ECN 190 Homework 2 Computer Problems

Kevin Chen (914861432) John Mayhew (914807483)

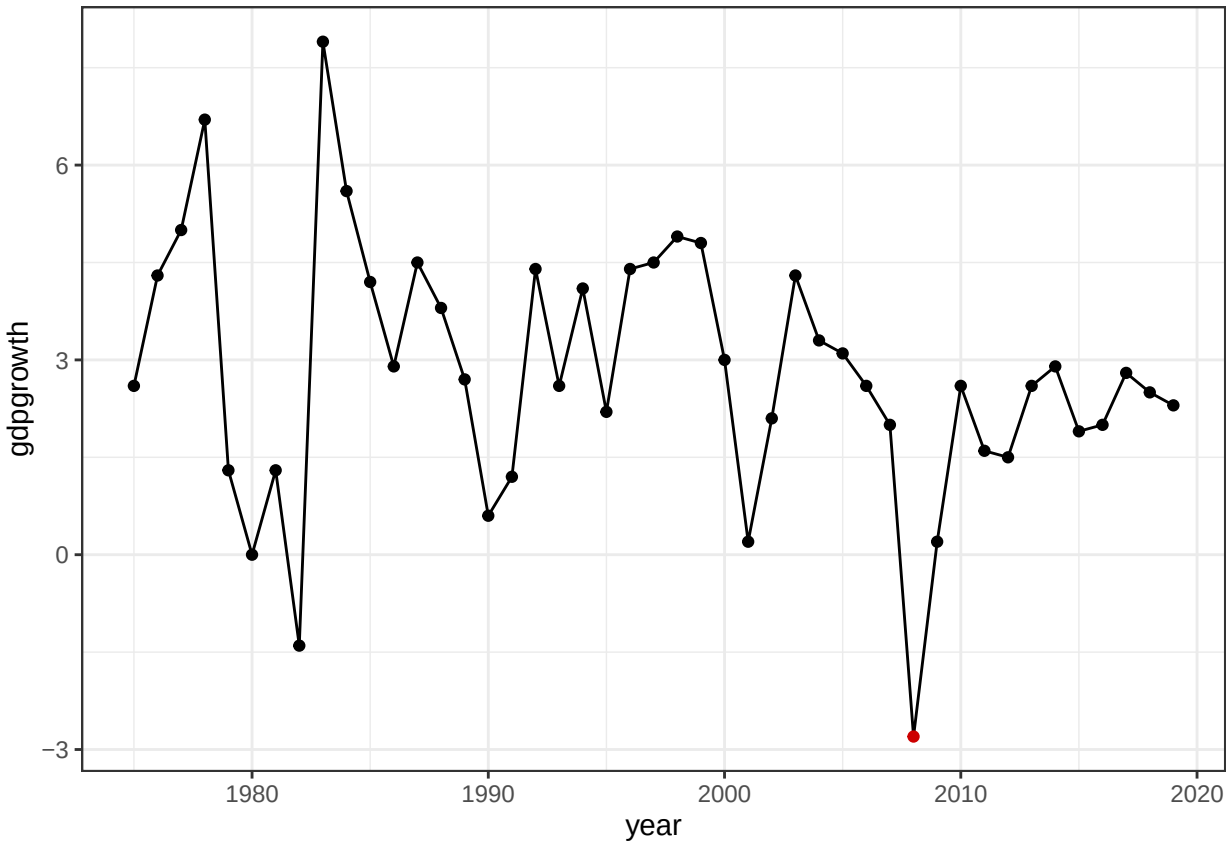
4/10/2020

1. Obtain time series data about the U.S. economony and summarize.

a. Create a data frame named "USeconomy" which includes three variables, "year" is 1975-2019, "gdpgrowth" is U.S. GDP for each year obtained from Table B-1, and "ue" is U.S. unemployment rate for each year obtained from Table B-27.

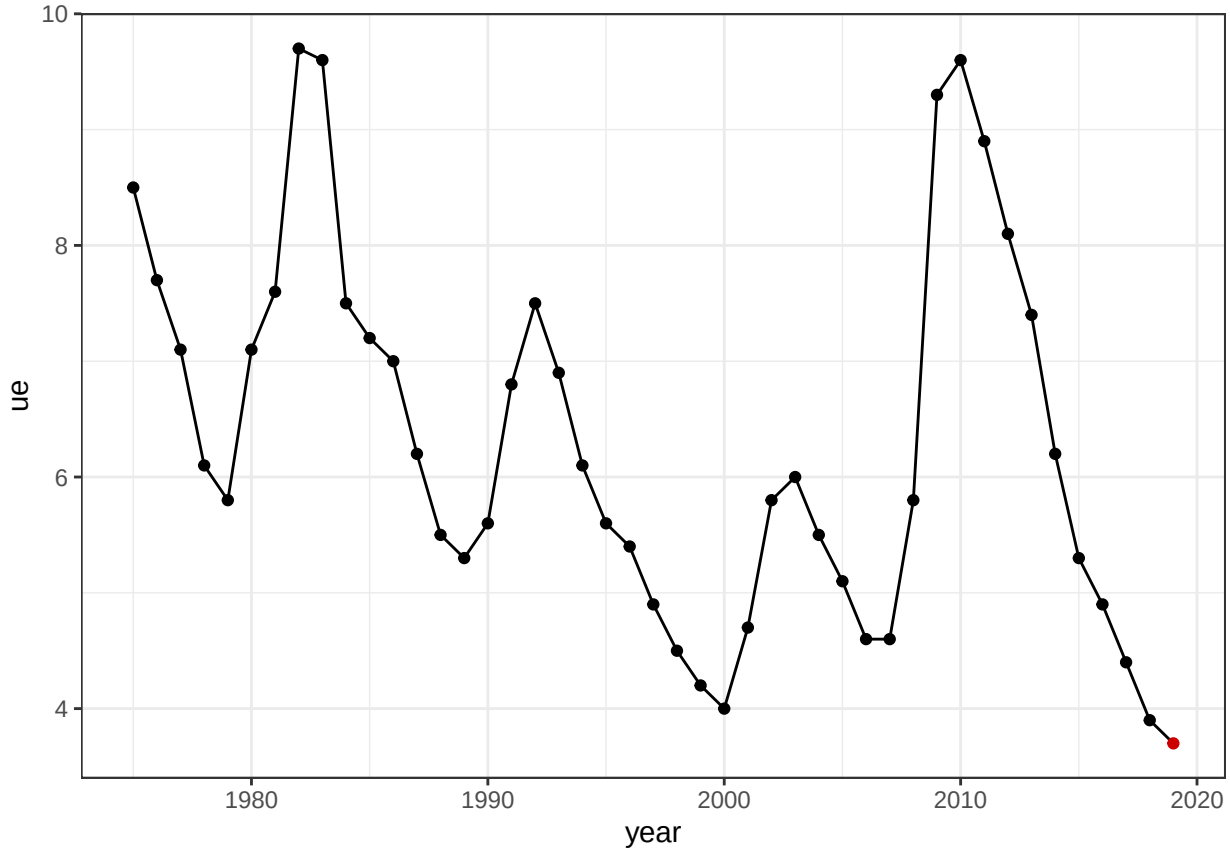
```
##    year gdpgrowth  ue
## 1  1975       2.6 8.5
## 2  1976       4.3 7.7
## 3  1977       5.0 7.1
## 4  1978       6.7 6.1
## 5  1979       1.3 5.8
## 6  1980       0.0 7.1
## 7  1981       1.3 7.6
## 8  1982      -1.4 9.7
## 9  1983       7.9 9.6
## 10 1984       5.6 7.5
```

b. Draw a line plot of gdpgrowth and year. What is the minimum value of this variable? In which year did this minimum value occur?



The minimum GDP growth is -2.8% and it happened in 2008.

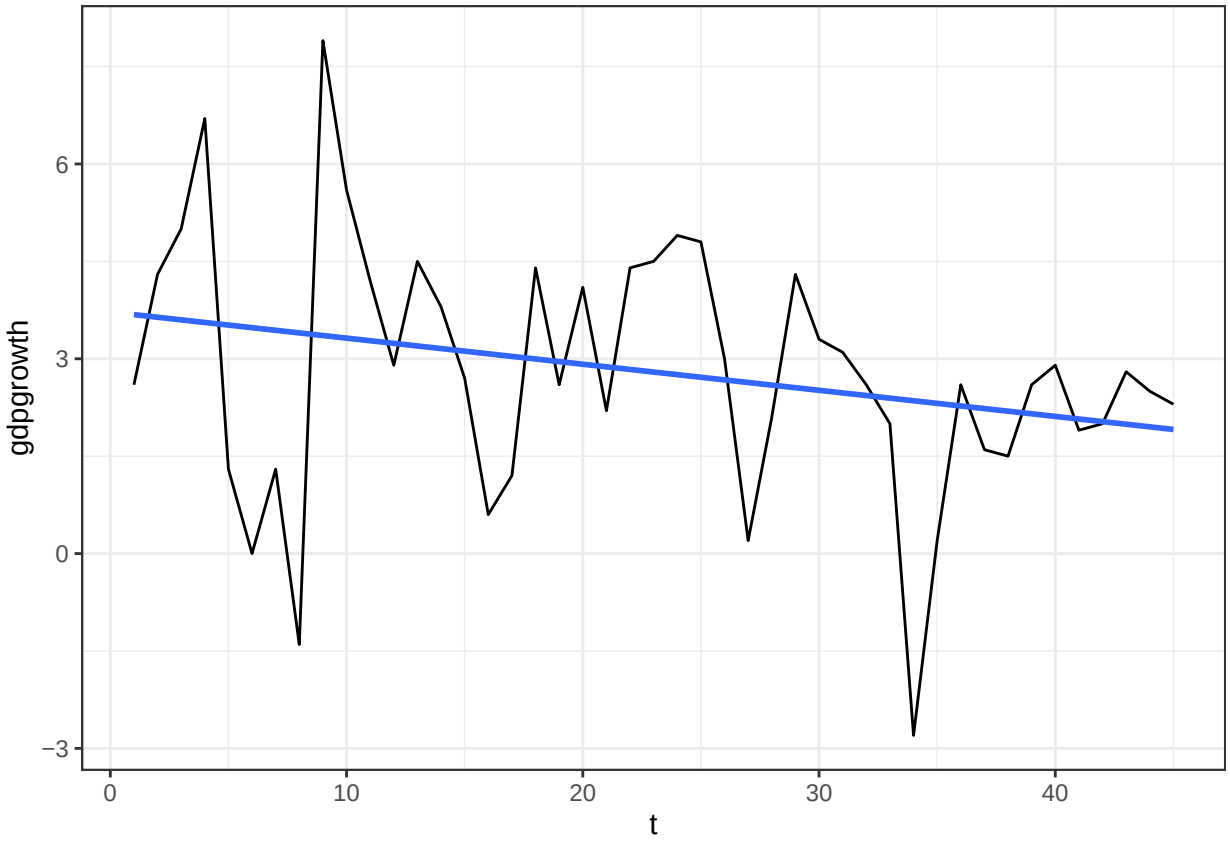
c. Draw a line plot of ue and year. What is the minimum value of this variable? In which year did this minimum value occur?



The minimum unemployment rate is 3.7% and it happened in 2019.

d. Generate a time period t variable that is equal to 1 for year 1975, 2 for year 1976, etc. Regress gdpgrowth on t. How would you interpret the intercept and the slope?

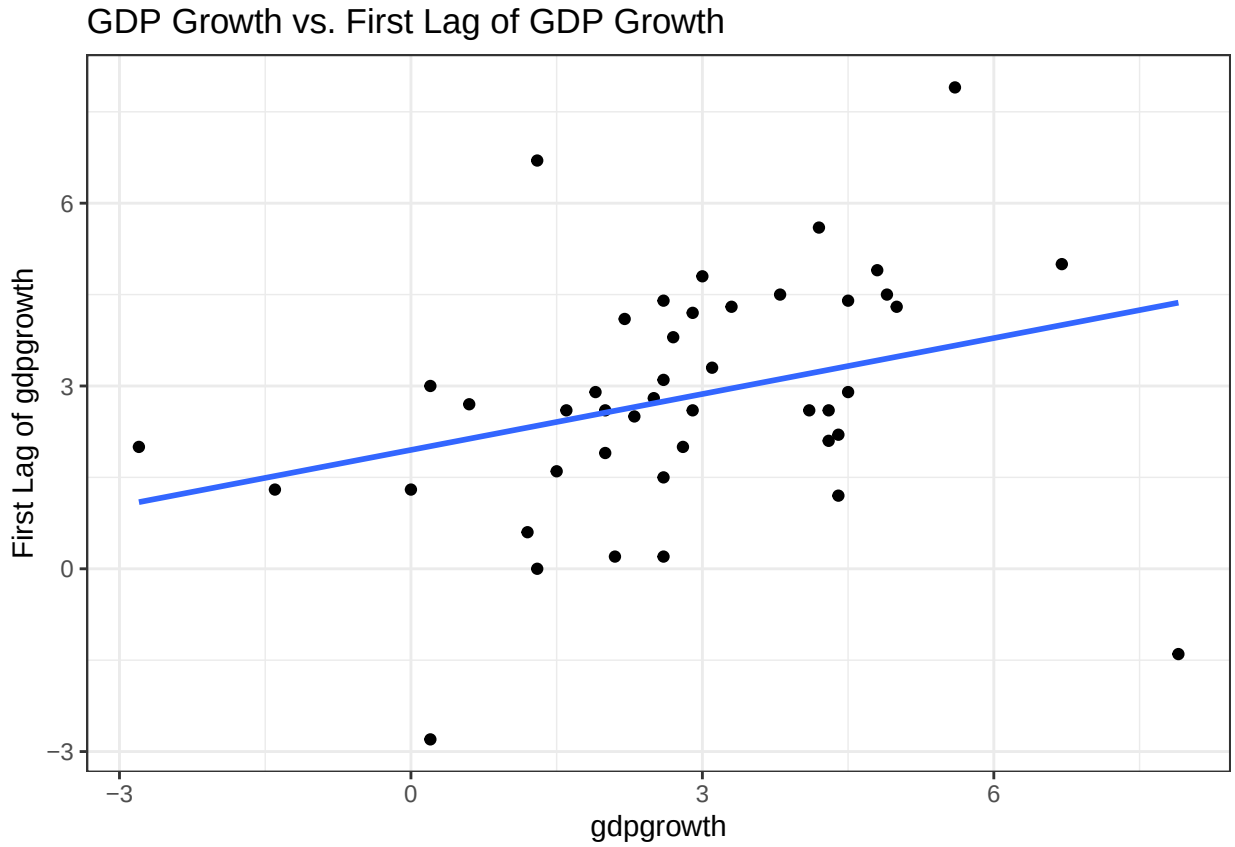
```
##
## Call:
## lm(formula = gdpgrowth ~ t, data = USeconomy)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.1535 -0.6759  0.3268  1.1839  4.5419
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.71980    0.58370   6.373 1.05e-07 ***
## t           -0.04018    0.02210  -1.818    0.076 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.925 on 43 degrees of freedom
## Multiple R-squared:  0.07141,    Adjusted R-squared:  0.04981
## F-statistic: 3.307 on 1 and 43 DF,  p-value: 0.07597
```



The intercept coefficient is 3.71980. This is the gdp growth rate our model predicts for time 0, the initial condition before the start of our time series, which is the year 1974.

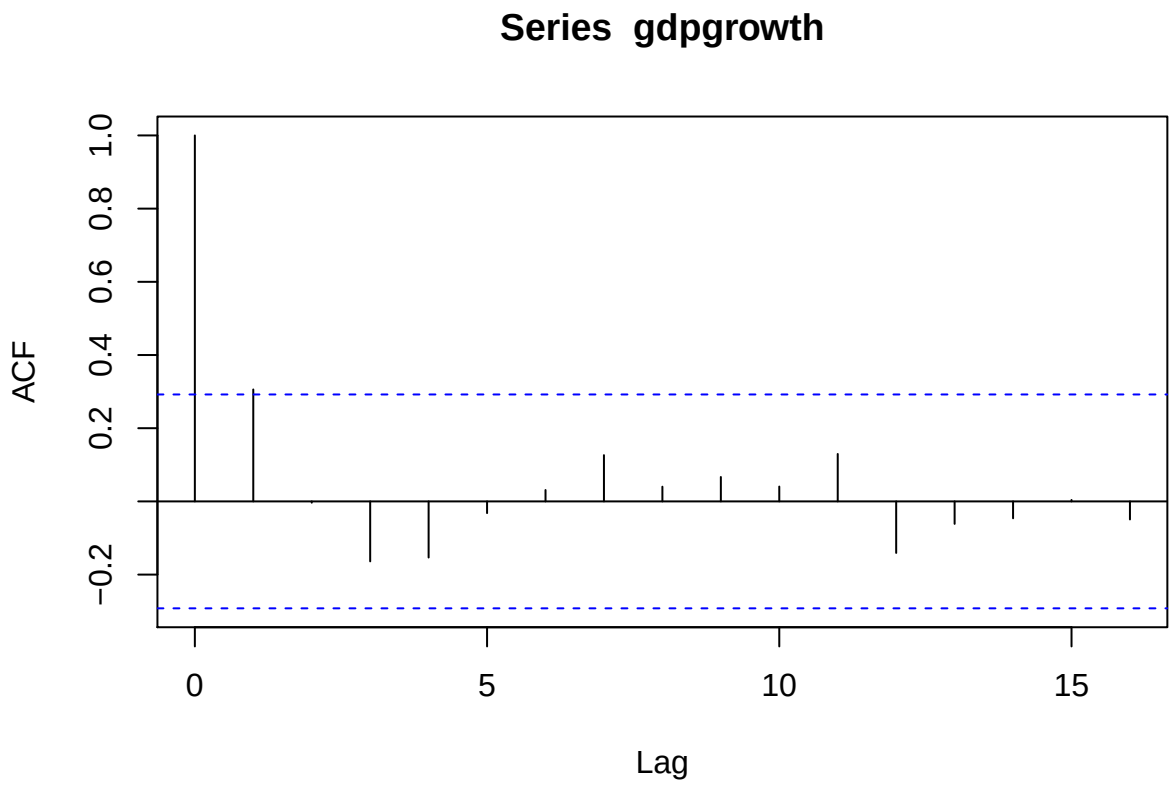
The slope coefficient is -0.04108. This measures the change in gdp growth from one period to the next. So if t increases by one unit (one year), gdp growth will decrease by 0.04%.

e. Now, draw a scatter plot gdpgrowth and its first lag. Do you think it implies a positive or negative association between the two variables? Would you say that the association is strong or weak?



The association between gdpgrowth and its first lag is positive, but the correlation between the two appears to be weak.

f. First draw the autocorrelation function (with the default maximum of 16 lags) of gdpgrowth. Then ask R to report those autocorrelation numbers. Do you think gdpgrowth is stationary or not?




```
##
## Autocorrelations of series 'gdpgrowth', by lag
##
##      0      1      2      3      4      5      6      7      8      9     10
##  1.000  0.306 -0.004 -0.164 -0.153 -0.032  0.031  0.126  0.040  0.067  0.041
##     11     12     13     14     15     16
##  0.130 -0.141 -0.061 -0.046  0.004 -0.049
```

Given that the ACF plot has a significant correlation between points of lag 1, the plot of gdpgrowth vs its first difference has a non-zero correlation, and the mean of the original time series appears to decrease over time, we conclude that the series gdpgrowth is not stationary.

2. Use the data in BARIUM for this exercise.

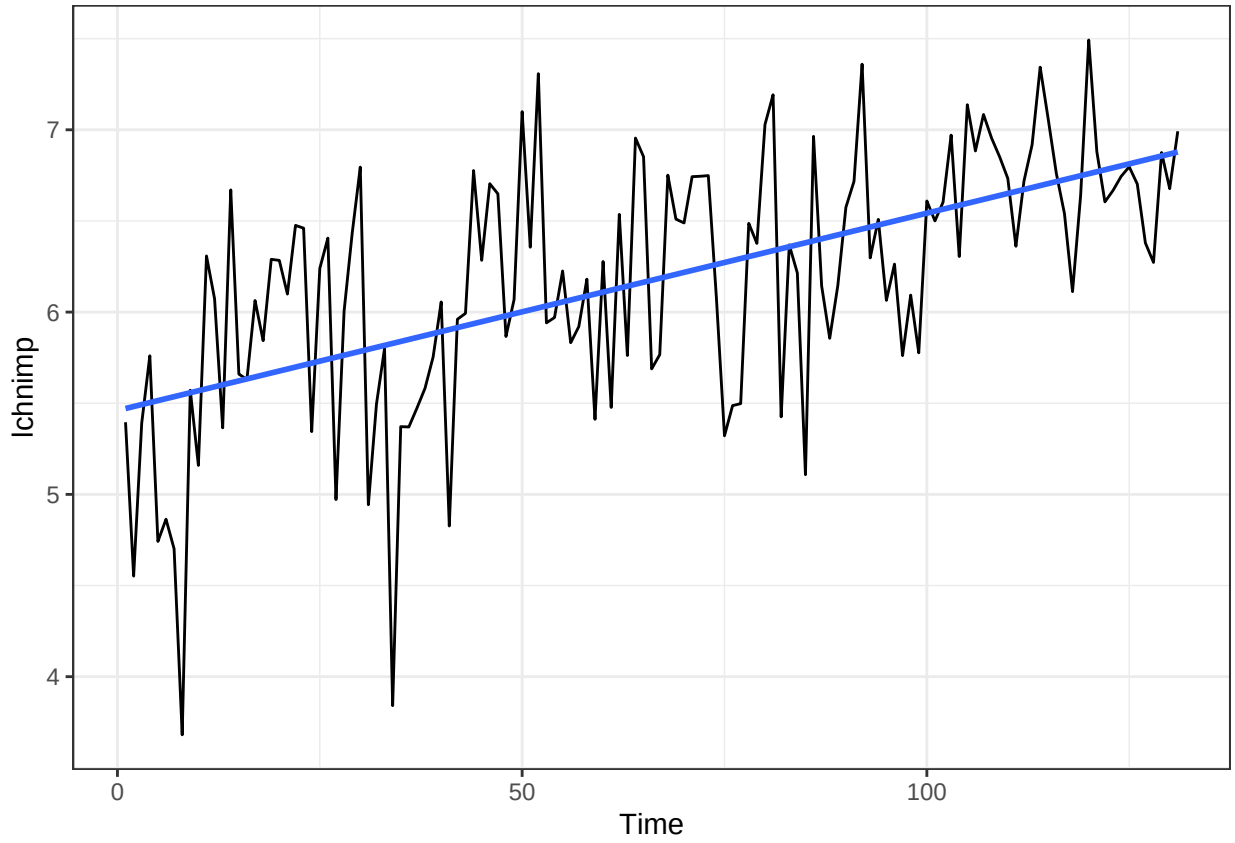
(i) Add a linear time trend to equation (10.22). Are any variables, other than the trend, statistically significant?

```
##
## Call:
## lm(formula = lchnimp ~ lchempi + lgas + lrtwex + befile6 + affile6 +
##     afdec6 + t, data = barium)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.94317 -0.31168  0.03172  0.36366  1.21218
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.367526  20.782165  -0.114  0.90949
## lchempi     -0.686236   1.239711  -0.554  0.58089
## lgas         0.465679   0.876178   0.531  0.59604
## lrtwex       0.078224   0.472440   0.166  0.86876
## befile6      0.090470   0.251289   0.360  0.71945
## affile6      0.097006   0.257313   0.377  0.70683
## afdec6      -0.351502   0.282542  -1.244  0.21584
## t            0.012706   0.003844   3.305  0.00124 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5748 on 123 degrees of freedom
## Multiple R-squared:  0.3616, Adjusted R-squared:  0.3252
## F-statistic: 9.951 on 7 and 123 DF,  p-value: 8.358e-10
```

Only the linear time trend (t) is a significant variable in the model that includes all variables from equation (10.22) and t. All other variables are insignificant.

(ii) In the equation estimated in part (i), test for joint significance of all variables except the time trend. What do you conclude?

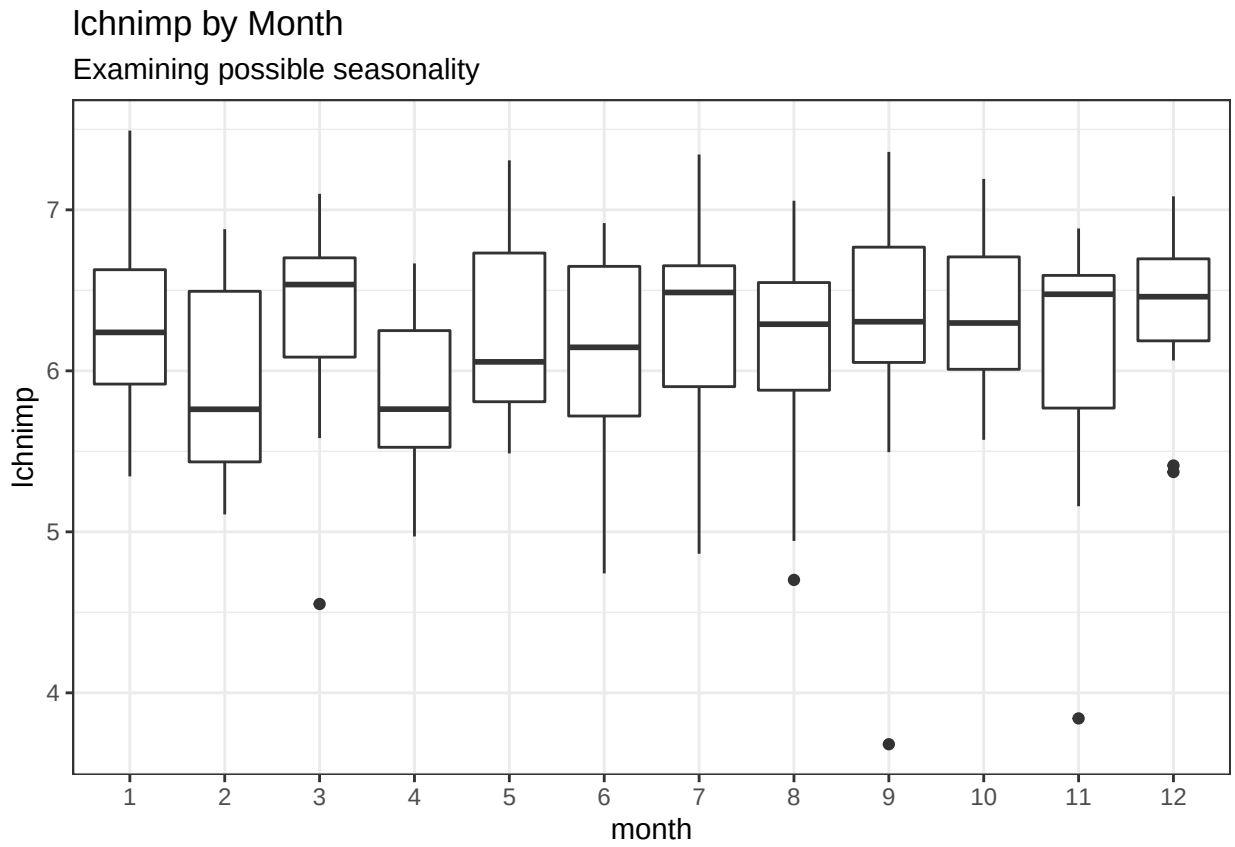
```
## Linear hypothesis test
##
## Hypothesis:
## lchempi = 0
## lgas = 0
## lrtwex = 0
## befile6 = 0
## affile6 = 0
## afdec6 = 0
##
## Model 1: restricted model
## Model 2: lchnimp ~ lchempi + lgas + lrtwex + befile6 + affile6 + afdec6 +
##     t
##
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    129 41.709
## 2    123 40.638  6     1.071 0.5402 0.7767
```



According to the joint significance hypothesis test on all variables other than t, the additional variables are not significant since the p-value (0.7767) indicates that the F-test is not significant at any reasonable level of alpha (0.10, 0.05, 0.01).

(iii) Add monthly dummy variables to this equation and test for seasonality. Does including the monthly dummies change any other estimates or their standard errors in important ways?

```
##
## Call:
## lm(formula = lchnimp ~ lchempi + lgas + lrtwex + befile6 + affile6 +
##     afdec6 + t + feb + mar + apr + may + jun + jul + aug + sep +
##     oct + nov + dec, data = barium)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.86054 -0.36284  0.02233  0.37155  1.09845
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 27.300074  31.397067   0.870  0.38643
## lchempi     -0.451656   1.271528  -0.355  0.72310
## lgas        -0.820624   1.345056  -0.610  0.54303
## lrtwex      -0.197141   0.529531  -0.372  0.71038
## befile6      0.164851   0.256979   0.641  0.52251
## affile6      0.153400   0.271986   0.564  0.57388
## afdec6      -0.295016   0.299428  -0.985  0.32662
## t            0.012339   0.003916   3.151  0.00209 **
## feb         -0.355415   0.293754  -1.210  0.22886
## mar          0.062566   0.254858   0.245  0.80652
## apr         -0.440615   0.258398  -1.705  0.09093 .
## may          0.031299   0.259200   0.121  0.90410
## jun         -0.200950   0.259213  -0.775  0.43984
## jul          0.011111   0.268378   0.041  0.96705
## aug         -0.127114   0.267792  -0.475  0.63594
## sep         -0.075193   0.258350  -0.291  0.77155
## oct          0.079763   0.257051   0.310  0.75691
## nov         -0.260303   0.253062  -1.029  0.30588
## dec          0.096533   0.261553   0.369  0.71277
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5788 on 112 degrees of freedom
## Multiple R-squared:  0.4106, Adjusted R-squared:  0.3158
## F-statistic: 4.334 on 18 and 112 DF,  p-value: 6.188e-07
```



There appears to be no seasonality. In examining the estimates and standard errors for the two models, the variables from the model without months appear unchanged when months are added to the overall model. From these analyses, we would conclude that months play no factor in predicting lchnimp, which means there is no significant seasonality to the time series data.

Appendix

```
library(ggplot2)
library(ggthemes)
library(dplyr)
library(wooldridge)
library(lmtest)
library(sandwich)
library(car)
tableb1 = read.csv("GDPdata.csv")[,-1]
tableb27 = read.csv("Unemployment.csv")[,-1]

USeconomy = cbind.data.frame(year = tableb1$Yr, gdpgrowth = tableb1$GDP,
                             ue = tableb27$Unemployment.Rate)
print(head(USeconomy, 10))
USeconomy$min = ifelse(USeconomy$gdpgrowth == min(USeconomy$gdpgrowth), "Min", "N")
USeconomy %>% ggplot(aes(x = year, y = gdpgrowth)) + geom_line() + theme_bw() +
  geom_point(aes(color = min)) + theme(legend.position = "none") +
  scale_color_manual(values = c("red3", "black"))
USeconomy$minUE = ifelse(USeconomy$ue == min(USeconomy$ue), "Min", "N")
USeconomy %>% ggplot(aes(x = year, y = ue)) + geom_line() + theme_bw() +
  geom_point(aes(color = minUE)) + theme(legend.position = "none") +
  scale_color_manual(values = c("red3", "black"))
cat("\\newpage")
USeconomy$t <- seq(1, nrow(USeconomy))
USeconomy = USeconomy %>% select(t, year, gdpgrowth, ue)
model <- lm(gdpgrowth ~ t, data = USeconomy)
summary(model)

USeconomy %>% ggplot(aes(x = t, y = gdpgrowth)) + geom_line() + theme_bw() +
  theme(legend.position = "none") +
  scale_color_manual(values = c("red3", "black")) + geom_smooth(method = "lm",
                                                                formula = "y~x",
                                                                se = F)

X = USeconomy$gdpgrowth
Xlag = X %>% lag(n = 1)
lag.df = cbind.data.frame(gdpgrowth = X, FirstDiff = Xlag)
lag.df %>% ggplot(aes(x = gdpgrowth, y = FirstDiff)) + geom_point() +
  geom_smooth(method = "lm", formula = "y~x", se = F) + theme_bw() +
  scale_y_continuous("First Lag of gdpgrowth") +
  ggtitle("GDP Growth vs. First Lag of GDP Growth")
par(mfrow = c(1, 1))
gdpgrowth = X
acf(gdpgrowth, lag.max = 16)
acf(gdpgrowth, lag.max = 16, plot = F)
cat("\\newpage")
data(barium)
barium = barium %>% select(t, everything())
lm.model2 = lm(lchnimp ~ lchempi + lgas + lrtwex + befile6 + affile6 + afdec6 + t,
               data = barium)
summary(lm.model2)
cat("\\newpage")
```

```r
linearHypothesis(lm.model2, c("lchempi=0", "lgas=0", "lrtwex=0",
                              "befile6=0", "affile6=0", "afdec6=0"))
barium %>% ggplot(aes(x = t, y = lchnimp)) + geom_line() +
  geom_smooth(method = "lm", formula = "y~x", se = F) + theme_bw() +
  scale_x_continuous("Time")

lm.model3 = lm(lchnimp ~ lchempi + lgas + lrtwex + befile6 + affile6 + afdec6 + t +
               feb + mar + apr + may + jun + jul + aug + sep + oct +
                 nov + dec, data = barium)
summary(lm.model3)

for (i in 1:nrow(barium)){
  if (barium$feb[i] == 1){
    barium$month[i] = "2"
  }else if (barium$mar[i] == 1){
    barium$month[i] = "3"
  }else if (barium$apr[i] == 1){
    barium$month[i] = "4"
  }else if (barium$may[i] == 1){
    barium$month[i] = "5"
  }else if (barium$jun[i] == 1){
    barium$month[i] = "6"
  }else if (barium$jul[i] == 1){
    barium$month[i] = "7"
  }else if (barium$aug[i] == 1){
    barium$month[i] = "8"
  }else if (barium$sep[i] == 1){
    barium$month[i] = "9"
  }else if (barium$oct[i] == 1){
    barium$month[i] = "10"
  }else if (barium$nov[i] == 1){
    barium$month[i] = "11"
  }else if (barium$dec[i] == 1){
    barium$month[i] = "12"
  }else{
    barium$month[i] = "1"
  }
}

barium$month = factor(barium$month, levels = c("1", "2", "3", "4", "5", "6", "7", "8",
                                               "9", "10","11", "12"))
barium %>% ggplot(aes(x = month, y = lchnimp)) + geom_boxplot() + theme_bw() +
  ggtitle("lchnimp by Month", subtitle = "Examining possible seasonality")
cat("\\newpage")
```