

## BEEM 011 Applied Econometrics 1

### Assignment 4 (August 2021 Reassessment)

#### Submission Deadline: 10<sup>th</sup> August before 12pm

Direction: Answer all the questions and submit your exercise in PDF. Include the necessary regression results, tables and figures as part of your main text. Also include your R scripts or computer codes as an appendix to your submission (You can copy your codes from the R script and paste them towards the end of the MS Word document). Please comment on each procedure so I would know what you are doing (or intend to do). Submission is via E-BART.

#### Tips:

- Install and load the following packages: “AER”, “sandwich” and “plm”.
- Label your plots and tables properly.
- Make sure you copy and paste all the necessary figures, tables and regression results.
- Be explicit about your assumptions. For example, what is your significance level when you do conduct hypothesis tests?
- Always make use of the evidence provided by your results (e.g., estimates, measures of fit, etc.).
- Be precise and concise.

(30 points)

1. The data for this exercise are drawn from the Panel Study of Income Dynamics (PSID). In R, type `data("Wages", package="plm")` to load the data. Load the “plm” package in the library. Then, type `wages.panel <- pdata.frame(Wages, index=595)` to convert the data frame into a balanced panel dataset and create *time* variable with values 1 to 7, pertaining to years 1976-1982. The individuals (*id*) in the sample are 595 heads of household between the ages of 18 and 65 in 1976, who report a positive wage (*lwage*, assumed that is measured in logarithm) in some private, non-farm employment for all 7 years. So, for each individual, we have seven annual observations on the following wage-determining characteristics: years of education (*ed*), years of full-time work experience (*exp*), weeks worked (*wks*), occupation (*blucol*, factor variable indicating if the individual has blue-collar occupation), industry (*ind*= 1, if the individual works in a manufacturing industry and zero otherwise), residence (*south* and *smsa* indicating if the individual resides in the south, or in a standard metropolitan statistical area. respectively), marital status (*married*, indicator variable that states if the individual is married), union coverage (*union*, indicator variable that states if the individual’s wage is set by a union contract), sex and race (*sex* and *black* which indicate if the individual is female or black, respectively.).
  - a. Estimate the model  $lwage_{it} = \beta_0 + \beta_1 exp_{it} + u_{it}$ . Print a summary of the model which reports standard errors that is robust to heteroskedasticity. Interpret the coefficients. Is the slope coefficient statistically significant? If years of work full-time experience

increased by twice the size of its standard deviation, by how much is wage predicted to change?

- b. Extend and estimate the model by including individual or entity fixed effects and time fixed effects. If possible, print a summary of the model which reports standard errors that is robust to heteroskedasticity and serial correlation. Describe your result. What happened to your estimated slope coefficient? Provide an intuition behind your observation.
- c. There are other variables in the dataset that could help better explain the variation of the dependent variable, in addition to what you have in (b). Briefly discuss why you included them in your model. Estimate and model and discuss your results. Make sure to provide an explanation behind the observed change in your estimated slope coefficient for *exp*.

(30 points)

2. Generate a subset of the data that includes only the first year for each individual (i.e.,  $time_i = 1$ ). Estimate the model  $lwage_i = \beta_0 + \beta_1 married_i + \beta_2 ind_i + \beta_3 ed_i + \beta_4 exp_i + \beta_5 (Married_i * ind_i) + \beta_6 (married_i * ed_i) + \beta_7 (ms_i * exp_i) + u_i$ . Based on your estimated regression, write two estimated regression functions: one for married and one for unmarried.

Then, test if all coefficients involving the variable  $married_i$  are zero. What can you conclude? Re-estimate the above regression but set  $\beta_5 = \beta_6 = \beta_7 = 0$ . Is the coefficient for  $married_i$  statistically significant?

Run another regression that allows the slopes to differ in addition but assume the intercept between married and unmarried workers to be the same. Test for the joint significance of  $\beta_5, \beta_6$ , and  $\beta_7$ . Do we have strong evidence to suggest that these slope coefficients are statistically different from 0? Looking at the tests you have performed, what is your conclusion?

(40 points)

3. Use the subset of the data you constructed in (2), and perform the following:

- a. Estimate the following regression model:

$$LWAGE_i = \beta_0 + \beta_1 WKS_i + u_i.$$

Interpret your estimated slope coefficient. Then, plot *LWAGE* against *WKS* and add the estimated regression line. What do you notice?

- b. Estimate an alternative model:

$$LWAGE_i = \beta_0 + \beta_1 \log(WKS_i) + u_i.$$

Interpret your estimated slope coefficient. Visualize your results using a scatterplot and add the regression line. In comparison to the previous exercise, what do you notice now?

- c. Using your estimated regression in (b), how much your predicted wage would increase if the number of weeks worked increased by 1 unit, from 10 to 11 weeks? And how your predicted wage would increase if the number of weeks worked increased by 1 unit, from 20 to 21 weeks? Briefly explain why or why not the change in predicted wage are different between two cases.

- d. Using the regression model in (b), determine the optimal order of a polylog model using sequential testing''
- i. Use a maximum polynomial order of 4 and the significance level of 0.05. Estimate a model which starts with the highest polynomial order.
  - ii. Assess the p-value (use robust standard errors) of the relevant parameter and compare it to the significance level.
  - iii. If you cannot reject the null, repeat steps (i) and (iii) for the next lowest polynomial order, otherwise stop and print the R-sq of the selected model.