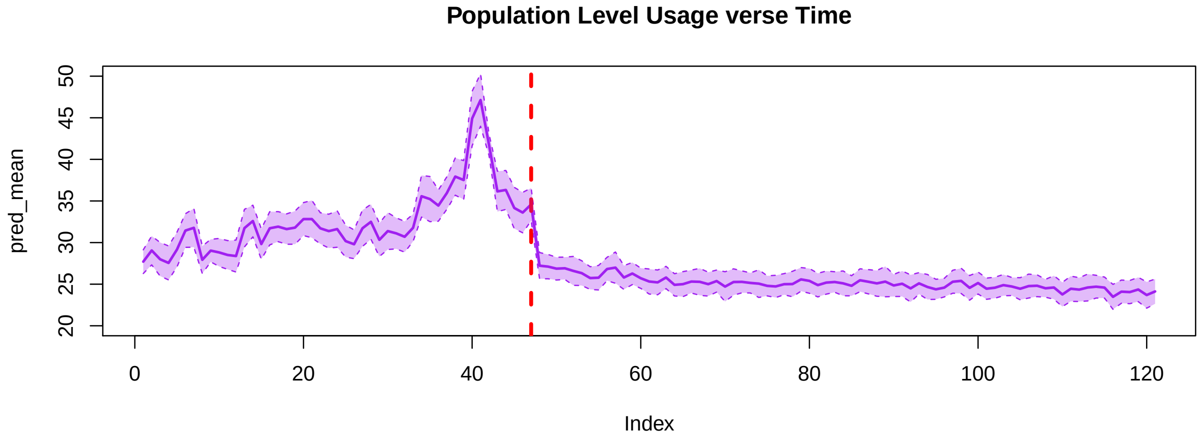
(b)



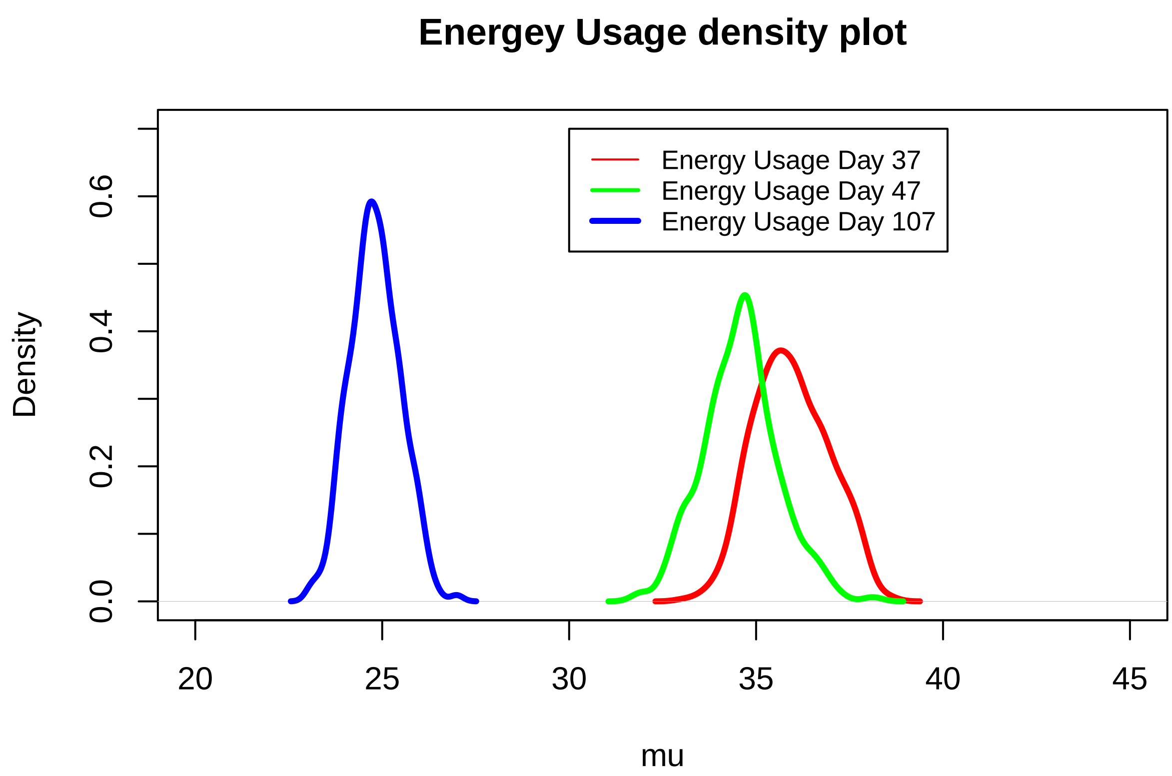
We plot regression cofficent β as the effect of number residents, since it reflects how much the comssumption increses with 1 unit increase of number of residents.

We plot the posterior mean as the solid line, the 95% credible interval as dotted line. The plots are information for posterior distributions of effect of number of residents at population lavel $\beta_t[2]$ as well as indifudual levels $\beta_{i,t}[2]$. Idividuals display similar patterns as poplulation level, slightly increasing with time. #83 with 4 residnets in house have higher effect with one unit increase of resident. Effect for #83 also increses the most along time.

(4)



From the above figure we can see that the usage of the electorcity in population level experince a sharp increasing from 30 to 45 and a sharp down from 45 to 32 during the two weeks before traiff structure changed. At the day the change executed, there is another sharp decreasing on the electorcity usage from aound 33 to 25. After the change, we see that population use less and less energy during the next 2.5 months which is not returning to the pattern but ues much less than the pattern before the policy change. From the figure below, we plot the density of the Energy use at day 37 which is 10 days before the policy, day 47 which is the day the policy executed and day 107 which is 2 month later from policy, we can see that there is a tendency people use less electorcity after the policy. By closley examining the posterior mean, we can conclude that the usage doesn't return to previous level.



(Problem 2)

5.

(a)

Here we use two methods to induce sparsity: the first is Bayesian Lasso; the second is Two-group mixture modeling with Spike-and-slab priors. Bayesian Lasso assign regression coefficients a double exponential prior that centered at 0. Spike-and-slab perform the variable selection with a Bernoulli latent auxiliary variable, to assign if the corresponding coefficient comes from zero.

Spike-and-slab provides this latent variable, which provides direct information on our variable selection problem. For Bayesian Lasso, we need extra testing procedures. Bayesian lasso does not put any point-mass on zero for the prior. Lasso shrinks all parameters with the Laplace prior.

Compared with frequentists' methods (i.e. traditional lasso), Bayesian methods require MCMC which is computationally more expensive. Both methods would preduce similar results, however, Bayesian method could generate posterior distributions for all parameters, which is more useful than point estimates. ### (b)

The selected variables for Bayesian Lasso are shown below. We set the threshold of probability being positive or negative at 0.54. In total, we selected 10 variables. It seems selecting variables with probability of being negative or positive is not ideal, since beta[76] has mean -0.0001 but with probability being positive 0.55.

```
var.select = (jags.fit.s$BUGSoutput$summary[1001:2000, 1] > 0.54) + (jags.fit.s$BUGSoutput$summary[1001
sum(var.select)
```

```
## [1] 10
```

```
jags.fit.s$BUGSoutput$summary[1:1000,][var.select>0, ]
```

```
##                     mean        sd       2.5%         25%           50%
## beta[47]    0.0231253095 0.3999496 -0.2083893 -0.01290803  0.002249475
## beta[76]   -0.0001070976 0.3561241 -0.2000443 -0.01488278  0.001750169
## beta[177]   0.0117851760 0.5512119 -0.1785933 -0.01780062 -0.001541818
## beta[275]  -0.0268908468 0.4327662 -0.2530532 -0.01879861 -0.001989726
## beta[304]  -0.0222961082 0.4034116 -0.2295827 -0.01796981 -0.001449790
## beta[410]   0.0111555582 0.3862873 -0.2089281 -0.01718777 -0.001955630
## beta[693]   0.0378286536 0.5709526 -0.1850640 -0.01420834  0.001390754
## beta[729]  -0.0151598928 0.3628338 -0.2129968 -0.02024956 -0.001454013
## beta[745]  -0.0106874288 0.2400337 -0.2600873 -0.01947399 -0.001517709
## beta[996]   0.0220663378 0.9908412 -0.1999478 -0.01955777 -0.001963354
##                   75%     97.5%
## beta[47]   0.02019221 0.2099449
## beta[76]   0.01755948 0.2820739
## beta[177]  0.01381253 0.1947565
## beta[275]  0.01405657 0.1631913
## beta[304]  0.01346258 0.1579782
## beta[410]  0.01315023 0.2207876
## beta[693]  0.01856673 0.3324403
## beta[729]  0.01484838 0.2050923
## beta[745]  0.01324924 0.1521403
## beta[996]  0.01348866 0.1758965
```

```
jags.fit.s$BUGSoutput$summary[1001:2000,][var.select>0, ]
```

```
##                 mean        sd 2.5% 25% 50% 75% 97.5%
## prob_neg[47]   0.453 0.4980352    0   0   0   1     1
## prob_neg[76]   0.450 0.4977427    0   0   0   1     1
## prob_neg[177]  0.548 0.4979397    0   0   1   1     1
## prob_neg[275]  0.552 0.4975375    0   0   1   1     1
## prob_neg[304]  0.545 0.4982201    0   0   1   1     1
## prob_neg[410]  0.545 0.4982201    0   0   1   1     1
## prob_neg[693]  0.458 0.4984822    0   0   0   1     1
## prob_neg[729]  0.546 0.4981286    0   0   1   1     1
## prob_neg[745]  0.541 0.4985655    0   0   1   1     1
## prob_neg[996]  0.545 0.4982201    0   0   1   1     1
```

```
jags.fit.s$BUGSoutput$summary[2001:3000,][var.select>0, ]
```

```
##                 mean        sd 2.5% 25% 50% 75% 97.5%
## prob_pos[47]   0.547 0.4980352    0   0   1   1     1
## prob_pos[76]   0.550 0.4977427    0   0   1   1     1
## prob_pos[177]  0.452 0.4979397    0   0   0   1     1
## prob_pos[275]  0.448 0.4975375    0   0   0   1     1
## prob_pos[304]  0.455 0.4982201    0   0   0   1     1
## prob_pos[410]  0.455 0.4982201    0   0   0   1     1
## prob_pos[693]  0.542 0.4984822    0   0   1   1     1
## prob_pos[729]  0.454 0.4981286    0   0   0   1     1
## prob_pos[745]  0.459 0.4985655    0   0   0   1     1
## prob_pos[996]  0.455 0.4982201    0   0   0   1     1
```

Variables selected by Spike-and-slab are shown below. Means of Bernoulli auxiliary variable for feature selection are shown by gamma[i]. We can see results are better than Bayesian lasso, since Spike-and-slab provides a more direct way for variable selection, unlike Bayesian lasso requires further tests.

```r
sum(jags.fit.s2$BUGSoutput$summary[1001:2000, 1] > 0.54)
```

```
## [1] 8
```

```r
jags.fit.s2$BUGSoutput$summary[1001: 2000,][jags.fit.s2$BUGSoutput$summary[1001:2000, 1] > 0.54, ]
```

```
##             mean        sd 2.5% 25% 50% 75% 97.5%
## gamma[172] 0.542 0.4984822    0   0   1   1     1
## gamma[209] 0.541 0.4985655    0   0   1   1     1
## gamma[356] 0.546 0.4981286    0   0   1   1     1
## gamma[491] 0.541 0.4985655    0   0   1   1     1
## gamma[517] 0.552 0.4975375    0   0   1   1     1
## gamma[770] 0.546 0.4981286    0   0   1   1     1
## gamma[782] 0.543 0.4983968    0   0   1   1     1
## gamma[834] 0.541 0.4985655    0   0   1   1     1
```

```r
jags.fit.s2$BUGSoutput$summary[1: 1000,][jags.fit.s2$BUGSoutput$summary[1001:2000, 1] > 0.54, ]
```

```
##                  mean       sd      2.5%        25%          50%       75%
## beta[172] -0.03540215 2.179206 -4.815912 -0.3598660 -0.010212572 0.3480818
## beta[209] -0.04529348 2.406991 -5.542167 -0.4180400 -0.002457397 0.2515986
## beta[356] -0.06631986 2.320785 -5.391325 -0.4863734 -0.011390059 0.2579368
## beta[491]  0.10697013 2.328480 -4.950863 -0.2183797  0.013449066 0.6367591
## beta[517] -0.10561471 2.463762 -5.554069 -0.6036541 -0.005907191 0.2532708
## beta[770]  0.05091301 2.295969 -4.978952 -0.4356859 -0.008919571 0.2490436
## beta[782] -0.05834215 2.413809 -5.607624 -0.3013522 -0.003259758 0.3703847
## beta[834] -0.08160203 2.327330 -5.682744 -0.3313229 -0.001022064 0.2226150
##              97.5%
## beta[172] 4.983304
## beta[209] 5.569559
## beta[356] 5.182293
## beta[491] 5.179847
## beta[517] 5.122940
## beta[770] 5.648752
## beta[782] 5.160353
## beta[834] 5.084725
```