Assignment Empirical Industrial Economics

Coraline Best

February 2023

1 Summary statistics

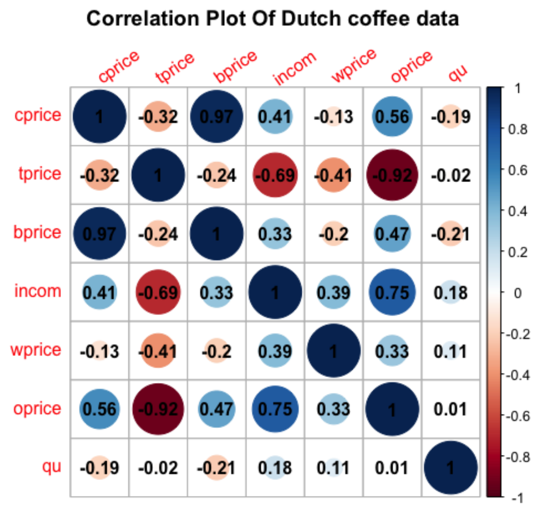
This is a panel data, recording monthly and extending for seven years from 1990 to 1996, which gives us 84 observations in total. We start this assignment by creating some summary statistics of the data that we are using. The following summary table we use mean and standard deviation to capture basic distribution of sampling. For instance, we see that the average price of roasted coffee per kg is 12.83€.

Descriptive statistics of the Dutch coffee market, 1990-1996

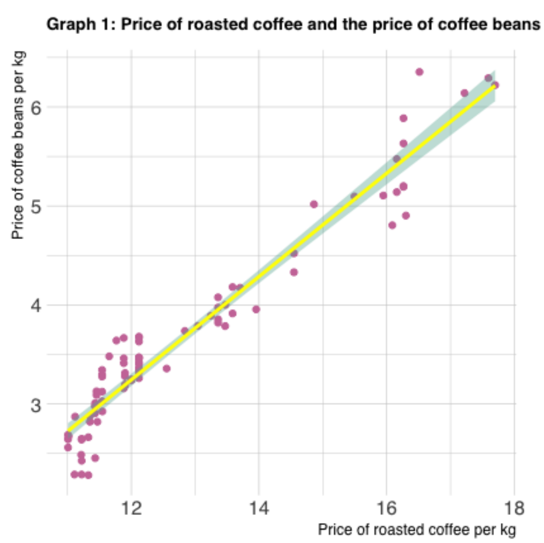
Variable	N	N = 84[1]
Per capita consumption of roasted coffee in kg	84	0.68 (0.08)
Price of roasted coffee per kg	84	12.83 (1.86)
Price of tea per kg	84	17.65 (0.55)
Income per capita	84	1,793.61 (85.54)
Price of coffee beans per kilo	84	3.68 (1.00)
Wage for an individual (monthly)	84	29.19 (0.38)
Price index for other goods	84	1.09 (0.06)

[1] Mean (SD)

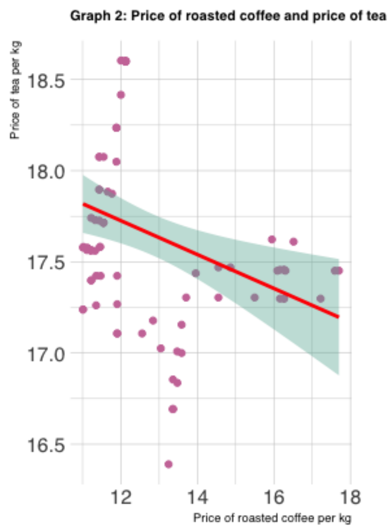
The following graph is a correlation plot of the data. It is very useful because it allows us to see the level of correlation between the variables. On the diagonal we see that the coefficient is equal to 1, and it is very intuitive because, for instance, the price of tea has a perfect correlation with itself. Observing this graph, we find out that there is quite a high correlation between the price of roasted coffee and the price of coffee beans (0.97) and it is very close to 1, meaning there is almost perfect correlation. We also see that the price of roasted coffee is moderately correlated with the price index of other goods and low correlation with the income per capita, but has negative and low correlation with the price of tea.

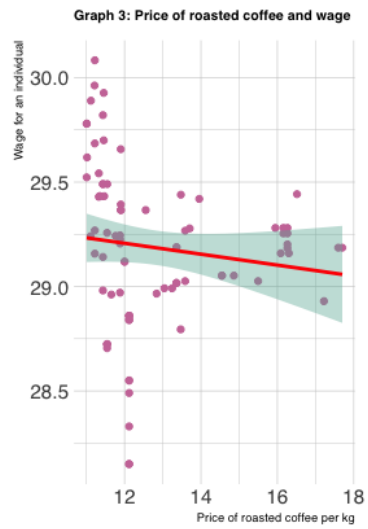


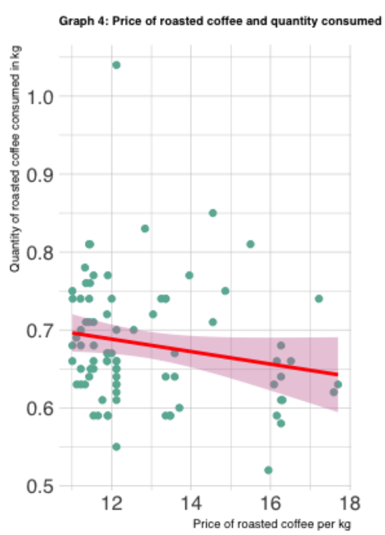
Because I found that there was a high correlation between the price of roasted coffee and the price of coffee beans I found it interesting to represent this relationship visually. We can see very well on the graph below that the two variables are positively correlated.

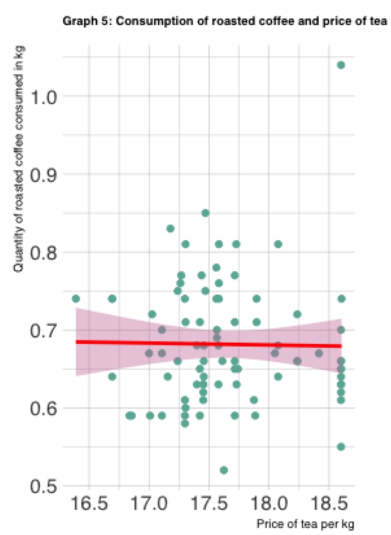


The following graphs represent relationships between the price of roasted coffee and the price of tea (graph 2), the price of roasted coffee and wages (graph 3), the price of roasted coffee and the quantity of roasted coffee consumed (graph 4), the quantity of roasted coffee consumed and the price of tea (graph 5).

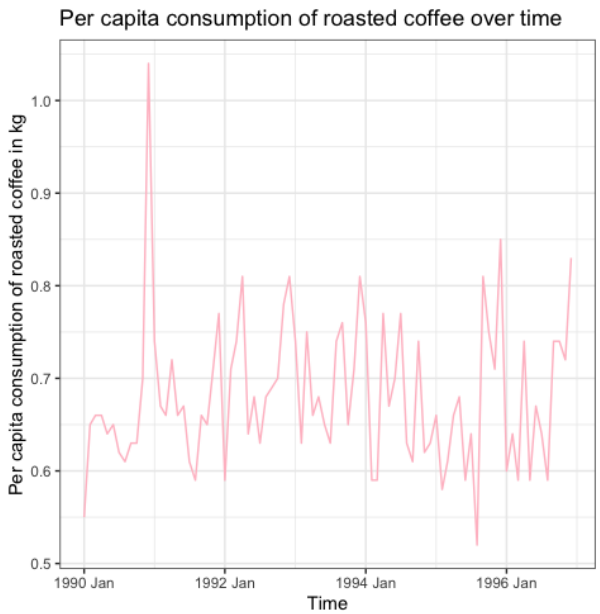


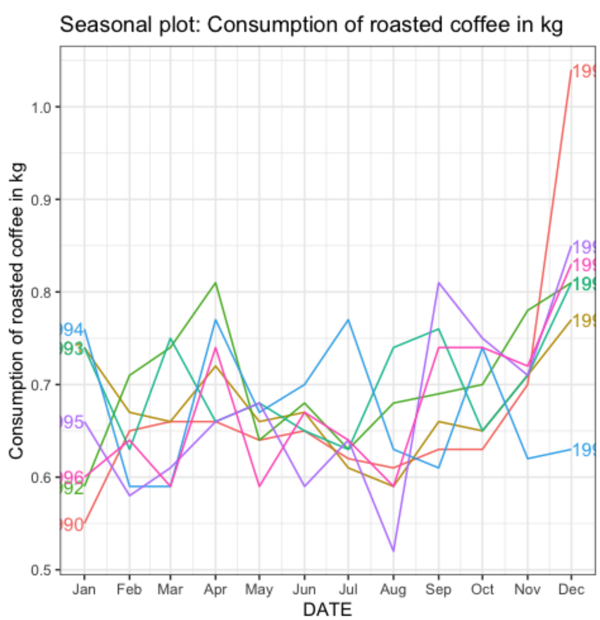






In order to see if there are any time trends in the data, I have tried to represent two graphs. The first one simply represents the amount of coffee consumed over time and the second one is a seasonal plot.





Looking at these two graphs, it seems like there are no long-time trends but strong seasonality within the data.

2 Simple Linear Regression

In this part I run the following simple linear regression:

$$log(CoffeeConsumption) = \beta_0 + \beta_1 log(CoffeePrice) + \epsilon$$

The results I obtain are presented in the following table:

Table 1: Estimates using simple OLS regression

	Dependent variable:
	Per capita consumption of roasted coffee in kg
Price of roasted coffee per kg	-0.165*
	(0.089)
(Intercept)	0.030
	(0.227)
Observations	84
R^2	0.040
Adjusted R^2	0.028
Residual Std. Error	0.110 (df = 82)
F Statistic	3.420* (df = 1; 82)
Note:	*p<0.1; **p<0.05; ***p<0.01

The logarithm could be a good specification because it indicates how much the quantity of coffee consumed would change given a one percent increase in price of coffee. Indeed β_1 in this specification can be interpreted as the elasticity of demand. It is therefore useful to use log transformation and it is also efficient in transforming a highly skewed variable to be more normalized.

I see from the above table that one percent increase in the price of coffee leads to a 0.165 percent decrease in the quantity of coffee consumed. However, this estimate is not significant at 5 percent level of significance. I do not have enough evidence in the sample to reject the null hypothesis. Its t-statistic = -1.849 and the corresponding p-value = 0.0680 inferior to 0.05 which further proves that $\beta_1 = -0.165$ is not significant at 5 percent significance level.

It is generally difficult to judge a causal relationship between two variables. No single method can establish a causal relationship and it is important to use multiple methods to prove that a relationship is causal. The ideal setting would be to conduct a randomized control trial to establish a causal relationship between two given variables, but this is not the case in our exercise. Indeed, we have simply run a simple linear regression without any control. It is then very likely that we have omitted variable bias, and that other variables affect the amount of roasted coffee consumed and are not accounted for in this model. As a result, β_1 is probably biased and does not tell that there is a causal relationship between coffee prices and quantities consumed.

Because I have strong seasonality, I need to account for that and that is why in the following table I conduct a linear regression adding quarter dummy variables.

We now run the following regression:

$$log(CoffeeConsumption) = \beta_0 + \beta_1 log(CoffeePrice) + \beta_2 Q1 + \beta_3 Q2 + \beta_4 Q3 + \epsilon$$

Table 2: Estimates using simple OLS regression adding quarter dummies

	Dependent variable:
	Per capita consumption of roasted coffee in kg
Price of roasted coffee per kg	-0.191**
	(0.081)
(Intercept)	-0.129***
	(0.031)
q2	-0.093***
	(0.031)
q3	-0.119***
	(0.031)
Constant	0.181
	(0.208)
Observations	84
R^2	0.250
Adjusted R^2	0.212
Residual Std. Error	0.099 (df = 79)
F Statistic	6.581*** (df = 4; 79)
Note:	*p<0.1; **p<0.05; ***p<0.01

I omit Quarter 4 in order to avoid perfect collinearity.

The effect of coffee price (log) becomes significant at 5 percent level after adding quarter dummies. All the dummy variables have significant implication at 1 percent level.

The outcome shows that there is significant quarter/seasonal variation of roasted coffee consumption. By adding quarter dummies to the model, the confounding effect is better controlled. Coffee price is significantly associated with different quarters as well as coffee consumption. In the first model, coefficient of coffee price reflects the effect of omitted quarter variables in addition to its own effect. Now that quarter variations are controlled, the coffee price variable no longer captures partial effect of quarters, now reflect its own significant correlation with coffee consumption.

By controlling for quarters, β_1 has decreased to $\beta_1 = -0.191$. It is now significant at 5 percent significance level. This indicates that the new β_1 is able to better capture the percentage change in quantity of roasted coffee consumed when there is a 1 percent increase in coffee price.

I still think the relationship is not causal because it is really likely that there are other omitted variables.

3 Factors that shift demand or supply

Apart from roasted coffee price, there are likely other factors that affect the quantity of coffee consumed. Some of these include consumers' income and price of tea (which can be seen as a substitute). In addition, the coffee bean price is likely to correlate with coffee price. The inflation and the price index of other goods are also very likely to shift the supply and demand.

4 Re-estimating the demand adding controls

In this part I run a model adding some controls. I run now the following regression:

$$log(Coffee) = \beta_0+\beta_1 log(CoffeePrice)+\beta_2 Q1+\beta_3 Q2+ \beta_4 Q3+\beta_5 log(wage)+\beta_6 log(incom)+\beta_7 log(tprice)+\epsilon$$

The following table displays the result:

Table 3: Estimates using OLS regression adding controls

	Dependent variable:
	Per capita consumption of roasted coffee in kg
Price of roasted coffee per kg	-0.234**
	(0.097)
Wage for an individual	0.609
	(1.030)
Income per capita	0.523
	(0.384)
Price of tea per kg	0.338
	(0.517)
q1	-0.113***
	(0.033)
q2	-0.096***
	(0.031)
q3	-0.108***
	(0.032)
Constant	-6.658
	(4.775)
Observations	84
R^2	0.285
Adjusted R^2	0.219
Residual Std. Error	0.099 (df = 76)
F Statistic	4.325*** (df = 7; 76)
Note:	*p<0.1; **p<0.05; ***p<0.01

In relation to 1.2, the estimate on the coefficient of logprice has become significant. Its standard error has also decreased slightly, which means the estimate is relatively better. The R-sqr has improved quite greatly, which means that the new model has a better fit. Adding control variables in the regression model allows omitted variable bias to be accounted for to a certain degree.

5 Capturing the causal effect using instrumental variable

The estimates are still not causal because I still have the presence of omitted variable bias and I am dealing with endogeneity problem, and β_1 is not causal because it is likely to be biased. In addition, the price of coffee beans, as I have seen in the summary statistics, is highly correlated with the price of roasted coffee. As a result, I thought it would be a good idea to run a regression model using an instrumental variable. Instrumental variable is commonly used to deal with endogeneity problem. In this case our instrument will be the price of coffee beans. Our instrument satisfies all the good conditions to be a good instrument variable: price of coffee been is highly correlated with the price of roasted coffee, and the price of coffee bean is not correlated to the disturbance term. We can see that with the first stage regression; the results are reported in the table below:

Table 4: First stage regression results

	Dependent variable:
	Per capita consumption of roasted coffee in kg
Price of coffee beans per kg	0.505***
	(0.018)
q1	-0.014
	(0.013)
q2	-0.017
	(0.013)
q3	-0.004
	(0.013)
Constant	1.910***
	(0.025)
Observations	84
R^2	0.907
Adjusted R^2	0.902
Residual Std. Error	0.042 (df = 79)
F Statistic	192.119*** (df = 4; 79)
Note:	*p<0.1; **p<0.05; ***p<0.01

That is why the instrument is relevant to our problem. The coefficient for the price of coffee beans is very significant (at the 1 percent level) and the value of the R squared is also pretty high (0.907).

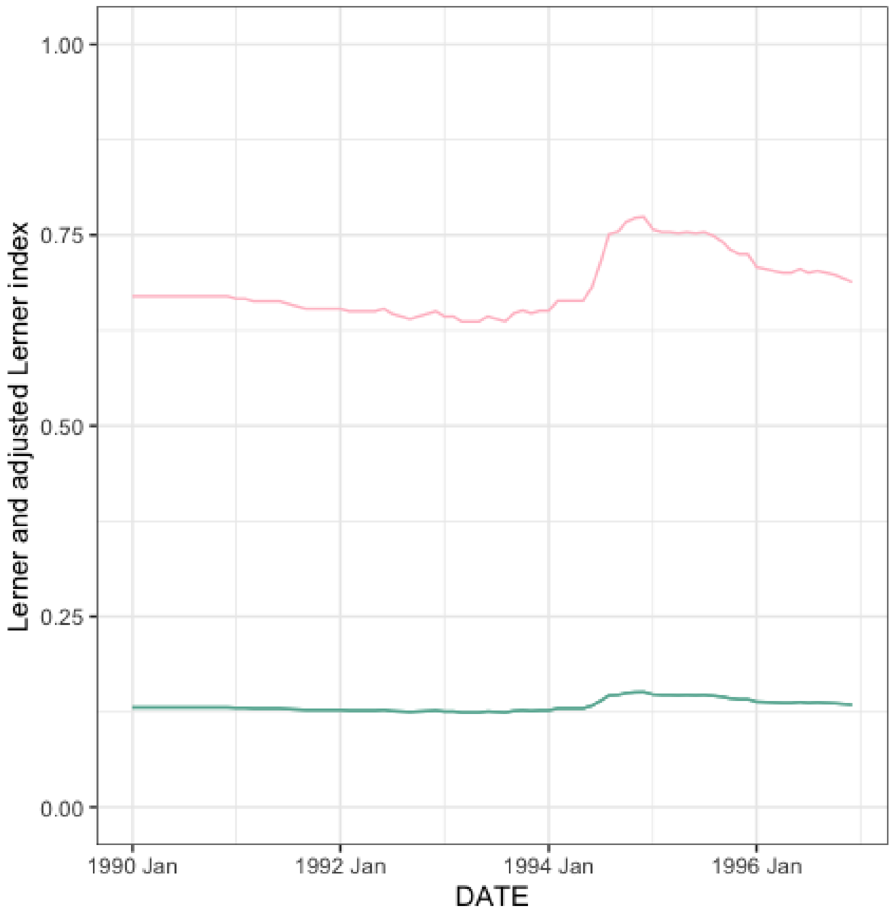
Table 5: Second stage regression results

	Dependent variable:
	Per capita consumption of roasted coffee in kg
Price of roasted coffee per kg	-0.195**
	(0.092)
Constant	0.106
	(0.235)
Observations	84
R^2	0.039
Adjusted R^2	0.027
Residual Std. Error	0.110 (df = 82)
Note:	*p<0.1; **p<0.05; ***p<0.01

β_1 is equal -0.195 and I can now say that β_1 is unbiased and that captures the causal effect of prices on quantity of roasted coffee. Because our β_1 can be interpreted in terms of price elasticity, we can say that the consumers are not very price elastic. Indeed, it is the contrary, the demand for roasted coffee is quite inelastic.

6 Lerner index

For this problem, I am going to use the elasticity I found in problem 1.5 as it is the best estimate. That is, for the adjusted Lerner index I am going to use the coefficient -0.195 for the elasticity.



The Lerner index is represented by the pink curve and the adjusted Lerner index is the green curve. As I know, the Lerner index is a measure of the market power. The adjusted Lerner index is just an extension of the Lerner index that account for elasticity. Therefore I can say that the adjusted Lerner index is displaying a more comprehensive measure of market power. If I look at the first graph more carefully, I can see after January 1994 the Lerner index rocketed and reached the highest level close to 1 in the seven years, which implies a peak level of market power in this market. The more the values of the adjusted Lerner index are close to 0, the more the market is competitive. So, in this case I can conclude confidently that the market is an oligopoly.

7 Estimating Lambda

In order to estimate the parameter lambda, I can rearrange the formula of the market price that is given. And I obtain the following:

$$\lambda \ = \eta - \ \eta * c/p$$

I then create a new variable "lambda" with the formula I just gave above. I am then able to estimate the parameter lambda. I decided to take the value of the elasticity in absolute value to estimate the lambda. Once the variable created I take the mean of the variable and I find 0.13. The result is very close to 0, but still superior to 0. So I have a lambda that is slightly superior to 0, and I can therefore conclude that I have an oligopoly.