

The assignment will count 40% to your final grade. Submission deadline is Friday, 3rd December, 20:00 PM! Please send your assignment per email before the deadline.

Instructions:

Every-one needs to submit their own solutions of the assignment. If responses are evidently identical in the submitted assignments of more than one person (copied text...), this content will not be graded. This also holds for any submitted code. However, working together and helping each other on the assignment is allowed. If you submit code, please include comments in the code, to make it easier for me to follow.

You can get full points by showing only the correct output (Graphs, Regression Tables) for each exercise, and providing the correct written answers. It does not matter with which program you create this output. But if the output is not correct, I will try to check Stata code, and sometimes also R code, but not code from other programs.

Total number of points is 100. Full Mark (A+) with 90 points or more. Pass (C-) with 50 points.

Download the Dataset “wber final5.dta” from CMS. It was created by the market research firm Nielsen to estimate the determinants for the demand of different brands of beer in the Chicago Metro Area. The variable *upc* (Unique Product Code) identifies the precise product in the data, e.g. “6-Pack Budwiser Light”. *Brand* identifies the broader brand, e.g. “Budwiser Light”. Observations in the dataset are identified by the variables *store*, *upc*, and *week*. Thus each observation records the sales of a given product in a given week in a given store. *Price* gives the average price of the product in that week at that store, *Sales* the amount of the product sold in the store in the week, and *margin* the average percent of the price that the store could keep from the price. The variable *poverty* shows the share of the population in the zip-code where the shop is located living under the official poverty line.

In graphs you show and interpret in your answers, cut out extreme outlier observations.

Exercise 1:

(32 Points)

Pick the product (upc), and the brand with the most observations in the dataset.

- a) For each of the two, plot (scatterplot) Prices (on y-axis) against Sales (on x-axis), and fit a straight line and a quadratic line into the data in the graph. (8)
- b) What is the main difference that you observe between the product and the brand? What do you think is the reason for the difference? (4)
- c) Demean the Prices and Sales data from the brand with the most observations at the product (upc) level. In other words, take the average of price and sales for each upc from that brand, and subtract them from the Price and Sales data from that brand. Plot the demeaned prices against the demeaned Sales. Does the graph differ from the graph for the largest brand in 1.a? Why do you think that could be the case? (6)
- d) Add a line to the graph from 1.c that shows the distribution of observations along the x-axis. How do you interpret the fitted quadratic line in the graph from 1.c? (6)
- e) Using again the data from the brand with the most observation, take the log of Price and the log of Sales, and regress log Sales on log Prices. How would you interpret the coefficient you estimate on log price? Is it a useful estimate? Can you think of a way to improve the estimation of the coefficient? If yes, show the regression output. (8)

Exercise 2:

(18 Points)

Pick the product (upc) with the highest and the lowest average price.

- a) For each of the two products, regress the log of Sales on the log of Price. What is the difference that you observe between the estimation outcomes for the two products? What do you think is the reason for the difference? (8)
- b) Consider a demand curve of the form $Q = AP^b$, where Q represents the quantity demanded, P the price of the product, and A and b are parameters. Derive the elasticity of Q with respect to P . (5)
- c) Show that you can rearrange the demand curve from 2.b to $\log(Q) = A + b \log(P)$. What does it say about the results we get from regressing $\log(Q)$ on $\log(P)$? (5)

Exercise 3:

(16 Points)

Take the median of the poverty variable in the dataset. Then take again the products with the highest and lowest average price.

- a) For each of the two products, create a graph which plots 1) Prices against Sales in zip-codes with above median poverty levels (in one color), and 2) Prices against Sales in zip-codes with below median poverty levels (in another color). For both type of zip-codes, fit a linear line into the data plot. (11)
- b) What are the main difference you observe between the two graphs? Why do you think this could be the case? (5)

Exercise 4: (34 Points)

- a) For each product (upc), obtain the average margin and average price, and plot the first against the latter. What do you see? Are you surprised? (8)

For each product in the dataset, estimate the product-specific demand elasticity (do not worry about fixed effects or other identification strategies).

- b) Plot average margins against the demand elasticities on the product level, and fit a linear line into the plot. What do you observe? Is the result consistent with micro-economic theory? Explain briefly why or why not. (20)
- c) Calculate the Lerner Index for each product based on its demand elasticity estimate, and plot it against the average margin of the product. Is the result as you expect? (6)

Hint 1: you can get product specific elasticities estimates β_i for each product $i = \{1, \dots, I\}$ from a single regression of the form $\log Q = a_i + \dots + a_I + \beta_1 \log P_1 + \dots + \beta_I \log P_I + \varepsilon$, where variable P_i contains the price of product i for all observations from product i , and 0 for observations from all other products, and $a_i - a_I$ are product fixed effects (dummies).

An alternative approach is to create a “loop” over all products:

```

levelsof upc, local(U)
foreach X of local U {
... code to estimate elasticity ... if upc == `X'
}

```

Hint 2: After running a regression, the estimated coefficients are stored in Stata in a “matrix” e(b), and can be incorporated into the dataset by

```

matrix b = e(b)
matrix b1 = b[1, "Name_coeff" .. "Name_coeff"]'
svmat b1

```