Homework Assignment #2

Quilvio Hernandez

4/15/2020

Worked with Aman Singh.

Question 1

Obtain time series data about the U.S. economy and summarize. For this question, you will need to first download the DGP growth data and the civilian unemployment data from 2020 Economic Report of the President (Table B-1 and B-27) at http://www.gpo.gov/fdsys/browse/collection.action?collectionCode=ERP (http://www.gpo.gov/fdsys/browse/collection.action?collectionCode=ERP). For example, for the GDP growth data, download Table B-1 in xls format (click the above link, go to 2020, scroll down to find Table B-1, and then load the numbers in the first two columns of the table into R. For the unemployment data, you will need the numbers in the first two columns of Table B-27.

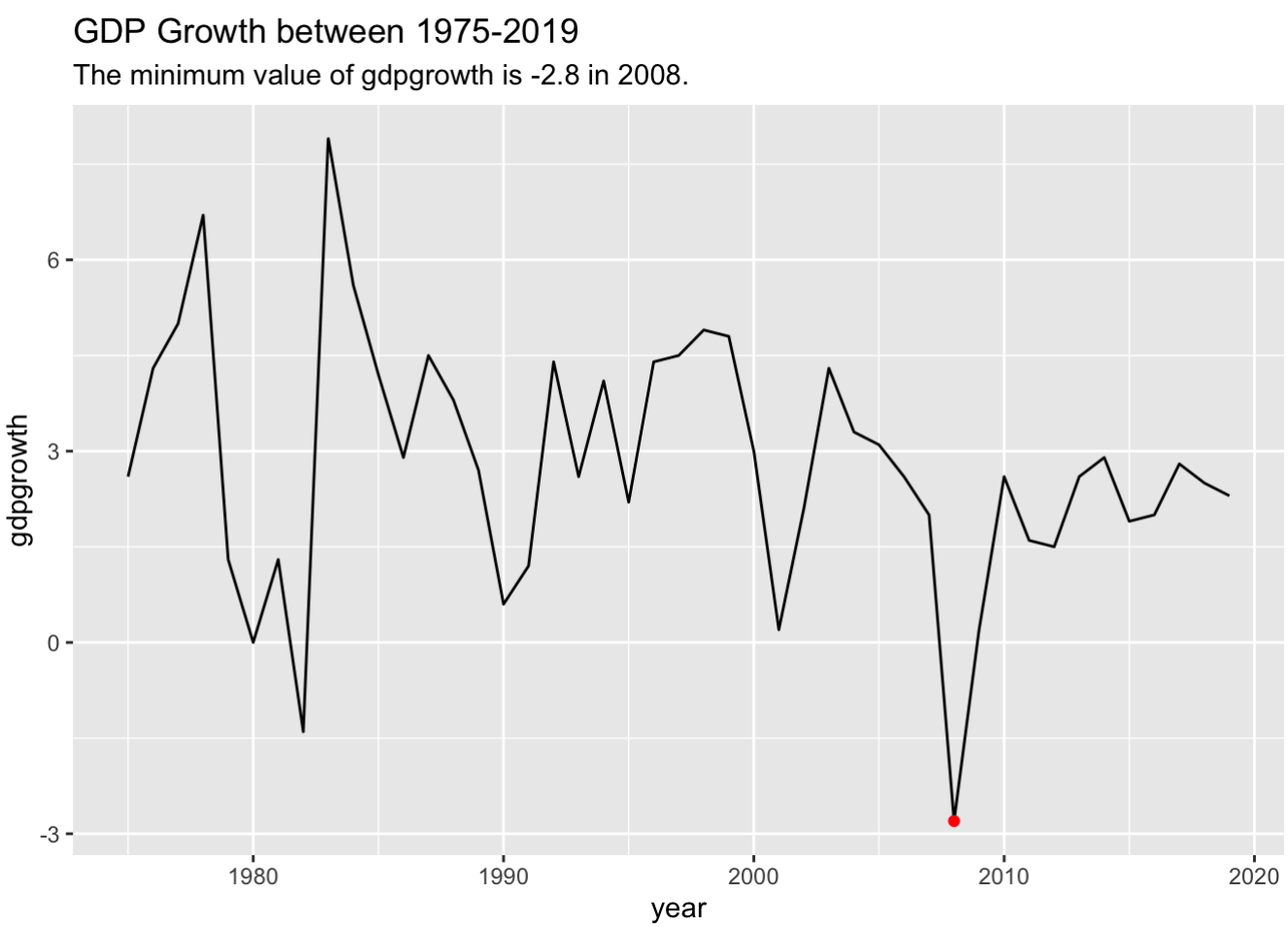
a. Create a data frame named “USeconomy” which includes three variables, “year” is 1975-2019, “gdpgrowth” is U.S. GDP for each year obtained from Table B-1, and “ue” is U.S. unemployment rate for each year obtained from Table B-27.

```
rm(list = ls())
library(tidyverse)
library(readxl)
setwd("//Users//quilviohernandez//Desktop//Spring2020//ECN 190//Homework//Homework 2")
USeconomy <- read_excel("ERP-2020-table1.xls", col_names = FALSE, skip = 12) %>%
  slice(1:45) %>%
  select("...1", "...2") %>%
  rename(year = ...1, gdpgrowth = ...2)
USeconomy$year <- USeconomy %>%
  pull(year) %>%
  substr(., 1, 4) %>%
  as.numeric(.)
unemployment <- read_excel("ERP-2020-table27.xls", col_names = FALSE, skip = 4) %>%
  rename(ue = ...2)
USeconomy$ue <- unemployment$ue[1:45]
head(USeconomy)
```

```
## # A tibble: 6 x 3
##    year gdpgrowth    ue
##   <dbl>     <dbl> <dbl>
## 1  1975       2.6   8.5
## 2  1976       4.3   7.7
## 3  1977       5     7.1
## 4  1978       6.7   6.1
## 5  1979       1.3   5.8
## 6  1980       0     7.1
```

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?

```
USeconomy %>%
  mutate(color = (min(gdpgrowth) == gdpgrowth)) %>%
  ggplot(aes(x =year, y = gdpgrowth)) +
  geom_line() +
  geom_point(aes(color = color), show.legend = FALSE) +
  scale_color_manual(values = c(NA, "red")) +
  labs(title = "GDP Growth between 1975-2019", subtitle = "The minimum value of gdpgrowt
h is -2.8 in 2008.")
```




```
min(USeconomy$gdpgrowth)
```

```
## [1] -2.8
```

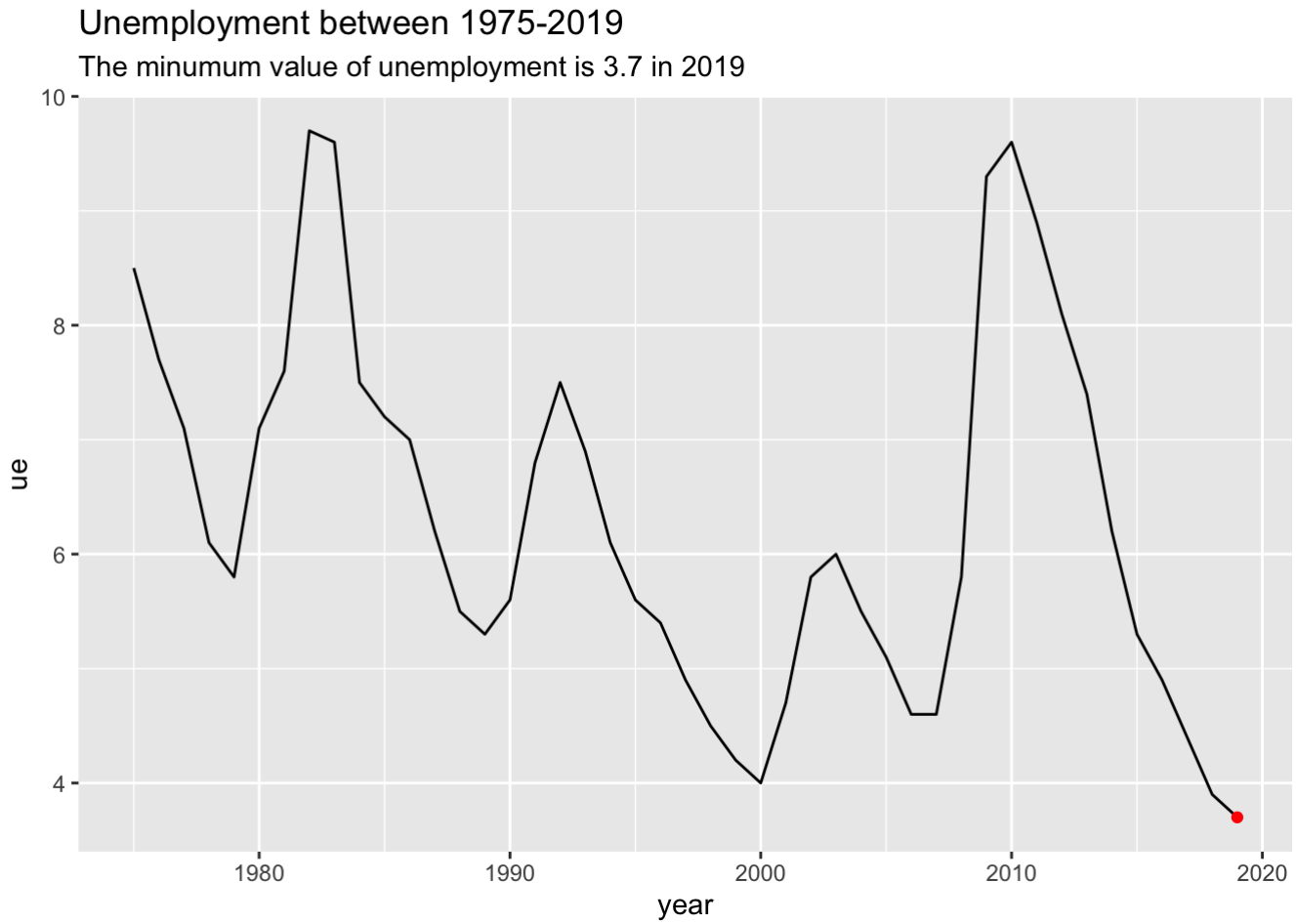
```
USeconomy$year[which.min(USeconomy$gdpgrowth)]
```

```
## [1] 2008
```

The minimum value of gdpgrowth is -2.8 which occured 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?

```
USeconomy %>%
  mutate(color = (min(ue) == ue)) %>%
  ggplot(aes(x =year, y = ue)) +
  geom_line() +
  geom_point(aes(color = color), show.legend = FALSE) +
  scale_color_manual(values = c(NA, "red")) +
  labs(title = "Unemployment between 1975-2019", subtitle = "The minumum value of unempl
oyment is 3.7 in 2019")
```




```
min(USeconomy$ue)
```

```
## [1] 3.7
```

```
USeconomy$year[which.min(USeconomy$ue)]
```

```
## [1] 2019
```

The minimum value of gdpgrowth is 3.7 which occured 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?

```
USeconomy <- USeconomy %>%
  mutate(t = year - 1974)
summary(lm(gdpgrowth ~ t, data = USeconomy))
```

```
##
## 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
```

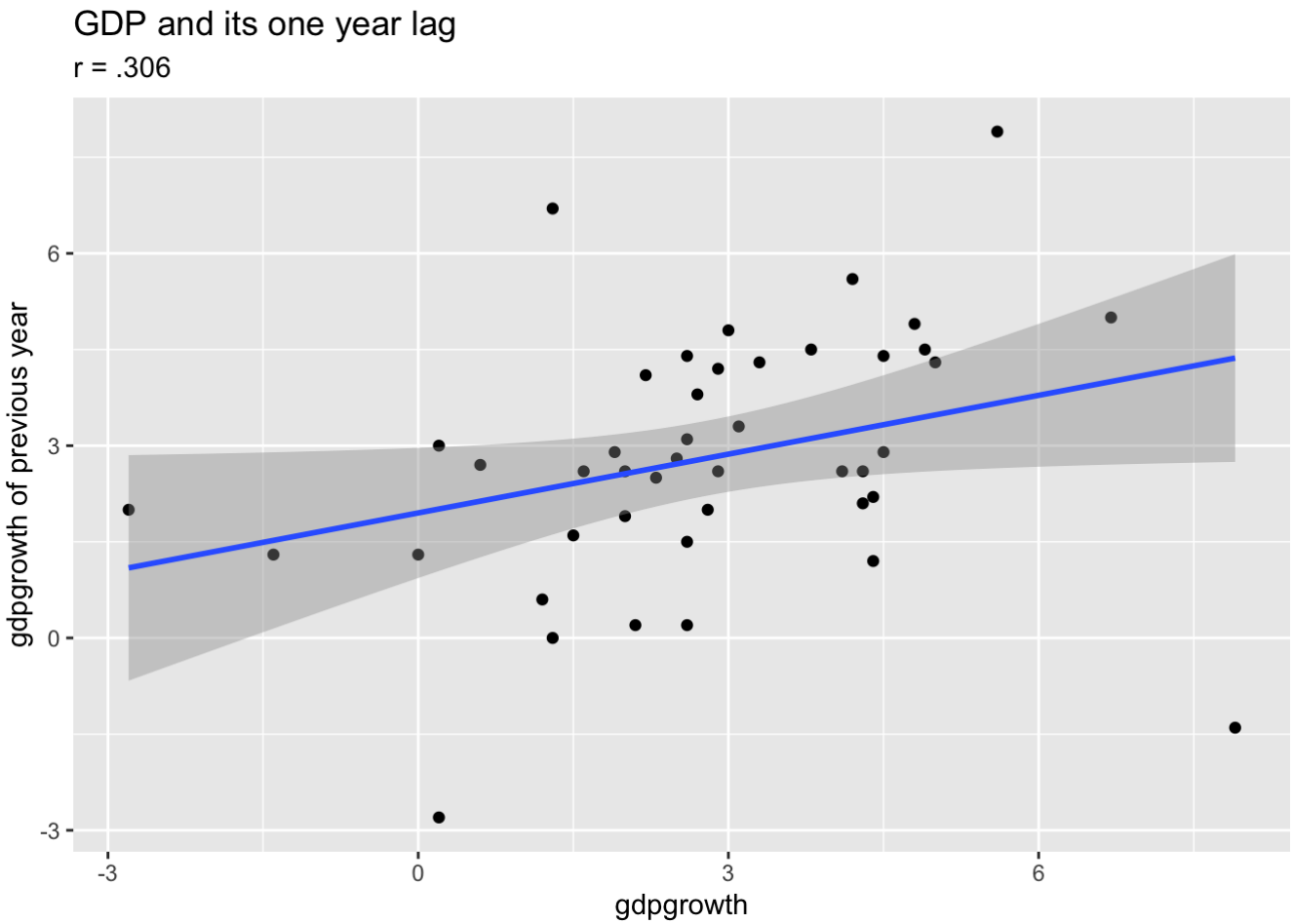
We can interpret the intercept as the predicted gdpgrowth (3.7) for t=0 (1974). We can interpret the slope coefficient for **t** as the average change in gdpgrowth per year. For each additional year (increase in **t** by one unit), we would expect to see a decline in gdpgrowth by .04, on average.

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

```
USeconomy_lag <- USeconomy %>%
  mutate(lag1 = lag(gdpgrowth))
cor(USeconomy_lag$gdpgrowth, USeconomy_lag$lag1, use = "complete.obs")
```

```
## [1] 0.3061333
```

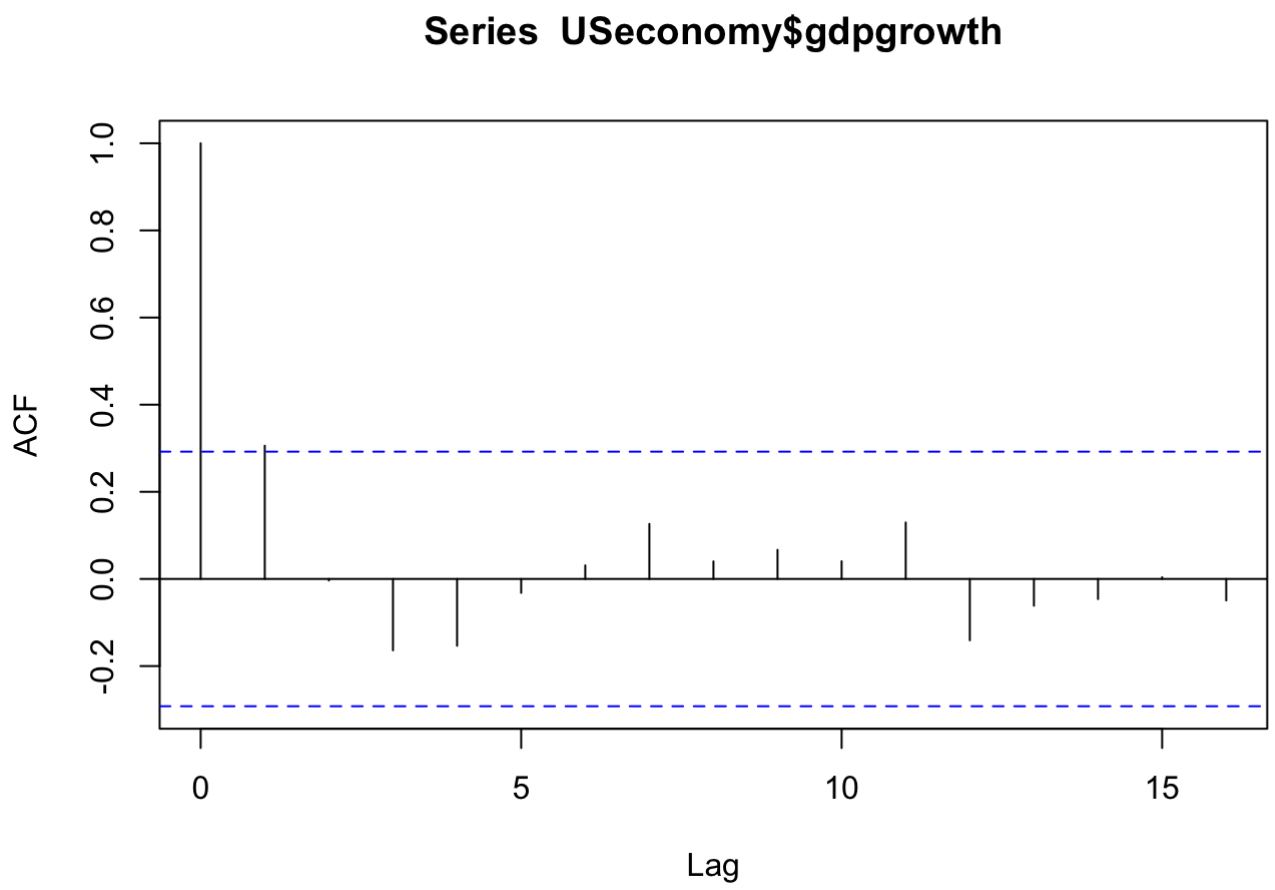
```
USeconomy_lag %>%
  ggplot(aes(x = gdpgrowth, y = lag1), colour = "blue") +
  geom_point() +
  labs(title = "GDP and its one year lag", y = "gdpgrowth of previous year", subtitle =
"r = .306") +
  geom_smooth(method = "lm")
```



There is a moderately weak (r=.306) positive relationship.

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?

```
acf(USeconomy$gdpgrowth)
```




```
acf(USeconomy$gdpgrowth, plot = FALSE)
```

```
##
## Autocorrelations of series 'USeconomy$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
```

gdpgrowth appears to be stationary since our autocorrelation stays within our confidence intervals.

Question 2

Wooldridge C10.2 Use the data in BARIUM.RAW for this exercise.

```
rm(list=ls())
library(haven)
barium <- read_dta("//Users//quilviohernandez//Desktop//Spring2020//ECN 190//Data//BARIUM.dta")
names(barium)
```

```
##  [1] "chnimp"   "bchlimp"  "befile6"  "affile6"  "afdec6"   "befile12"
##  [7] "affile12" "afdec12"  "chempi"   "gas"      "rtwex"    "spr"
## [13] "sum"      "fall"     "lchnimp"  "lgas"     "lrtwex"   "lchempi"
## [19] "t"        "feb"      "mar"      "apr"      "may"      "jun"
## [25] "jul"      "aug"      "sep"      "oct"      "nov"      "dec"
## [31] "percchn"
```

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

```
summary(lm(lchnimp ~ chempi + lgas + lrtwex + befile6 + affile6 + afdec6 + t, data = bar
ium))
```

```
##
## Call:
## lm(formula = lchnimp ~ chempi + lgas + lrtwex + befile6 + affile6 +
##     afdec6 + t, data = barium)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.9433 -0.3233  0.0372  0.3538  1.1964
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.449476  20.574819  -0.265  0.79156
## chempi      -0.007675   0.010457  -0.734  0.46437
## lgas         0.509496   0.879473   0.579  0.56343
## lrtwex       0.006867   0.492782   0.014  0.98890
## befile6      0.089250   0.251051   0.356  0.72282
## affile6      0.102729   0.256888   0.400  0.68992
## afdec6      -0.337303   0.283068  -1.192  0.23571
## t            0.013530   0.004073   3.322  0.00118 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5743 on 123 degrees of freedom
## Multiple R-squared:  0.3628, Adjusted R-squared:  0.3265
## F-statistic:    10 on 7 and 123 DF,  p-value: 7.503e-10
```

No. The only statistically significant variable is the trend. This can be seen by the t values.

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

```
part1 <- lm(lchnimp ~ chempi + lgas + lrtwex + befile6 + affile6 + afdec6 + t, data = ba
rium)
qf(0.95, 6, 123)
```

```
## [1] 2.173112
```

```
library(car)
```

```
## Loading required package: carData
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```
## The following object is masked from 'package:purrr':
##
##     some
```

```
linearHypothesis(part1,c("chempi=0","lgas=0","lrtwex=0","befile6=0","affile6=0","afdec6=
0"))
```

```
## Linear hypothesis test
##
## Hypothesis:
## chempi = 0
## lgas = 0
## lrtwex = 0
## befile6 = 0
## affile6 = 0
## afdec6 = 0
##
## Model 1: restricted model
## Model 2: lchnimp ~ chempi + lgas + lrtwex + befile6 + affile6 + afdec6 +
##     t
##
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    129 41.709
## 2    123 40.562  6    1.1474 0.5799 0.7458
```

With a p-value of .746, greater than any significance level, we fail to reject the null hypothesis and conclude the explanatory variables, besides time trend, are jointly insignificant. We find that besides the time trend (seasonality), none of the other variables have good predictive power. We conclude the original analysis did not include a statistically significant variable.

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

```
summary(lm(lchnimp ~ chempi + lgas + lrtwex + befile6 + affile6 + afdec6 + t +
             feb + mar + apr + may + jun + jul + aug + sep + oct + nov + dec, data = bar
ium))
```

```
##
## Call:
## lm(formula = lchnimp ~ chempi + 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
## -1.85775 -0.36687  0.03422  0.36966  1.08613
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 24.684122  31.758926   0.777  0.43866
## chempi      -0.005458   0.010755  -0.508  0.61279
## lgas        -0.762972   1.350495  -0.565  0.57323
## lrtwex      -0.251166   0.547251  -0.459  0.64715
## befile6      0.163784   0.256834   0.638  0.52497
## affile6      0.158978   0.271724   0.585  0.55968
## afdec6      -0.282421   0.300249  -0.941  0.34892
## t            0.013025   0.004160   3.131  0.00222 **
## feb         -0.348298   0.294076  -1.184  0.23877
## mar          0.063504   0.254722   0.249  0.80358
## apr         -0.438208   0.258331  -1.696  0.09261 .
## may          0.031216   0.259023   0.121  0.90429
## jun         -0.199855   0.259049  -0.771  0.44204
## jul          0.011494   0.268177   0.043  0.96589
## aug         -0.125417   0.267672  -0.469  0.64030
## sep         -0.070569   0.258467  -0.273  0.78534
## oct          0.082381   0.256998   0.321  0.74915
## nov         -0.259676   0.252897  -1.027  0.30673
## dec          0.094889   0.261316   0.363  0.71720
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5784 on 112 degrees of freedom
## Multiple R-squared:  0.4113, Adjusted R-squared:  0.3166
## F-statistic: 4.346 on 18 and 112 DF,  p-value: 5.869e-07
```

```
part2 <- lm(lchnimp ~ chempi + lgas + lrtwex + befile6 + affile6 + afdec6 + t +
             feb + mar + apr + may + jun + jul + aug + sep + oct + nov + dec, data = bar
ium)
linearHypothesis(part2,c("feb=0", "mar=0", "apr=0", "may=0", "jun=0", "jul=0", "aug=0",
"sep=0", "oct=0", "nov=0", "dec=0"))
```

```
## Linear hypothesis test
##
## Hypothesis:
## feb = 0
## mar = 0
## apr = 0
## may = 0
## jun = 0
## jul = 0
## aug = 0
## sep = 0
## oct = 0
## nov = 0
## dec = 0
##
## Model 1: restricted model
## Model 2: lchnimp ~ chempi + lgas + lrtwex + befile6 + affile6 + afdec6 +
##     t + feb + mar + apr + may + jun + jul + aug + sep + oct +
##     nov + dec
##
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    123 40.562
## 2    112 37.475 11    3.0869 0.8387  0.602
```

Including the monthly dummy variables does NOT change any estimate or standard error in important ways. This fact can further be seen by the results of our linearHypothesis command where we find a p-value of .6 and conclude the monthly dummy variables are jointly insignificant.