89/100

20+5+10+10+6+5+8+15+10

Assignment 2

Tai Lo Yeung, Pan Yiming, Qiwei Liu

Barcelona Graduate School of Economics

Universitat Pompeu Fabra

Nov 7, 2020

gasoline.dta

a

Compute the regression of per capita consumption of gasoline (G/pop) on all other explanatory variables, including the time trend (you can use either Year directly or (Year- 1960)). Comment on whether the signs of the estimates agree with your expectations (i.e. comment on your expectations, too!).

By running code in R, we obtained:

```
Call:
lm(formula = g ~ year + pg + y + pnc + puc + ppt + pd + pn +
    ps, data = gasoline)

Residuals:
      Min        1Q    Median        3Q       Max
-0.036578 -0.014761  0.002717  0.010443  0.033908

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.146e-04  2.106e-01   0.001  0.99957
year         9.402e-03  8.799e-03   1.069  0.29507
pg          -1.207e-01  2.616e-02  -4.613 9.33e-05 ***
y            1.114e-04  3.234e-05   3.444  0.00196 **
pnc          6.350e-02  1.325e-01   0.479  0.63572
puc         -4.085e-02  2.880e-02  -1.418  0.16802
ppt          5.888e-02  4.083e-02   1.442  0.16122
pd           2.896e-01  3.557e-01   0.814  0.42290
pn           5.424e-01  4.430e-01   1.224  0.23177
ps          -8.770e-01  3.637e-01  -2.411  0.02326 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02133 on 26 degrees of freedom
Multiple R-squared:  0.9829,Adjusted R-squared:  0.977
F-statistic: 166.5 on 9 and 26 DF,  p-value: < 2.2e-16
```

Most signs of the estimates agree with our expectations besides the negative coefficient from variable 'ps'. In specific, all signs are positive except price index for gasoline, price index for used cars and aggregate price index for consumer services. It's easy to interpret this result because when all other equal: for 'year', the later it is, the more the consumption of gasoline in US due to the development; for 'pg', the surge in the price of gasoline will bring down the consumption of gasoline; for 'y', with more income, individual can consume more gasoline; for

’pnc’, relatively speaking, expensive cars tend to consume more gasoline; for ’puc’, expensive used car may be in better condition so it cost less gasoline; for ’ppt’, if the public transportation is expensive, individual will prefer to drive more; for ’pd’ and ’pn’, if the CPI grow, the total expenditure of gasoline will grow too which leads to growth in gasoline consumption; for ’ps’, we thought it should be positive too.

b

Test the hypothesis that consumers do not differentiate between changes in the prices of new and used cars.

```
Linear hypothesis test

Hypothesis:
pnc - puc = 0

Model 1: restricted model
Model 2: g ~ year + pg + y + pnc + puc + ppt + pd + pn + ps

  Res.Df      RSS Df  Sum of Sq    F Pr(>F)
1     27 0.012110
2     26 0.011833  1 0.00027759 0.61 0.4419
```

We failed to reject the hypothesis that consumers do not differentiate between changes in the prices of new and used cars.

c

Find the value of the elasticity of gasoline demand to its own price, to income and the cross-price elasticity with respect to changes in the price of public transportation from the results of this linear model (you will have to do a bit of thinking here, and remember the concept of elasticity).

By doing simple computation, we obtained the three elasticity respectably as -0.278, 1.02, and 0.161, please refer to the Code section for formula.

d

Reestimate the above regression in logarithms (except for the time trend). All the coefficients are now directly elasticities. How do the estimates compare with the results in the previous analysis?

By doing a logged regression, we have:

```
Call:
lm(formula = g ~ log(pg) + year + log(y) + log(pnc) + log(puc) +
    log(ppt) + log(pd) + log(pn) + log(ps), data = gasoline)

Residuals:
      Min        1Q    Median        3Q       Max
-0.036731 -0.010979  0.002899  0.011391  0.038455

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -8.900022   1.851553  -4.807 5.60e-05 ***
log(pg)     -0.580511   0.068259  -8.505 5.51e-09 ***
year        -0.008532   0.007027  -1.214  0.23562
log(y)       1.167884   0.190788   6.121 1.80e-06 ***
log(pnc)     0.198881   0.221662   0.897  0.37783
log(puc)    -0.097155   0.071692  -1.355  0.18702
log(ppt)     0.089685   0.085969   1.043  0.30645
log(pd)      0.789716   0.313865   2.516  0.01837 *
log(pn)      1.377616   0.294919   4.671 8.01e-05 ***
log(ps)     -1.287363   0.397212  -3.241  0.00325 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0194 on 26 degrees of freedom
Multiple R-squared:  0.9859,Adjusted R-squared:  0.981
F-statistic: 201.9 on 9 and 26 DF,  p-value: < 2.2e-16
```

We noticed that the sign of elasticities are the same with what we calculated in part c, but the value is quite different for pg and ppt, the elasticity of y fit well. Also, there are more significances in this model.

e

Do the two regression models imply the same regarding the elasticities? Which one do you think is more correct: the linear model or the log-log model? Why?

Although the elasticities for some variables seems different, the intuition remain the same because the signs of elasticity are all the same. I prefer the log-log model because this is more straightforward. We need to calculate the elasticity afterward using the mean of variable if using linear model, but we can get the elasticity directly through log-log model.

Production.dta

a

Use nonlinear least squares to get estimates of the elasticities of output to labor and capital (see below how to do this with Stata and R).

By running the code, we have:

```
Formula: valueadd ~ (alpha * labor^beta * capital^gama)

Parameters:
      Estimate Std. Error t value Pr(>|t|)
alpha  1.30403    0.26389   4.942 5.21e-06 ***
beta   0.82880    0.03189  25.990  < 2e-16 ***
gama   0.22150    0.03123   7.093 9.00e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 79.16 on 69 degrees of freedom

Number of iterations to convergence: 5
Achieved convergence tolerance: 1.498e-06
```

As shown above, the elasticities of output to labor and capital are 0.82880 and 0.22150 respectively.

b

Transform now this nonlinear model by taking logs of value added, labor and capital. Notice that the model now is perfectly linear, so it can be estimated by OLS. Do it, and compare the estimates of A, β and γ, and . Have these changed much?

```
Call:
lm(formula = log(valueadd) ~ log(labor) + log(capital), data = Production)

Residuals:
     Min       1Q   Median       3Q      Max
-0.44625 -0.14466  0.03981  0.12703  0.45446

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.54032    0.15151   3.566 0.000663 ***
log(labor)    0.76531    0.02735  27.980  < 2e-16 ***
log(capital)  0.23234    0.02882   8.061 1.55e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2173 on 69 degrees of freedom
Multiple R-squared:  0.9496,Adjusted R-squared:  0.9481
F-statistic: 649.6 on 2 and 69 DF,  p-value: < 2.2e-16
```

The β and γ havn't change much compared to the nonlinear one, but A changed a lot.

c

Test the following hypotheses:
a. Labor elasticity is equal to 0.6.
b. Constant returns to scale ($\beta+\gamma=1$).
c. Labor elasticity is equal to 0.6 and the production function has constant returns to scale.

For a,

```
Linear hypothesis test

Hypothesis:
log(labor) = 0.6

Model 1: restricted model
Model 2: log(valueadd) ~ log(labor) + log(capital)

  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1     70 4.9816
2     69 3.2572  1    1.7243 36.528 6.833e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

We strongly reject the null hypothesis that Labor elasticity is equal to 0.6.

for b,

```
Linear hypothesis test

Hypothesis:
log(labor)  + log(capital) = 1

Model 1: restricted model
Model 2: log(valueadd) ~ log(labor) + log(capital)

  Res.Df    RSS Df  Sum of Sq      F Pr(>F)
1     70 3.2575
2     69 3.2572  1 0.00029407 0.0062 0.9373
```

We cannot reject the hypothesis of Constant returns to scale.

for c,

```
Linear hypothesis test

Hypothesis:
log(labor) = 0.6
log(labor)  + log(capital) = 1

Model 1: restricted model
Model 2: log(valueadd) ~ log(labor) + log(capital)

  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1     71 5.5638
2     69 3.2572  2    2.3066 24.431 9.505e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

We strongly reject the joint hypothesis of a and b.

d

Test whether the translog function is more appropriate than the Cobb-Douglas function (so, in a sense, test the linearity of lny or that the Cobb-Douglas function is enough to account for output). Are the coefficients of the translog function as easy to interpret as those of the Cobb- Douglas function?

Follow the setting given, we obtained:

```
Call:
lm(formula = lny ~ lnl + lnk + lnl2 + lnk2 + lnl * lnk)

Residuals:
     Min       1Q   Median       3Q      Max
-0.44894 -0.13240  0.03271  0.11202  0.43813

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.33234    0.93944   1.418   0.1608
lnl          0.43582    0.16903   2.578   0.0122 *
lnk          0.23429    0.27508   0.852   0.3975
lnl2        -0.02042    0.04470  -0.457   0.6494
lnk2        -0.07201    0.04729  -1.523   0.1326
lnl:lnk      0.08009    0.04313   1.857   0.0678 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2123 on 66 degrees of freedom
Multiple R-squared:  0.9539,Adjusted R-squared:  0.9504
F-statistic: 273.3 on 5 and 66 DF,  p-value: < 2.2e-16
```

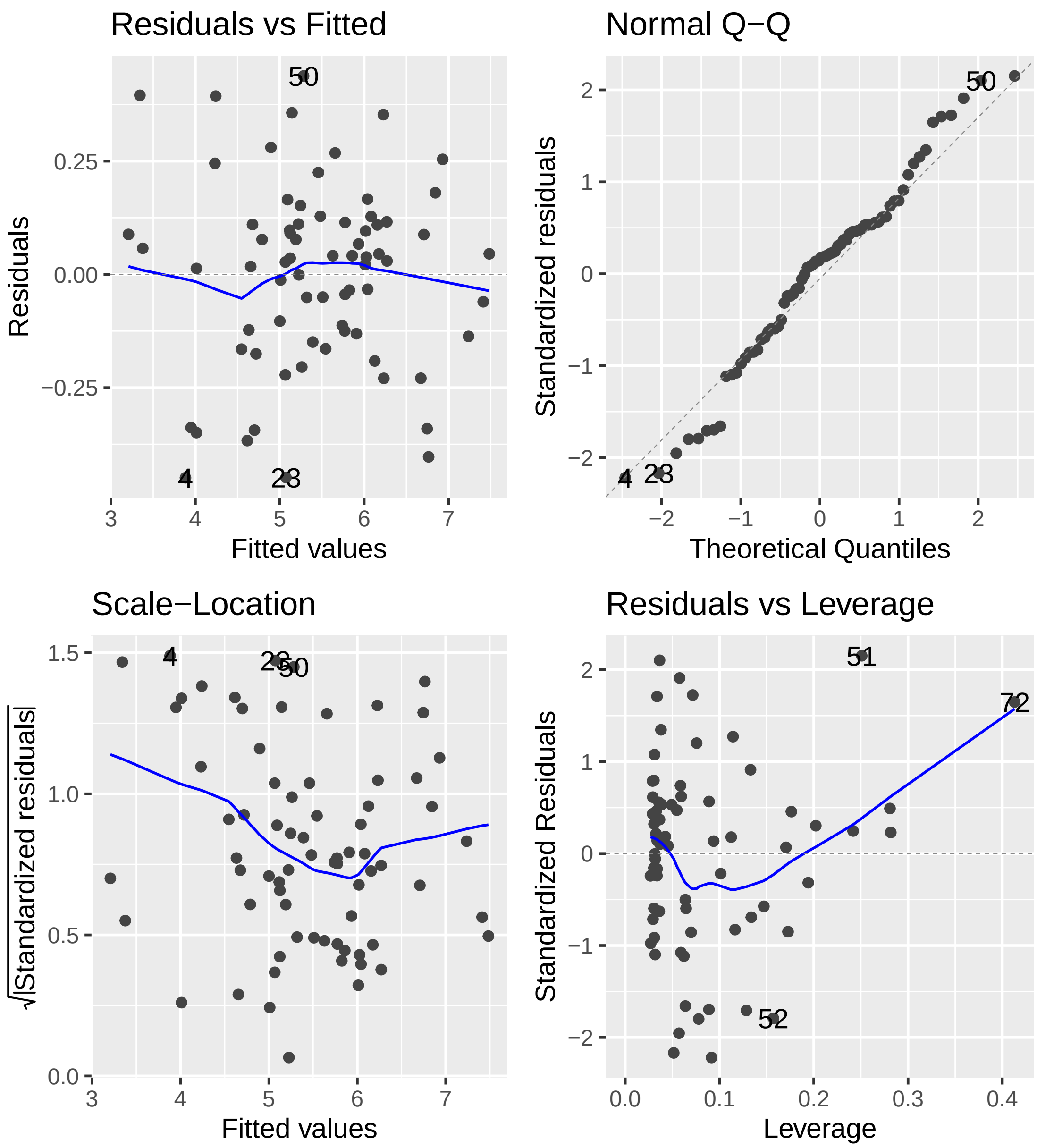
And we further test the linearity of lny, as below:
1. Fitted:

```
RESET test

data:  trans_log
RESET = 2.2191, df1 = 1, df2 = 65, p-value = 0.1411
```

2. Power of Xs:

```
RESET test

data:  trans_log
RESET = 4.5629, df1 = 4, df2 = 62, p-value = 0.002703
```

We cannot reject the hypothesis that there is no omitted variables with fitted test but we strongly reject with the power of regressor variables test. Therefore we can say there is nonlinearity in this model.



The coefficients of the translog function are more complex compared to Cobb-Douglas function which only have three variables. So in this case, since the original Cobb-Douglas function works well enough as a description of the data, the translog function is not necessary even though it is more flexible.

code

```
library(sandwich)
library(ggplot2)
library(lmtest)
library(zoo)
library(corrplot)
library(car)

# 1

# a
gasoline[,1] <- gasoline[,1] - 1960
gasoline[,2] <- gasoline[,2]/gasoline[,11]
```

```r
reg_g <- lm(g ~ year + pg + y + pnc + puc + ppt + pd + pn + ps, data = gasoline)
summary(reg_g)

# b
H_0 <- c('pnc = puc')
linearHypothesis(reg_g, H_0)

# c

ela_pg <- coef(reg_g)[3] * mean(gasoline$pg)/mean(gasoline$g)
ela_y <- coef(reg_g)[4] * mean(gasoline$y)/mean(gasoline$g)
ela_ppt <- coef(reg_g)[7] * mean(gasoline$ppt)/mean(gasoline$g)


# d
log_g <- lm(g ~ log(pg) + year + log(y) + log(pnc) + log(puc)
           + log(ppt) + log(pd) + log(pn) + log(ps), data = gasoline)
summary(log_g)

# 2

# a
reg_va <- nls(valueadd ~ (alpha * labor^beta * capital^gama),
             data = Production, start=list(alpha=1.5, beta=0.5,gama=0.5))
summary(reg_va)

# b
log_va <- lm(log(valueadd) ~ log(labor) + log(capital), data = Production)
summary(log_va)

# c
H_1 = c('log(labor) = 0.6')
H_2 = c('log(labor) + log(capital) = 1')
H_3 = c('log(labor) = 0.6', 'log(labor) + log(capital) = 1')

linearHypothesis(log_va, H_1)
linearHypothesis(log_va, H_2)
linearHypothesis(log_va, H_3)

# d
lny <- log(Production$valueadd)
lnl <- log(Production$labor)
lnk <- log(Production$capital)
lnl2 <- 0.5 * lnl^2
lnk2 <- 0.5 * lnk^2

trans_log <- lm(lny ~ lnl + lnk + lnl2 + lnk2 + lnl * lnk)
summary(trans_log)

library(ggfortify)
autoplot(trans_log)

resettest(trans_log, power = 3, type = 'fitted')
resettest(trans_log, power = 3, type = 'regressor')
```