Microeconometrics
Spring Semester 2019/20
Problem set 1
Due: February 20^{th}

Note: Send a copy of your solutions to Sónia Félix at sfelix@novasbe.pt and to me (teresa.molina@novasbe.pt). The solutions should include a .pdf or .doc document with the answers to each one of the questions ***and*** *the log files or do files of the exercises done using stata. Late problem sets will not be accepted.*

Exercise 1 (Ordinary Least Squares | Generalized Least Squares)

For this exercise please create a do-file and corresponding log-file to hand in with your solution.

Use the dataset **wage1.dta** to estimate the determinants of wages. Consider the following regression model:

$$lwage_i = \beta_0 + \beta_1 exper_i + \beta_2 exper_i^2 + \beta_3 married_i + \beta_4 female_i + \beta_5 married_i \times female_i + u_i,$$

where the dependent variable $lwage$ is the logarithm of hourly wages. The independent variables are the individuals' experience (in years) and squared experience ($exper$ and $exper^2$, respectively), a dummy variable equal to 1 for married individuals and 0 for single individuals ($married$), a dummy variable equal to 1 for females and 0 for males ($female$), and an interaction term between married and female ($married \times female$).

(a) Estimate this regression model by OLS. Provide a rigorous interpretation of the regression coefficients.

(b) Assuming the Gauss-Markov assumptions hold in the theoretical model and holding the other factors constant, what is the number of years of experience that maximizes the average logarithm of wages?

(c) Assuming the Gauss-Markov assumptions hold in the theoretical model, does experience have a statistically significant impact on the logarithm of wages? State clearly the test behind the estimation output above, the null and alternative hypotheses and the decision rule you use.

(d) Perform a heteroskedasticity test in the estimated regression model. State clearly the null and alternative hypotheses, the test statistic you consider, and the rejection rule. What do you conclude?

(e) One possible heteroskedasticity test is to regress the squared residuals on the set of explanatory variables and the squared fitted values of the initial regression model. Someone suggests adding the fitted values of the initial regression model to this heteroskedasticity test. Is it a good idea? Explain.

(f) Estimate the model using a feasible generalized least squares estimator. Explain the steps of the estimation procedure. Why is it important to correct for the problem of heteroskedasticity?

(g) Consider now a simpler version of the initial model:

$$lwage_i = \beta_0 + \beta_1 exper_i + \beta_2 married_i + \beta_3 female_i + u_i,$$

Rewrite the regression model given above in a way that allows you to test directly (i.e. by performing a simple t-test on a coefficient) whether the impact of one additional year of experience on the logarithm of wages is double the impact of being married. Perform the test describing carefully the null and alternative hypotheses of the test, the test statistic, and the rejection rule.

Exercise 2 (Instrumental variables)

For this exercise please create a do-file and corresponding log-file to hand in with your solution.

Consider the dataset **QOB.dta** that is comprised of a subset of the data used in the seminal paper by Angrist and Krueger (1991) "Does compulsory school attendance affect schooling and earnings?". In the paper the quarter of birth of individuals is used as an instrument for education in order to estimate the impact of compulsory school on earnings. The authors use samples from Census data for men born in 1920s, 1930s, and 1940s.

(a) Compute some descriptive statistics in order to get a general idea of the data. In particular, calculate the average number if years of schooling by each quarter of birth for men born in the 1920s, the 1930s, and the 1940s. What do you observe?

For the next parts of the exercise consider the sub-sample of men born in 1930-1939.

(b) Estimate the returns to education by OLS using age and squared age as control variables. Interpret the results and explain why the estimated returns to education might not have a causal interpretation.

(c) Explain why the quarter of birth might be a good instrument for education when estimating returns to education.

(d) Construct a dummy variable, *first_qob*, which equals one for men born in the first quarter of the year and zero otherwise. Compute the IV estimate $IV = (Z'X)^{-1}Z'Y$ of returns to education considering *first_qob* as instrumental variable. Compare the estimate with the OLS returns to education estimated in the regression of *lwage* on a constant and years of education.

(e) Suppose in the following that we have three instrumental variables, $Z1$, $Z2$, and $Z3$ representing dummy variables for first-, second-, and third-quarter births, respectively. Generate these three instrumental variables.

  (i) Describe and estimate the first-stage equation with multiple instruments. Include the following explanatory variables as additional control variables: *age*, *agesq*, *race*, *married*, and *smsa*.

  (ii) Compute an F-test under the null hypothesis that the quarter of birth dummy variables have no effect on the total years of education. Are these instruments valid?

  (iii) Estimate the returns to education by 2SLS and compare the results to standard OLS estimates [consider the same set of control variables as in part (ii)].

  (iv) Are the instruments exogenous?