AREC/ECON 335
Introduction of Econometrics
Instructor: Lackson D. Mudenda
Instructions for Individual Final Projects

CAUTION: No credit will be given for late projects. Projects are to be completed individually. Your submissions should be your own work and not the work of ANY OTHER person. Plagiarizing, copying, or any other evidence of academic dishonesty will result in a score of ZERO. Please *start early* and stop by office hours if things come up.

Your final project for this course is a chance for you to demonstrate what you have learned about econometrics. To do this, you will be creating and executing your own econometric study! So, take some time to come up with a question that interests you (it could come from econ or another subject or theoretical basis) *and* has easily accessible data. Have fun and be creative!

You will have two submissions related to your project. The first will be ONLY a few tables of summary statistics and regression results with a line sentence describing each regression. After approval of Part I, you will go on to do the final write up of your paper as Part II. By me looking at your regressions before your final submission, you get a chance to correct what you may have done wrong. Note that, you can get approval early on by submitting tables days before the deadline(You don't have to wait for Part Ideadline). However, each submission (parts I and II respectively) should *not exceed 5 double spaced pages*. Longer submissions are not necessarily better and will lose points if over the page limit.

Cover Page (**Both Submissions**)
Please include and sign the following:
"Honor Pledge: "I did not give, receive, or use any unauthorized assistance on this project."
Signed_____."
This cover page does not count toward the page limit.

"Unauthorized assistance" includes submitting a paper or part of a paper from another class this term or any past term by you or any writing by anyone else. You may, however, use *data* that you have used in another context if your research question and major analysis is different and your submission to this class is unique. Please be advised that the online submission will be checked automatically using the plagiarism detection software available at CSU.

Appendices (**Both Submissions**)
Attach copies of all RStudio (or other computer program) original output, even though you have also formatted this information into tables in the main text. These output appendices do not count toward the page limit (though the formatted versions of the tables within the text do). To clarify, it is required to have *both* (1) nicely formatted tables within the text and (2) RStudio/computer output files/printouts at the end. Original figures from your computer output may appear either in the main body or in the appendix depending on your preference. Your appendices will be different across Parts I and II (though you should have an appendix section in both submissions).

Data Guidelines
Your data set must have at least *50* observations and might have a lot more. It is up to you as to where you find data. Lots of economic data are available on the web. (Your data does not need to be strictly "economic." You should NOT simply use data that came with the book (or other textbooks) or that is downloadable from the Statistical Software websites. Instead, you should look for a topic which is relevant to you and find data to test it. This is your opportunity to do something new and creative. Note that you will likely want to concentrate on cross-sectional data since we focused on this type for most of the class (and since methods vary for time-series and for panels). If you do still choose an alternate type of data, you should be careful in your write-up to critically list any interpretation problems that may remain especially in your conclusions section in the second submission.)

The Yale Law Library has a great list of U.S. data sources at: https://library.law.yale.edu/news/75-sourceseconomic-data-statistics-reports-and-commentary, and international organizations such as the World Bank

(https://data.worldbank.org/) and the International Monetary Fund (https://www.imf.org/en/Data) also have public use data. So do state and local government websites such as https://data.colorado.gov/ and https://opendata.fcgov.com/. There's also lots of data on other things in life (e.g., financial data at https://www.wsj.com/market-data; sports data at https://www.espn.com/ (pick a sport and then select "stats")). There are all kinds of other sources out there too. It is up to you where you locate data.

Instructions:
Part I of your paper should for the most part be formatted tables of results with appendices of RStudio (or other software) output. Basically, in the first submission include scatter plots of your Y and the main X variable, tables of summary statistics and regression results. The regression results should have at least three (3) specifications.

After approval of Part I based on your submissions, you can then start writing your final paper. Your final paper (i.e., Part II) should have the following, clearly labeled sections:

1. *Introduction and Statement of Research Question*
    - Pose and motivate the question to be investigated with your data and econometric model.
        - Good econometric questions are generally based on economic theory; however, econometrics can be used to analyze all kinds of cause and effect relationships even if they don't directly relate to previous theory courses. You can study just about anything that interests you. Since the goal of the project is causal identification of the effect of one variable on another, a good question should **specify the one primary X variable and one primary Y variable** that you are interested in.

2. *Formulation of the Model*
    - Express the question both verbally and in the form of an equation to be estimated. Be sure to include the names and units of the dependent and independent variables.
    - You must include **at least three independent variables**, but can include as many as you'd like beyond that.
    - Explain the mechanisms you believe pertain to your study question and why you selected the particular X variables that you did and why you didn't select others.
    - You may also want to briefly cite relevant literature (e.g. previous studies) that you come across while researching your topic.

3. *Data Description*
    - Describe the data used to estimate the model. Your data set must have at least **40 observations**.
        - You should **construct a table that provides the mean, minimum, maximum and standard deviation** (and any other summary statistics you feel are relevant) of **all variables** in the model.
            - Describe the content of the table(s) in the text of your paper.
            - You might also want to show and discuss key scatterplots or graphs of interest.
    - It is up to you as to where you find the data. Lots of economic data are available on the web (e.g. government agency websites).
    - Your data does not need to be strictly economic though.
        - Look up Freakonomics and other work by Steve Levitt; you can study whatever interests you. If you like sports, pull some data off ESPN; if you like comic books, key in data from a price guide; if you raise horses, you may know where to get sales data; etc.
    - You should **not** simply use data that came with the book.
    - This is your opportunity to do something new and creative.
    - You may want to focus on cross-sectional data since we discuss this most in class, however if you do choose an alternate type of data analysis you should be careful in your write- up and interpretation.
    - To use R, it may be necessary to import your data from another format (e.g., Excel, ASCII, CSV, etc.). To do this, use *the steps we have followed in Problem sets)*. If data you're interested in is in a different format other than .csv, let me know so I can show you how to load into RStudio.

4. *Empirical Results*
    - Present the results of the OLS estimation (and/or other appropriate technique learned in class) **in the form of a table(s).**
        - Tables should be formatted as easy-to-read tables rather than cut and pasted out of RStudio.
        - You may want to, for example, examine your key parameter estimates with and without additional regressors to compare any differences (i.e., run a single regressor model and then a multiple regression model).
    - You should show **at least three specifications of your design**.
        - These will likely include various non-linear versions of a baseline model, for example by adding a polynomial or interaction variable, or by creating relevant categorical or binary variables to analyze group effects.
    - Explain the **meaning of estimated coefficients** in your model and whether or not coefficients are **statistically significant** and at what level.
    - Also discuss the **goodness of fit** measure for the model (e.g. R-squared, SER, adjusted R-squared).

5. *Summary and Discussion*
    - **Summarize your main results and discuss limitations** to your approach.
        - Carefully explain caveats to your study, especially whether or not certain parameter estimates may be biased.
        - Suggest what other independent variables, possible functional forms, or statistical tests might be appropriate to include and any interesting follow-up questions or extensions that may have come to mind.
        - Address all issues of internal and external validity.

Bibliography

- All cited sources should be reported in a bibliography in **MLA format**.
    - One useful reference is your textbook.
- Also include anything relevant that is not within the main document at the end, such as **attachments** of key R output (Here I want regression or summary statistics outputs, and not a series of lines of code)

Formatting:

- Your report must be typed and proofread.
- There is **no set minimum length** of your paper as long as you cover all of the requirements.
    - However, your paper should **not exceed five(5) double spaced pages with 12-point font and one-inch margins**.
    - This five-page maximum does not include attachments, graphs, tables etc.
- Projects are to be turned in via Canvas and **must be completed individually**.
    - Your paper should be only your own work and not that of **any other** person.
    - Plagiarizing, copying, working with other students, or any other evidence of academic dishonesty will result in a score of **zero** on this project.
- Please **include the following in your paper and sign by typing your name:**
  "Honor Pledge: 'I did not give, receive, or use any unauthorized assistance on this project.' Signed ----------"
    - "Unauthorized assistance" includes submitting a paper or part of a paper from another class this term or any past term by you or by anyone else. You may, however, use data that you have used in another context as long as your research question and major analysis is different and your submission to this class is therefore unique.

This looks like a lot but is really your opportunity to show that you can use basic applied econometrics outside of just reciting info on an exam. This project should illustrate why this toolset should now appear on your resume!

These projects also make excellent discussion items at job interviews. Start early, do your best, and be creative!