ECON 4400-Project

Spring 2020

Due April 10 by 5:00 pm The paper must be uploaded to Carmen in a .pdf file. **The turn-it-in feature is enabled, meaning any paper that plagiarizes from another person who has ever completed this paper in this class will be flagged.**

1 Basic Description

Objective:
The purpose of this class is to apply the analytical and quantitative skills which should be acquired in this course. The final project should be a professional looking manuscript with easily interpreted graphics and charts.

Description:
The project requires using statistical techniques learned in this course (OLS) to analyze data from the Current Population Survey (CPS). You must determine the specific topic that you want to examine using the data set that is posted on Carmen.

Formatting:
All aspects of your final paper must be typed. The paper should be double spaced with font of 11 point. There is no minimum number of pages. **There is a strict maximum of 12 pages including title page, tables, graphs, and appendices.** Anything beyond 12 pages will not be graded. The paper should include the following sections: introduction, data, empirical methodology, results, and conclusions.

Data:
You will use the data set cps_march_2019.dta found on Carmen. The data is a sub-sample of the Community Population Survey (CPS).

- The CPS, sponsored jointly by the U.S. Census Bureau and the U.S. Bureau of Labor Statistics (BLS), is the primary source of labor force statistics for the population of the United States. The CPS is one of the oldest, largest, and most well-recognized surveys in the United States. It is immensely important, providing information on many of the things that define us as individuals and as a society - our work, our -earnings, and our education.

To avoid the complexity of working with panel data, I have taken the time aspect is taken out. There are literally thousands of data points available, but I have narrowed it for you. The data set that you will work with includes 99 variables and for some variable over 200k observations. **In your paper, you will not work with all 99 variables and 200k observations. Instead, after you decide on your topic given the variables in the data set, you will narrow down the number of variables and limit the sample size.**

2 Variable definitions

You will be able to find document with variable description and detail information about each variable in the data set on Carmen in the folder Project\Data\Code Book. In addition, you may see short description of variables in STATA after you load the data set; use command: *describe* to get variable description. When choosing the variables for your project make sure that the variable described in the code-book is present in the data-set. Many of the variables are categorical and many that seem ordinal are not. Simply type "tab variablename", i.e tab citizen, and Stata will display how many observations belong to each category and what those categories indicate.

3 Working in a team

This is a group project. Besides applying the tools learned in the course, the goal of the project is to develop your team skills such as learning to listen, division of labor, solving disagreements, etc. **In case your partner drops from the class during the semester, then you will need to complete the project on your own.**
The paper must include all names. In order to get some bonus points there should be more work. This could include, but is not limited to: a literature review section, additional specifications included in output tables, use of interaction terms, linear probability models, other techniques covered in the text that we have yet to discuss, etc.

4 Tables and Figures

As described in the rubric, there are tables and figures required in the data and results sections. Sections on Data and Result are description of the tables and graphs in your paper. All tables and figures should follow the same general format:

- All objects should be labeled, numbered, and titled. For example: Table 1: Summary Statistics.
- All objects should provide insight into your paper. If your paper is about gender wage differences, then the figure that you include should be some sort of graphic that shows some detailed differences in wages by gender.
- Summary statistics tables should include means and standard deviations, min, max, number of observations for each variable.
- Regression results tables should include standard errors and stars for 1%, 5%, and 10% significance levels.
- Similar regressions should all be in the same table.
- Summary statistics tables should only include variables used in your analysis.
- Categorical variables that are non-ordinal must be broken up into dummies in the summary statistics tables.

- Ordinal categorical variables should have the categories described in the or table footnotes. For example, mother's education could have a mean of 3.8. Explain that.
- The main idea of a table or figure should be plainly understood without having to read the text.

5 Originality

All papers are turned in through Carmen using the turnitin app. I expect that the similarity scores will not be 0% since everyone is using the same data set and the data descriptions will be similar. A similarity score above 20% is problematic. **If the paper shows up as not searchable pdf, it will not be accepted.**

6 Commonly missed Points

Grammar: Do not simply write your paper in another language and use google translate to change it to English. Those papers are nearly impossible to read. Reading the paper out loud to yourself before you hand it in will help tremendously.

Following Directions: Include everything that is on the attached grading rubric and in the correct section.

Potential Problems with your regression: All methodologies have some potential problems. When you address those problems in the methodology section do not include problems that you can fix with your current skills and data, specifically discuss which coefficients could be affected and the direction of the affect.

Equation: The equation that is written out in the empirical methodology section must be the same equation used to generate the tables. Write out β_1 rather than B1 or $\beta 1$. If the equation has perfect multicollinearity or treats non-ordinal categorical variables as continuous, there will be large deductions.

Tables and graphs checklist: Missing any part of the checklist will result in significant deductions.

- Tables and graphs must be understandable without reading the text. This means there should be a clear title that explains the contents of the table, such as "Summary Statistics".
- Do not use variable names in tables and figures.For example, rather than momedu my table should say "Mother's Education".
- There should also be a footnote which adds some detail to the and explains what is shown. For example "Each entry is an OLS coefficient with standard errors in parentheses. Stars represent p-values as ..."
- If plotting distribution in a figure, use percent as the y-axis.

- All figures and tables should be referenced in the text.
- Figures must show more than just means.
- Cut and pasting from the stata output window will get 0 points. Use the estout program to create the tables or type them in a word processing program.

Flow of your essay: The project is to be a coherent essay just as if you were writing an essay for a literature class. This means that the paragraphs should have more than one sentence. In fact, a well-written English paragraph typically has about 6 sentences. A series of 1–2 sentence paragraphs is nothing more than a list.

7 Grading Rubrics

A condensed version of this form will be used to grade papers.

1. _____**Formatting 5pts**: The paper conforms to the formatting standards outlined in the assignment: Double spaced, font, margins, less than 12 pages, and labeling sections. The file is a .pdf. [1]
2. _____**Organization 5pts**:
   - Ideas are presented clearly, free of spelling and grammatical errors. Any references are cited. (-1 for each instance[2])
   - The paper is not cut and pasted from homeworks and transitions well between sections. (5) (**Not a list of tasks.**[3])
3. _____**Introduction 5pts**: The question to be analyzed is described. The importance of the question is discussed. The data source is mentioned. The empirical methodology is mentioned. (OLS) A preview of results is given
4. _____**Data Section 15pts**:
   - The source of the data is mentioned. There is a brief discussion of means, standard deviations, etc. of the wage variable and whatever other variable(s) on which you choose to focus. (5)
   - A summary table with means and standard deviations which is referenced in the text. All variables included in your regression must be in the summary statistics and categories such as race must be broken into dummies. (5)
   - At least one graphic (histogram(don't use density), scatter plot, etc.) which describes data and is referenced in the text. The graphic must be relevant to your topic. (5)
5. _____**Empirical Methodology 15pts**
   - The question being analyzed is clearly described and the estimated equation is written out. (5)
   - The inclusion of each variable is supported by theoretical reasoning and predicted signs of each variable are discussed. If your equation has perfect multicollinearity or uses non-ordinal categorical variables as continuous variables you will lose all 5 of these points. Your regression equation is correctly specified (i.e. you have used logs or polynomial if necessary). (5)
   - Potential concerns with sample selection, omitted variables, etc. are discussed along with the consequences of those problems. Discuss specifically which coefficients are affected and how. (5)
6. _____**Results 30pts**
   - Results table. Must be clean, understandable without reading text. (Cut and paste from STATA output is a zero.) (7)

[1] Negative scores are possible for this section.

[2] Up to 20 points can be deducted

[3] If your paper is just a list of tasks and does not resemble a research paper, you will be deducted 20 points.

- Interpretation of coefficients (8) (You must interpret at least 4 coefficients including at least one 1 dummy variable and 1 continuous variable.)
- Hypothesis testing: discuss which coefficients are statistically significant at a 5% significance level; discuss the R^2 of the regression; conduct an F-test on some aspect of the regression results. (5)
- Use your estimates to predict the outcomes of two hypothetical people. (Unless you understand how to calculate predicted values with a logged dependent variable, use a level-level for this part.) (3)
- Discuss the consistency of your results with your predictions (i.e. you should have some expectations about the results before your run your first regression). (2)
- Estimate an alternative specification, present (in a table) and discuss. (5)

7. _____**Conclusion 5 pts** The conclusion summarizes results and the significance of those results.
8. _____ **Overall Quality 10 pts** This is the only subjective portion of your grade. The average score will be 6/10.
9. _____ **Overall Difficulty 10 pts** Difficulty points are granted for the relative econometric rigor of your paper. This may include working without a partner, creating more complex graphics, using interaction terms, developing a unique question, running additional regressions, citations and descriptions of relative economic literature (must be available on EconLit) etc.