

# **ECON 124: Problem Set #1**

Due on Jun 15, 2025

*Dr. Deniz Baglan*

**Alejandro Ouslan**

## Problem 1

Consider two least-square Regressions

$$y = X_1\beta_1 + \epsilon$$

and

$$y = X_1\beta_1 + X_2\beta_2 + \epsilon$$

Let  $R_1^2$  and  $R_2^2$  be the R-squared from the two regressions, respectively. Show that  $R_2^2 \geq R_1^2$

## Problem 2

Use the cps09mar data for this question. The data set and the description file is attached under the assignment of Canvas. Estimate a log wage regression for the subsample of white male Hispanics. In addition to education, experience, and its square, include a set of binary variables for Northeast, South and West so that Midwest is the excluded group. For marital status, create variables for married, widow or divorced, and separated, so that single (never married) is the excluded group.

## Problem 3

The data koop tobias subsample is extracted from koop and tobias (2004) study of the relationship between wages and education, ability, and family characteristics. Let  $X_1$  equal a constant, education, experience, and ability. Let  $X_2$  contain the mothers education, the fathers education, and the number of siblings. Let  $y$  be the log of the hourly wage. Show your regression output

1. Compute the least square regression coefficients in the regression of  $y$  on  $X_1$ . Report the coefficients.
2. Compute the least square regression coefficients in the regression of  $y$  on  $X_1$  and  $X_2$ . Report the coefficients.
3. Regress each of the three variables in  $X_2$  on all the variables in  $X_1$  and compute the residuals from each regression. Arrange these new variables in the  $(15 \times 3)$  matrix  $X_2^*$ . What are the sample means of these three variables?
4. Compute the  $R^2$  for the regression of  $y$  on  $X_1$  and  $X_2$ . Repeat the computation for the case in which the constant term is omitted from  $X_1$ . What happens to  $R^2$ ?
5. Compute the adjusted  $R^2$  for the full regression including the constant term.

Suppose that you have two independent unbiased estimators of the same parameter  $\theta$ , say  $\hat{\theta}_1$  and  $\hat{\theta}_2$ , with different variances  $v_1$  and  $v_2$ . What linear combination  $\hat{\theta} = c_1\hat{\theta}_1 + c_2\hat{\theta}_2$  is the minimum variance unbiased estimator of  $\theta$ ?

Analyze the properties of the Least Squares (LS) estimator of the slope coefficient in a simple linear regression model using the Monte Carlo method. Specifically, the true model is given by

$$y_i = 1 + 2x_i + \epsilon_i$$

Where the intercept is 1, and the slope coefficient is 2. The errors  $\epsilon_i$  are normally distributed with mean zero and variance 3,  $\epsilon \sim N(0, 3)$ . The independent variable  $x$  is uniformly distributed between 0 and 6,  $X \sim U(0, 6)$ .

Conduct the following experiment: Draw 1000 samples of sizes 25, 50, and 100 from the distribution of the error term and the predictor variable. For each sample, estimate the slope coefficient using the LS estimator. Then, analyze the

1. Unbiasedness, variance, and the shape of the distribution of the LS estimator of the slope coefficient by plotting the distribution for each sample size.
2. Compare the variance of the estimated slope coefficient obtained from the Monte Carlo simulation with the variance obtained using the asymptotic formula and the variance obtained using a bootstrap method