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>

# 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

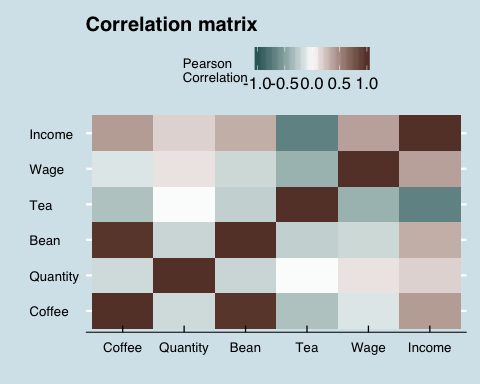
We begin by showing som esumamry statistics,

### 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  
## % Tue Apr 2 22:37:44 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}

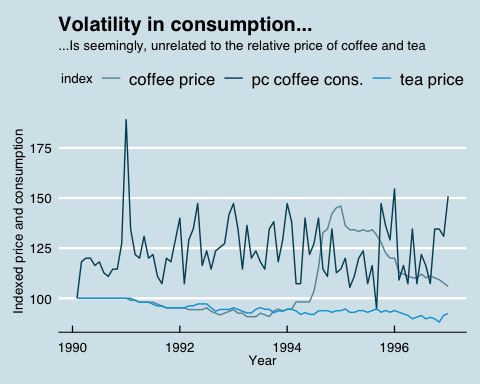
cormatdf <- df %>% select(Coffee, Quantity, Bean, Tea, Wage, Income)  
cormat <- round(cor(cormatdf),2)  
  
melted\_cormat <- melt(cormat)  
  
melted\_cormat %>% ggplot(aes(x=Var1, y=Var2, fill=value)) +   
 geom\_tile() + theme\_economist() + scale\_colour\_economist() +  
 labs(title = "Correlation matrix", y = " ", x = " ") +  
 scale\_fill\_gradient2(low = "#336666", high = "#664033", mid = "white",   
 midpoint = 0, limit = c(-1,1), space = "Lab",   
 name="Pearson\nCorrelation")



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

# 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)

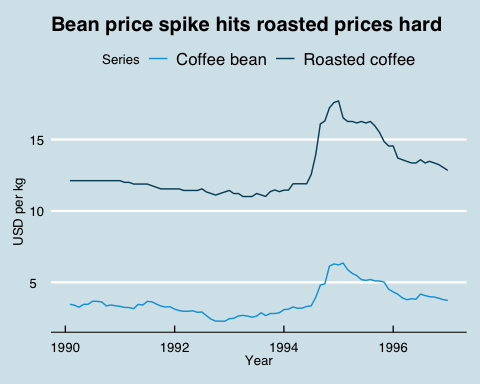
## [1] -0.186854

cor(df$Tea, df$Quantity)

## [1] -0.01671327

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 = "USD per kg", x = "Year",   
 title = "Bean price spike hits roasted prices hard")

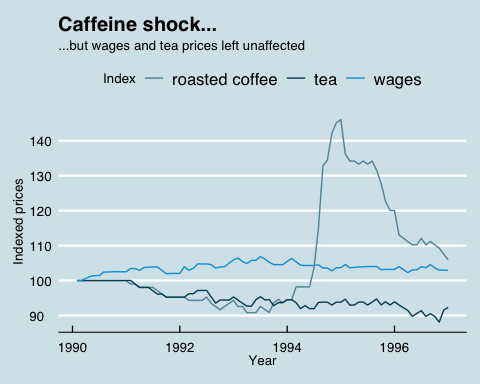
## Warning: attributes are not identical across measure variables;  
## they will be dropped



### 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



## Regressions and demand estimation

We begin by inflation adjust the variables.

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)

First, we run a regression without any controls

Second

Where are quarterly dummies.

Third,

where denotes tea.

Finally, we run the regression

Note to ourselves: following probably unecesary

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

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)  
  
# seasonal controls summary

## Supply and demand shifters

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:

## One major concern

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)  
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)

##   
## \begin{table}[!htbp] \centering   
## \caption{}   
## \label{}   
## \normalsize   
## \begin{tabular}{@{\extracolsep{2pt}}lccccc}   
## \\[-1.8ex]\hline   
## \hline \\[-1.8ex]   
## & \multicolumn{5}{c}{\textit{Dependent variable:}} \\   
## \cline{2-6}   
## \\[-1.8ex] & \multicolumn{5}{c}{log(Quantity)} \\   
## \\[-1.8ex] & \multicolumn{4}{c}{\textit{OLS}} & \textit{instrumental} \\   
## & \multicolumn{4}{c}{\textit{}} & \textit{variable} \\   
## \\[-1.8ex] & (1) & (2) & (3) & (4) & (5)\\   
## \hline \\[-1.8ex]   
## log(Coffee) & $-$0.238$^{\*\*}$ & $-$0.254$^{\*\*\*}$ & $-$0.255$^{\*\*\*}$ & $-$0.270$^{\*\*\*}$ & $-$0.288$^{\*\*\*}$ \\   
## & (0.104) & (0.094) & (0.095) & (0.095) & (0.101) \\   
## q1 & & $-$0.127$^{\*\*\*}$ & $-$0.127$^{\*\*\*}$ & $-$0.111$^{\*\*\*}$ & $-$0.111$^{\*\*\*}$ \\   
## & & (0.030) & (0.031) & (0.033) & (0.033) \\   
## q2 & & $-$0.092$^{\*\*\*}$ & $-$0.092$^{\*\*\*}$ & $-$0.092$^{\*\*\*}$ & $-$0.093$^{\*\*\*}$ \\   
## & & (0.030) & (0.031) & (0.030) & (0.030) \\   
## q3 & & $-$0.118$^{\*\*\*}$ & $-$0.118$^{\*\*\*}$ & $-$0.106$^{\*\*\*}$ & $-$0.106$^{\*\*\*}$ \\   
## & & (0.030) & (0.030) & (0.031) & (0.031) \\   
## log(Tea) & & & $-$0.015 & 0.200 & 0.201 \\   
## & & & (0.133) & (0.205) & (0.205) \\   
## log(Income) & & & & 0.513 & 0.521 \\   
## & & & & (0.374) & (0.374) \\   
## Constant & 0.196 & 0.319 & 0.365 & $-$4.051 & $-$4.067 \\   
## & (0.257) & (0.234) & (0.458) & (3.246) & (3.247) \\   
## \hline \\[-1.8ex]   
## Observations & 84 & 84 & 84 & 84 & 84 \\   
## R$^{2}$ & 0.060 & 0.265 & 0.265 & 0.282 & 0.282 \\   
## \hline   
## \hline \\[-1.8ex]   
## \textit{Note:} & \multicolumn{5}{r}{$^{\*}$p$<$0.1; $^{\*\*}$p$<$0.05; $^{\*\*\*}$p$<$0.01} \\   
## \end{tabular}   
## \end{table}

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

## Degree of competition in the Dutch coffee market

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

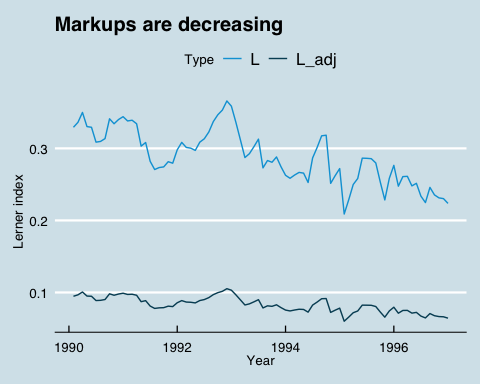
### Adjusted and unadjusted Lerner indices

We calculate:

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

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  
## % Tue Apr 2 22:37:49 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}

## 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 ~ c + 0, data = df) # plus 0 for no intercept  
dummies <- lm(Coffee ~ c + 0 + q1 + q2 + q3 + q4 +   
 log(Tea) + log(Income), data = df) # add controls  
# obtain estimate of b  
b <- dummies$coefficients[1]

We obtain the following estimate for : 1.23 which we use to plug into the following formula:

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

We estimate $= $ 0.053, which means the market is composed of

18.73 equally sized firms.