Assignment 3

Tai Lo Yeung, Pan Yiming, Qiwei Liu

Barcelona Graduate School of Economics

Universitat Pompeu Fabra

Nov 13, 2020

Tests of structural change

a

Estimate the model using the full sample, and then use two subperiods: 1960-1973 and 1974-1995. Perform a test of equality of the coefficients of the two equations using the sums of squares of these three regressions.

In order to test the equality of the coefficients, we need to use this formula $F_{(k,n_1+n_2-2k)} = \frac{[RSS_1-(RSS_2+RSS_3)]/k}{[(RSS_2+RSS_3)/(n_1+n_2-2k)]}$, where RSS_1 is the sum of squares from the full regression and RSS_2, RSS_3 are sum of squares from two separate period regression, n is the sample sizes, k is the number of parameters estimated. By running the code, we obtained:

```
> show(F_vb)
[1] 14.95803
> show(p_vb)
[1] 4.594551e-07
```

We strongly reject the hypothesis that there is no difference between the coefficients of two periods. Therefore, there is a structural change in our case.

b

Now imagine that you know the intercepts to be different, so you want to test for a change in the other five coefficients beyond a simple shift in the constant term. How would you do this? (Hint: still you may want to run the two separate regressions for the two subperiods –"unrestricted" model-, but modify the full-sample regression in order to accommodate in that regression the differing intercepts). Do it.

We can add a dummy variable d here, which is 0 before year 1974 and 1 on and after 1974, since we know the intercepts are different. And we run a regression with it and do the test using this restricted regression with the 2 separate periods regression again, we have:

```
> show(F_d1)
[1] 9.249545
> show(p_d1)
[1] 4.594551e-07
```

As shown above, we cannot reject the hypothesis thus there are a change in the other five coefficients beyond a simple shift in the constant term.

c

Now imagine that you know that the price elasticities on automobile prices and the coefficient of the time trend do not change, but the other three coefficients could change or not. How would you test for this change? Do it.

Similar to what we have done above, this time we only put dummy variables on inputs we want to test:

```
Call:
lm(formula = lng ~ d + I((d * lny)) + I((d * lnpg)) + gasoline$year +
    lny + lnpg + lnpnc + lnpuc)

Residuals:
      Min        1Q    Median        3Q       Max
-0.029130 -0.008614 -0.000336  0.012614  0.028771

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   11.717257   6.125905   1.913 0.066443 .
d              7.866236   2.437511   3.227 0.003269 **
I((d * lny))  -0.853410   0.269992  -3.161 0.003859 **
I((d * lnpg)) -0.384158   0.220767  -1.740 0.093229 .
gasoline$year -0.012906   0.003675  -3.512 0.001583 **
lny            1.517165   0.191995   7.902  1.7e-08 ***
lnpg           0.091661   0.228314   0.401 0.691234
lnpnc          0.593333   0.135015   4.395 0.000155 ***
lnpuc         -0.108076   0.044716  -2.417 0.022680 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01724 on 27 degrees of freedom
Multiple R-squared:   0.99,Adjusted R-squared:  0.9871
F-statistic: 335.2 on 8 and 27 DF,  p-value: < 2.2e-16
```

From the result above, under the assumption that the price elasticities on automobile prices and the coefficient of the time trend do not change, both coefficients of intercept, lny are changed after 1973 at 0.01 significance level, but the coefficient of lnpg doesn't changed because lnpg itself isn't significant at all.

We also run a likelihood test:

```
> show(F_vc)
[1] 21.37061
> show(p_vc)
[1] 1.33705e-07
```

We cannot reject the hypothesis thus this further proved that there are a change in the coefficients beyond a simple shift in the constant term.

On instrumental variables

a

Why do we include the seasonal variables and why do we include only twelve of them?

We include the seasonal variable because grain is a seasonal product, as most of the farm crop. If we do not include the seasonal variables we might misread the trend. We don't include the last seasonal variable because we want to set it as the base.

b

Estimate this demand equation by OLS. What is the estimated value of the price elasticity of demand?

```
Call:
lm(formula = log(quantity) ~ log(price) + ice + seas1 + seas2 +
    seas3 + seas4 + seas5 + seas6 + seas7 + seas8 + seas9 + seas10 +
    seas11 + seas12, data = JEC)

Residuals:
     Min       1Q   Median       3Q      Max
-1.39102 -0.24296  0.06575  0.28284  1.05884

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.861233   0.171361  51.711  < 2e-16 ***
log(price)  -0.638885   0.082389  -7.755 1.26e-13 ***
ice          0.447754   0.119604   3.744 0.000216 ***
seas1       -0.132822   0.110959  -1.197 0.232197
seas2        0.066888   0.111298   0.601 0.548286
seas3        0.111436   0.111308   1.001 0.317527
seas4        0.155422   0.110743   1.403 0.161477
seas5        0.109658   0.129918   0.844 0.399282
seas6        0.046833   0.159596   0.293 0.769377
seas7        0.122552   0.160041   0.766 0.444397
seas8       -0.235008   0.159856  -1.470 0.142533
seas9        0.003561   0.160021   0.022 0.982262
seas10       0.169247   0.161295   1.049 0.294849
seas11       0.215184   0.160096   1.344 0.179890
seas12       0.219633   0.159136   1.380 0.168524
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3973 on 313 degrees of freedom
Multiple R-squared:  0.3126,Adjusted R-squared:  0.2819
F-statistic: 10.17 on 14 and 313 DF,  p-value: < 2.2e-16
```

So the estimated value of the price elasticity of demand is -0.638885, which means when the price of grain increase by 1%, the total tons transported in week i will decrease by 0.6%.

c

Interpret and explain the coefficient β_2.

β_2 is the coefficient of Ice, that is when the Great Lakes are frozen, the total tons transported by railroad in week i increase by approximately 0.45%. This is intuitive because when there is no alternative shipping method, the railroad should carry more grain.

d

Explain why this OLS estimator is probably biased (think “supply and demand”).

The reason OLS estimator is probably biased is because it didn’t consider the fact that both the supply and demand of grain and the variable ice are highly affected by season, which will cause endogeneity. For example, in certain seasons the grain supply is abundant, so the price will be lower compare to other season. And the variable ’ice’ will be 1 only during the cold seasons. Moreover, the cartel will affect the market price too in this dataset. All factors mentioned will lead to a biased result.

e

You are thinking of using the Cartel variable as an IV for ln(P). Use your economic common sense to examine if this variable satisfies the conditions to be a valid instrument.

A valid instrument variable should satisfies the following two requirements:
a. The instrument must be correlated with the endogenous explanatory variables, conditionally on the other covariates. If this correlation is strong, then the instrument is said to have a strong first stage.
b. The instrument cannot be correlated with the error term in the explanatory equation, conditionally on the other covariates. If this condition is met, then the instrument is said to satisfy the exclusion restriction.
In this case, the Cartel is highly correlated with the endogenous variable 'price', and Cartel is not directly related to the total tons transported in week i, it affect the dependent variable through the endogenous variable 'price'. Hence we believe Cartel is a valid instrument and we will test it later.

f

Estimate the first stage regression. Is Cartel a weak instrument? (Present a more or less formal test).

```
Call:
ivreg(formula = log(quantity) ~ log(price) + ice + seas1 + seas2 +
    seas3 + seas4 + seas5 + seas6 + seas7 + seas8 + seas9 + seas10 +
    seas11 + seas12 | ice + seas1 + seas2 + seas3 + seas4 + seas5 +
    seas6 + seas7 + seas8 + seas9 + seas10 + seas11 + seas12 +
    cartel, data = JEC)

Residuals:
     Min       1Q   Median       3Q      Max
-1.38295 -0.27275  0.07318  0.27703  1.09320

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.573535   0.216445  39.611  < 2e-16 ***
log(price)  -0.866587   0.132123  -6.559 2.24e-10 ***
ice          0.422934   0.121569   3.479 0.000575 ***
seas1       -0.130973   0.112307  -1.166 0.244420
seas2        0.090952   0.113167   0.804 0.422181
seas3        0.135872   0.113194   1.200 0.230912
seas4        0.152511   0.112094   1.361 0.174632
seas5        0.073562   0.132494   0.555 0.579148
seas6       -0.006064   0.163277  -0.037 0.970397
seas7        0.060232   0.164392   0.366 0.714319
seas8       -0.293599   0.163930  -1.791 0.074259 .
seas9       -0.058372   0.164343  -0.355 0.722689
seas10       0.085811   0.167514   0.512 0.608831
seas11       0.151791   0.164530   0.923 0.356941
seas12       0.178656   0.162119   1.102 0.271306

Diagnostic tests:
                 df1 df2 statistic p-value
Weak instruments   1 313   207.222  <2e-16 ***
Wu-Hausman         1 312     5.124  0.0243 *
Sargan             0  NA        NA      NA
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4021 on 313 degrees of freedom
Multiple R-Squared: 0.2959,Adjusted R-squared: 0.2644
Wald test: 8.807 on 14 and 313 DF,  p-value: 3.5e-16
```

From the result above, we can tell that the instruments variable is not weak because we reject the F-test on the instruments in the first stage.

g

Estimate the demand equation using instrumental variables. What is the estimated value of demand elasticity now?

After the introduction of instrument variables, the estimated value of demand elasticity now become -0.866587, the absolute value is larger than the one we had before as -0.638885.

h

Carry out and comment on the Hausman test for comparison of the two sets of estimates. Is there evidence of a problem of endogeneity?

Since in the test we done before, we rejected the Hausman test at a 5% significance level, it essentially means that the OLS and IV estimates are not similar, so the endogeneity is a problem according to this test.

i

Do your estimates suggest that the cartel was charging a price that maximized profits (that is, the price that a monopolist would charge? (Hint: what should a monopolist do if price elasticity is lower than one?)

For a monopolist, he don't wants the demand change too much when he lift the price, which means that he needs the price elasticity of demand to be lower than 1. In our test, if we don't include the Cartel instrument, the estimates price elasticity is indeed lower than the one we estimates with the Cartel instrument. So our estimates do suggest that the cartel was charging a price that increased profits but not maximized. If the cartel want to maximized profits it need to adjust the elasticity equal to 1.

code

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

# 1
lng <- log(gasoline$g/gasoline$pop)
lny <- log(gasoline$y)
lnpg <- log(gasoline$pg)
lnpnc <- log(gasoline$pnc)
lnpuc <- log(gasoline$puc)

# a
reg_full <- lm(lng ~ gasoline$year + lny + lnpg + lnpnc + lnpuc)
reg_1973 <- lm(lng ~ gasoline$year + lny + lnpg + lnpnc + lnpuc, subset = (gasoline$year <= 1973))
reg_1974 <- lm(lng ~ gasoline$year + lny + lnpg + lnpnc + lnpuc, subset = (1974 <= gasoline$year))
SSR_0 <- sum(reg_full$residuals^2)
SSR_1 <- sum(reg_1973$residuals^2)
SSR_2 <- sum(reg_1974$residuals^2)

F_vb <- ((SSR_0 - (SSR_1 + SSR_2))/6)/((SSR_1 + SSR_2)/(36-12))
p_vb <- pf(F_vb, 6, 24, lower.tail=F)
show(F_vb)
show(p_vb)

# b
d <- as.numeric(gasoline$year >= 1974)
```

```
reg_d <- lm(lng ~ d + gasoline$year + lny + lnpg + lnpnc + lnpuc)
summary(reg_d)

SSR_d <- sum(reg_d$residuals^2)
F_d1 <- ((SSR_d - (SSR_1 + SSR_2))/6)/((SSR_1 + SSR_2)/(36-12))
p_d1 <- pf(F_vb, 6, 24, lower.tail=F)
show(F_d1)
show(p_d1)

# c
reg_dd <- lm(lng ~ d + I((d*lny)) + I((d*lnpg)) + gasoline$year + lny + lnpg + lnpnc + lnpuc)
summary(reg_dd)

SSR_3 = sum(resid(reg_dd)^2)
F_vc = ((SSR_0 - SSR_3)/3)/(SSR_3/(36-6))
p_vc = pf(F_vc, 3, 30, lower.tail=F)
show(F_vc)
show(p_vc)

# 2

# b
reg_unsea <- lm(log(quantity) ~ log(price) + ice + seas1 + seas2 + seas3 + seas4 + seas5 + seas6 + se
summary(reg_unsea)


# g
library(AER)
reg_sea <-ivreg(log(quantity) ~ log(price) + ice + seas1 + seas2 + seas3 + seas4 + seas5 + seas6 + se
summary(reg_sea, diagnostics=TRUE)
```