Competition in the Dutch coffee market

David Hakula, Christopher Hayes, Filip Mellgren

2019-04-02

library(rio)  
library(tidyverse)  
library(AER)  
library(ggthemes)  
library(stargazer)

df <- import("dutch\_coffee.dta")

df %>% head() %>% as\_tibble()

## # A tibble: 6 x 14  
## maand year month qu cprice tprice incom q1 q2 q3 q4  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 29 1990 1 0.55 12.1 18.6 1641. 1 0 0 0  
## 2 22 1990 2 0.65 12.1 18.6 1539. 1 0 0 0  
## 3 50 1990 3 0.66 12.1 18.6 1681. 1 0 0 0  
## 4 1 1990 4 0.66 12.1 18.6 1656. 0 1 0 0  
## 5 57 1990 5 0.64 12.1 18.6 1701. 0 1 0 0  
## 6 43 1990 6 0.65 12.1 18.6 1733. 0 1 0 0  
## # … with 3 more variables: bprice <dbl>, wprice <dbl>, oprice <dbl>

# Create time variable  
df <- df %>% mutate(time = year + month/12)  
  
# rename variables  
#df <- df %>% rename(Quantity = qu,  
 # `Coffee price` = cprice,  
 # `Tea price` = tprice,  
 # `Wage level` = wprice,  
 # `Bean price` = bprice)  
df <- df %>% rename(`Coffee price` = cprice)

## Summary statistics

We begin by showing som esumamry statistics,

### Summary Table

#summary(df$`Coffee price`)  
sapply(select(df,qu, `Coffee price`, tprice, wprice, bprice), summary)

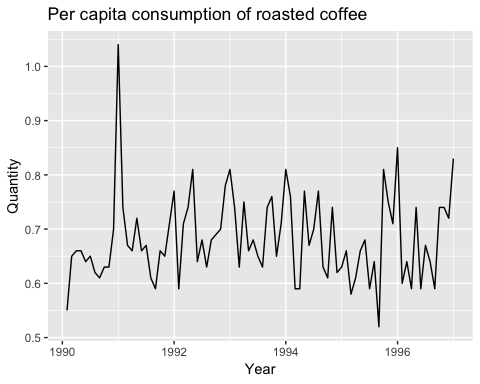
## qu Coffee price tprice wprice bprice  
## Min. 0.5200000 11.00917 16.38983 28.15000 2.280374  
## 1st Qu. 0.6300000 11.45455 17.29825 28.97850 3.002358  
## Median 0.6600000 12.12000 17.56278 29.19489 3.405882  
## Mean 0.6815476 12.82877 17.64858 29.18545 3.676011  
## 3rd Qu. 0.7400000 13.50216 17.93442 29.43054 3.980769  
## Max. 1.0400000 17.69912 18.60396 30.08333 6.353982

cormatdf <- df %>% select(`Coffee price`, qu, bprice, tprice, wprice)

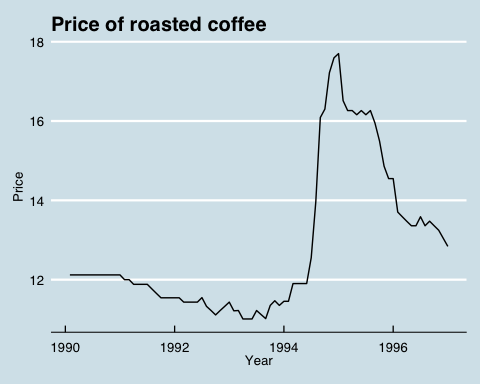
cormat <- round(cor(),2)

### Per capita consumption of roasted coffe and price of roasted coffee

# descriptives are important!  
# Plot all the time series in separate graphs  
# a) per capita consumption of roasted coffee and price of roasted coffee  
df %>% ggplot(aes(x = time, y = qu)) +  
 geom\_line() +  
 labs(y = "Quantity", x = "Year", title = "Per capita consumption of roasted coffee")

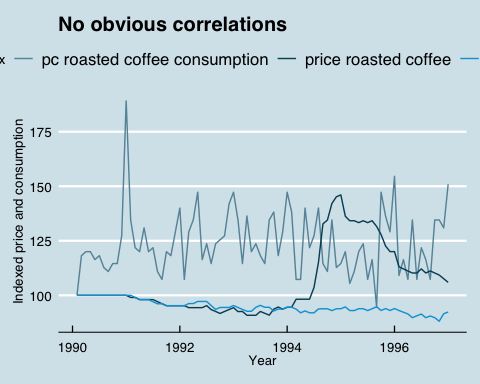


df %>% ggplot(aes(x = time, y = `Coffee price`)) +  
 geom\_line() +  
 theme\_economist() + scale\_colour\_economist() +  
 labs(y = "Price", x = "Year", title = "Price of roasted coffee")



# Coffee and tea and quantity indexed  
df <- df %>% mutate("price roasted coffee" = `Coffee price`/`Coffee price`[1]\*100,  
 "price tea" = tprice/tprice[1]\*100,  
 "pc roasted coffee consumption" = qu/qu[1]\*100,  
 "price coffee beans" = bprice/bprice[1]\*100)   
  
df %>% gather(`price roasted coffee`, `price tea`, `pc roasted coffee consumption`, key = "index", value = "price") %>%   
 ggplot(aes(x = time, y = price, color = index )) +  
 geom\_line() +  
 theme\_economist() + scale\_colour\_economist() +  
 labs(y = "Indexed price and consumption", x = "Year", title = "No obvious correlations")

## Warning: attributes are not identical across measure variables;  
## they will be dropped



# Calculate correlations  
cor(df$`Coffee price`, df$tprice)

## [1] -0.3161684

cor(df$`Coffee price`, df$qu)

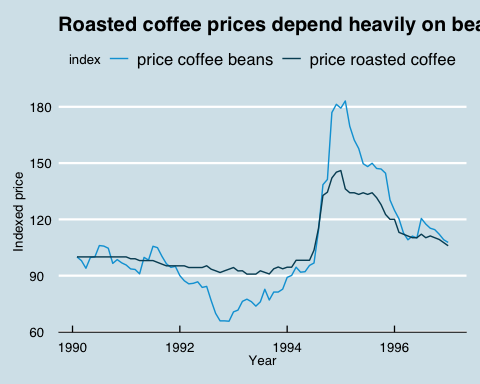
## [1] -0.186854

cor(df$tprice, df$qu)

## [1] -0.01671327

df %>% gather(`price roasted coffee`, `price coffee beans`, key = "index", value = "price") %>%   
 ggplot(aes(x = time, y = price, color = index )) +  
 geom\_line() +  
 theme\_economist() + scale\_colour\_economist() +  
 labs(y = "Indexed price", x = "Year", title = "Roasted coffee prices depend heavily on bean prices")

## Warning: attributes are not identical across measure variables;  
## they will be dropped



### Consumption of roasted coffee and price of tea

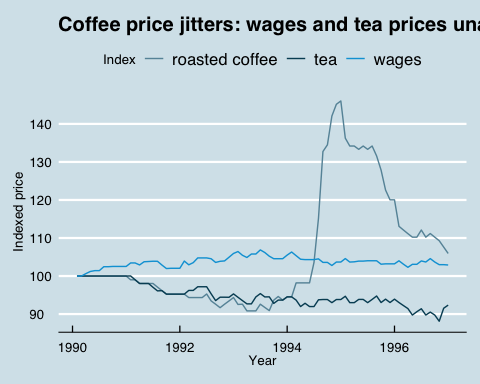
# consumption of roasted coffee and price of tea  
df %>% ggplot(aes(x = time, y = tprice)) +  
 geom\_line() +  
 labs(y = "Price", x = "Year", title = "Price of tea")



### c and d

df <- df %>% mutate("roasted coffee" = `Coffee price`/`Coffee price`[1]\*100,  
 tea = tprice/tprice[1]\*100,  
 wages = wprice/wprice[1]\*100)   
  
df %>% gather(`roasted coffee`, tea, wages, key = "Index", value = "price") %>%   
 ggplot(aes(x = time, y = price, color = Index)) +  
 geom\_line() +  
 theme\_economist() + scale\_colour\_economist() +   
 labs(y = "Indexed price", x = "Year",   
 title = "Coffee price jitters: wages and tea prices unaffected")

## Warning: attributes are not identical across measure variables;  
## they will be dropped



# price of roasted coffee and price of labor  
df %>% ggplot(aes(x = time, y = wprice)) +  
 geom\_line() +  
 labs(y = "Price", x = "Year", title = "Wages over time")



# price of roasted coffee and price of tea

## Regress

df <- df %>% mutate(`Coffee price` = `Coffee price` / oprice,  
 wprice = wprice / oprice,  
 tprice = tprice / oprice)  
  
no\_controls <- lm(log(qu) ~ log(`Coffee price`), data = df)  
quarter\_controls <- lm(log(qu) ~ log(`Coffee price`) + q1 + q2 + q3, data = df)

stargazer(no\_controls, quarter\_controls, header = FALSE)

## Supply and demand shifts

Start from the data that we have.

Valid instrument, corr(zx) > 0, E(z epsilon) = 0. How to find instruments?

Supply shift: wages, prices of beans

Demand shift:

## Log linear demand estimation

$\beta\_1 = \frac{dQ}{dP}\frac{P}$

tea\_control <- lm(log(qu) ~log(`Coffee price`) + q1 + q2 + q3 + log(tprice), data = df)  
income\_control <- lm(log(qu) ~log(`Coffee price`) + q1 + q2 + q3 + log(tprice) +  
 log(incom), data = df)  
  
# seasonal controls summary  
stargazer(no\_controls, quarter\_controls, tea\_control, income\_control,   
 header = FALSE, type = "text")

##   
## ==========================================================================================================  
## Dependent variable:   
## --------------------------------------------------------------------------------------  
## log(qu)   
## (1) (2) (3) (4)   
## ----------------------------------------------------------------------------------------------------------  
## log(`Coffee price`) -0.238\*\* -0.254\*\*\* -0.255\*\*\* -0.270\*\*\*   
## (0.104) (0.094) (0.095) (0.095)   
##   
## q1 -0.127\*\*\* -0.127\*\*\* -0.111\*\*\*   
## (0.030) (0.031) (0.033)   
##   
## q2 -0.092\*\*\* -0.092\*\*\* -0.092\*\*\*   
## (0.030) (0.031) (0.030)   
##   
## q3 -0.118\*\*\* -0.118\*\*\* -0.106\*\*\*   
## (0.030) (0.030) (0.031)   
##   
## log(tprice) -0.015 0.200   
## (0.133) (0.205)   
##   
## log(incom) 0.513   
## (0.374)   
##   
## Constant 0.196 0.319 0.365 -4.051   
## (0.257) (0.234) (0.458) (3.246)   
##   
## ----------------------------------------------------------------------------------------------------------  
## Observations 84 84 84 84   
## R2 0.060 0.265 0.265 0.282   
## Adjusted R2 0.048 0.228 0.218 0.227   
## Residual Std. Error 0.109 (df = 82) 0.098 (df = 79) 0.099 (df = 78) 0.098 (df = 77)   
## F Statistic 5.219\*\* (df = 1; 82) 7.112\*\*\* (df = 4; 79) 5.621\*\*\* (df = 5; 78) 5.052\*\*\* (df = 6; 77)  
## ==========================================================================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## One major concern

ivreg

IV\_spec <- ivreg(log(qu) ~ log(`Coffee price`) + q1 + q2 + q3 + log(tprice) +  
 log(incom) | q1 + q2 + q3 + log(tprice) +  
 log(incom) + log(bprice) + log(wprice), data = df)  
  
stargazer(IV\_spec, income\_control, header = FALSE, type = "text")

##   
## ================================================================  
## Dependent variable:   
## ----------------------------------  
## log(qu)   
## instrumental OLS   
## variable   
## (1) (2)   
## ----------------------------------------------------------------  
## log(`Coffee price`) -0.288\*\*\* -0.270\*\*\*   
## (0.101) (0.095)   
##   
## q1 -0.111\*\*\* -0.111\*\*\*   
## (0.033) (0.033)   
##   
## q2 -0.093\*\*\* -0.092\*\*\*   
## (0.030) (0.030)   
##   
## q3 -0.106\*\*\* -0.106\*\*\*   
## (0.031) (0.031)   
##   
## log(tprice) 0.201 0.200   
## (0.205) (0.205)   
##   
## log(incom) 0.521 0.513   
## (0.374) (0.374)   
##   
## Constant -4.067 -4.051   
## (3.247) (3.246)   
##   
## ----------------------------------------------------------------  
## Observations 84 84   
## R2 0.282 0.282   
## Adjusted R2 0.226 0.227   
## Residual Std. Error (df = 77) 0.098 0.098   
## F Statistic 5.052\*\*\* (df = 6; 77)  
## ================================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

summary(IV\_spec, diagnostics = TRUE)

##   
## Call:  
## ivreg(formula = log(qu) ~ log(`Coffee price`) + q1 + q2 + q3 +   
## log(tprice) + log(incom) | q1 + q2 + q3 + log(tprice) + log(incom) +   
## log(bprice) + log(wprice), data = df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.190660 -0.074534 -0.007248 0.060874 0.329964   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -4.06707 3.24656 -1.253 0.21409   
## log(`Coffee price`) -0.28771 0.10053 -2.862 0.00542 \*\*  
## q1 -0.11077 0.03262 -3.396 0.00108 \*\*  
## q2 -0.09255 0.03046 -3.038 0.00325 \*\*  
## q3 -0.10600 0.03148 -3.367 0.00119 \*\*  
## log(tprice) 0.20076 0.20530 0.978 0.33121   
## log(incom) 0.52098 0.37386 1.394 0.16747   
##   
## Diagnostic tests:  
## df1 df2 statistic p-value   
## Weak instruments 2 76 345.070 <2e-16 \*\*\*  
## Wu-Hausman 1 76 0.314 0.577   
## Sargan 1 NA 0.141 0.707   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.09823 on 77 degrees of freedom  
## Multiple R-Squared: 0.2821, Adjusted R-squared: 0.2262   
## Wald test: 5.081 on 6 and 77 DF, p-value: 0.0001933

## Degree of competition in the Dutch coffee market

c0 = 4  
h = 1.19  
df <- df %>% mutate(c = c0 + h\*bprice) # we know the cost already

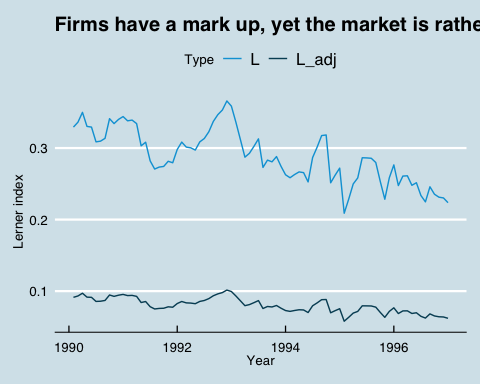
### Lerner index

df <- df %>% mutate(L = (`Coffee price` - c)/`Coffee price`)

### Adjusted Lerner index

eta = -0.27737 # The price elastsicity of demand from the IV  
df <- df %>% mutate(L\_adj = -eta \* L)

# Summary statistics for both and seasonal variation (plot)  
  
df %>% gather(L, L\_adj, key = Type, value = Lerner\_index) %>%  
 ggplot(aes(x = time, y = Lerner\_index, color = Type)) +  
 geom\_line() +  
 theme\_economist() + scale\_colour\_economist() +  
 labs(title = "Firms have a mark up, yet the market is rather competitive",  
 y = "Lerner index", x = "Year")



quarterly\_table <- df %>% group\_by(q1, q2, q3, q4) %>%   
 summarize("mean unadjusted" = mean(L),  
 "mean adjusted" = mean(L\_adj),  
 "std unadjusted" = sqrt(var(L)),  
 "std adjusted" = sqrt(var(L\_adj))) %>%  
 mutate(quarter = case\_when(  
 q1 == 1 ~ "Q1",  
 q2 == 1 ~ "Q2",  
 q3 == 1 ~ "Q3",  
 q4 == 1 ~ "Q4")) %>%  
 as\_tibble()  
  
quarterly\_table <- quarterly\_table %>% select(-q1, -q2, -q3, -q4) %>%  
 select(quarter, everything()) %>%  
 arrange(quarter)  
quarterly\_table

## # A tibble: 4 x 5  
## quarter `mean unadjusted` `mean adjusted` `std unadjusted` `std adjusted`  
## <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 Q1 0.291 0.0808 0.0417 0.0116   
## 2 Q2 0.288 0.0800 0.0275 0.00762  
## 3 Q3 0.290 0.0804 0.0344 0.00955  
## 4 Q4 0.286 0.0794 0.0457 0.0127

## Conduct parameter

Estimate for the entire period , estimate from a regression and solve for lambda:

We estimate the following regression: where is a vector of four quarterly dummies (including all four because we don’t include any intercept).

no\_dummies <- lm(`Coffee price` ~ c + 0, data = df) # plus 0 for no intercept  
q\_dummies <- lm(`Coffee price` ~ c + 0 + q1 + q2 + q3 + q4, data = df) # add controls  
# obtain estimate of b  
b <- no\_dummies$coefficients[1]

We obtain the following estimate for : 1.4 which we use to plug into the following formula:

lambda = -eta \* (b-1)/b

We estimate to be equal to 0.08, which means the market is composed of $ = $ 12.65 equally sized firms.