

Microeconometrics

Problem set 1

Note. Upload a copy of your solutions on Moodle. The solutions should include a .pdf document with the answers to each one of the questions **including the screenshot of command used to generate the output and the screenshot of results or the output** generated from the statistical software (feel free to use STATA or R, or other softwares). Please check the submission deadline on moodle, late problem sets will not be accepted.

1. (100%) **Instrumental variables.** Consider the dataset **QOB** 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. IMPORTANT: focus on the sub-sample of men born in 1930-1939. Year of birth is the variable YOB and includes only the last 2 digits.
 - (a) Compute the non-parametric cdf and pdf of the (log-)wage and interpret the results.
 - (b) Compute the non-parametric pdf of the (log-)wage for individuals with less or equal than 8 years of education and for individuals with more than 8 years of education and compare the two.
 - (c) Estimate the returns to education by OLS using age and squared age as control variables.¹ Interpret the results and explain why the estimated returns to education might not have a causal interpretation.
 - (d) Explain why the quarter of birth might be a good instrument for education when estimating returns to education.
 - (e) 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.
 - (f) 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 (i)].
 - iv. Are the instruments exogenous?
 - (g) Consider a world where quarter of birth is not a valid instrument. Propose an alternative IV strategy that could estimate the causal effect of education on earnings using a different exogenous source of variation. Justify your choice.

¹Returns to education are estimated using a linear regression of wages (generally in logs) on years of schooling.