

# ECON 2123B: Assignment 2

Due April 10 before midnight on OWL

Emmanuel Murray Leclair

## Problem 1

To complete this problem, open the ceosal dataset included with the assignment on OWL (*A2.ceosal.xls*). This is dataset containing information on the salary of CEOs and other relevant variables in different industries. Your primary interest is to study how the salary of CEOs affect sales. You think that higher paid CEOs will put more effort into their business, which should result in higher sales. There are 4 possible industries: finance, consumer product, industrial and transport/utilities. The main model of interest is as follows:

$$\begin{aligned} lsales_i = & \beta_0 + \beta_1 utility_i + \beta_2 consprod_i + \beta_3 indus_i + \beta_4 lsalary_i + \beta_5 (lsalary_i * utility_i) \\ & + \beta_6 (lsalary_i * consprod_i) + \beta_7 (lsalary_i * indus_i) + \beta_8 lroe_i + \beta_9 ros_i + u_i \end{aligned} \quad (1)$$

The variables in the dataset are as follows:

- $lsales_i$ : natural log of sales
- $lsalary_i$ : natural log of salary
- $lroe_i$ : natural log of return on equity
- $ros_i$ : return on stock price
- $finance_i$ : indicator for finance industry
- $indus_i$ : indicator for manufacturing industry
- $consprod_i$ : indicator for consumer product industry
- $utility_i$ : indicator for transport/utilities industry

You are asked answer the following questions:

1. Explain why  $finance_i$  was removed from the model in (1). Write down one alternative version of the same model that would include  $finance_i$ .

2. What is the interpretation of  $\beta_0, \beta_1, \beta_2$  and  $\beta_3$ ?
3. Provide a summary statistics of the data. In particular, show the average value for each variables.
4. Estimate the model in (1), and show regression results in a concise table.
5. Construct a graph of the predicted relationship between  $lsalary_i$  and  $lsales_i$  **for each industry** (you may construct 4 different graphs, or plot all 4 relationships in the same graph). Then, provide a detailed explanation of the relationship between  $lsalary_i$  and  $lsales_i$  **across** industries.
6. Perform a joint hypothesis test of the following null hypothesis that tests whether we should include the different industries in the model:

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_5 = \beta_6 = \beta_7 = 0$$

Do you reject the null at the 90% confidence level? What about the 95% and 99% confidence level?

## Problem 2

To complete this problem, open the wage dataset included with the assignment on OWL (*A2.wage.xls*). This is a dataset containing information on the average wages and education level across U.S. states between 1950 and 1990. You are interested in estimating the returns to schooling,  $\beta_1$ , in the following model:

$$\ln wage_{it} = \beta_0 + \beta_1 educ_{it} + \beta_2 age_{it} + \beta_3 age_{it}^2 + u_{it} \quad (2)$$

The variables in the dataset are as follows (you will have to create  $age_{it}^2$ )

- $\ln wage_{it}$  is the average (log) wage in state  $i$  at year  $t$
- $educ_{it}$  is the average number of years of education for the population in state  $i$  and at  $t$
- $age_{it}$  and  $age_{it}^2$  are the average age and squared of age for the population in state  $i$  at year  $t$
- $ca_{it}$  is the compulsory number of years of education in state  $i$  at year  $t$
- $cl_{it}$  is the minimum age required for a child to start working in state  $i$  at year  $t$
- $year_t$  is the year of observation
- $state_i$  is the state of observation

You are asked to answer the following questions:

1. Estimate  $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$  using the OLS method. Report the results. What is interpretation of  $\hat{\beta}_1$ ?
2. What does the estimate of  $\hat{\beta}_2$  and  $\hat{\beta}_3$  suggest on the relationship between age and (log) wages? Plot the **predicted** relationship between age and (log) wage. What happens as age increases?
3. You are concerned that some unobserved variables (e.g. ability) affect both education and wages, creating endogeneity. Fortunately, you have access to two variables ( $ca_{it}$  and  $cl_{it}$ ) that may be used as an instrument for education:
  - (a) Explain why  $ca_{it}$  may be a valid instrument.
  - (b) Explain why  $cl_{it}$  may be a valid instrument.
4. Estimate  $\hat{\beta}_{1,IV}$  using  $ca_{it}$  as an instrument. Is it statistically significant at the 95% confidence level? *Note: you should follow the two-step method seen in class*

**first stage:**  $educ_{it} = \alpha_0 + \alpha_1 ca_{it} + \alpha_2 age_{it} + \alpha_3 age_{it}^2 + \epsilon_{it}$

**second stage:**  $lnwage_{it} = \beta_0 + \beta_1 \widehat{educ}_{it} + \beta_2 age_{it} + \beta_3 age_{it}^2 + u_{it}$

5. Estimate  $\hat{\beta}_{1,IV}$  using  $cl_{it}$  as an instrument. Is it statistically significant at the 95% confidence level?
6. Explain your findings. From the relationship between both  $\hat{\beta}_{1,IV}$  and  $\hat{\beta}_{1,OLS}$ , what can you conclude about the returns to schooling?
7. Perform the Durbin-Wu-Hausman test. Do you conclude that  $educ_{it}$  is endogenous at the 95% confidence level?