Econometrics
Problem Set 1.

Instructor: Alexei Onatski

1. Suppose that a farm's production of Y units of an agricultural product depends only on the amount of rain $X \in [0, 2]$ and the quality of soil $\varepsilon \in [-1, 1]$. The dependence is described by the following structural equation

$$Y = -X^2 + (2 + \varepsilon) X. \tag{1}$$

Further, suppose that the joint density of X and ε is given by

$$f_{X\varepsilon}(x, z) = \frac{1 + x - z}{8} \text{ for } x \in [0, 2] \text{ and } z \in [-1, 1]$$

(a) What's the causal effect of a marginal change in rain X on farm's production? How is it different for farms with very good soil ($\varepsilon = 1$) and very bad soil ($\varepsilon = -1$)?

(b) Find the average causal effect ACE of X on Y as a function of X. Plot it. Comment.

(c) What's $E(Y)$? What's $E(Y|X)$?

(d) How is $\frac{\partial}{\partial X} E(Y|X)$ different from the ACE that you found in (b)? Comment on the regression interpretation in this setting.

(e) What's the BLP of Y given X? What does OLS of Y on constant and X estimate?

2. Suppose that Y and X are $n \times 1$ vectors of data, and the following conditions hold: (1) $Y = X\beta + \varepsilon$, (2) $rank(X) = 1$, (3) $E(\varepsilon|X) = 0$ and (4) $Var(\varepsilon|X) = \sigma^2 I_n$. Let $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ and $\bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$.

(a) Consider the estimator $\tilde{\beta} = \bar{Y}/\bar{X}$. Show that $\tilde{\beta}$ is linear and conditionally unbiased. Calculate its conditional variance and compare it to the conditional variance of the OLS estimator.

(b) Suppose that you decide to use the first m $(< n)$ observations and do OLS. Show that this estimator $\hat{\beta}_{(m)}$ is linear and conditionally unbiased, but not minimum conditional variance.

(c) Could you suggest a minimum conditional variance, linear estimator (not necessarily unbiased)?

3. Consider the following linear regression model

$$y = \beta x + \varepsilon,$$

where

$$x = \begin{cases} 1/5 \text{ with probability } 1/2 \\ 7/5 \text{ with probability } 1/2 \end{cases},$$

and ε is independent from x, and has mean zero and variance 1.

(a) Suppose that only one observation (y, x) is available. What is the OLS estimator $\hat{\beta}$? What is its unconditional mean and unconditional variance?

(b) Consider another estimator $\tilde{\beta} = xy$. Is $\tilde{\beta}$ unconditionally unbiased for β? What is its unconditional variance?

(c) Suppose that the true β equals zero. Which estimator has smaller unconditional variance: $\hat{\beta}$ or $\tilde{\beta}$? Does your answer contradict the Gauss-Markov theorem? Explain.

4. Imagine that you are an expert economist and that you have to testify in court in about 30 min. During your testimony, you will be using a simple homoskedastic regression model

$$y_i = \alpha + \beta x_i + \varepsilon_i,$$

where x_i is a scalar explanatory variable. Your assistants have prepared the following OLS results for you:

$$\hat{\alpha}_{OLS}, \quad \widehat{SE}\left(\hat{\alpha}_{OLS}\right), \quad \hat{\beta}_{OLS}, \quad \widehat{SE}\left(\hat{\beta}_{OLS}\right), \quad R^2, \quad \bar{R}^2, \quad \begin{array}{c}\text{Sum of squared} \\ \text{residuals, } SSR\end{array}$$

Taking a final look at the notes that the assistants gave you, you notice that they made a terrible mistake! Instead of regressing y on constant and x they regressed x on constant and y!!! However, you are cool, calm and collected and in 30 min you are getting the true values of all of the above indicators by manipulating the false values. How do you do this?

5. File PS1.dta contains data on wage, education, and age for 1500 individuals. These data are taken from a 1990 cross-section of the National Longitudinal Survey of Yourth (NLSY) dataset. The variables in the file are:
w0: earnings (in dollars)
ed0: education (in years)
a0: age (in years).

   (a) Open the file in STATA. Create the experience variable by subtracting (education +6) from age. Regress log(wage) on constant, education, experience, and experience squared. Report the results.
   (b) By using 'predict' command, create variable equal to the fitted values from the above regression. Regress log(wage) on constant, education, experience, and the created fitted values. Explain theoretically the obtained estimates of the coefficients.
   (c) Now partial out log(wage) and experience with respect to constant, education, and experience squared. Regress partialled out log(wage) on partialled out experience. Discuss the coefficient and standard error estimates on partialled out experience in relation to the first wage regression. Explain the constant term estimate.
   (d) Finally, regress log(wage) on partialled out experience. How do the results compare to the previous 'partialled out' regression? Why?