

# Introduction to Econometrics

## Problem Set

Annalisa Scognamiglio

Due on April 20

### Problem 1

Let  $\hat{\beta}_0$  and  $\hat{\beta}_1$  be the intercept and the slope from the regression of  $y_i$  on  $x_i$ , using  $n$  observations.

1. Let  $c_1$  and  $c_2$ , with  $c_2 \neq 0$ , be constants. Let  $\tilde{\beta}_0$  and  $\tilde{\beta}_1$  be the intercept and the slope from the regression of  $c_1 y_i$  on  $c_2 x_i$ . Show that  $\tilde{\beta}_1 = (c_1/c_2)\hat{\beta}_1$  and  $\tilde{\beta}_0 = c_1\hat{\beta}_0$ .
2. Now let  $\tilde{\beta}_0$  and  $\tilde{\beta}_1$  be the intercept and the slope from the regression of  $(c_1 + y_i)$  on  $(c_2 + x_i)$ . Show that  $\tilde{\beta}_1 = \hat{\beta}_1$  and  $\tilde{\beta}_0 = \hat{\beta}_0 + c_1 - c_2\hat{\beta}_1$ .

### Problem 2

Suppose that average worker productivity at manufacturing firms (*avgprod*) depends on two factors, average hours of training (*avgtrain*) and average worker ability (*avgabil*):

$$avgprod = \beta_0 + \beta_1 avgtrain + \beta_2 avgabil + u$$

Assume that this equation satisfies the Gauss-Markov assumptions. If grants to fund workers' training programs have been given to firms whose workers have less than average ability, what is the likely bias in  $\tilde{\beta}_1$  obtained from a regression of *avgprod* on *avgtrain*?

### Problem 3

Use the data in DISCRIM to answer this question. These are zip code-level data on prices of various items at fast-food restaurants, along with characteristics of the zip code population, in New Jersey and Pennsylvania. The idea is to see whether fast-food restaurants charge higher prices in areas with a larger concentration of blacks.

1. Find the average value of *prpbck* and *income* in the sample, along with their standard deviations. What are the units of measurement of *prpbck* and *income*?
2. Consider a model to explain the price of soda, *psoda*, in terms of the proportion of the population that is black and median income:

$$psoda = \beta_0 + \beta_1 prpbck + \beta_2 income + u$$

Estimate this model by OLS and report the results. Interpret the coefficient on *prpbck*. Is this coefficient statistically significant at the 5% level?

3. Compare the estimates from part (2) with the simple regression estimate from *psoda* on *prpblck*. Is the coefficient on *prpblck* larger or smaller when you control for income? Discuss.
4. A model with a constant price elasticity with respect to income may be more appropriate. Modify the model above accordingly and report both the model and the estimates you obtain.

## Problem 4

Use the data in WAGE2 for this exercise.

1. Estimate the model

$$\log(\text{wage}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{tenure} + \beta_4 \text{married} + \beta_5 \text{black} + \beta_6 \text{south} + \beta_7 \text{urban} + u$$

and report the results. Holding other factors fixed, what is the approximate difference in monthly salary between blacks and nonblacks? Is this difference statistically significant at the 5% level?

2. Interpret the estimates for  $\beta_6$  and  $\beta_7$ .
3. Add the variables  $\text{exper}^2$  and  $\text{tenure}^2$  to the equation and show that they are jointly insignificant even at the 20% level.
4. Extend the original model to allow the return to education to depend on race and test whether the return to education does depend on race.
5. Can you interpret the coefficient on the dummy variable *black* as evidence (against or in favor) of race discrimination in the labor market? Discuss.

## Problem 5

Using a dataset of your choice,

1. Choose the outcome variable  $y$  and explanatory variable  $x$  and explain why you think it would be interesting to estimate the relationship between such two variables.
2. Estimate a simple linear regression model of the form

$$y = \beta_0 + \beta_1 x + u$$

and report the estimates.

3. Discuss whether your setting satisfies the Gauss-Markov assumptions.