Case Study I

WS 2020/2021

Deadline: October 18th, 2020@ 23:55

Engel's law is a well known economic theory which states that the higher the household income, the proportion of income spent on food decreases. One could assume that an increase in income frees up the possibility to spend more money on other goods and services, while the expenditure of a poor family would likely center around food as there is a limitation in which one can be economical in basic necessities.

To confirm this theory in a quantitative manner, we try to fit a simple linear regression to a dataset called **engel** obtained from Koenker and Bassett (1982). The data is provided in two formats: the Eviews workfile **engel.wf1** and the CSV-file **engel.csv**, both available on Learn@WU. Please read the appropriate file to your system (either EViews or R) before you work with the following tasks.

The original source of this data is Ducpetiaux (1855) which was then used by none other than Engel himself to make his case in his seminal paper Engel (1857). Here are descriptions of this data:

- The data is collected in 19th century Belgium.
- There are 2 variables (*income* and *foodexp*) and 235 observations in total.
- The variable *income* refers to the household income and the unit of this value is 1 franc per value (100 = 100 francs).
- The variable *foodexp* refers to the houseold food expenditure. Again, the unit is 1 franc per value.

1 Data Analysis

1.1

Create a scatter plot of income (the *income* variable) on the x-axis versus food expenditure (the *foodexp* variable) on the y-axis and comment on any features in the plot such as potential outliers, non-linearity etc.

1.2

Transform the data by taking log transformations of both variables and create a scatter plot as in 1.1. How does the plot compare to the previous one obtained with the raw data?

1.3

Create a histogram and a table of descriptive statistics (mean, median, minimal and maximal observations, ...) for each of the variables, *income* and *foodexp*.

1.4

Now, consider a simple linear regression where the dependent variable, Y, is household food expenditure and the explanatory variable, X, is household income which we refer to as Model

1.

That is, the Model 1 regresses *foodexp* on *income* in the following manner:

$$Y = \beta_0 + \beta_1 X + u, \ \mathbb{E}[u|X] = 0.$$

Calculate and report the estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ for Model 1.

1.5

Using the estimate of $\hat{\beta}_1$ obtained in 1.4, quantify how the change of *income* affects the *foodexp* in expectation. The explanation should refer to not only the value, but the appropriate unit of change as well.

1.6

This time, consider a simple linear regression where both the dependent variable as well as the explanatory variable are logged which we refer to as Model 2.

That is, it is fitted with the following formula often referred to as log-log model:

$$\log(Y) = \beta_0 + \beta_1 \log(X) + u, \ \mathbb{E}[u|X] = 0$$

where $Y = foodexp$ and $X = income$ as before.

Report the estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ for Model 2.

1.7

Again, using the estimate of $\hat{\beta}_1$ obtained in 1.6, quantify how the change of *income* affects the *foodexp* in expectation. Make sure the report includes the appropriate unit of change in addition to the value.

1.8

What is the expected food expenditure for a household with an income of 10,000 francs? Report the values obtained from both models.

2 Theory

2.1

Now let's consider exchanging the dependent variable and the explanatory variable considered in Model 1.

That is, we instead estimate

$$X = \gamma_0 + \gamma_1 Y + v, \ \mathbb{E}[v|Y] = 0$$

where $Y = foodexp$ and $X = income$ as before.

Is it true that $\gamma_1 = \frac{1}{\beta_1}$ (β_1 in Model 1 to be precise)? Provide a theoretical argument and check the assertion on the given data set.

(Hint: when regressing Y on X we have $\beta_1 = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}$)

2.2

In this exercise, you are asked to show that the two standard forms of our most important assumption are actually equivalent. To this end, show step by step (mention properties such as linearity of expectation) that

$$Y = \beta_0 + \beta_1 X + u, \quad \mathbb{E}[u|X] = 0 \tag{1}$$

implies

$$\mathbb{E}[Y|X] = \beta_0 + \beta_1 X \tag{2}$$

and vice versa.