Universidad Carlos III de Madrid, Department of Economics
ECONOMETRICS I Fall 2020
Assingment 4, Due January 11

1. We will analyze the reemployment bonus experiments conducted in Pennsylvania by the United States Department of Labor between July 1988 and October 1989. This experiment involved the randomized assignment of new claimants for unemployment insurance (UI) benefits into one of several treatment groups or a control group. Claimants assigned to the control group were handled according to the usual procedures of the unemployment insurance system, while claimants assigned to treatment were awarded cash bonuses if they were able to demonstrate full-time reemployment within a specified qualifying period. The corresponding data have been previously analyzed by Billias (2000), Koenker and Billias (2001) and Koenker and Xiao (2002), among others. Here we focus solely on a single treatment group (the most generous treatment, treatment 4; claimants for unemployment benefits that were assigned to this treatment were offered a bonus equal to six times the usual weekly benefit if they secured full-time employment within 12 weeks). The corresponding dataset is penn_jae.data and will be posted in Aula Global, together with papers with related applications. The variable Y is the log of duration of unemployment for the UI claimants ("inuidur1"), D is the treatment indicator (=1 if treated, 0 otherwise, "tg"), and X a p-dimensional vector of pre-treatment variables. The vector X includes an intercept and some demographic and socioeconomic control variables listed in e.g. Koenker and Xiao (2002). I recommend you use R markdown to create the output file.

   (a) Read the data (Penn<- as.data.frame(read.table("penn_jae.dat", header=T)); and estimate the ATE using the standard difference of sample means and a linear regression using as controls

   $$X <-\text{"}female+black+othrace+dep+q2+q3+q4+q5+q6+agelt35+agegt54+durable+lusd+husd\text{"}.$$

   That is, using the model

   $$Y_i = \alpha_0 + \tau_0 D_i + \beta_0' X_i + \varepsilon_i, \tag{1}$$

   where we assume $E[\varepsilon_i \mid X_i, D_i] = 0$. Interpret the results, provide standard errors for the estimates, and evaluate the assumptions underlying your interpretations.

   (b) One way to evaluate if the randomization is sucessful is to test the significance of θ_0 in a Probit specification of the propensity score $p(x) = \Phi(x'\theta_0)$. Run such a test and interpret the results. Discuss the type of test, critical value, etc.

   (c) Estimate the ATE by DML based on Lasso.

   (d) Construct 95% CI for the ATE using the previous estimates.

2. Using the "quantreg" package to run quantile regressions, solve the following questions:

   (a) Suppose now that (1) holds with ε_i independent of D_i and X_i. What is the corresponding quantile regression of Y given X_i, D_i associated to (1)? How would you test (visually) the previous independence assumption using quantile regression?

   (b) Estimate for a grid of 300 points between [0.2,0.8] the quantile coefficients of Y given D and X (use alphagrid <-c(seq(0.2,0.8,length.out=300)) and plot the coefficients (similarly to Figure 1 in Koenker and Xiao). Is the independence of the error a realistic assumption in this application? Interpret the coefficient of treatment in the quantile regressions.

   (c) Explain how you would compute bootstrap standard errors for the QRE estimator. What is the advantage of doing bootstrap for computing standard errors over estimating the asymptotic variance for quantile regression? Compute a bootstrap confidence interval for the median treatment effect.

   (d) How would you allow for quantile treatment effects that vary with some covariates? Read Escanciano and Goh (2019), and investigate how treatment effects may vary by age using your preferred method.

3. This problem illustrates the possible inconsistency of the bootstrap. Suppose we observe $X_1, ..., X_n$ from a $U[0, \theta]$.

   (a) Prove that the MLE of θ is $\widehat{\theta}_n = \max\{X_1, ..., X_n\}$.

   (b) Prove that $T_n(\theta) = n(\widehat{\theta}_n - \theta) \to_d -\theta Z$, where $Z \sim \exp(1)$. Recall that $Z \sim \exp(1)$ if $\Pr(Z \leq x) = 1(x \geq 0)(1 - \exp(-x))$. Hint: $\{T_n(\theta) \leq x\} = \cap_{i=1}^{n} \{X_i \leq (x/n) + \theta\}$.

   (c) Let $\widehat{\theta}_n^* = \max\{X_1^*, ..., X_n^*\}$ for a bootstrap sample, $X_1^*, ..., X_n^*$, from the ecdf F_n. Prove that $\Pr(n(\widehat{\theta}_n^* - \widehat{\theta}_n) = 0) = 1 - \left(1 - \frac{1}{n}\right)^n \to 1 - e^{-1} > 0$.

   (d) Conclude from the previous point that the nonparametric bootstrap is inconsistent for approximating the distribution of $T_n(\theta)$.

4. Do a Monte Carlo simulation study comparing the empirical size and power performance of the robust LM (discussed in class) and the Wald test based on DML in the setting used in Practice 12 (December 15th). For size and power you can set the ATE to $\tau_0 \in \{-2, -1, -0.5, -0.25, 0, 0.25, 0.5, 1, 2\}$. The sample sizes are $n = 100$ and $n = 300$, and the Monte Carlo simulations are based on 1000 replications. Report plots and/or tables.