

Applied Regression and Time Series Analysis

Homework 3

Jeffrey Yau

January 1, 2016

Instructions

The weekly assignment serves two purposes: (1) Extend the materials taught in the asynchronous materials; some new concepts or techniques are introduced in the weekly assignment. (2) Ensure that you have learned the concepts, techniques, theories, statistical models covered in a specific week. Below are some guidelines:

- **Submission: Submit 2 files. Missing one of the two files will result in a 50% reduction in grade.**
 1. A report (in pdf format) detailing your answers and all the steps to arrive at your answers
 2. A well-documented R-script, jupyter notebook, or Rmd file detailing all of the codes used to arrive at your answers.
- Late submission will not receive any credit.
- Answers need to be typed. Latex, which you will likely have to use in R markdown, is great for typesetting documents with mathematical symbols.
- All the steps used to arrive at your final answers need to be shown clearly. These steps are as important as the final answer.
- The final answer of each question needs to be very easy identified; the use of bold fonts, highlights, or circling will help.
- This is a group project. Form a group with 3 or 4 people.
- Although this is a group project, we encourage you to attempt all of the exercises before discussing with your teammates. Do not use the "division-of-labor" approach. Each of the students in a group is expected to make sufficient contribution to the lab. If any of your teammate does not make sufficient contribution, please contact your instructor.
- **DO NOT copy and paste or even leverage on the solutions we gave to the students in previous semesters. Violation will be reported to the Director of the MIDS program and the Office that oversees UC Berkeley Academic Integrity. In any case, the lab has various subtle changes that make those answers not directly applicable.**

Exercises

1. Load the `twoyear.RData` dataset and describe the basic structure of the data
2. Typically, you will need to thoroughly analyze each of the variables in the data set using univariate, bivariate, and multivariate analyses before attempting any model. For this homework, assume that this step has been conducted. Estimate the following regression:

$$\log(\text{wage}) = \beta_0 + \beta_1 jc + \beta_2 univ + \beta_3 exper + \beta_4 black + \beta_5 hispanic + \beta_6 AA + \beta_7 BA + \beta_8 exper * black + \epsilon$$

Interpret the coefficients $\hat{\beta}_4$ and $\hat{\beta}_8$.

3. With this model, test that the return to university education is 7%.
4. With this model, test that the return to junior college education is equal for black and non-black.
5. With this model, test whether the return to university education is equal to the return to 1 year of working experience.
6. Test the overall significance of this regression.
7. Including a square term of working experience to the regression model built above, estimate the linear regression model again. What is the estimated return to work experience in this model?
8. Provide the diagnosis of the homoskedasticity assumption. Does this assumption hold? If so, how does it affect the testing of no effect of university education on salary change? If not, what potential remedies are available?