ECON 861
Empirical MicroEconomics II
Winter Term 2012/13

Assignment 1
Due: In Class (DH 213) on ***January 29th*** *2012*

Rules of the 'game':
1) you can work in groups (actually, I encourage you to do so), but you MUST submit your own set of answers.
2) you MUST list the people you worked with in the front page.
3) late submissions will NOT be accepted. If you cannot make it to the class, you can pop into my office (DH306) and submit the assignment whenever you are ready, before the 29th.
4) I will keep the set of answers you are going to hand in, so please make copies if you think you are going to need them.
5) you have to dowload the datasets for the exercises from the webpage of the course:
http://qed.econ.queensu.ca/pub/faculty/cozzi/Webpage/Cozzi_ECON861.htm

Devote some time to give the graphs, plots and tables a format easy to understand. Also the way you present your answers matter for the final grade. Even if a question is mainly analytical (which is not the case for this assignment), briefly explain what you are doing, stressing the economic meaning of the various steps.

Ex.1: Exploring the Dataset and The Linear Classical Model.

Download the first dataset: PS1_1.dta. It includes selected variables for a subsample of 1,000 males from the 1988 Current Population Survey (http://www.census.gov/cps/). We are going to make the assumption that the sample is representative of the male U.S. workforce.

The individual series are:
AGE - self-explanatory
P_EXP - Years of potential work experience
WAGE - Hourly wage in dollars
GRADE - Highest educational grade completed
OCC_CAT - Indicator variable for occupational category:
1 Managers and administrators
2 Professionals
3 Nurses and other non-doctors
4 Clerical
5 Sales people
6 Service workers
7 Manual workers
8 Craft workers

IND_CAT - Indicator variable for industrial category:
1 Natural resources
2 Durables
3 Non-durables
4 Construction
5 Transportation (including Post Office)
6 Communication and utilities
7 Wholesale trade
8 Retail trade
9 FIRE (Finance, Insurance, Real Estate)
10 Education

11 Health and welfare
12 Business services
13 Personal and other services
14 Public administration

D_MARRIED - Dummy for marital status: 1 if married, 0 otherwise
D_PART - Dummy for part-time status: 1 if part-time worker, 0 otherwise
D_UNION - Dummy for union status: 1 if union member, 0 otherwise

a) Compute the percentage of male workers:
- employed in each industry.
- in each occupation.
- who are married.
- who work part-time.
- who are unionized.
- who are between the age of 25 and 55, married, unionized, and with at most a high school degree (grade 12 and below).

Report these values in a set of tables, with a short comment for each.

b.1) Estimate the non-parametric Kernel density of wages by union status (use the command *kdensity* in STATA) and present them in a graph.

b.2) Compute, report and briefly discuss the following statistics for wages: Mean, Range, Standard Deviation, Interquartile range, 1st, 10th, 25th, 50th, 75th, 90th, 99th percentiles (use the command *tabstat* in STATA).

b.3) Compute the same statistics for wages in point b.2) by Union status, Marital Status and Part-time status. Do you find anything interesting? Comment.

b.4) Plot the distributions of age, potential experience and education. For these variables study the same statistics in point b.2). Do you find anything weird? Comment.

c) Compute the average wage for each occupation. Can you get them all at once running a simple regression? Comment on the idea of reference group in a linear regression with (at least) a dummy variable among the covariates.

d.1) We are going to estimate a simple econometric model for wages: a 'Mincerian' equation. Before doing that, compute, report and discuss the correlation matrix for wages, potential experience, its square, and education.

d.2) Run a regression with the log of wages (why?) as the dependent variable and potential experience, its square, education and a constant as covariates. How much are the returns to education?

d.3) What is the estimated Variance/Covariance Matrix of the error term (using the unbiased estimator of the variance)? What is the Variance/Covariance Matrix of the estimators?

d.4) Test the hypothesis that the sum of the coefficients on experience and education is equal to zero.

d.5) Consider the variable grade_new. Plot wages Vs. grade, and after that wages Vs. grade_new: do you notice anything? What happens if you run the same regression as before substituting grade with grade_new? How can you deal with this issue?

Ex.2: Problems with The Linear Classical Model: Heteroskedasticity.

Download the second dataset: PS1_2.dta. It includes several variables on a sample of 706 individuals. The labels in the dataset describe each variable.

Consider the following model to explain sleeping behavior:

$$sleep = \beta_0 + \beta_1 totwrk + \beta_2 educ + \beta_3 age + \beta_4 age^2 + \beta_5 yngkid + \beta_6 male + u$$

1) Write down a model that allows the variance of u to differ between men and women. The variance should not depend on other factors.

2) Use the data to estimate the parameters of the model for heteroskedasticity. Is the estimated variance higher for men or for women?

3) Is the variance statistically different for men and for women?

4) Estimate the equation for sleeping behavior with the White robust standard errors. What happens to the estimated Standard Errors? Does this affect your inferences?

Ex.3: More Problems with The Linear Classical Model.

Download the third dataset: PS1_3.dta. It includes a set of variables on a sample of 88 houses sold in 1990 in the Boston MA area. The labels in the dataset describe each variable.

Consider the following model to explain the selling price:

$$price = \beta_0 + \beta_1 lotsize + \beta_2 sqrtft + \beta_3 bdrms + u$$

1) Estimate the model and compute the Breusch-Pagan test for heteroskedasticity (the F version of the test). What do you conclude?

2) Apply the White test for heteroskedasticity, using the Chi-square form of the test statistic. (Don't use the command is STATA: build the test yourselves!) What do you conclude by computing the p-value?

3) Estimate the equation for pricing behavior with the White robust standard errors. Discuss any important differences with the usual standard errors.

4) Transform the model to explain the selling price by considering the log of all variables in the equation above. Redo the steps 1, 2 and 3 and briefly discuss any notable differences.

Ex.4: Problems with The Linear Classical Model: Omitted Regressors.

Use the data from the PS1_4.dta. It includes selected variables for a subsample of 540 males and females of one wave of the NLSY79 (http://www.bls.gov/nls/nlsy79.htm). The only variables you are going to use are:

EARNINGS - Current hourly earnings in $ reported at the interview
S - Years of schooling (highest grade completed)
ASVABC - Scaled score on a set of IQ tests.

More in detail, ASVABC is the composite of ASVAB2 (with double weight), ASVAB3 and ASVAB4, where
ASVAB2 Arithmetic reasoning
ASVAB3 Word knowledge
ASVAB4 Paragraph comprehension
ASVAB5 Numerical operations (speed test)
ASVAB6 Coding speed (speed test)

Regress the logarithm of earnings on:
1) S and ASVABC
2) S only
3) ASVABC only

a) Compute the correlation between S and ASVABC. Compare the coefficients on S in the regressions 1) and 2). Give both an intuitive and a statistical explanation of the direction of the change.

b) Repeat the exercise for the coefficients on ASVABC.

c) What is the problem with OLS in this case?