ECON 4650–001: Principles of Econometrics

Marcio Santetti | Spring 2021

Final Exam

Instructions: This Final Exam is worth 50 points, and it covers ***all*** contents of our course. Carefully read the following problems, and prepare your answers in an R script. That is, you will submit a *single* .R file with all your answers.

This assignment's page has a **template** for your answer script. Please consider using it.

This exam is **due 05/03** (Monday, 3pm). As per our class syllabus, late submissions **will not be accepted**. Upload your R script with your name (mine would be `marcio.R`) to this exam's page.

Contents

Problem 1

Suppose you are working with an econometric model and, after your first regression, you notice that the *standard errors* of your estimated coefficients ($\hat{\beta}_s$) do not look appropriate, either appearing to be too *low* or too *large*. Based on this problem, answer the following questions:

(a) Give *two* possible violations of the Classical Assumptions that may have originated such problem.

(b) Select one of your answers from (a) and explain the possible *origin(s)* of such problem. Feel free to use equations and/or other mathematical expressions to help in your answer.

(c) Given that the regression returned either too low or too large standard errors, what is the main *consequence* associated with this fact to your overall results?

(d) What *test(s)* can you conduct to confirm/deny your suspicion? If you select statistical tests, make sure to indicate null and alternative hypotheses.

(e) In case a violation is confirmed, what are possible *remedies* to solve this problem? Explain.

Problem 2

The `liquor5.csv` file contains observations on annual expenditure on liquor (*liquor*) and annual income (*income*), (both in thousands of dollars) for 40 randomly selected households for three consecutive years. This file is available at this exam's page, as well as a text file describing all its variables.

(a) Set up these data as a *panel data set*.

(b) Estimate a *fixed effects* model, with *liquor* as the dependent variable, and *income* and the first lag of *liquor* as independent variables.

(c) Interpret the *effects* of the independent variables from (b)'s model.

(d) Now, estimate a *random effects* model, with the same specification as before. Interpret the *slope* coefficients.

(e) Lastly, run a *Hausman* test comparing both models. Which one is preferred? Explain.

Important: After you are done, make sure to *detach* the `plm` package.

Problem 3

Based on the `OECDGrowth` data set from the `AER` package, let us estimate a classical textbook Solow growth model for OECD countries, based on data from Nonneman and Vanhoudt (1996):

$$log(gdp85/gdp60) = \beta_0 + \beta_1 log(gdp60) + \beta_2 log(invest) + \beta_3 log(popgrowth + 0.05) + u_t$$

where the dependent variable is the GDP growth rate between 1960 and 1985, *invest* is a measure of the accumulation of physical capital, and *popgrowth* is population growth (with 5% added in order to account for labor productivity growth). For the following questions, assume a significance level (α) of 5%.

(a) Run a Durbin-Watson test for first degree serial correlation in this model's residuals. Report the *null* and *alternative hypotheses*, the *p-value*, and your *inference* from this test.

(b) If we want to estimate an *augmented* version of the Solow model, it is possible to add "human capital" variables, such as education and research and development (R&D) covariates. The variables *school* and *randd* fill this gap. Add the *logs* of these two variables to the first model and run *again* the same test for serial correlation. Report the *p-value* and your *inference* from this test.

(c) Compare your results from parts (a) and (b). Do you believe this is a case of *pure* or *impure* serial correlation? Explain your reasoning.

(d) Interpret the *adjusted* R^2 from (b)'s model.

(e) Also from (b)'s model, interpret the effect of *invest* on the dependent variable.

Problem 4

Some more OECD data! Download the `oecd.csv` data set (available at this assignment's page), used by Everaert and Pozzi (2014) to examine the predictability of consumption growth in 15 OECD countries. For this exercise, however, we will only concentrate on US data.

(a) After importing the data set, filter only observations for the US economy.

(b) Estimate the following model:

$$csumptn_t = \beta_0 + \beta_1 hours_t + \beta_2 gov_t + \beta_3 r_t + \beta_4 inc_t + u_t$$

(c) Run a *RESET* test (using second and third powers) to verify whether (b)'s model is well specified. Assume $\alpha = 0.05$. What do you conclude?

(d) Next, estimate a new model. *Remove* the statistically *insignificant* variable(s) from (b)'s model, also adding the first *lag* of the dependent variable on the right-hand side.

(e) Then, estimate a *RESET* test for (d)'s model. What do you conclude? Explain.

Problem 5

Let us use the data set from Mroz (1987) one more time. It is also available at this exam's page, as well as a `.txt` file describing its variables.

Before we begin, run the following line of code (I'm declaring the data set as `mroz_data`):

```
mroz_data_sample <- mroz_data %>% filter(wage > 0)
```

This way, we exclude individuals who do not receive any wage. This sample will be used in parts (a), (b), and (c).

(a) Estimate the following regression model:

$$log(wage_i) = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + \beta_3 exper_i^2 + u_i$$

And interpret *all* slope coefficients.

(b) From (a)'s model, run *Breusch-Pagan* and *White* tests for heteroskedasticity (for the White test, make sure to include interactions). What do you conclude at $\alpha = 0.05$?

(c) In case you have rejected the null hypothesis of homoskedasticity in *at least* one of the tests from (b), estimate *Eicker-Huber-White* robust standard errors. Are the statistical significances (at $\alpha = 0.05$) the same compared to (a)'s model?

(d) Now, using the *full* data set, estimate the following *logit* model:

$$\mathbb{E}(lfp_i = 1) = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + u_i$$

(e) Estimate *Average Marginal Effects* (AME) for (d)'s model. Interpret the *slope* coefficients.