Introductory Applied Econometrics
EEP/IAS 118
Spring

Daily Assignment Lecture 6

Solution

1 Practice Exercise after Lecture 6

In order to test whether the retail price of gasoline varies across areas in ways that suggest price discrimination, we use ZIP code-level data on the price of gasoline in South California over time, and estimate the following regressions:

$$\widehat{\ln(R)} = -0.32 + 0.008prpblack + 0.011\ln(income) + 0.075MC \qquad n = 1462 \quad R^2 = .678 \tag{1}$$

$$\widehat{\ln(R)} = -0.010 + 0.004prpblack + 0.083MC \qquad n = 1462 \quad R^2 = .608 \tag{2}$$

where R is the retail price of gasoline, *prpblack* is the proportion of the ZIP code population that is black, *income* the average income in the ZIP code, and MC a dummy variable for a Major City in South CA.

1. Interpret the estimated coefficients on *prpblack* and on MC in model 1 (Remember to discuss sign and size).
   Sol.
   Interpreting *prpblck*
   - Sign: The sign on *prpblck* is positive: a higher percentage of black in a given ZIP is associated with higher gasoline prices. This suggests discrimination on the basis of race
   - Size: A one unit increase in the proportion of blacks is associated with a 0.8% increase in the predicted price of gasoline, holding income and Major City constant

   Interpreting MC
   - Sign: Living in MC is associated with a positive effect on gasoline prices, suggesting that individuals in MC pay higher prices for gasoline than in non Major cities in South California.
   - Size: Living in MC is associated with a 7.5% increase in the predicted price of gasoline, holding income and proportion of black constant

2. Comparing the results from the two estimations, is discrimination larger or smaller when you control for the income variable? What do you infer from this on the correlation between the average income and the racial characteristics of the zip code (justify your claim)? **Sol.**
   (a) First, we see that when we control for income the coefficient on proportion black increases, suggesting higher levels of discrimination. This means that our coefficient in the second regression exhibits downward bias.
   (b) Second, we see that the covariance between our ommitted variable income and our y variable (R) is positive:
   $$cov(y, x_{om}) = cov(ln(R), ln(income)) > 0$$
   So the price of gasoline is higher in Zip codes with higher income.
   (c) We can infer then that
   $$cov(x, x_{om}) = cov(prpblck, ln(income)) < 0$$
   Concluding sentence: ZIP codes with lower average income are charged lower prices for their gasoline. Moreover areas with a higher proportion of blacks have lower income . Hence, if we fail to include income it will look like areas with higher proportion of blacks are charged relatively less high gasoline prices than they are in reality because we are confounding the effect of income and black.

2 R applied Exercise

The following table contains the quantity and the price of a barrel of oil for twelve periods. Price is in dollars and quantity is in thousand of barrels. This table is in the Daily Assigment Folder saved as two file formats, oilDemand.dta and oilDemand.xlsx.

period	World Q	oil price
1	61440	145.43
2	62083	145.21
3	62769	134.41
4	64494	121.29
5	66023	114.24
6	67769	107.88
7	69652	103.73
8	70206	94.62
9	73530	86.70
10	74540	75.07
11	76258	73.26
12	75502	67.35

(i) USING R, load either oilDemand.dta or the excell equivalent

```
#--------------------------------------------
#set your working directory
#--------------------------------------------
#setwd("/Users/sberto/Desktop/")
setwd("/Users/sofiavillas-boas/Dropbox/EEP118_Spring2021/Daily Assignments/6-DA-Lecture6")

#--------------------------------------------
#1. Read in data and see the top rows to see column names etc
#--------------------------------------------
#read in DA6 data set
#read in an excell dataset
my_data <- read_excel("oilDemand.xlsx")
head(my_data)
```

(ii) Estimate the relationship between Quantity Q and price using OLS; that is, obtain the intercept and slope estimates in the equation

$$\widehat{Q} = \hat{\alpha}_0 + \hat{\alpha}_1 price$$

Comment on the direction of the relationship. Does the intercept have a useful interpretation here? Explain. How much higher is the Quantity predicted to be if the price is increased by 25 dollars?

```
#--------------------------------------------------------------------------------------------------------------
#run regressions
#--------------------------------------------------------------------------------------------------------------
reg1<-lm(WorldQ ~ oilPrice,my_data)
summary(reg1)
```

```
#output is

lm(formula = WorldQ ~ oilPrice, data = my_data)

Residuals:
    Min      1Q  Median      3Q     Max
-1167.0  -662.5  -234.4   696.4  1229.1

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 89317.681   1086.032   82.24 1.73e-15 ***
oilPrice     -195.043      9.967  -19.57 2.66e-09 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 904.4 on 10 degrees of freedom
Multiple R-squared:  0.9746,Adjusted R-squared:  0.972
F-statistic: 382.9 on 1 and 10 DF,  p-value: 2.656e-09


alpha zero hat  is total quantity purchased when price is zero.  if oil were free,
this is how much Q would be consumed per period.\\


How much higher is the Quantity predicted to be if the price is increased by 25 dollars?
\\

quantity increases by  (25 * -195.0426 ), that is quantity drops by -4876.066049 thousand
barrels of oil, holding everything else contant (ceteris paribus)
```

(iii) Create a new variable that consists of the fitted values and residuals for each observation.

```
#fitted values
my_data$Q_hat <-reg1$fitted.values

#residuals
myData$e<-my_data$WorldQ-my_data$Q_hat
```

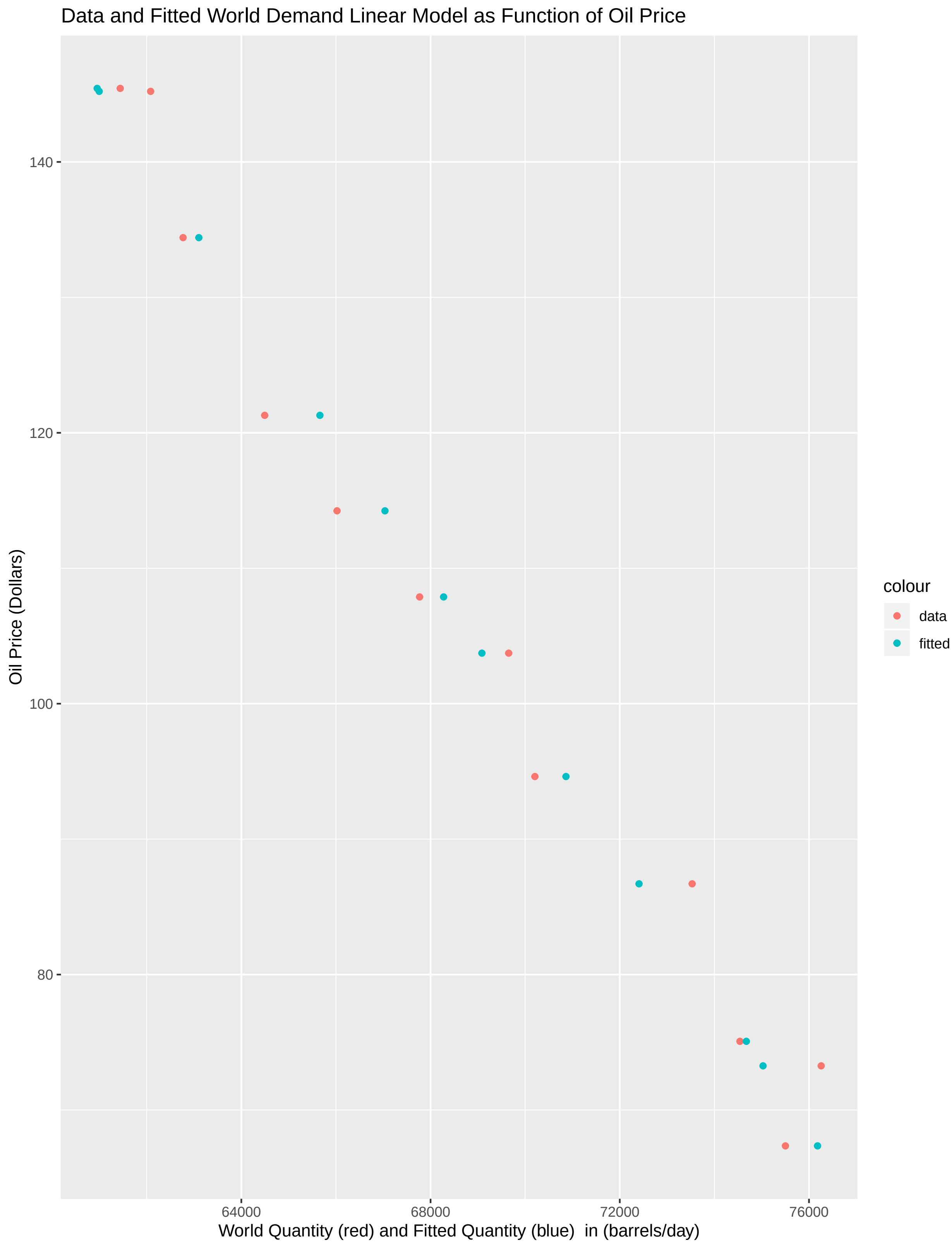
(iv) How much variation in Quantity for these twelve periods is explained by price? Explain.
R squared is 97.46 percent.

(v) . See solutions pdf and also DA6.R code file for commands to solve the above.

3 Replication Lecture 6

Feel free to replicate all we did in Lecture 6, by running in R the code Lecture6.R that uses Lecture6.dta.

Figure 1: Quantity and Price and Fitted Quantity and Price



Note: output from DA6.R code