Introductory Econometrics
Home Assignment 2

by Hieu Nguyen

Fall 2024

Solution of the assignment is to be delivered electronically to **254279@muni.cz** by December 11, 2024 23:59:59 the latest. Late submissions will not be accepted, resulting in zero points.

Form teams of two people, please. Only one team member is supposed to submit the solution with both team members' names and email addresses on the first page of the document. Teams are required to work independently, and any form of plagiarism will be treated accordingly. Please understand that the main advantage of teamwork is the synergy from solving the problems together and the possibility to share and discuss your econometric knowledge with your teammate. It is not about a pure division of tasks. So, please, do cooperate and make sure you both understand all solutions completely.

The text itself can be written in any software of your choice (MS Word, LaTeX, Pages etc.), but the .pdf format [5 MB max, .xls(x) can be attached in .zip] of the final document is required.

Please, name the file ECONOMETRICS_Surname1_Surname2_HA02.pdf.

In your report, please, be clear and reasonably concise, but do explain all essential steps (e.g., important matrices) of your solution/reasoning. Keep in mind that not only the correctness of your answers and interpretations is assessed, but also the text-editing quality is an integral part of your output.

Fingers crossed!
Hieu Nguyen

Problem 1: Wage equation for young males

(4 points: 4 · up to 1 point based on the quality and completeness of the analysis)

Dataset `wage4c.gdt` was used in the 1990s to study the interindustry wage differentials estimated for individuals. It contains 935 observations of monthly wages, study and occupational experience, and family and personal characteristics of young males. Find a specific description of the variables in the dataset.

1. Check the dataset, report the main summary descriptive statistics of the original variables used in the model equation below, and briefly discuss whether everything seems all right. Then construct the dependent variable. State and explain your working hypotheses about signs of the slope coefficients first, then estimate this model in Gretl and report the results:

$$\ln(\text{wage}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{tenure} + \beta_4 \text{married} + \epsilon.$$

   Finally, interpret the estimated coefficients $\hat{\beta}_1$ and $\hat{\beta}_4$.

2. State the null hypothesis that another year of general workforce experience has the same impact on wage as another year of tenure with the current employer. Test this hypothesis at the 10% significance level. What do you conclude? *You can do the test manually or via Gretl.*

3. Test the joint hypothesis that $\beta_2 = \beta_4 = 0$ at the 5% significance level. *First, do the test manually. Then, check your results by conducting the same but automated test in Gretl.* **Note: If you only show the test in Gretl, you won't get any point of this question.**

4. As economists, we might be interested in a potential issue of racial discrimination in the labor market. Enrich the model with three additional dummy variables, estimate it again, and report the results in the usual form:

$$\ln(\text{wage}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{tenure} + \beta_4 \text{married} + \beta_5 \text{darkskin} + \beta_6 \text{south} + \beta_7 \text{urban} + \nu.$$

   Holding other factors fixed, what is the estimated difference in monthly wage between dark skin and non-dark skin individuals? Is it statistically significant? Show clearly your test *either manually or via Grel.*

Problem 2: Wage Equation and Return to Education in 70s

(6 points)

One of the most recent Nobel laureates in Economic Sciences from 2021, David Card, used in his working paper from 1993 wage and education data for a sample of men in the U.S. in 1976 to estimate the return to education. Dataset `card.csv` contains 3,010 observations of hourly wages, schooling and occupational experience, family and personal characteristics, and potential proxies for unobserved personal qualities. Please find a specific description of the variables and the units of measurement in the attached `.txt` file. Be aware that there are missing values in the dataset.

The task of this creative empirical exercise is to develop your own explanatory/predictive model for the determination of individual wage ('wage equation'). The goal is **not to develop the best possible model** based on the given dataset but to successfully create a **relatively simple but useful and intuitive empirical model that includes the essential variables** while following general suggestions in parts 1. to 6. You should also carefully report the progress of your analysis step by step. If you would like to extend your analysis even further (either in an individual part or in general, e.g., with the multicollinearity analysis or by considering more variables), you are more than welcome to do so.

1. (1.5 pts) Suggest a few (two or three) intuitively the most important explanatory variables for the determination of wage and report and briefly describe their main summary statistics (also include the dependent wage). Discuss suitable functional forms of the variables and estimate the resulting model with OLS. Comment on and interpret the important findings from the result of OLS.

2. (1.5 pt) Suggest two additional potentially important explanatory variables and, taking advantage of the four important variable selection criteria, analyze if they belong to the model. Also, compare your new model to two alternative models with different functional forms between variables.

3. (2 pts) Add from the dataset two or three potentially important intercept dummies and explain why your selection makes sense. Re-estimate the extended model, interpret the newly estimated coefficients, and decide whether the new dummies should remain in the model. Next, add a slope dummy interacting with one of the included quantitative variables, explain your motivation, and interpret the newly added estimated coefficient. Finally, add another interaction term (between two quantitative variables or two dummies), re-estimate, and interpret. What is your resulting model after this step of the analysis? Interpret the overall significance of the regression.

4. (1 pt) Apply the White test for heteroskedasticity. Should we re-calculate the model to obtain heteroskedasticity-robust standard errors? If yes, please do so and interpret your results.

Attached:

1. Dataset: `card.csv`, `wage4c.gdt`

2. Description of variables: `card_description2.txt`