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)  
library(reshape2)  
library(xtable)

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>

# Introduction

The purpose of this report is to evaluate the Dutch coffee market, on behalf of the Directorate-General for Competition at the European Commission. Through the report, we use monthly data for the Dutch coffee market during the period 1990-1996 as basis for the analysis.

Let us first note that we only have aggregate time-series data. However, we can nonetheless address the question of market concentration indirectly by exploiting the negative relationship it has with price elasticity of demand, and its positive relationship with markups (even in a market with homogenous products).

The Dutch coffee market is dominated by

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

## Summary statistics

To introduce the dataset, we provide summary statistics over the most relevant variables in Table 1 below.

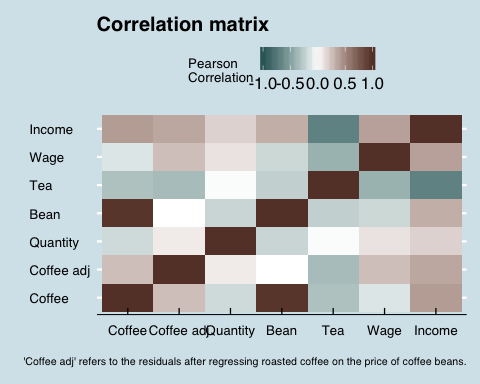
We observe that both roasted coffee prices and quantity consumed skew right, with outlying maximum values.

### Summary Table

#summary(df$Coffee)  
xtable(sapply(select(df,Quantity, Coffee, Tea, Wage, Bean), summary)%>% t())

## % latex table generated in R 3.5.2 by xtable 1.8-3 package  
## % Thu Apr 4 15:56:05 2019  
## \begin{table}[ht]  
## \centering  
## \begin{tabular}{rrrrrrr}  
## \hline  
## & Min. & 1st Qu. & Median & Mean & 3rd Qu. & Max. \\   
## \hline  
## Quantity & 0.52 & 0.63 & 0.66 & 0.68 & 0.74 & 1.04 \\   
## Coffee & 11.01 & 11.45 & 12.12 & 12.83 & 13.50 & 17.70 \\   
## Tea & 16.39 & 17.30 & 17.56 & 17.65 & 17.93 & 18.60 \\   
## Wage & 28.15 & 28.98 & 29.19 & 29.19 & 29.43 & 30.08 \\   
## Bean & 2.28 & 3.00 & 3.41 & 3.68 & 3.98 & 6.35 \\   
## \hline  
## \end{tabular}  
## \end{table}

tmp <- lm(Coffee ~ Bean, data = df)  
df$`Coffee adj` <- tmp$residuals  
cormatdf <- df %>% select(Coffee, `Coffee adj`, Quantity, Bean, Tea, Wage, Income)  
cormat <- round(cor(cormatdf),2)  
  
melted\_cormat <- melt(cormat)  
  
melted\_cormat %>% ggplot(aes(x=Var1, y=Var2, fill=value)) +   
 geom\_tile() + theme\_economist() + scale\_colour\_economist() +  
 labs(title = "Correlation matrix", y = " ", x = " ",  
 caption = "'Coffee adj' refers to the residuals after regressing roasted coffee on the price of coffee beans.") +  
 scale\_fill\_gradient2(low = "#336666", high = "#664033", mid = "white",   
 midpoint = 0, limit = c(-1,1), space = "Lab",   
 name="Pearson\nCorrelation")

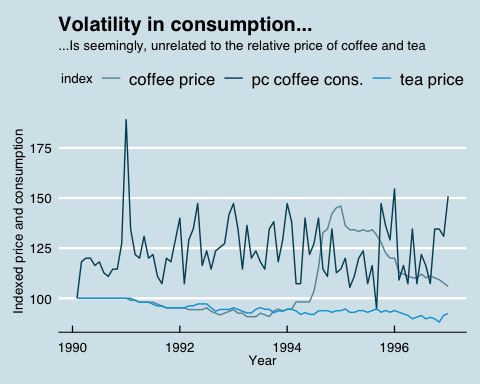
 Look at legend value separation

### Consumption of roasted coffee

The consumption of roasted coffee is largely independent of the prices of coffee and tea.

# Coffee and tea and quantity indexed  
df <- df %>% mutate("coffee price" = Coffee/Coffee[1]\*100,  
 "tea price" = Tea/Tea[1]\*100,  
 "pc coffee cons." = Quantity/Quantity[1]\*100,  
 "price coffee beans" = Bean/Bean[1]\*100)   
  
df %>% gather(`coffee price`, `tea price`, `pc coffee cons.`, 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 = "Volatility in consumption...", subtitle = "...Is seemingly, unrelated to the relative price of coffee and tea")

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



# Calculate correlations  
cor(df$Coffee, df$Tea)

## [1] -0.3161684

cor(df$Coffee, df$Quantity)

## [1] -0.186854

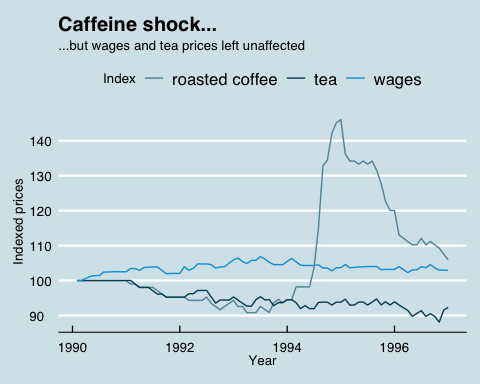
cor(df$Tea, df$Quantity)

## [1] -0.01671327

### c and d

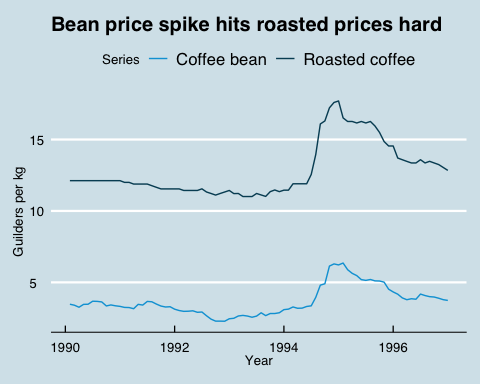
df <- df %>% mutate("roasted coffee" = Coffee/Coffee[1]\*100,  
 tea = Tea/Tea[1]\*100,  
 wages = Wage/Wage[1]\*100)   
  
df %>% gather(`roasted coffee`, tea, wages, key = "Index", value = "price") %>%   
 ggplot(aes(x = time, y = price, color = Index)) +  
 geom\_line() +  
 theme\_economist() + scale\_colour\_economist()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

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



df %>% rename(`Roasted coffee` = Coffee, `Coffee bean` = Bean) %>% gather(`Roasted coffee`, `Coffee bean`, key = "Series",   
 value = "price") %>%   
 ggplot(aes(x = time, y = price, color = Series )) +  
 geom\_line() +  
 theme\_economist() + scale\_colour\_economist() +  
 labs(y = "Guilders per kg", x = "Year",   
 title = "Bean price spike hits roasted prices hard")

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



# Percentage change in prices  
max(df$Coffee)/df$Coffee[55]

## [1] 1.268267

max(df$Bean)/df$Bean[54]

## [1] 1.892676

# Regressions and demand estimation

We begin by inflation adjust the variables before running the regressions.

df <- df %>% mutate(Coffee = Coffee / oprice,  
 Wage = Wage / oprice,  
 Tea = Tea / oprice)  
  
no\_controls <- lm(log(Quantity) ~ log(Coffee), data = df)  
quarter\_controls <- lm(log(Quantity) ~ log(Coffee) + q1 + q2 + q3, data = df)

We specify four different OLS regressions, each with a separate set of control variables specified in table (XXX). We regress logged values against logged values which allows us to directly interpret the coefficient estimates as elasticities. Note that this specification assumes the elasticity of demand to be constant.

First, we run a regression without any controls

Second,

Where are quarterly dummies.

Third,

where denotes tea.

Finally, we run the regression

tea\_control <- lm(log(Quantity) ~log(Coffee) + q1 + q2 + q3 + log(Tea), data = df)  
income\_control <- lm(log(Quantity) ~log(Coffee) + q1 + q2 + q3 + log(Tea) +  
 log(Income), data = df)

## Shift in supply and demand

From the variables visible to us, we would intuitively expect the demand curve to be shifted by the price of substitutes (tea) and consumer income. We would expect the supply curve to be shifted by the prices of inputs (beans and labour) and the season, which might affect yield.

Supply shifts

* Wages
* Prices of beans

Demand shifts

* Price of substitutes, tea
* Income of consumers

We expect that tea and coffee are substitutes, meaning that… Additionally, we anticipate that the level of income can have an effect on coffe consumption. We expect the wages of roasters constitute a small share of the overall income…

Comment on correlation between these two and the adjusted coffee

Question how exogenous they are

Economic theory

## Overcoming endogeneity

Describe problem of omitted variable bias (what variables are omitted).

What have we observed?

Things that increase both quantity demanded and price for consumers (quality)

Tea being endogenous

Complement

Also, reverse causality

To overcome the plausible issue of having endogeneity in the error term, we use the logarithm of bean prices together with the logarithm of wages, potential supply shifters, as instruments for the logarithm of the price of roasted coffee.

These two instruments are likely to fulfill both the relevance, and the validity conditions required of excluded instruments. That the bean prices is a relevant instrument is clear by figure ???, whereas there is only a moderate correlation between wages and the price of roasted coffee beans. Either way, it is noted that the first stage F statistic is 345 which is well above the requirements.

Arguing for validity is more involved. Having valid instruments means the instruments are exogenous to the error term, conditional on the set of control variables included in the regression specification. In our case, it means that the bean price and the wage must be independent of the quantity of coffee consumed. This could be violated if, for instance, customer taste for coffee increased which lead to an increase in coffee consumed; or if higher wages of workers was associated with a change in coffee consumption (for example if coffee workers comprise a large proportion of the consumer base). On the other hand, investigating what caused the spike in the bean prices, it becomes evident that this event was driven by a series of severe frosts in Brazil (normally accounting for 25-35% of the world’s coffee) that June and July 1994, which is independent of quantity consumed. It could still be the case that the instrument is exogenous… but with the Sargan statistic being non significant, both instruments lead to similar estimates, strengthening the case that they indeed provide consistent estimates of elasticity of demand.

Note that we do not use the demand shifters as instruments since these shift the demand curve that we want to pin down.

Mention that elasticity is constant?

Idea: show graph of wage correlation with roasted coffe prices controlling for the controls in th IV specification

First stage F statistic: 345 > 10: strong first stage.

Wu Hausman: 0.577. Means we cannot reject the null hypothesis of the OLS and IV estimates being the same.

Sargan: 0.707 Insignificant meaning we cannot reject the null that the instruments lead to the same estimate.

Interpret the result. Mention Wu Hausman and interpret the elasticity and its magnitude.

Mention that its inelastic.

IV\_spec <- ivreg(log(Quantity) ~ log(Coffee) + q1 + q2 + q3 + log(Tea) +  
 log(Income) | q1 + q2 + q3 + log(Tea) +  
 log(Income) + log(Bean) + log(Wage), data = df)  
summary(IV\_spec, diagnostics = TRUE)

##   
## Call:  
## ivreg(formula = log(Quantity) ~ log(Coffee) + q1 + q2 + q3 +   
## log(Tea) + log(Income) | q1 + q2 + q3 + log(Tea) + log(Income) +   
## log(Bean) + log(Wage), 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) -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(Tea) 0.20076 0.20530 0.978 0.33121   
## log(Income) 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

stargazer(no\_controls, quarter\_controls, tea\_control,   
 income\_control, IV\_spec,   
 column.sep.width = "2pt",   
 font.size = "normalsize",   
 omit.stat = c("f", "ser", "adj.rsq"),   
 no.space = TRUE, header = FALSE)

## Degree of competition in the Dutch coffee market

Next, we move on to estimate the degree of competition. Since the technology of the process of converting raw beans into roasted coffee

Emphasize the log log specification and constant elasticty of demand

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

### Adjusted and unadjusted Lerner indices

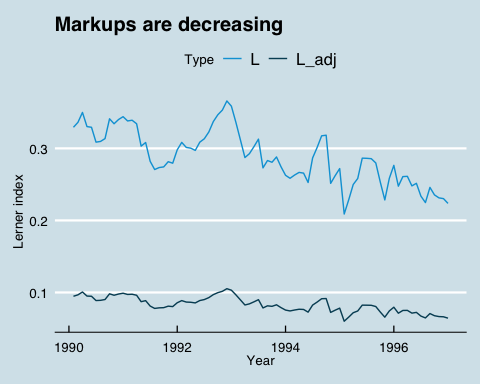
We calculate:

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

We use the estimate of the elasticity from the IV specification despite the fact that the Wu Hausman indicates that the efficient estimate from the OLS is not significantly different. We do this because of the concern that there still may be a slight bias in the OLS and because the standard errors are not very different.

eta = IV\_spec[[1]][2] # The price elastsicity of demand from the IV  
df <- df %>% mutate(L\_adj = -eta \* L) #invert the sign

# 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 = "Markups are decreasing",  
 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)  
library(xtable)  
xtable(quarterly\_table, digits = 3)

## % latex table generated in R 3.5.2 by xtable 1.8-3 package  
## % Thu Apr 4 15:56:10 2019  
## \begin{table}[ht]  
## \centering  
## \begin{tabular}{rlrrrr}  
## \hline  
## & quarter & mean unadjusted & mean adjusted & std unadjusted & std adjusted \\   
## \hline  
## 1 & Q1 & 0.291 & 0.084 & 0.042 & 0.012 \\   
## 2 & Q2 & 0.288 & 0.083 & 0.027 & 0.008 \\   
## 3 & Q3 & 0.290 & 0.083 & 0.034 & 0.010 \\   
## 4 & Q4 & 0.286 & 0.082 & 0.046 & 0.013 \\   
## \hline  
## \end{tabular}  
## \end{table}

From table XXX, we conclude that there are no seasonal effects in the Lerner index

## Conduct parameter

Estimate for the entire period

Next,

* ,
* , estimate from a regression and solve for lambda:

We estimate the following regression: where is a vector of four quarterly dummies (including three so as to not include an intercept).

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

We obtain the following estimate for : 1.38 which we use to plug into the following formula:

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

We estimate 0.079, which means the market is composed of

12.69 equally sized firms.

Index of market power. Assuming identical firms, lambda is the inverse of number of identical firms.

This market can then be said to be equivalent to N Cournot competitors.