Advanced Economies

Muhammad Tamjid Rahman

Loading R packages

Part 1)

01)

```
data <-read.table("https://www.ssc.wisc.edu/~bhansen/econometrics/cps09mar.txt", header = FALSE, sep = 
colnames(data) <-c("age","female","hisp","edu","earn","hours","week","union","uncov","region","race","ma


edu<- data$edu>=12
data1<-data[edu,]

tot <- data1$hours*data1$week
wage <- data1$earn/tot
Y<-data1$log_wage <-as.matrix(log(wage))
data1$exp <- data1$age-data1$edu-6
data1$exp2 <- (data1$exp^2)/100

data1$female_union <- ifelse(data1$female == 1 & data1$union ==1, 1, 0)
data1$male_union <- ifelse(data1$female == 0 & data1$union ==1, 1, 0)
data1$female_married <- ifelse(data1$female == 1 & data1$marital <=3, 1, 0)
data1$male_married <- ifelse(data1$female == 0 & data1$marital <=3, 1, 0)
data1$female_exmarried <- ifelse(data1$female == 1 & (data1$marital==4|data1$marital==5|data1$marital==
data1$male_exmarried <- ifelse(data1$female == 0 & (data1$marital==4|data1$marital==5|data1$marital==6 
data1$black<-ifelse(data1$race == 2, 1, 0)
data1$Am_Ind<-ifelse(data1$race == 3, 1, 0)
data1$Asian<-ifelse(data1$race == 4, 1, 0)
data1$mixed<-ifelse(data1$race >=6, 1, 0)

x1<-data1[-c(1,5:13)]

x2<-x1[,c('edu','exp','exp2','female','female_union',
        'male_union','female_married','male_married',
        'female_exmarried','male_exmarried','hisp','black','Am_Ind',
        'Asian','mixed')]

intercept <- rep(1, nrow(x2))

X <- as.matrix(cbind(x2,intercept))
xx<-t(X)%*%X
xxi<-solve(xx)
xy<-t(X)%*%Y
```

```r
beta<-xxi%*%xy

n<-nrow(Y)
k<-ncol(X)

### Sigma
e <- Y-X%*%beta
sigma2<-as.numeric(1/(n-k)*(t(e)%*%e))
sigma<-as.numeric( sqrt(sigma2))

### S(beta)
u1 <- X*(e%*%matrix(1,1,k))

v1 <- xxi %*% (t(u1)%*%u1) %*% xxi

s1 <- sqrt(diag(v1))
```

Using OLS the estimator was calculated. The formula for $b\hat{e}ta$ is,

$$\hat{\beta} = (X'X)^{-1}(XY)$$

Then $\hat{\sigma}$ was calcilated by,

$$\hat{\sigma} = \sqrt{\frac{\sum e'e}{n-k}}$$

e=Y-X$\hat{\beta}$

n= number of row of Y

k= number of column of X

Standard errors of the estimators are heteroskedasticity-consistent and calculated by Horn-Horn-Duncan formula.

$$s(\hat{\beta}) = \sqrt{diag(X'X)^{-1}(\sum_{i=1}^{n}(1-h_{ii})^{-1}x_i x_i' \hat{e_i}^2))(X'X)^{-1}}$$

Where, $h_{ii} = x_i(X'X)x_i'$

```r
cbind(beta,s1)
```

```
##                                   s1
## edu                0.11669830 0.0012820059
## exp                0.03315647 0.0009517888
## exp2              -0.05643068 0.0020751922
## female            -0.09825725 0.0110300255
## female_union       0.02285629 0.0196045724
## male_union         0.09518739 0.0202969553
## female_married     0.01615988 0.0095395061
## male_married       0.21112049 0.0096989277
## female_exmarried -0.00642132 0.0118587152
## male_exmarried     0.08289541 0.0145582227
## hisp              -0.10813640 0.0081195128
## black             -0.09553060 0.0083352591
## Am_Ind            -0.13743914 0.0263682529
```

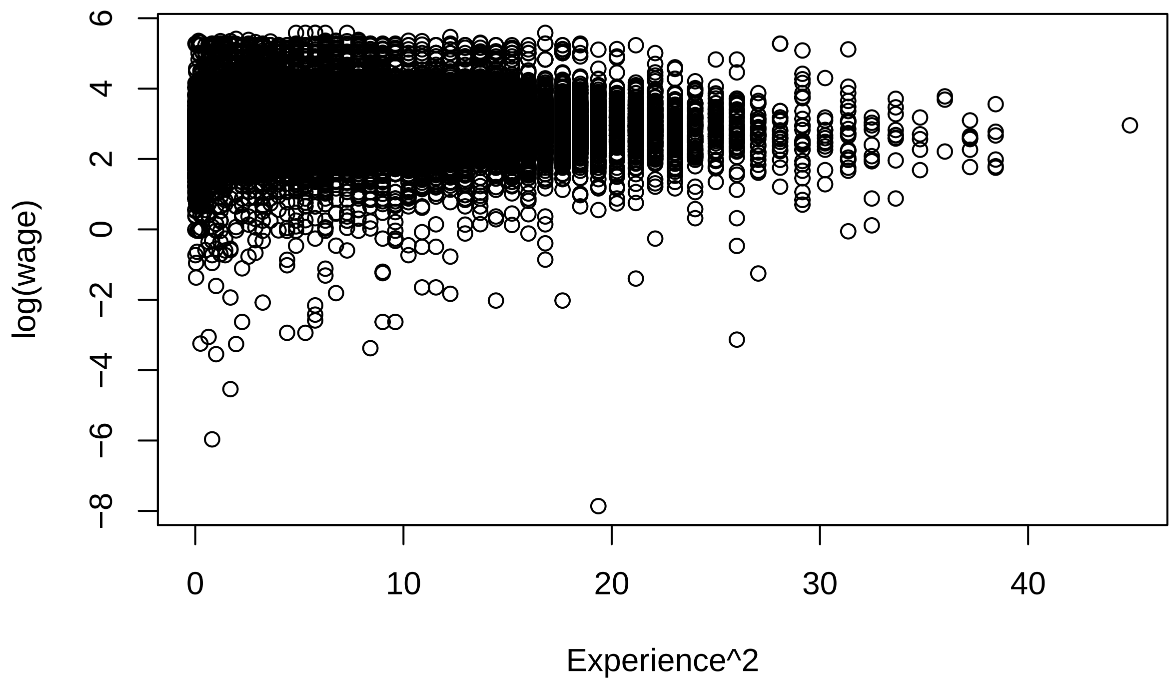
```
## Asian              -0.03842176 0.0133545758
## mixed              -0.04128004 0.0208495047
## intercept           0.90850729 0.0212522346
```

Table 1: OLS Estimates of Linear Equation for Log(Wage)

	$\hat{\beta}$	$s(\hat{\beta})$
Education	0.117	0.001
Experience	0.033	0.001
$Experience^2/100$	-0.056	0.002
Female	-0.098	0.011
Female Union Member	0.023	0.020
Male Union Member	0.095	0.020
Married Female	0.016	0.010
Married Male	0.21	0.010
Formerly Married Feale	-0.006	0.012
Formerly Married Male	0.083	0.015
Hispanic	-0.108	0.008
Black	-0.096	0.008
American Indian	-0.137	0.027
Asian	-0.038	0.013
Mixed Race	-0.041	0.021
Intercept	0.909	0.0201
σ^2	0.565	
Sample Size	46,943	

02)

```
plot(X[,3],Y, xlab = 'Experience^2', ylab = 'log(wage)')
```



If the residuals depend on variables, then Heteroscedasticity exist. Otherwise homoscedastic. From the plot of $Experience^2$ agaist log(wage), we can see there is no exact pattern. Residual will increase by decreasing the value of the variable. So, we can guess heteroscedasticity is present in the data. We can do some statistic tests to be confirmed.

Heteroscedasticity test

```
# Heteroscedasticity test

# e as dependent variable and X as independent variable
E<-data.matrix(e)
xe<-t(X)%*%E
xe<-t(X)%*%E

beta_e<-xxi%*%xe #OLS estimator

e_e <- E-(X%*%beta_e)
sigma2_e<-as.numeric(1/(n-k)*(t(e_e)%*%e_e))
sigma_e<-as.numeric( sqrt(sigma2_e))
ve<-(t(E-mean(E))%*%(E-mean(E)))/(n-1) # variance of E

#R square
R2 <- as.numeric(1-(sigma2_e/ve))

p<- ncol(x2) #number of variable
```

```r
(R2/p)/((1-R2)/n-p-1)
```

Breusch-Pagan test

```
## [1] 1.331858e-06
```

```r
qf(.95, df1=p, df2=(n-p-1)) # F distn with df1=k & df2=n-k-1
```

```
## [1] 1.666599
```

For Breusch-Pagan test the calculated value is 1.331858e-06.And the tabulated value is 1.666599 .

```r
n*R2
```

LM statistic test

```
## [1] -15.00511
```

```r
qchisq(.95, df=p) #Chi square with k df
```

```
## [1] 24.99579
```

For LM statistic test the calculated value is -15.00511 and the tabulated value is 24.99579 .

In both tests the calculated value is less than the tabulated value. So we can conclude that heteroscedasticity is present.

But for large number of observations, the effect of heteroscedasticity is minor.

03

Variance of estimated partial effect

```r
##Delta method
v0 <- xxi*sigma2

# covariance matrix for exp and exp2
v0a<-data.frame(v0[c(2,3),c(2,3)])
V<-matrix(c(v0a[1,1],v0a[1,2]/100,v0a[2,1]/100,v0a[2,2]/10000),2,2)
```

```r
print(V)
```

```
##               [,1]          [,2]
## [1,]  7.858481e-07 -1.543066e-08
## [2,] -1.543066e-08  3.294452e-10
```

```r
exp_max<- -beta[2]/(2*beta[3]/100)
print(exp_max)
```

```
## [1] 29.37805
```

Delta method was used to find the variance of the estimated partial effect with respect to experience when log(wage) as dependent variable.

The estimated partial effect with respect to experience when log(wage) as dependent variable.

$$\frac{\delta E(log(wage))}{\delta Experience} = \beta_{Experience} - \frac{2\beta_{Experience^2} \times Experience}{100}$$

$$\therefore Experience_{max} = \frac{\beta_{Experience \times 100}}{-2\beta_{Experience^2} \times Experience} = 29.37805$$

```
dexp_max_dbeta_exp <- -1/(2*beta[3]/100)
dexp_max_dbeta_exp2 <- beta[2]/(2*(beta[3]/100)^2)
G<- data.matrix(cbind(dexp_max_dbeta_exp,dexp_max_dbeta_exp2))
print(G)
```

```
##      dexp_max_dbeta_exp dexp_max_dbeta_exp2
## [1,]           886.0429            52060.43
```

$$\frac{Experience_{max}}{\delta\beta_{Experience}} = \frac{-1}{2 \times \beta_{Experience^2}} = 886.0429$$
$$\frac{Experience_{max}}{\delta\beta_{Experience^2}} = \frac{\beta_{Experience}}{2 \times \beta^2_{Experience^2}} = 52060.43$$
$$\therefore G = (886.0429 \;\; 52060.43)$$

Covariance matrix for Experience and Experience^2

```
v0 <- xxi*sigma2 #

# covariance matrix for exp and exp2
v0a<-data.frame(v0[c(2,3),c(2,3)])
V<-matrix(c(v0a[1,1],v0a[1,2]/100,v0a[2,1]/100,v0a[2,2]/10000),2,2)

print(V)
```

```
##               [,1]          [,2]
## [1,]  7.858481e-07 -1.543066e-08
## [2,] -1.543066e-08  3.294452e-10
```

$$V = \begin{pmatrix} 7.858481e-07 & -1.543066e-08 \\ -1.543066e-08 & 3.294452e-10 \end{pmatrix}$$

```
var_pexp<-G%*%V%*%t(G) #variance of partial effect
print(var_pexp)
```

```
##            [,1]
## [1,] 0.08627495
```

$$\therefore GVG' = 0.08627495$$

So, the variance of the estimated partial effect with respect to experience when log(wage) as dependent variable is 0.08627495

04

leverage and influence

```
leverage <- rowSums(X*(X%*%xxi))
r <- e/(1-leverage) # \tilde{e}
d <- leverage*e/(1-leverage) # h_ii \tilde{e}
print(max(abs(d)))
```

```
## [1] 0.01332381
```

```r
# which has the max value?
ind <- which(abs(d)==max(abs(d)))
print(X[ind,])
```

```
##               edu               exp              exp2            female
##             18.00             15.00              2.25              0.00
##      female_union        male_union    female_married      male_married
##              0.00              0.00              0.00              1.00
## female_exmarried   male_exmarried              hisp             black
##              0.00              0.00              0.00              0.00
##            Am_Ind             Asian             mixed         intercept
##              1.00              0.00              0.00              1.00
```

```r
print(leverage[ind])
```

```
##       42575
## 0.002185683
```

```r
x_i <- X[-ind,]
y_i <- Y[-ind]
xx_i <- t(x_i)%*%x_i
xy_i <- t(x_i)%*%y_i
beta_i <- solve(xx_i,xy_i)
betas <- cbind(beta,beta_i)
print(betas)
```

```
##                         [,1]         [,2]
## edu               0.11669830  0.116797267
## exp               0.03315647  0.033160099
## exp2             -0.05643068 -0.056452631
## female           -0.09825725 -0.098351234
## female_union      0.02285629  0.022929406
## male_union        0.09518739  0.095036669
## female_married    0.01615988  0.016275905
## male_married      0.21112049  0.211502373
## female_exmarried -0.00642132 -0.006308417
## male_exmarried    0.08289541  0.082945807
## hisp             -0.10813640 -0.108314453
## black            -0.09553060 -0.095429412
## Am_Ind           -0.13743914 -0.124776661
## Asian            -0.03842176 -0.038562632
## mixed            -0.04128004 -0.041214438
## intercept         0.90850729  0.907000290
```

```r
u1 <- X*(e%*%matrix(1,1,k))
u2 <- X*((e/sqrt(1-leverage))%*%matrix(1,1,k))
u3 <- X*((e/(1-leverage))%*%matrix(1,1,k))
XXi <- solve(t(X)%*%X)

v1 <- XXi %*% (t(u1)%*%u1) %*% XXi
v1a <- n/(n-k) * XXi %*% (t(u1)%*%u1) %*% XXi
v2 <- XXi %*% (t(u2)%*%u2) %*% XXi
v3 <- XXi %*% (t(u3)%*%u3) %*% XXi

s1 <- sqrt(diag(v1)) # HC0
s1a <- sqrt(diag(v1a)) # HC1
```

```
s2 <- sqrt(diag(v2)) # HC2
s3 <- sqrt(diag(v3)) # HC3
s4<-cbind(s1,s1a,s2,s3)
```

```
print(s4)
```

```
##                            s1         s1a           s2           s3
## edu              0.0012820059 0.001282224 0.0012822458 0.0012824859
## exp              0.0009517888 0.000951951 0.0009520861 0.0009523836
## exp2             0.0020751922 0.002075546 0.0020759054 0.0020766191
## female           0.0110300255 0.011031906 0.0110326882 0.0110353523
## female_union     0.0196045724 0.019607914 0.0196294151 0.0196542922
## male_union       0.0202969553 0.020300415 0.0203152446 0.0203335518
## female_married   0.0095395061 0.009541132 0.0095416515 0.0095437981
## male_married     0.0096989277 0.009700581 0.0097010045 0.0097030824
## female_exmarried 0.0118587152 0.011860737 0.0118616096 0.0118645058
## male_exmarried   0.0145582227 0.014560704 0.0145624229 0.0145666256
## hisp             0.0081195128 0.008120897 0.0081212145 0.0081229172
## black            0.0083352591 0.008336680 0.0083370361 0.0083388138
## Am_Ind           0.0263682529 0.026372748 0.0263979144 0.0264276106
## Asian            0.0133545758 0.013356852 0.0133585841 0.0133625943
## mixed            0.0208495047 0.020853059 0.0208645353 0.0208795776
## intercept        0.0212522346 0.021255857 0.0212566374 0.0212610430
```

The four covariance estimators HC0, HC1, HC2, and HC3. ## 05)

Part 2)

Jackknife Variance

```
jackknife_variance <- function(iN) {

#iN = 50 # sample size
ik = 1 # no. of regressors
sigma = 1 # standard deviation of the errors

beta = 1
set.seed(63)
# we make the variables
mX = matrix(rnorm(iN*ik), iN, ik)
ve = rnorm(iN) * sigma
vy = c(mX %*% beta + ve)

hbeta <- function(vy, mX){

  return(c(sum(mX^3*vy)/sum(mX^4)))
}

### Jackknife

jackknife <- function(some_statistic, vy, mX){

  ftmp <- function(iter){
    return(some_statistic( vy = vy[-iter], mX = mX[-iter,] ))
```

```
  }

  return(sapply(1:length(vy), ftmp))
}

ret = jackknife(hbeta, vy, mX)
ret

tmp = ret - mean(ret)
return(crossprod(tmp) * (iN - 1)/iN)
}
```

Bootstrap variance

```
## Bootstrap
bootstrap_variance<- function(iN){
# a comarison or a small experiment
set.seed(63)
ik = 1 # no. of regressors
sigma = 1 # standard deviation of the errors

beta = 1
set.seed(63)
# we make the variables
mX = matrix(rnorm(iN*ik), iN, ik)
ve = rnorm(iN) * sigma
vy = c(mX %*% beta + ve)

hbeta <- function(vy, mX){

  return(c(sum(mX^3*vy)/sum(mX^4)))
}

nonparametric_bootstrap <- function(some_statistic, vy, mX, iB=1000){

  ftmp <- function(vs) return(some_statistic(vy[vs], mX[vs,]))

  resamples = matrix(sample(1:length(vy), length(vy)*iB, replace=TRUE), length(vy), iB)

  return(apply(resamples, 2, ftmp))

}

iB = 500
ret = nonparametric_bootstrap(hbeta, vy, mX, iB)
ret

tmp = ret - mean(ret)
crossprod(tmp)/(iB - 1)
}
```

From the result, we see that the variance decreases with increase of sample size in both cases.For a small sample size Bootstrap gave small variance but for large sample size both gave almost same result.

Table 2: Jakknife Variance and Bootstrap Variance.

Sample Size	**Jakknife Variance**	**Bootstrap Variance**
50	0.0658	0.0310
100	0.0057	0.0072
500	0.0051	0.0051