Week2 Assignment Econometrics

Pradeepta Das

11 November 2020

Questions

(a) Prove:

$$E(b_R) = \beta_1 + P\beta_2$$

First, it is important to express b_R in terms of ϵ:

$$\begin{aligned} b_R &= (X_1'X_1)^{-1}X_1'y \\ &= (X_1'X_1)^{-1}X_1'(X_1\beta_1 + X_2\beta_2 + \epsilon) \\ &= (X_1'X_1)^{-1}X_1'X_2\beta_2 + (X_1'X_1)^{-1}X_1'\epsilon \\ &= \beta_1 + P\beta_2 + (X_1'X_1)^{-1}X_1'\epsilon \\ so, E(b_R) &= \beta_1 + P\beta_2 \\ as E(\epsilon) &= 0 \end{aligned}$$

(b) Prove:

$$var(b_R) = \sigma^2(X_1'X_1)^{-1}$$

First, it is important to express b_R in terms of ϵ:

$$\begin{aligned} b_R - E(b_R) &= \beta_1 + P\beta_2 + (X_1'X_1)^{-1}X_1'\epsilon - \beta_1 + P\beta_2 \\ &= (X_1'X_1)^{-1}X_1'\epsilon \\ var(b_R) &= E[((X_1'X_1)^{-1}X_1'\epsilon)((X_1'X_1)^{-1}X_1'\epsilon)'] \\ &= E[((X_1'X_1)^{-1}X_1'\epsilon)(\epsilon' X_1((X_1'X_1)^{-1})')] \\ &= E[((X_1'X_1)^{-1}X_1'\epsilon)(\epsilon' X_1(X_1'X_1)^{-1})] \\ &= E[(X_1'X_1)^{-1}X_1'\epsilon\epsilon' X_1(X_1'X_1)^{-1}] \\ &= (X_1'X_1)^{-1}X_1'E[\epsilon\epsilon']X_1(X_1'X_1)^{-1}(X_1 is fixed) \\ &= (X_1'X_1)^{-1}X_1'E[\sigma^2 I]X_1(X_1'X_1)^{-1} \\ &= \sigma^2(X_1'X_1)^{-1}X_1'X_1(X_1'X_1)^{-1} \\ &= \sigma^2(X_1'X_1)^{-1} \end{aligned}$$

(c) Prove:

$$b_R = b_1 + Pb_2$$

$$b_R = (X_1'X_1)^{-1}X_1'y$$
$$= (X_1'X_1)^{-1}X_1'(X_1b_1 + X_2b_2 + res)$$
$$= b_1 + Pb_2 + (X_1'X_1)^{-1}X_1'res$$
$$= b_1 + Pb_2$$
$$\text{as} X_1'res = 0 \text{because of orthogonality}$$

(d) Argue that the columns of the (2×3) matrix P are obtained by regressing each of the variables ‘Age’, ‘Educ’, and ‘Parttime’ on a constant term and the variable ‘Female’.

So,

$$P = (X_1'X_1)^{-1}X_1'(X2)$$

is a (2×3) matrix, with columns as such:

Column 1:

$$(X_1'X_1)^{-1}X_1'(Age) \text{the OLS formula for regressing ‘Age’ on X1.}$$

Column 2:

$$(X_1'X_1)^{-1}X_1'(Educ) \text{the OLS formula for regressing ‘Educ’ on X1.}$$

Column 3:

$$(X_1'X_1)^{-1}X_1'(Parttime) \text{the OLS formula for regressing ‘Parttime’ on X1.}$$

(e) P matrix value.

```
file25 = "C:\\Users\\Gannon_chef\\Documents\\Pradeepta\\Coursera\\Econometrics\\TrainExer21.txt"
new_logwage<-read.table(file25, header = TRUE, sep = "", dec = ".")
str(new_logwage)
```

```
## 'data.frame':    500 obs. of  7 variables:
##  $ Observ. : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Wage    : int  66 34 70 47 107 188 123 57 42 200 ...
##  $ LogWage : num  4.19 3.53 4.25 3.85 4.67 ...
##  $ Female  : int  0 1 1 0 1 1 1 1 1 0 ...
##  $ Age     : int  49 42 42 38 54 54 47 39 25 59 ...
##  $ Educ    : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ Parttime: int  1 1 1 0 1 0 0 1 0 0 ...
```

```
X1 = new_logwage['Female']
X1['Constant']<-rep(1, length(new_logwage['Female']))
X1 <- X1[,c(2,1)]
X2 = new_logwage[c('Age', 'Educ', 'Parttime')]
str(X1)
```

```
## 'data.frame':    500 obs. of  2 variables:
##  $ Constant: num  1 1 1 1 1 1 1 1 1 1 ...
##  $ Female  : int  0 1 1 0 1 1 1 1 1 0 ...
```

```
str(X2)
```

```
## 'data.frame':    500 obs. of  3 variables:
##  $ Age     : int  49 42 42 38 54 54 47 39 25 59 ...
##  $ Educ    : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ Parttime: int  1 1 1 0 1 0 0 1 0 0 ...
```

Now, value of P matrix can be computed as:

```
x1_mat = as.matrix(sapply(X1, as.numeric))
x2_mat = as.matrix(sapply(X2, as.numeric))
P<-solve(t(x1_mat)%*%(x1_mat))%*%t(x1_mat)%*%x2_mat
P
```

```
##                 Age       Educ  Parttime
## Constant 40.0506329  2.2594937 0.1962025
## Female   -0.1104155 -0.4931893 0.2494496
```

(f) Check the numerical validity of the result in part (c). Note: This equation will not hold exactly because the coefficients have been rounded to two or three decimals; preciser results would have been obtained for higher precision coefficients.

```
model <- lm(LogWage~Female+Age+Educ+Parttime, data=new_logwage)
summary(model)
```

```
##
## Call:
## lm(formula = LogWage ~ Female + Age + Educ + Parttime, data = new_logwage)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.73799 -0.15705 -0.00388  0.16495  0.77827
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.052697   0.055333  55.170   <2e-16 ***
## Female      -0.041073   0.024710  -1.662   0.0971 .
## Age          0.030606   0.001273  24.041   <2e-16 ***
## Educ         0.233163   0.010660  21.873   <2e-16 ***
## Parttime    -0.365407   0.031570 -11.575   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2452 on 495 degrees of freedom
## Multiple R-squared:  0.704,  Adjusted R-squared:  0.7016
## F-statistic: 294.3 on 4 and 495 DF,  p-value: < 2.2e-16
```

```
b1=as.matrix(model$coeff[1:2])
b2=as.matrix(model$coeff[3:5])
b1+P%*%b2
```

```
##                   [,1]
## (Intercept)  4.7336234
## Female      -0.2505962
```

Now, we should be able to replicate this value if we regress only using the Female and intercept! If we do that :

```
x1_model <- lm(LogWage~Female, data=new_logwage)
x1_model$coeff
```

```
## (Intercept)      Female
##   4.7336234  -0.2505962
```

We can see that the values match exactly!