STATS 305A HW#8

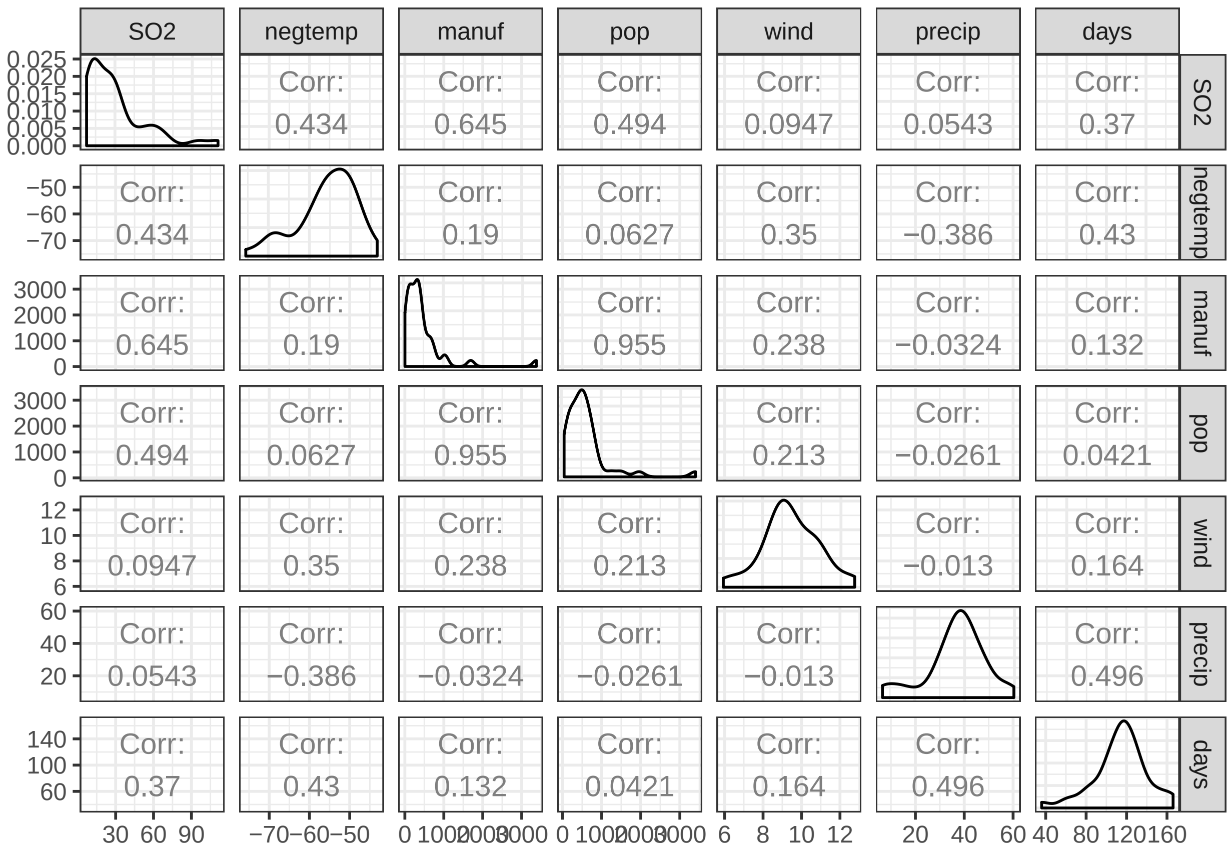
Adhitya Venkatesh

December 7, 2017

1.

a.), b.)

```
data(usair, package = "brinla")
library(ggplot2, GGally)
pairs.chart = GGally::ggpairs(usair, lower = list(continuous = "cor"), upper = list(continous = "points
print(pairs.chart)
```



There are no glaring problems of multicollinearity bar perhaps the ecological predictors "pop" and "manuf" with a nearly perfect correlation of 0.955. All the predictors are mutually correlated to various degress with "manuf", "pop", "negtemp", and "day" being the most highly correlated with Sulfur Dioxide levels.

c.)

```
usair.formula1=SO2~negtemp+manuf+wind+precip+days
usair.lm1 = lm(usair.formula1, data = usair)
summary(usair.lm1)
```

```
##
## Call:
## lm(formula = usair.formula1, data = usair)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -20.439  -8.719  -3.198   7.170  58.024
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 135.771432  50.060984   2.712  0.01029 *
## negtemp       1.771413   0.636641   2.782  0.00864 **
## manuf         0.025573   0.004604   5.554 2.99e-06 ***
## wind         -3.737852   1.944409  -1.922  0.06273 .
## precip        0.625897   0.388500   1.611  0.11615
## days         -0.057060   0.174775  -0.326  0.74601
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.79 on 35 degrees of freedom
## Multiple R-squared:  0.604,  Adjusted R-squared:  0.5475
## F-statistic: 10.68 on 5 and 35 DF,  p-value: 2.75e-06
```

Based on the summary, "negtemp" and "manuf" are statistically significant predictors whereas "wind", "precip", and "days" aren't.

d.)

```r
library("INLA")
```

```
## Loading required package: sp

## Loading required package: Matrix

## This is INLA_17.06.20 built 2017-06-20 03:44:36 UTC.
## See www.r-inla.org/contact-us for how to get help.
```

```r
usair.inla1 = inla(usair.formula1, data = usair, control.compute = list(dic=TRUE, cpo = TRUE))
usair.inla1$summary.fixed
```

```
##                   mean          sd  0.025quant     0.5quant  0.975quant
## (Intercept) 135.48860403 50.108848092 36.51592366 135.49520429 234.2740740
## negtemp       1.76896164  0.637544242  0.50979835   1.76901270   3.0259234
## manuf         0.02556704  0.004612736  0.01645768   0.02556708   0.0346623
## wind         -3.72286652  1.944199432 -7.56070749  -3.72335627   0.1118341
## precip        0.62484935  0.389129822 -0.14365564   0.62486760   1.3920816
## days         -0.05665577  0.175068837 -0.40233177  -0.05667279   0.2885867
##                   mode          kld
## (Intercept) 135.51175951 1.017835e-13
## negtemp       1.76916256 9.396106e-14
## manuf         0.02556756 8.662473e-14
## wind         -3.72409100 1.154894e-13
## precip        0.62493523 9.045826e-14
## days         -0.05668889 8.932813e-14
```

```r
usair.inla1$summary.hyperpar
```

```
##                                             mean           sd
## Precision for the Gaussian observations 0.004241498 0.0009810649
##                                          0.025quant    0.5quant
## Precision for the Gaussian observations 0.002542247 0.004165219
##                                          0.975quant        mode
## Precision for the Gaussian observations 0.006377441 0.004011763
```

```r
summary(usair.inla1)
```

```
##
## Call:
## c("inla(formula = usair.formula1, data = usair, control.compute = list(dic = TRUE, ",  "    cpo = TRU
##
## Time used:
##  Pre-processing    Running inla Post-processing           Total
##          0.9861          0.1259          0.2450          1.3569
##
## Fixed effects:
##                 mean      sd 0.025quant 0.5quant 0.975quant     mode kld
## (Intercept) 135.4886 50.1088    36.5159 135.4952   234.2741 135.5118   0
## negtemp       1.7690  0.6375     0.5098   1.7690     3.0259   1.7692   0
## manuf         0.0256  0.0046     0.0165   0.0256     0.0347   0.0256   0
## wind         -3.7229  1.9442    -7.5607  -3.7234     0.1118  -3.7241   0
## precip        0.6248  0.3891    -0.1437   0.6249     1.3921   0.6249   0
## days         -0.0567  0.1751    -0.4023  -0.0567     0.2886  -0.0567   0
##
## The model has no random effects
##
## Model hyperparameters:
##                                           mean    sd 0.025quant 0.5quant
## Precision for the Gaussian observations 0.0042 0.001     0.0025   0.0042
##                                         0.975quant  mode
## Precision for the Gaussian observations     0.0064 0.004
##
## Expected number of effective parameters(std dev): 5.996(0.0011)
## Number of equivalent replicates : 6.838
##
## Deviance Information Criterion (DIC) ...: 350.71
## Effective number of parameters .........: 7.201
##
## Marginal log-Likelihood:  -208.73
## CPO and PIT are computed
##
## Posterior marginals for linear predictor and fitted values computed
```

It's clear from comparing the summaries from parts c.) and d.) that the Bayesian and frequentist estimates are roughly equivalent. Using the "usair.inla2" model from the next part, we obtain:

```r
usair.inla2 = inla(usair.formula1, data = usair, control.compute = list(dic = TRUE, cpo = TRUE), control
usair.inla2$summary.fixed
```

```
##                     mean          sd   0.025quant     0.5quant  0.975quant
## (Intercept) 123.60987606 9.561056494 104.83343099 123.61116908 142.3619733
## negtemp       1.54918826 0.212739638   1.13000614   1.54950512   1.9661902
## manuf         0.02614260 0.004011081   0.01824695   0.02613783   0.0340534
## wind         -4.36101924 0.805865485  -5.95030618  -4.35885836  -2.7851629
## precip        0.50744470 0.246845115   0.01978107   0.50792232   0.9918840
## days          0.02701884 0.090149214  -0.14900073   0.02652620   0.2056115
##                     mode          kld
## (Intercept) 123.61458234 0.000000e+00
## negtemp       1.55016355 1.717160e-12
## manuf         0.02612882 1.686401e-12
```

```
## wind         -4.35443065 1.600150e-13
## precip        0.50888825 2.547683e-12
## days          0.02557473 3.263329e-12
```

```r
usair.inla2$summary.hyperpar
```

```
##                                               mean          sd  0.025quant
## Precision for the Gaussian observations 0.00534061 0.001071889 0.003452831
##                                            0.5quant  0.975quant
## Precision for the Gaussian observations 0.005267273 0.007648044
##                                                mode
## Precision for the Gaussian observations 0.005118375
```

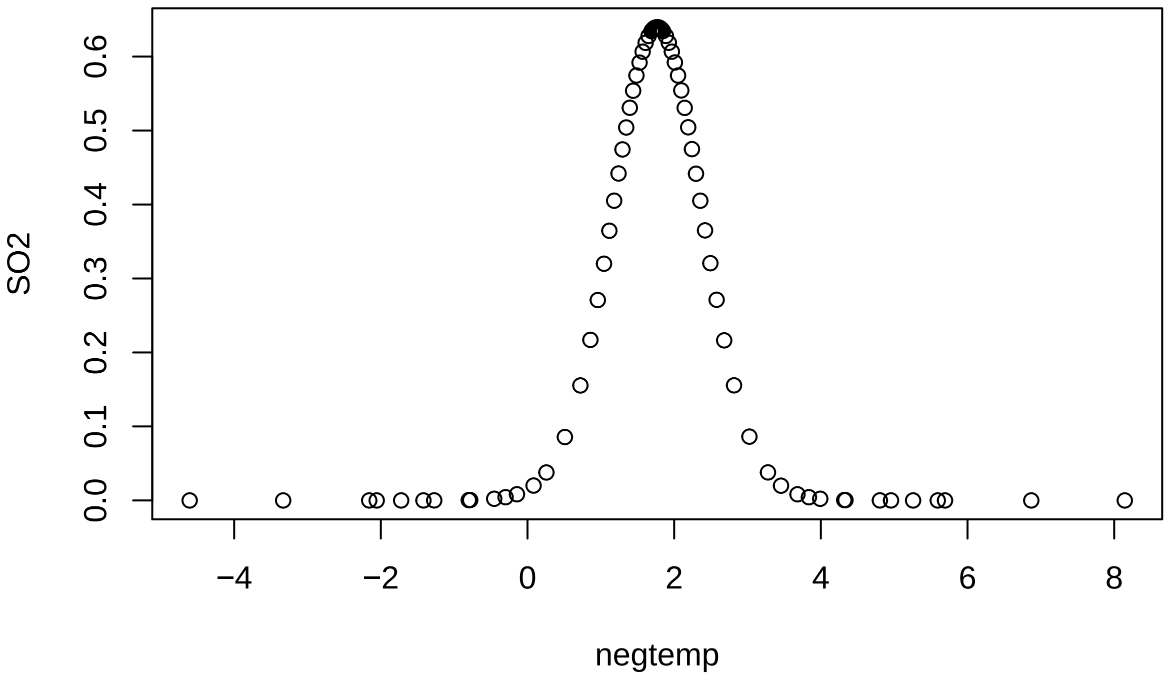
```r
summary(usair.inla2)
```

```
##
## Call:
## c("inla(formula = usair.formula1, data = usair, control.compute = list(dic = TRUE, ",  "    cpo = TRU
##
## Time used:
##  Pre-processing    Running inla Post-processing           Total
##          1.9686          0.2697          1.7493          3.9876
##
## Fixed effects:
##                 mean     sd 0.025quant 0.5quant 0.975quant     mode kld
## (Intercept) 123.6099 9.5611   104.8334 123.6112   142.3620 123.6146   0
## negtemp       1.5492 0.2127     1.1300   1.5495     1.9662   1.5502   0
## manuf         0.0261 0.0040     0.0182   0.0261     0.0341   0.0261   0
## wind         -4.3610 0.8059    -5.9503  -4.3589    -2.7852  -4.3544   0
## precip        0.5074 0.2468     0.0198   0.5079     0.9919   0.5089   0
## days          0.0270 0.0901    -0.1490   0.0265     0.2056   0.0256   0
##
## The model has no random effects
##
## Model hyperparameters:
##                                           mean     sd 0.025quant 0.5quant
## Precision for the Gaussian observations 0.0053 0.0011     0.0035   0.0053
##                                         0.975quant   mode
## Precision for the Gaussian observations     0.0076 0.0051
##
## Expected number of effective parameters(std dev): 4.327(0.0732)
## Number of equivalent replicates : 9.476
##
## Deviance Information Criterion (DIC) ...: 347.49
## Effective number of parameters .........: 5.233
##
## Marginal log-Likelihood:  -198.63
## CPO and PIT are computed
##
## Posterior marginals for linear predictor and fitted values computed
```

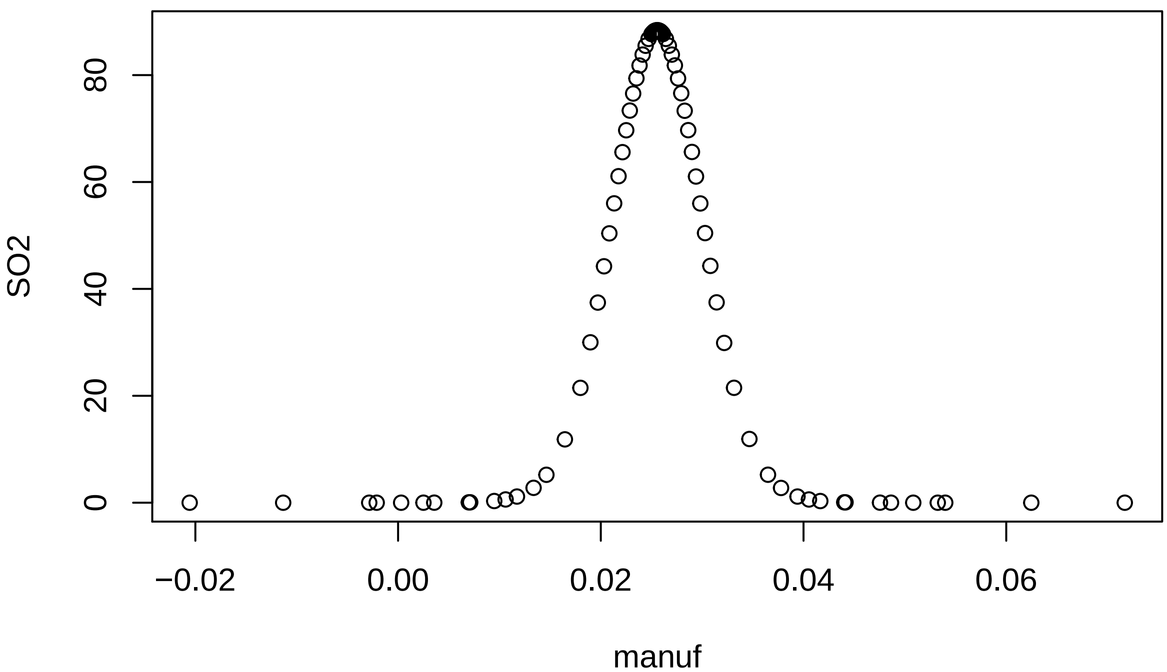
We clearly see the strong influence choice of prior has on the posterior estimates.

e.)

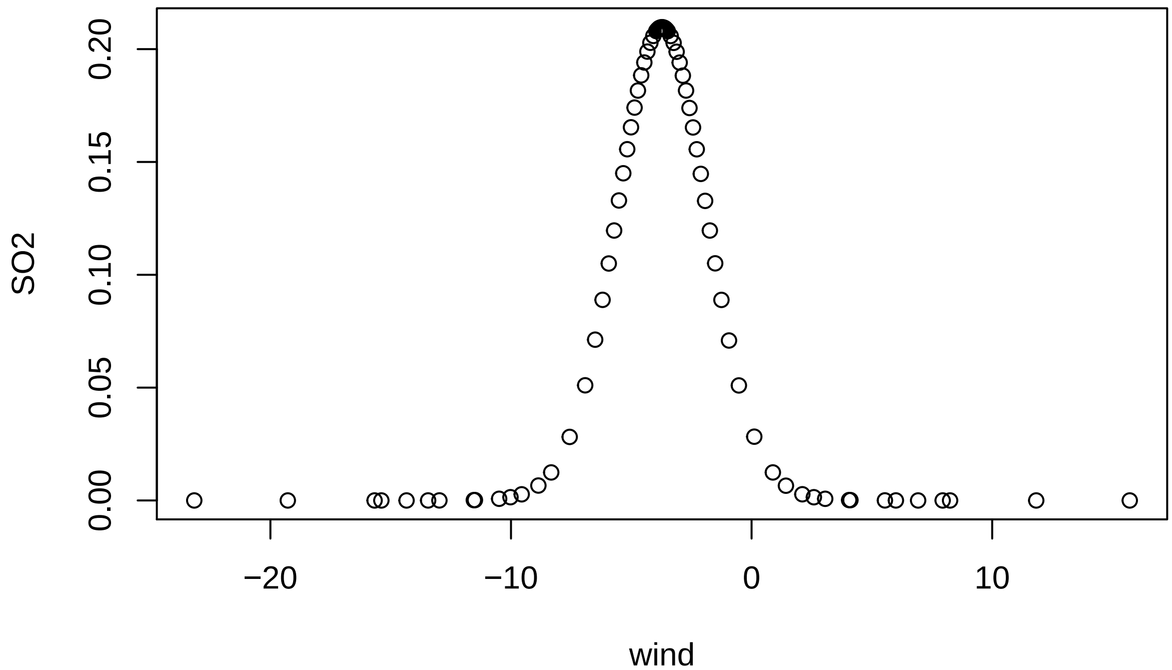
```r
plot(usair.inla1$marginals.fixed$negtemp, xlab = "negtemp", ylab = "SO2")
```



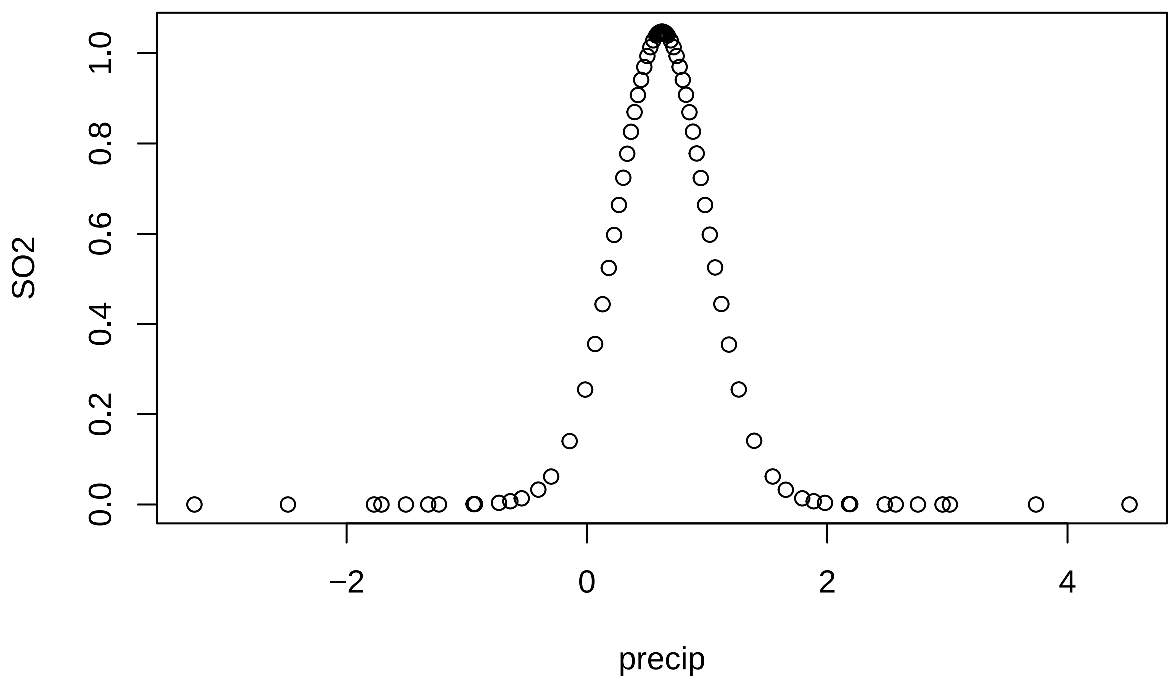

```r
plot(usair.inla1$marginals.fixed$manuf, xlab = "manuf", ylab = "SO2")
```



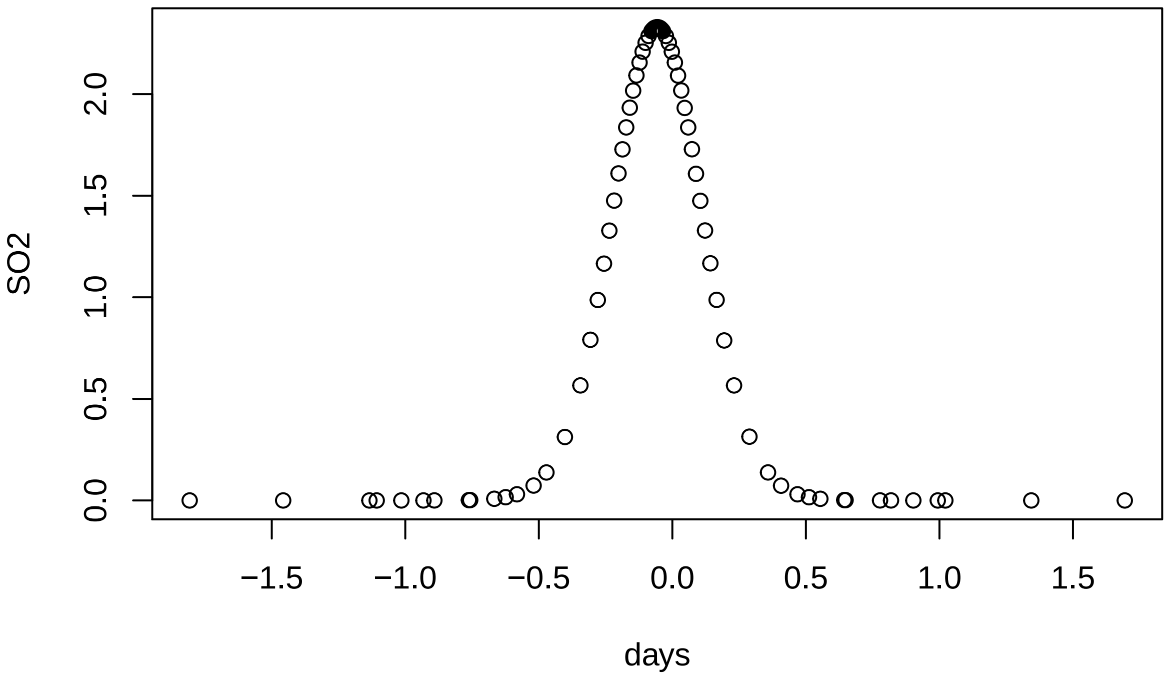

```r
plot(usair.inla1$marginals.fixed$wind, xlab = "wind", ylab = "SO2")
```




```
plot(usair.inla1$marginals.fixed$precip, xlab = "precip", ylab = "SO2")
```




```
plot(usair.inla1$marginals.fixed$days, xlab = "days", ylab = "SO2")
```



All plots indicate the distributions are normal as expected.

f.)

```
usair.inla2 = inla(usair.formula1, data = usair, control.compute = list(dic = TRUE, cpo = TRUE), control
summary(usair.inla2)
```

```
##
## Call:
## c("inla(formula = usair.formula1, data = usair, control.compute = list(dic = TRUE, ",  "    cpo = TRU
##
## Time used:
##  Pre-processing    Running inla Post-processing           Total
##          1.0017          0.2733          0.1109          1.3859
##
## Fixed effects:
##                 mean     sd 0.025quant 0.5quant 0.975quant     mode kld
## (Intercept) 102.3903 9.5635    83.6079 102.3920   121.1461 102.3963   0
## negtemp       1.3844 0.2127     0.9657   1.3846     1.8018   1.3850   0
## manuf         0.0255 0.0040     0.0176   0.0255     0.0334   0.0255   0
## wind         -2.9520 0.8034    -4.5299  -2.9521    -1.3749  -2.9523   0
## precip        0.4408 0.2469    -0.0468   0.4412     0.9254   0.4420   0
## days          0.0406 0.0900    -0.1360   0.0404     0.2181   0.0400   0
##
## The model has no random effects
##
## Model hyperparameters:
##                                          mean     sd 0.025quant 0.5quant
## Precision for the Gaussian observations 0.0053 0.0011     0.0035   0.0053
##                                         0.975quant   mode
## Precision for the Gaussian observations     0.0076 0.0051
##
## Expected number of effective parameters(std dev): 4.327(0.0729)
## Number of equivalent replicates : 9.476
```

```
##
## Deviance Information Criterion (DIC) ...: 347.51
## Effective number of parameters .........: 5.236
##
## Marginal log-Likelihood:  -197.51
## CPO and PIT are computed
##
## Posterior marginals for linear predictor and fitted values computed
```

We can see from the summary that the prior information definitely affected the parameter estimates, notably for "negtemp" that dropped from 1.77 in the frequentist model to 1.38 in this Bayesian model.

g.)

```r
newdata = data.frame(negtemp = c(-50, -60, -40), manuf=c(150,100,400), pop = c(200, 100, 300), wind = c

#frequentist prediction
predict(usair.lm1, newdata)
```

```
##        1        2        3
## 33.72743 18.94993 55.47696
```

```r
#Bayesian prediction
usair.combined = rbind(usair, data.frame(SO2 = c(NA, NA, NA), newdata))
usair.link = c(rep(NA, nrow(usair)), rep(1, nrow(newdata)))
usair.inla1.pred = inla(usair.formula1, data = usair.combined, control.predictor = list(link = usair.lin
usair.inla1.pred$summary.fitted.values[(nrow(usair) + 1):nrow(usair.combined),]
```

```
##                          mean       sd 0.025quant 0.5quant 0.975quant
## fitted.Predictor.42 33.65311 14.94934   4.172056 33.65537   63.12978
## fitted.Predictor.43 18.92731  5.33361   8.409584 18.92800   29.44426
## fitted.Predictor.44 55.40395 17.65764  20.583297 55.40619   90.22212
##                          mode
## fitted.Predictor.42 33.65962
## fitted.Predictor.43 18.92931
## fitted.Predictor.44 55.41043
```

Comparing the predicted values in the frequentist case with the mean column of the Bayesian predictions, it is clear that they yield extremely close results.

h.)

```r
#DIC

L_0 = logLik(usair.lm1)[1] #log-liklihood when model accounts for all predictors

DIC_0 = -2*(L_0 -df.residual(usair.lm1))

#From here, we run regressions of all possible 5-predictor combinations of the original 6 and compare t


#AIC
library(MASS)
stepAIC(usair.lm1)
```

```
## Start:  AIC=231.78
## SO2 ~ negtemp + manuf + wind + precip + days
##
```

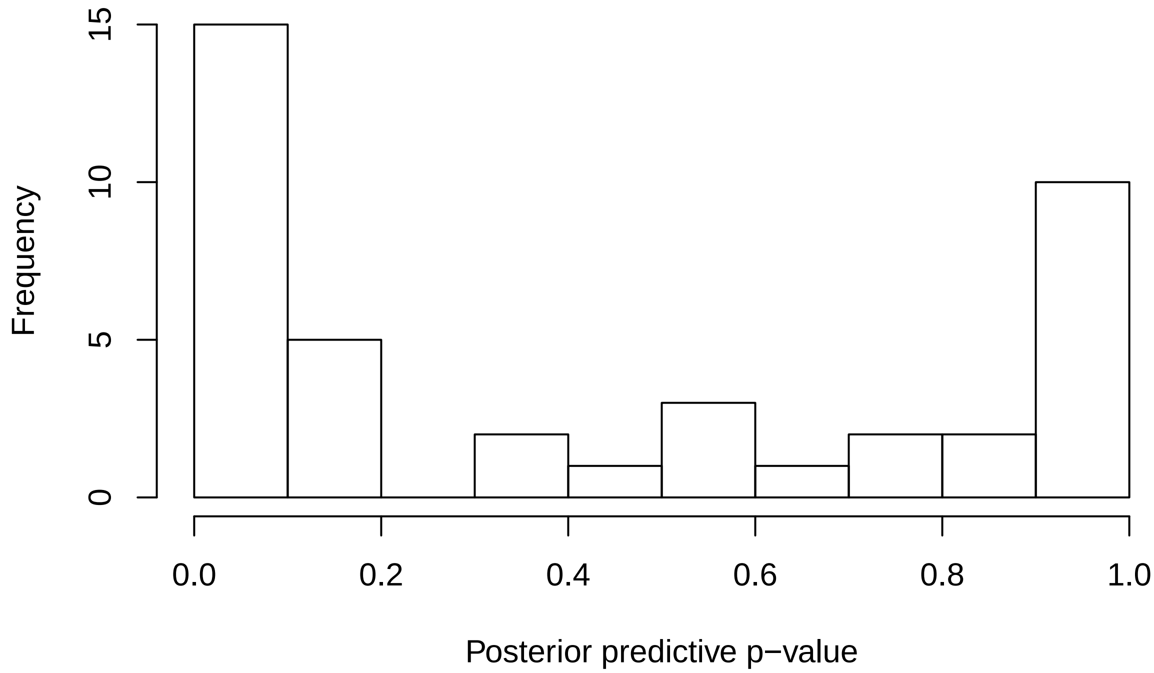
```
##           Df Sum of Sq     RSS    AIC
## - days     1      26.6  8752.9 229.91
## <none>                  8726.3 231.78
## - precip   1     647.1  9373.4 232.72
## - wind     1     921.4  9647.7 233.90
## - negtemp  1    1930.3 10656.6 237.97
## - manuf    1    7692.0 16418.4 255.70
##
## Step:  AIC=229.91
## SO2 ~ negtemp + manuf + wind + precip
##
##           Df Sum of Sq     RSS    AIC
## <none>                  8752.9 229.91
## - wind     1     894.8  9647.7 231.90
## - precip   1    1269.7 10022.6 233.46
## - negtemp  1    3919.0 12671.9 243.08
## - manuf    1    7665.8 16418.7 253.70
```

```
##
## Call:
## lm(formula = SO2 ~ negtemp + manuf + wind + precip, data = usair)
##
## Coefficients:
## (Intercept)      negtemp        manuf         wind       precip
##   123.11833      1.61144      0.02548     -3.63024      0.52423
```

Under AIC, the best model uses covariates "negtemp", "manuf", "wind", and "precip".

i.)

```r
usair.inla3.pred = inla(usair.formula1, data = usair, control.predictor = list(link = 1, compute = TRUE
post.predicted.pval = vector(mode = "numeric", length = nrow(usair))
for(i in (1:nrow(usair))){
 post.predicted.pval[i] = inla.pmarginal(q = usair$SO2[i], marginal = usair.inla3.pred$marginals.fitted
}
hist(post.predicted.pval, main = "", breaks = 10, xlab = "Posterior predictive p-value")
```



From the histogram , it's clear that the distribution of p-values is bimodal with peaks at p-value = 0 and 1. This suggests there may be samples that are fundamentally different from the others.

j.)

```
hist(usair.inla2$cpo$pit, main = "", breaks = 10, xlab = "PIT")
qqplot(qunif(ppoint(length(usair.inla2$cpo$pit))), usair.inla2$cpo$pit, main = "Q-Q plot for Unif(0,1)"
qqline(usair.inla2$cpo$pit, distribution = function(p) qunif(p), prob = c(0.1, 0.9))
```

2.)

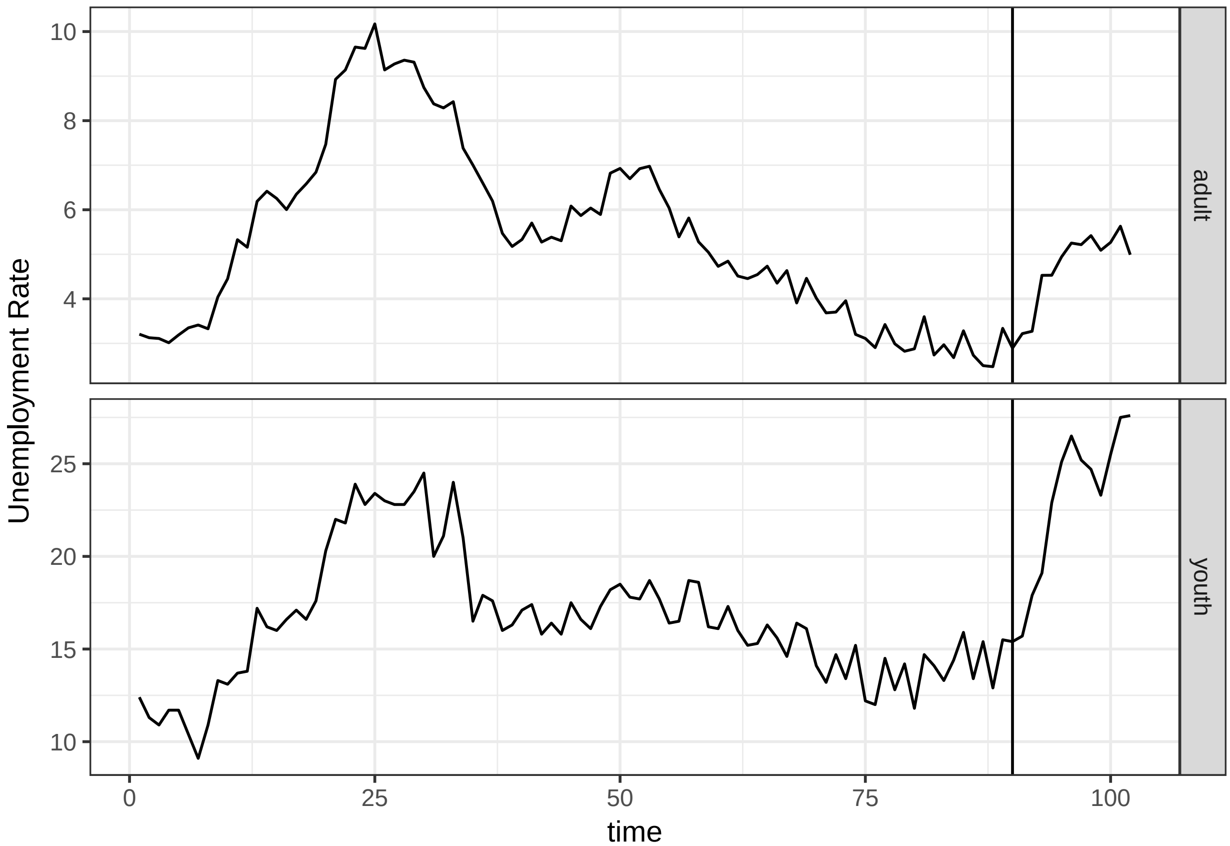
a.)

```
library(tidyr)
```

```
##
## Attaching package: 'tidyr'

## The following object is masked from 'package:Matrix':
##
##     expand
```

```
data(nzunemploy, package = "brinla")
nzunemploy$time = 1:nrow(nzunemploy)
qplot(time, value, data = gather(nzunemploy[,c(2,3,5)], variable, value, -time), geom = "line") + geom_
```



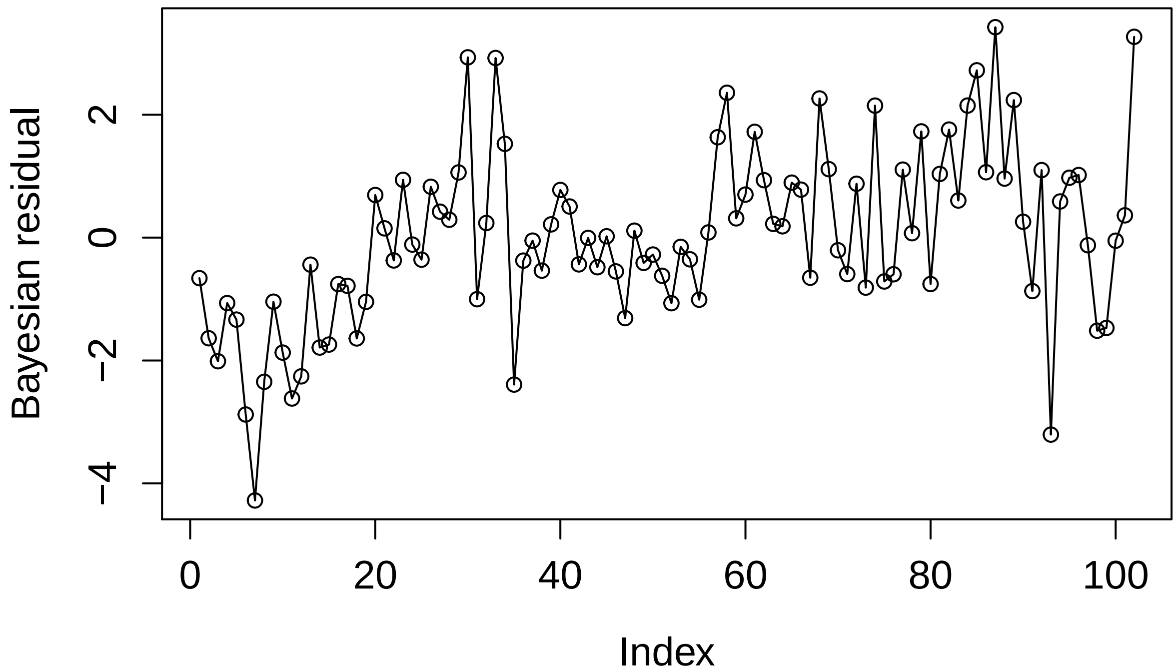
Seems evident from the stark contrast in the unemployment trends after the minimmum wage was abolished in that for the youth the rate significantly picked up (nearly exponential uptick) post-abolishment whereas from the time series for the adult unemployment rate, abolishment had little effect on the trend dynamics.

b.)

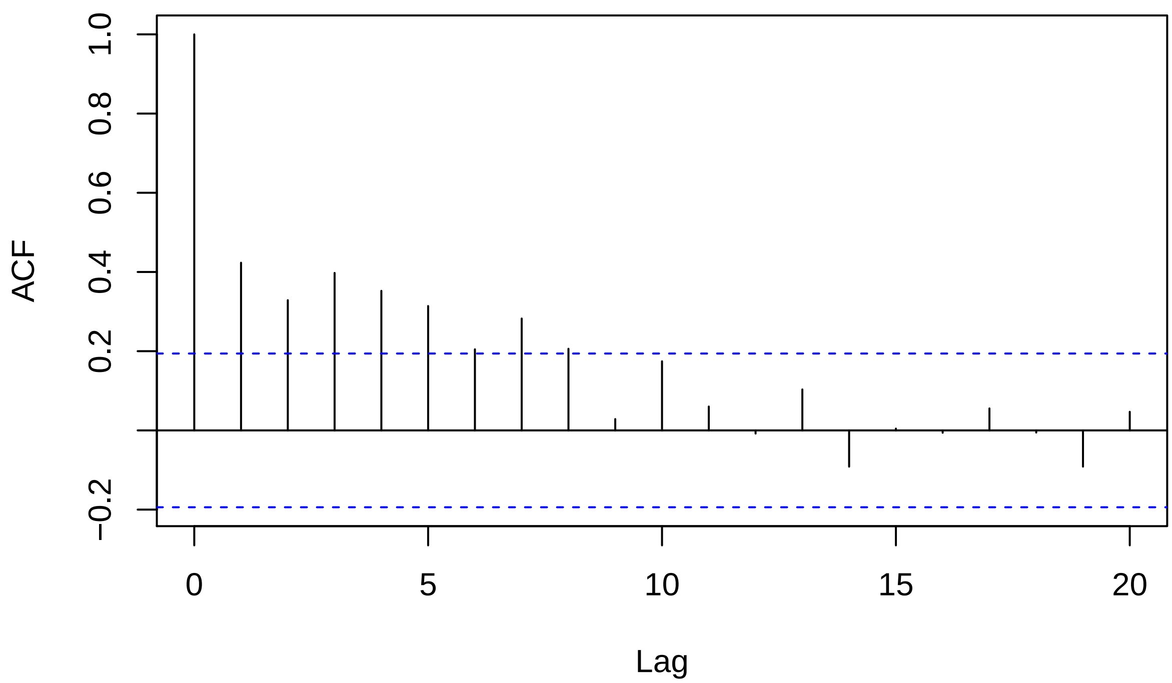
```
nzunemploy$centeredadult = with(nzunemploy, adult - mean(adult))
formula1 = youth~centeredadult*policy
nzunemploy.inla1 = inla(formula1, data = nzunemploy)
round(nzunemploy.inla1$summary.fixed, 4)
```

```
##                              mean     sd 0.025quant 0.5quant 0.975quant
## (Intercept)               16.2823 0.1536    15.9800  16.2823    16.5843
## centeredadult              1.5333 0.0751     1.3855   1.5333     1.6810
## policyEqual                9.4417 0.5266     8.4055   9.4417    10.4766
## centeredadult:policyEqual  2.8533 0.4622     1.9437   2.8533     3.7618
##                              mode kld
## (Intercept)               16.2823   0
## centeredadult              1.5333   0
## policyEqual                9.4418   0
## centeredadult:policyEqual  2.8533   0
```

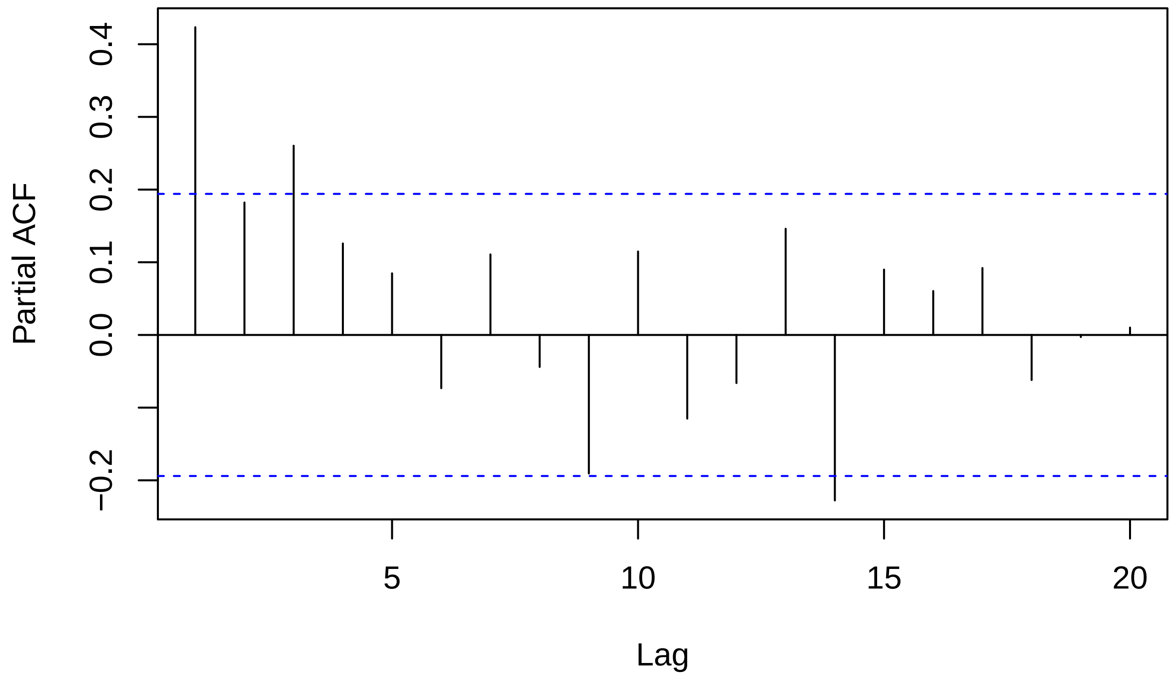
```
nzunemploy.res1 = brinla::bri.lmresid.plot(nzunemploy.inla1, type = "o")
```




```
acf(nzunemploy.res1$resid, main = "")
```




```
acf(nzunemploy.res1$resid, type = "partial", main = "")
```




```r
summary(nzunemploy.inla1)
```

```
##
## Call:
## "inla(formula = formula1, data = nzunemploy)"
##
## Time used:
##  Pre-processing    Running inla Post-processing           Total
##          0.8858          0.2619          0.0889          1.2366
##
## Fixed effects:
##                              mean     sd 0.025quant 0.5quant 0.975quant
## (Intercept)               16.2823 0.1536    15.9800  16.2823    16.5843
## centeredadult              1.5333 0.0751     1.3855   1.5333     1.6810
## policyEqual                9.4417 0.5266     8.4055   9.4417    10.4766
## centeredadult:policyEqual  2.8533 0.4622     1.9437   2.8533     3.7618
##                              mode kld
## (Intercept)               16.2823   0
## centeredadult              1.5333   0
## policyEqual                9.4418   0
## centeredadult:policyEqual  2.8533   0
##
## The model has no random effects
##
## Model hyperparameters:
##                                           mean     sd 0.025quant 0.5quant
## Precision for the Gaussian observations 0.4875 0.0685     0.3622   0.4843
##                                         0.975quant   mode
## Precision for the Gaussian observations     0.6309 0.4778
##
## Expected number of effective parameters(std dev): 4.00(1e-04)
## Number of equivalent replicates : 25.50
##
```

```
## Marginal log-Likelihood:  -207.72
```

From the graph, it is clear that both the "policyEqual" and interaction term (between "cenetedadult" and "policyequal") are significant. From the acf and partial acf graphs, there are certainly time differentials for which autocorrelation is significant.

c.)

```
formula2 = youth~centeredadult + f(time, model = "ar1")
nzunemploy.inla2 = inla(formula2, data = nzunemploy, control.family = list(hyper = list(prec = list(ini
summary(nzunemploy.inla2)
```

```
##
## Call:
## c("inla(formula = formula2, data = nzunemploy, control.family = list(hyper = list(prec = list(initia
##
## Time used:
##  Pre-processing    Running inla Post-processing           Total
##          1.2506          0.9820          0.1061          2.3388
##
## Fixed effects:
##                  mean     sd 0.025quant 0.5quant 0.975quant    mode   kld
## (Intercept)   18.4846 3.2403    12.9487  18.2028    25.8372 17.9322 2e-04
## centeredadult  1.6116 0.2974     1.0276   1.6111     2.1975  1.6100 0e+00
##
## Random effects:
## Name   Model
##  time   AR1 model
##
## Model hyperparameters:
##                        mean     sd 0.025quant 0.5quant 0.975quant   mode
## Precision for time 0.0663 0.0306     0.0197   0.0624     0.1364 0.0505
## Rho for time       0.9327 0.0316     0.8613   0.9366     0.9809 0.9491
##
## Expected number of effective parameters(std dev): 102.00(0.00)
## Number of equivalent replicates : 1.00
##
## Marginal log-Likelihood:  -199.29
```

This model yields similar conclusions to that of the previous part and reinforces the hypothesis in part a.).