ECON490 **Department of Economics**
Topics in Econometrics **UIUC**

Problem Set II
Due: Wednesday February 17th, 2016

Use the same dataset as Problem set I. The data set Lalonde.csv contains information on 445 men. There are twelve variables. The outcome is `re78`, earnings in 1978. The treatment indicator is `treat`. The ten covariates are `black, hispanic, education, married, nodegree, age, re74` (earnings in 1974), `re75` (earnings in 1975), and two indicators for zero earnings in 1974 and 1975, `u74` and `u75`.

1. Calculate the difference in averages (of `re78`) between men in the treated and control groups, and the standard error for the difference, and construct a 95% confidence interval for the difference using OLS assuming conditional homoskedasticity.

2. Calculate the difference in averages (of `re78`) between men in the treated and control groups, and the standard error for the difference, and construct a 95% confidence interval for the difference using OLS assuming conditional heteroskedasticity.

3. Consider the econometric model

$$y_i = \alpha + \beta x_i + u_i \ .$$

State whether each of the following statements is true or false. You need to explain your answer.

i) **TRUE or FALSE**. If $\mathrm{E}(u_i|X_i) = 0$, then $\mathrm{Cov}(X_i, u_i) = 0$. Explain.

ii) **TRUE or FALSE**. If $\mathrm{E}(u_i X_i) = 0$, then $\mathrm{E}(u_i^2 X_i^2) = 0$. Explain.

iii) **TRUE or FALSE**. If $\mathrm{E}(u_i|X_i) = 0$, then $\mathrm{E}(u_i X_i^3) = 0$. Explain.

4. To study the relationship between fuel consumption (Y) and flight time (X) of an airline, the following model is formulated:

$$Y_i = \alpha + \beta X_i + u_i \ ,$$

where Y is expressed in thousands of pounds and X in hours, using fractions of an hour as units of low-order decimal.
The statistics of "Flight times and fuel consumption" of an airline provides data on flight times and fuel consumption of 24 different trips made by an aircraft of the company. From these data the following statistics were drawn:

$$\sum Y_i = 219.719 \ ; \quad \sum X_i = 31.470 \ ; \quad \sum X_i^2 = 51.075 \ ;$$

$$\sum X_i Y_i = 349.486 \ ; \quad \sum Y_i^2 = 2396.504 \ .$$

i) Estimate α and β by OLS.

ii) Calculate the sums of squared residuals $\sum_{i=1}^{N} \hat{u}_i^2$.

iii) Estimate total consumption, in thousands of pounds, for a flight program consisting of 100 one hour flights.

5. Consider a linear regression function with the indicator function D_i only (no covariates) under a randomized experiment. We specify a linear regression function for the observed outcome Y_i as

$$Y_i = \alpha + \beta D_i + u_i \ ,$$

where the unobserved residual u_i captures unobserved determinants of the outcome. The ordinary least squares estimators for α and β are based on minimizing the sum of squared residuals over α and β, i.e.,

$$\left(\hat{\alpha}, \hat{\beta}\right) = \text{argmin}_{\alpha,\beta} \underbrace{\sum_{i=1}^{N} (Y_i - \alpha - \beta D_i)^2}_{SSR} \ .$$

i) Derive an OLS estimator for α and β and show that

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{D} \ .$$

$$\hat{\beta} = \frac{\sum_{i=1}^{N} Y_i D_i - \frac{1}{N}\sum_{i=1}^{N} Y_i \sum_{i=1}^{N} D_i}{\left(\sum_{i=1}^{N} D_i^2 - \frac{1}{N}\left(\sum_{i=1}^{N} D_i\right)^2\right)} \ .$$

(We basically solved the same problem in class where we had X_i instaed of D_i. I just want you to exercise this problem again on your own.)

ii) Show that solutions above can be expressed as

$$\hat{\beta} = \frac{\sum_{i=1}^{N}(D_i - \bar{D})(Y_i - \bar{Y})}{\sum_{i=1}^{N}(D_i - \bar{D})^2} \ ,$$

and

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{D} \ ,$$

where

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} Y_i \quad \text{and} \quad \bar{D} = \frac{1}{N}\sum_{i=1}^{N} D_i = \frac{N_1}{N} \ .$$

iii) Show that the OLS estimator $\hat{\beta}$ obtained above is identical to the difference in average outcomes by treatment status, i.e.,

$$\hat{\beta} = \frac{1}{N_1}\sum_{i=1}^{N} Y_i D_i - \frac{1}{N_0}\sum_{i=1}^{N} Y_i(1 - D_i) \ ,$$

where $N_0 = \sum_{i=1}^{N}(1 - D_i)$ and $N_1 = \sum_{i=1}^{N} D_i$.