

Econ3338.01: Introduction to Econometrics I

Project Instructions

You will be working on a project where you use multivariate regression analysis to analyze economic data. You will be responsible for determining the research question, formulating the regression model, finding the relevant data and papers, performing the analysis and discussing the results. Chapter 19 in Wooldridge's "Introductory Econometrics" has many useful examples and suggestions for carrying out an empirical project.

Technical Details-Formats

Your final project (in pdf) must be uploaded to a designated folder on Brightspace by April 14, 2020.

Please submit a pdf file with your first draft for a format check by April 5, 2020. If any important part of the paper is missing or is not properly presented, you will receive an email within 3-4 days. Do not submit docx files. Brightspace automatically converts docx into pdf and it may lead to some distortions in the lines that contain special symbols and equations.

1. **Cover page.** The cover page should include your name, student number, date, project title, and the course number.
2. **Length.** The length of the paper, including figures, tables and references, should not exceed 12 pages.
3. **Font size and space.** The text should be double-spaced, with size 12 font.
4. **Equations.** Use an equation editor (built-in in MS Word) to specify your model(s).
5. Include page number.

Proposed Outline

1. Introduction

In this section, describe the research question and explain why it is important. Focus on the dependent variable. Provide a brief description of what you will do in your project (in each section), without getting into detail.

2. Literature review

Provide a short review of a few journal articles and/or books that are closely related to your project. Briefly discuss the data and estimation methods that were used in these studies, as well as the

results that were obtained. Include the complete reference for each reviewed study in the reference section.

3. Methodology

This section must discuss in detail what you will do in this project. You should mention the questions that you will answer and how you plan to do so. For example, you write that you will investigate the effect of education and experience on wages. This will be done by considering a multivariate linear model, to be estimated by OLS. If there is a similar paper in the literature, you must explain the difference between your work and the cited paper. Is it in the methodology? Do you include more independent variables in your analysis? Do you use a different estimation technique? Do you have a different data set?

Specify the regression(s) that you plan to run using the following format:

$$wage_i = \alpha + \beta edu_i + u_i$$

Focus on the independent variables. Explain why you have included them in the model and whether you expect them to have a positive or negative impact on the dependent variable.

4. Description of the data

In this section, you describe your data set in detail: the variables, their nature (continuous, categorical, or binary 0/1), time period that they span, the number of observations, and the source of the data.

Summary statistics should be provided either in tables or figures, depending on the type of data. For continuous variables, report their mean, variance (or standard deviation), minimum and maximum values. Binary and categorical variables should not be treated as continuous in the table of summary statistics. For binary and categorical variables, please report the frequency of each category.

Provide some discussion of the descriptive statistics of the dependent and independent variables. If you notice interesting or unusual patterns in your data, mention them here. You can also include some preliminary analysis about the relationship between variables of interest, using scatterplots between pairs of variables.

5. Results

In section 3 you have explained your methodology. In this section, you should estimate the models and report the results. The regression outputs and specification tests must be provided and discussed. In this section, you will also discuss the model specification and potential biases. Your results would be reliable if OLS assumptions are satisfied. You may consider additional independent variables or use different nonlinear transformations of existing variables.

If you have regressed the same variable of interest on different independent variables, you should discuss which resulting model is better in terms of the goodness of fit (*adjusted R²*).

After running the regressions, you should test for functional misspecification and heteroskedasticity.

Interpret the coefficients (economic and statistical interpretation) in the final model.

6. Conclusion

This is the final section of your project. You should provide a summary of what you have done. In one or at most two paragraphs, state the questions that you wished to answer and your main findings (independent variables that have some effects on the dependent variables and magnitude of each effect). How can you use these results for policy making (practical purposes)? You may also provide suggestions for further research (i.e., including other variables, considering different functional forms, using different estimators, etc.).

7. References

Please use the Chicago Author-Date style for references:

https://www.chicagomanualofstyle.org/tools_citationguide/citation-guide-2.html ,

https://economics.uwo.ca/undergraduate/wuer/WUER_Citation_Guide.pdf

or any other consistent reference style.

Appendix

Including tables and figures in the main text may lead to some difficulties regarding the layout of your work. Instead, you may place all your tables and figures at the end of the file. All tables and figures should be labeled (e.g. Table 1, Figure 5, etc.) and must have a title. When you discuss the results in the text, use table and figure numbers to refer to them. In Section 5, refer to relevant tables and figures when discussing the results as follows:

“The results of running the regression. can be found in Table 2. The coefficient for education is statistically insignificant....”

If you move all tables and figures to the Appendix, the Appendix should have two separate sections: one for tables and one for figures. Do not include Stata commands or screenshots for the tables. Instead, create standard tables of summary statistics and the regression results in Excel, using command `outreg2`. Relevant Stata commands will be posted later on Brightspace.

Instructions for Section 5 “Results”

Note: If the dependent variable is binary, estimate all models using robust standard errors.

What independent variables should you include in your models?

Sometimes you have a theory that determines which independent variables you should include in your model (for example, you may try to quantify the parameters in a Cobb-Douglas production function: the independent variables are labor and physical capital). In other cases, you do not have such a theory but you have a large set of independent variables that you believe can be used to explain your dependent variable. In such a case you are not sure which variables you should include and which not. You can proceed in the following way:

- Begin with all independent variables that you think are relevant and that do not have near perfect collinearity issues. If applicable, you may consider cross products of these variables (you need to justify why you did so).
- Estimate your model with all of them.
- Check the output. Some of the parameter estimates may be statistically insignificant. Consider the hypothesis that they are jointly insignificant (F-test). If the null hypothesis is not rejected, remove the independent variables associated with these parameters. Otherwise, consider all the hypotheses that are related to subgroups of these coefficients. Estimate your model again and repeat the hypothesis testing. The coefficient estimates should not be considerably different (especially their sign), compared to the previous models, or else you may have an omitted variables bias. Examine again if the OLS assumptions are satisfied.

This procedure is called *general-to-specific* approach.

In Section 5:

1. Estimate your regression model as discussed above.
2. What is the adjusted R-squared? Could you add more independent variables? If you decide to explore the case of using more independent variables by including them in the “updated” model, use again the general-to-specific approach.
3. Perform misspecification testing (RESET, Breusch-Pagan and White’s tests) in the following order:
 - If functional form is found to be a problem when using RESET, change the specification by applying logarithms to suitable variables or by adding squared terms of some of the independent variables.
 - Estimate the “updated” model and check again for functional form using RESET. Select the model that looks less misspecified.
 - Check the “updated” model for heteroskedasticity (B-P and White’s tests). Note: if the dependent variable is binary, do not test for heteroscedasticity and always use robust standard errors.

- If heteroskedasticity is not an issue then you are ready to discuss your regression output results.
 - If heteroskedasticity is found then re-estimate your “updated” model using robust standard errors.
4. Discuss the regression output.
- Report the result of the goodness-of-fit test and the adjusted R-squared from the Stata regression output.
 - Individual parameter estimates: statistical interpretation
 - If you found a variable to be significant, report it and mention the significance level. In the regression output, the null is that the individual parameter is equal to zero (statistically insignificant).
 - If a variable of interest is insignificant, report it as well (do not delete it from the model). It means that based on the data set, you have found no statistical evidence of an effect of this independent variable on the dependent variable.
 - Reminder: if you have two or more independent variables that are individually insignificant and you consider removing them from the model, check whether they are jointly insignificant. If the F-test of this restriction does not reject the null, you can remove them.
 - Individual parameter estimates: economic interpretation
 - You can now interpret the parameter estimates in economic terms, i.e., what effects the independent variables have on the dependent variable. Do you think that the effects are large?
 - Provide economic interpretation for both statistically significant and insignificant parameter estimates. Interpret them accordingly.

Important: whenever you perform hypothesis testing, the p-value is relevant to the null hypothesis.

- a. $p\text{-value} < 1\%$: reject the null hypothesis at the 1% significance level
- b. $1\% < p\text{-value} < 5\%$: reject the null hypothesis at the 5% significance level
- c. $5\% < p\text{-value} < 10\%$: reject the null hypothesis at the 10% significance level
- d. $P\text{-value} > 10\%$: do not reject the null hypothesis.

Project Rubric

		Approximate weight (%)
Technical	Model description	10
	Data description	15
	Selection of independent variables (including F-tests)	15
	Misspecification testing	10
	Statistical significance of the results (discussion)	10
Qualitative	Introduction and motivation of the research question	10
	Literature review	10
	Economic significance of the results (discussion)	10
	Layout / format	10