Problem Set 2
Quantitative and Statistical Methods II

General Instructions

- The problem set is due by February 12th at noon;
- You should send it by email *bruno.conte@barcelonagse.eu*; it must include your *unique* Stata code, *a unique* log file, and your answer sheet. Tidiness is appreciated – e.g. material correcly labeled and organized in a zip file;
- Working in teams is allowed and strongly recommended (keeping the same groups as in the presentations is encouraged);
- The datasets needed to solve the computational questions are uploaded in the Classroom material.

"Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records", Angrist, American Economic Review, 1990

This paper aims at studying the effects of military services during the Vietnam War on long-term earnings of the veterans. The author implements the Wald estimator by instrumenting military service with a draft-eligibility dummy that was determined by a lottery over birthdays. In particular, from 1970 to 1972, random sequence numbers were randomly assigned to each birth date in cohorts of 19 years old men. Men with lottery number below a given eligibility threshold number were eligible for the draft, while men above the threshold could not be drafted.

1. The dataset `problemset2.dta` contains information on cohorts of 1950, 1951 and 1952. Run a regression of "future earnings" (i.e. earnings in years 1981-1984) on "veteran status" dummy. Discuss the potential bias of the estimated impact of veteran status on future earnings and its sign. Run an IV estimation instrumenting "veteran status" with "eligibility" dummy, as done in class (using the `ivregress` command). How does the point estimate of the variable "veteran status" compare to the one obtained before?
2. Explain why this bias disappears if "veteran status" dummy is instrumented with the "eligibility" dummy described above. State which are the necessary identifying assumptions and interpret them in this context.

3. Describe which are the structural, reduced form and first stage equations in this setup. Analytically show that, together, they provide the 'formula' for the IV estimator, i.e.

$$\hat{\beta}_D = \hat{\rho}_Z / \hat{\alpha}_Z, \tag{1}$$

where β_D, ρ_Z and α_Z are, respectively, the coefficients that multiply Z_i on the structural, reduced form and first stage equations, respectively. Use the `problemset2.dta` data to estimate the necessary equations to show that, indeed, estimating β_D as in (1) provides the same (or very close) estimates from the IV regression you made on exercise 1.

4. Show analytically that this estimator is equivalent to the Wald Estimand, i.e.

$$\hat{\beta}_D = \hat{\rho}_Z / \hat{\alpha}_Z \equiv \frac{\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]}{\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0]}. \tag{2}$$

Use the `problemset2.dta` data to check whether the direct calculation of the Wald Estimator provides the same (or very similar) estimates of $\hat{\beta}_D$ from Question 3.

5. Consider what happens to this analysis if the "eligibility" dummy has a direct impact on the outcome "future earnings". In particular, assume that the structural and first stage equations are, respectively,

$$Y_i = \beta_0 + \beta_D D_i + \beta_z Z_i + U_i, \quad \beta_Z \neq 0 \tag{3}$$

$$D_i = \alpha_0 + \alpha_Z Z_i + V_i, \tag{4}$$

where Y_i, D_i and Z_i stand, respectively, for "future earnings", "veteran status" and "eligibility". A naive econometrician decides to estimate treatment effects of D_i using Z_i as an IV.

   a. Which is the crucial assumption violated in this setup? Suggest a scenario in which this might happen in this application.

   b. Provide an expression of the resulting bias of the IV estimator. What does the size of the bias depend on?

6. In practice, many eligible-draft men were still exempted from service for health or other reasons, while many men who were draft-exempt nevertheless volunteered for service. Hence, veteran status was not completely determined by randomized draft-eligibility. Taking this fact into account, which additional assumption is needed in order to interpret the Wald estimator as the effect of veteran status on those whose treatment status can be changed by the instrument? In other words, which extra assumption is needed in order to interpret β_D as the Local Average Treatment Effect? Discuss its meaning in the present context.

7. Now go back to the setup of Question 5. Suppose, however, that $\beta_Z = 0$, but $\alpha_Z \to 0$.

   a. Do the necessary assumption for the estimation of $\hat{\beta}_D$ in (3) hold? Justify your answer.

   b. Is the estimator $\hat{\beta}_D$ subject to any issue? Which is it, if so?