HUDM5126 Linear Models and Regression Analysis Homework 11

Yifei Dong

11/11/2020

0. Data Prepartation

```r
setwd("~/Documents/Teachers College/Linear Models and Regression/Week 11/hw11/hw11_R")
getwd()
```

```
## [1] "/Users/yifei/Documents/Teachers College/Linear Models and Regression/Week 11/hw11/hw11_R"
```

```r
# Load packages
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(lmtest)
```

```
## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```r
library(orcutt)
```

1. KNNL 12.13

```r
# Load dataset
data <- read.table(paste("http://users.stat.ufl.edu/~rrandles/sta4210/Rclassnotes/data/",
                         "textdatasets/KutnerData",
                         "/Chapter%2012%20Data%20Sets/CH12PR13.txt", sep = ""))
```

```r
# Rename variables
data <- data %>%
  dplyr::select("Y" = V1,
```

```r
                  "X" = V2)
head(data, 5)
```

```
##       Y     X
## 1 220.4 2.521
## 2 203.9 2.171
## 3 207.2 2.234
## 4 221.9 2.524
## 5 211.3 2.305
```

a) Fit a simple linear regression model by ordinal least squares and obtain the residuals. Also obtain $s\{b_0\}$ and $s\{b_1\}$.

```r
attach(data)
reg <- lm(Y ~ X)
summary(reg)
```

```
##
## Call:
## lm(formula = Y ~ X)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.55515 -0.23700  0.05229  0.56250  0.80657
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  93.6865     0.8229   113.8   <2e-16 ***
## X            50.8801     0.2634   193.1   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.631 on 18 degrees of freedom
## Multiple R-squared:  0.9995, Adjusted R-squared:  0.9995
## F-statistic: 3.73e+04 on 1 and 18 DF,  p-value: < 2.2e-16
```

```r
# SE(b0)
summary(reg)$coefficient[3]
```

```
## [1] 0.8229241
```

```r
# SE(b1)
summary(reg)$coefficient[4]
```

```
## [1] 0.2634323
```

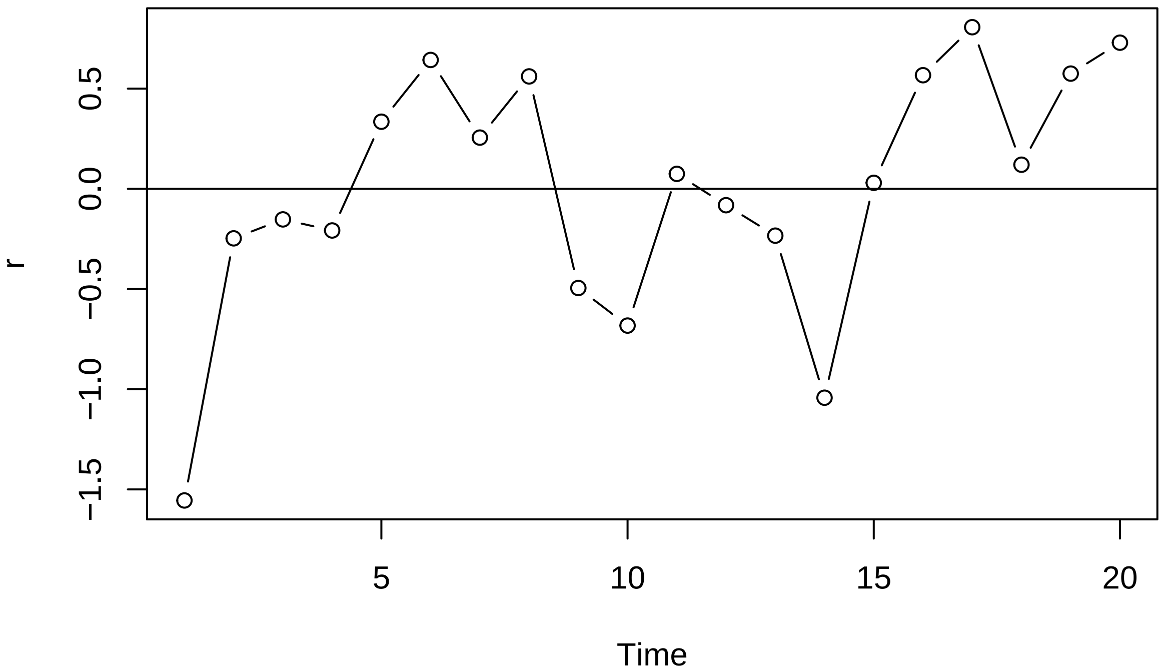
The estimated regression equation is:

$$\widehat{Y} = 93.6865 + 50.8801X$$

$s\{b_0\} = 0.8229$ and $s\{b_1\} = 0.2634$

b) Plot the residuals against time and explain whether you find any evidence of positive autocorrelation.

```r
r <- residuals(reg)
ts.plot(r, type = "b")
abline(h = 0)
```



I do observe a "sine" type of curve, which indicates a potential autocorrelation problem.

c) Conduct a formal test for positive autocorrelation using $\alpha = .01$. State the alternatives, decision rule, and conclusion. Is the residual analysis in part (b) in accord with the test result?

```
dwtest(Y ~ X)
```

```
##
##  Durbin-Watson test
##
## data:  Y ~ X
## DW = 0.97374, p-value = 0.002891
## alternative hypothesis: true autocorrelation is greater than 0
```

$$H_0 : \rho = 0$$

$$H_\alpha : \rho > 0$$

D-W Statistic is 0.97374 and p-value = 0.002891, which is smaller than 0.01. Therefore, we should reject H_0 and conclude that there is a strong evidence of positive autocorrelation. The residual analysis in part (b) is in accord with the test result.

2. KNNL 12.14

a) Obtain a point estimate of the autocorrelation parameter. How well does the approximate relationship (12.25) hold here between the point estimate and the Durbin-Waton test statistic?

```
# Obtain a point estimate of the autocorrelation parameter
# Manual estimation of rho
numerator = 0
n = nrow(data)
for (i in 2:n) numerator = numerator + r[i]*r[i-1]
rho = numerator/sum(r[1:(n-1)]^2)
rho
```

```
##        2
## 0.331904
```

Therefore, $\rho = 0.331904$

Recall 12.25, there exists an approximate relation between the Durbin-Watson test statistic D in (12.14) and the estimated autocorrelation parameter r in (12.22):

$$D \approx 2(1 - r) \tag{1}$$

```
2*(1-0.331904)
```

```
## [1] 1.336192
```

Durbin-Waton test statistic is 0.97374, which is smaller than 1.336192 using the estimated autocorrelation parameter.

b) Use one iteration to obtain the estimate b_0' and b_1' of the regression coefficients β_0' and β_1' in transformed model (12.17) and state the estimated function. Also obtain $s\{b_0'\}$ and $s\{b_1'\}$.

```
# Compute transformed variables:
n = nrow(data)
Yprime = 1:n
for (i in 2:n) Yprime[i] = Y[i]- rho*Y[i-1]
Yprime = Yprime[-1]

Xprime = 1:n
for (i in 2:n) Xprime[i] = X[i]- rho*X[i-1]
Xprime = Xprime[-1]

Regprime = lm(Yprime ~ Xprime)
summary(Regprime)
```

```
##
## Call:
## lm(formula = Yprime ~ Xprime)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.95813 -0.29553 -0.02312  0.34451  0.60490
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  63.3840     0.5592   113.4   <2e-16 ***
```

```
## Xprime       50.5470     0.2622   192.8   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4546 on 17 degrees of freedom
## Multiple R-squared:  0.9995, Adjusted R-squared:  0.9995
## F-statistic: 3.715e+04 on 1 and 17 DF,  p-value: < 2.2e-16
```

```r
# SE(b0)
summary(Regprime)$coefficient[3]
```

```
## [1] 0.5591553
```

```r
# SE(b1)
summary(Regprime)$coefficient[4]
```

```
## [1] 0.2622327
```

$b_0' = 63.3840$ and $b_1' = 50.5470$

The estimated function is:

$$\widehat{Y'} = 63.3840 + 50.5470X'$$

$s\{b_0'\} = 0.5592$ and $s\{b_1'\} = 0.2622$

c) Test whether any positive autocorrelation remains after the first iteration using $\alpha = .01$. State the alternatives, decision rule, and conclusion.

```r
dwtest(Yprime ~ Xprime)
```

```
##
##  Durbin-Watson test
##
## data:  Yprime ~ Xprime
## DW = 1.7612, p-value = 0.2337
## alternative hypothesis: true autocorrelation is greater than 0
```

$$H_0 : \rho = 0$$

$$H_\alpha : \rho > 0$$

D-W statistic is 1.7612 and p-value = 0.2337, which is greater than $\alpha = .01$. Therefore, we could not reject H_0 and conclude that no autocorrelation remains after the first iteration.

d) Restate the estimated regression function obtained in part (b) in terms of the original variables. Also obtain $s\{b_0\}$ and $s\{b_1\}$. Compare the estimated regression coefficients obtained with the Cochrane-Orcutt procedure and their estimated standard deviation with those obtained with ordinary least squares in Problem 12.13a.

```r
# Back to original model:
(b0 = Regprime$coef[1]/(1-rho))
```

```
## (Intercept)
##    94.87257
```

```r
(b1 = Regprime$coef[2])
```

```
##   Xprime
## 50.54696
```

```r
cochrane.orcutt(reg, convergence = 0) # results are the same
```

```
## Cochrane-orcutt estimation for first order autocorrelation
##
## Call:
## lm(formula = Y ~ X)
##
##  number of interaction: 1
##  rho 0.331904
##
## Durbin-Watson statistic
## (original):    0.97374 , p-value: 2.891e-03
## (transformed): 1.76117 , p-value: 2.337e-01
##
##  coefficients:
## (Intercept)           X
##    94.87257    50.54696
```

```r
# SE(b0)
(summary(Regprime)$coefficient[3]/(1-rho))
```

```
##         2
## 0.8369385
```

```r
# SE(b1)
(summary(Regprime)$coefficient[4])
```

```
## [1] 0.2622327
```

```r
# Compare to OLS
# coefficients
(summary(reg)$coefficient[1])
```

```
## [1] 93.68648
```

```r
(summary(reg)$coefficient[2])
```

```
## [1] 50.88007
```

```r
# OLS Standard deviation
(summary(reg)$coefficient[3])
```

```
## [1] 0.8229241
```

```r
(summary(reg)$coefficient[4])
```

```
## [1] 0.2634323
```

Back transform to:

$$\widehat{Y} = 94.87257 + 50.54696X$$

$s\{b_0\} = 0.8370$ and $s\{b_1\} = 0.2622$

Compare those estimations obtained with OLS, the coefficient of intercept is higher than that of OLS, but the coefficient of β_1 is smaller than OLS estimates. The standard deviation estimated with the Cochrane-Orcutt procedure is very close to OLS. However, it is necessary to point out that $s\{b_1\} = 0.2622 < 0.2634$, which is smaller than OLS estimates.

e) Based on the results in parts (c) and (d), does the Cochrane-Orcutt procedure appear to have been effective here?

We know that the estimated standard deviations $s\{b_k\}$ calculated according to ordinary least squares may seriously underestimate the true standard deviation $\sigma\{b_k\}$ when positive autocorrelation is present. In this case, with the Cochrane-Orcutt procedure, the estiamted standard deviation of $s\{b_1\}$ is smaller than than OLS estimate. Therefore, the Cochrane-Orcutt approach does not always work properly. A major reason is that when the error terms are positively related. the estimated r tends to **understimate** the autocorrelation parameter ρ (as we see in part a). When this bias is serious, it can significantly reduce the effectiveness of the Cochrane-Orcutt approach.

f) Staff time in month 21 is expected to be 3.625 thousand hours. Predict the amount of billings in constant dollars for month 21, using a 99 percent prediction interval. Interpret your interval (skip the CI).

```
# Forecasting
# Obtain last residual
e20 = Y[20]-(b0 + b1*X[20])

# X21 = 3.625
# Obtain Y.hat
Y.hat21 = b0 + b1*3.625

# Adjust with correlated residual
(F21 = Y.hat21 + rho*e20)
```

```
## (Intercept)
##    278.3537
```

The predicted amount of billings in thousands of constant dollars for month 21 is 278.3537.

g) Estimate β_1 with a 99 percent confidence interval. Interpret your interval.

```
# SE(b1)
(se1 <- (summary(Regprime)$coefficient[4]))
```

```
## [1] 0.2622327
```

```
confint(Regprime, level = 0.99)
```

```
##                0.5 %   99.5 %
## (Intercept) 61.76342 65.00455
## Xprime      49.78695 51.30697
```

Therefore,

$$49.78695 \leq \beta_1 \leq 51.30697$$

We predict with approximately 99 percent confidence that the true β_1 will be between 49.78695 and 51.30697, using Cochrane-Orcutt approach to eliminate the problem of autocorrelated errors.