Økonometri I

Problem set 2

Agenda
Recap of STATA exercises from problem set 1
Review of written response to question 6
Introduction of this week's exercises
Introduction of STATA's linear regression command: **regress**
STATA group work and exercises

An econometric analysis of Engel curves for U.S. households
In problem set 1, we undertook the initial analysis of the data and discussed regression models of Engel curves for food, clothes and alcohol. The objective of this week's problem set is to estimate a simple regression model of the Engel curve using the OLS estimator.

The starting point is a regression model with one explanatory variable. Specifically, let us consider the case where the dependent variable represents food expenditures, while the explanatory variable is total expenditure:

$$xfath_i = \beta_0 + \beta_1 xtot_i + u_i \tag{1}$$

In the consumption literature, it is common to use expenditure *share*, $xfath/xtot$, as the dependent variable instead of using total food expenditures. Furthermore, the logarithm of total expenditure deflated by an individual "consumer price index" is often used as the explanatory variable. In this case, the regression model is:

$$\frac{xfath_i}{xtot_i} = \delta_0 + \delta_1 log\left(\frac{xtot_i}{price_i}\right) + v_i \tag{2}$$

Group work

a. Discuss model (2). What is the interpretation of δ_1 when $\delta_1 > 0$ and $\delta_1 < 0$?
   (Hint: luxury versus necessity goods)

b. What is likely to be included in the error term in model (2)?
(Hint: What other variables may influence the dependent variable, $xfath/xtot$, besides the explanatory variable?)

STATA exercises

One group is asked to hand in a brief (max 1 page) written response to question 5. The answer should be uploaded to Absalon 24 hours before next week's class.

1. Results from STATA can be transferred to a log file which can be opened directly in e.g. Word. This is done using STATA's log command. The following program creates a log file with the results created in your STATA do-file:

   cap log close

   log using PS2, replace text

   —- Your relevant STATA code —-

   log close

   Try to use the log command when you answer the questions below.

2. Estimate the simple regression model (1) for men's consumption of food by OLS. In STATA, the OLS estimator is implemented using the **regress** command. The following program code may be helpful:

   regress xfath xtot if dmale==1

   Answer the following questions using your regression results from STATA.

   i. What is the interpretation of the slope β_1 and intercept β_0 in model (1)?
   ii. What is the estimate of the slope? And the intercept?
   iii. What is the total variation in the dependent variable, SST? The explained variation in the dependent variable, SSE? The variation in the residuals, SSR?
   iv. Find the coefficient of determination, R^2. How can it be calculated from the three measures from the previous question? How would you interpret the calculated R^2?
   v. What is the estimate of the variance of the error term, σ^2?

vi. Illustrate the estimated Engel curve in a graph together with the actual data observations. This code should be useful:

```
twoway (scatter xfath xtot if dmale==1) (lfit xfath xtot if dmale==1)
```

vii. Perform scatterplots of the residuals from the regression against *xtot* and the predicted value of food consumption, separately. Hint: You can calculate residuals and predicted values in STATA using the following code:

```
regress xfath xtot if dmale==1
predict yhat if e(sample)==1, xb
predict uhat if e(sample)==1, residuals
```

3. Construct the variables which are needed to estimate model (2).

4. Estimate model (2) by OLS for the budget shares of food, clothing and alcohol for men and women, separately. Provide a table with the estimated slope parameters (for men and women in all three expenditure categories).

5. Interpret the estimation results in light of the discussion on luxury versus necessity goods (see group work). Which parameter is central to the analysis? What conclusions can be drawn on the basis of the analysis?

Homework
Complete the STATA exercises.

Solve the following theoretical exercises (estimated time for the exam is 30 minutes):

1. Write up the linear regression model with a constant term and an explanatory variable in matrix form for n observations.

2. Write the OLS estimator in matrix form. Show that when one calculates the OLS estimator, then:

$$\widehat{\beta}_0 = \bar{y} - \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\bar{x}, \qquad \widehat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

Hint: Use the following rule for inverting a matrix:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

as well rules (A.7) and (A.8) in Math Refresher A in the textbook.