ps4_Aline

helloooo

2024-02-29

Exercise 1

Q1. Estimate the model via ols

```
my_data <- read_dta("/Users/alineabayo/Desktop/ARE_212/Problem\ Set\ 4/pset4_2024.dta")
my_data$lprice <- log(my_data$price)
my_data$lqu <- log(my_data$qu)
x1 <- cbind(1, my_data$fuel, my_data$lprice, my_data$weight)
y1 <- my_data$lqu
n1 <- length(y1)
Df<- nrow(x1)-ncol(x1)
#coefficients
b1 <- solve(t(x1) %*% x1) %*% t(x1) %*% y1
b1
```

```
##              [,1]
## [1,] 13.021293997
## [2,] -0.091952918
## [3,] -0.797538193
## [4,] -0.001074254
```

```
# projection matrix of reg y on X
P <- x1%*%solve(t(x1)%*%x1)%*%t(x1)
#Residuals
M <- diag(n1)-P
e <- M%*%y1
#Standard error
s1 <- as.numeric(t(e)%*%e)/Df
vb <- s1*solve(t(x1)%*%x1)
se <- sqrt(diag(vb))
se
```

```
## [1] 0.2233526284 0.0256164221 0.1488903142 0.0002400311
```

```


reg_lqu <- lm(y1~x1)
#bptest(reg_lqu)
```

Q2. BP Test Yes, we do have heteroskedasticity

```r
#inv(X'X) (\sum Xi' ei ei Xi) inv(X'X)
#inv(X'X) (\sum Xei Xei') inv(X'X)
Xe<-cbind(e,my_data$fuel*e,my_data$lprice*e, my_data$weight*e)

Vb_whiteRobust<-solve(t(x1) %*% x1) %*% t(Xe) %*%  Xe %*% solve(t(x1) %*% x1)
Vb_whiteRobust
```

Q3. White robust standard errors.

```
##               [,1]          [,2]          [,3]          [,4]
## [1,]  5.012006e-02  3.555884e-04 -2.265791e-02  1.405071e-05
## [2,]  3.555884e-04  5.402540e-04 -1.719085e-03  9.378235e-07
## [3,] -2.265791e-02 -1.719085e-03  2.207311e-02 -2.904649e-05
## [4,]  1.405071e-05  9.378235e-07 -2.904649e-05  6.228930e-08
```

```r
seb_whiteRobust<-sqrt(diag(Vb_whiteRobust))
seb_whiteRobust
```

```
## [1] 0.2238751058 0.0232433645 0.1485702322 0.0002495783
```

```r
se
```

```
## [1] 0.2233526284 0.0256164221 0.1488903142 0.0002400311
```

Yes, white se is the appropriate way to solve heteroskedaskicity

```r
#gamma=
b1_2<- b1[[2]]
b1_3<- b1[[3]]
b1_4<- b1[[4]]
```

Exercise 2

```r
x2 <- cbind(1, my_data$fuel, my_data$lprice, my_data$year, my_data$weight, my_data$luxury)
y2 <- my_data$lqu
n2 <- length(y2)
df2<- nrow(x2)-ncol(x2)
#coefficients
b2 <- solve(t(x2) %*% x2) %*% t(x2) %*% y2
b2
```

```
##               [,1]
## [1,] 68.7789505837
## [2,] -0.1524851356
## [3,] -1.4418206274
## [4,] -0.0274178787
## [5,] -0.0002186054
## [6,]  0.9853449176
```

```r
# projection matrix of reg y on X
P2 <- x2%*%solve(t(x2)%*%x2)%*%t(x2)
#Residuals
M2 <- diag(n2)-P2
e2 <- M%*%y2
#Standard error
s2 <- as.numeric(t(e2)%*%e2)/df2
vb2 <- s1*solve(t(x2)%*%x2)
se2 <- sqrt(diag(vb2))
se2
```

```
## [1] 15.563114942  0.034900407  0.191480541  0.007745049  0.000391238
## [6]  0.136641088
```

100% increase in quantity results in 144.18% decrease in price,holding all other factors constant.

Omitted Variable Bias Performance and Features: Cars with better performance metrics (such as acceleration, power) and advanced features (like autonomous driving capabilities, high-end entertainment systems) are generally priced higher. These features enhance the desirability and decisions to buy a car

Omitting variables that are correlated with both the dependent variablea and the independent variables (e.g., log price) can lead to omitted variable bias. This bias occurs because the OLS assume that all other variables are held constant, including those not included in the model. If the omitted variables are correlated with both the price and the quantity demanded, their effects are attributed to the variables included in the model, distorting the estimated coefficient of log price. In this case, we might overestimate the true impact of lprice on lqu.

```r


cor1<- cor(my_data$lprice,my_data$fuel)
cor1
```

If we omit fuel from equation (eq.2) how does your OLS estimate of the log price change?

```
## [1] 0.6340064
```

```r
x3 <- cbind(1,my_data$lprice, my_data$year, my_data$weight, my_data$luxury)
y3 <- my_data$lqu
ols_sans_fuel<- lm(y3~x3)
summary(ols_sans_fuel)
```

```
##
## Call:
## lm(formula = y3 ~ x3)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5189 -0.9579  0.0827  1.0149  3.1533
##
## Coefficients: (1 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) 23.1508250 11.4167490   2.028  0.04272 *
## x31                 NA         NA      NA       NA
## x32         -1.3808903  0.1889483  -7.308 3.99e-13 ***
## x33         -0.0046579  0.0056707  -0.821  0.41152
## x34         -0.0009875  0.0003457  -2.856  0.00433 **
## x35          0.9964319  0.1351698   7.372 2.52e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.389 on 1873 degrees of freedom
## Multiple R-squared:  0.1965, Adjusted R-squared:  0.1947
## F-statistic: 114.5 on 4 and 1873 DF,  p-value: < 2.2e-16
```

Exercise 3

```r
my_data$laverage_o1 <- log(my_data$average_o1)

x3 <- cbind(1,my_data$fuel, my_data$year, my_data$weight, my_data$luxury, my_data$laverage_o1)
y3 <- my_data$lprice
b3 <- solve(t(x3) %*% x3) %*% t(x3) %*% y3
b3[[6]]
```

```
## [1] 0.7720019
```

100% increase in price leads to 77.2% increase in this this specific country, controlling for other factors.

Exercise 4

```r
x4 <- cbind(1,my_data$fuel, my_data$year, my_data$weight, my_data$luxury, my_data$laverage_o1)
y4 <- my_data$lqu
b4 <- solve(t(x4) %*% x4) %*% t(x4) %*% y4
b4
```

```
##              [,1]
## [1,] 26.176493455
## [2,] -0.146165397
## [3,] -0.006309850
## [4,] -0.001169946
## [5,]  0.936857361
## [6,] -0.927842632
```

Exercise 5

```r
x5 <- cbind(1,my_data$fuel, my_data$year, my_data$weight, my_data$luxury, my_data$lprice)
y5 <- my_data$lqu
z5 <- cbind(1,my_data$fuel, my_data$year, my_data$weight, my_data$luxury, my_data$laverage_o1)
b5 <- solve(t(z5) %*% x5) %*% t(z5) %*% y5
b5
```

a)

```
##                 [,1]
## [1,] 56.8578477573
## [2,] -0.1492998550
## [3,] -0.0215690704
## [4,] -0.0006194225
## [5,]  0.9432700790
## [6,] -1.2018656767
```

100% increase in lprice leads to 120.1% decrease in lqu

```
#first stage
y6 <- my_data$lqu
z6 <- cbind(1,my_data$fuel, my_data$year, my_data$weight, my_data$luxury, my_data$laverage_o1)
a6<- solve(t(z6) %*% z6) %*% t(z6) %*% my_data$lprice
my_data$lpricecoef <- z6 %*% a6
#2nd stage
z7 <- cbind(1,my_data$fuel, my_data$year, my_data$weight, my_data$luxury, my_data$lpricecoef)
b7 <- solve(t(z7) %*% z7) %*% t(z7) %*% y6
b7
```

b)2SLS

```
##                 [,1]
## [1,] 56.8578477484
## [2,] -0.1492998550
## [3,] -0.0215690704
## [4,] -0.0006194225
## [5,]  0.9432700789
## [6,] -1.2018656766
```

100% increase in lprice leads to 120.1% decrease in lqu

```
my_data$first_stageresid <- my_data$lprice-my_data$lpricecoef
z8 <- cbind(1,my_data$fuel, my_data$year, my_data$weight, my_data$luxury, my_data$lprice, my_data$first_
b8 <- solve(t(z8) %*% z8) %*% t(z8) %*% y6
b8
```

c)

```
##                 [,1]
## [1,] 56.8578477605
## [2,] -0.1492998550
## [3,] -0.0215690704
## [4,] -0.0006194225
## [5,]  0.9432700790
## [6,] -1.2018656767
## [7,] -0.4543582724
```