**Economics 465 – Econometrics II**

**Final Project Guidelines**

**Fall 2015**

**Background**

The econometric tools you have been learning and practicing are regularly applied to solve important real-world problems. This project will give you the opportunity to apply your skills in such a setting. You will solve/analyze an econometric problem/dataset. There are many problems posted on kaggle.com, a platform by which firms, government agencies, and other organizations present unsolved problems to researchers, frequently offering a cash prize to the authors of the best solution. Econometric datasets, specifically time series in nature, are available on various commodities oriented web sites such as:

<http://www.indexmundi.com/commodities/>

<http://www.worldbank.org>

<http://research.stlouisfed.org/fred2/tags/series>

<https://www.quandl.com/c/markets/commodities>

Ultimately, it is up to you what problem/dataset you wish to tackle. You will determine the most appropriate methods to prepare and analyze your data and thoroughly describe your approach and findings. This project replaces a final exam in the course. While you might not use every econometric method covered in class, you will demonstrate your ability to perform the six course objectives:

1. Formulate and estimate an appropriate regression model for a given policy question and then interpret and explain the model’s results clearly using non-technical language.

2. Critically assess a model’s results using statistics and tests appropriate for the data, methods, and policy question.

3. Use appropriate econometric techniques to identify and (if possible) rectify violations of the classical Gauss-Markov model assumptions that may be present in your model; including:

a. heteroscedasticity

b. measurement error

c. simultaneity bias

4. Identify and implement appropriate methods for analyzing panel data and then interpret the results.

5. Identify and implement appropriate methods for analyzing time-series data, including:

a. dynamic models with lagged variables

b. autoregressive models

c. models subject to non-stationarity

**Authorized Resources**

Any Econ 465 (or archived Econ 465) materials on the Dropbox folder, your notes, your textbook, the materials on the textbook website, any other notes or textbooks from other statistics classes you have taken, background information and data on the Kaggle website, the “Forum” (discussion board) portion of the Kaggle website, and any other sources with relevant information on the topic or econometric methods that you choose. Details of winning entries to the competition you choose (if available) or other sources of “solutions” to your specific problem are **not** authorized. You may use Stata and other software for data processing as needed. You will work in groups of 2 or 3 on this assignment. You must have the members of your group identified to your instructor by **lesson 4**. You are responsible for ensuring that all parts of the submission reflect the understanding of all group members.

**Assignment**

The final project consists of several steps that will culminate in a report due at the beginning of class on **lesson 38** and a group presentation to be scheduled during the last three lessons of the course (M38 through M40). Your grade (discussed in more detail below) is based directly off of your final submission, but of course this will reflect your work on all steps. As with work you are likely to face as an Air Force officer, this assignment is more open-ended than most classroom assignments. Prudent individuals will review all of the requirements for the final deliverables and intermediate milestones, enabling them to plan ahead on their work and ask questions as needed.

**Milestones and Peer Review**

You will submit three “milestones” before the final project is due. These will help you to stay on schedule for completing the project and give you an opportunity to receive feedback on the direction of your work. You will also review one Milestone #2 submission and one Milestone #3 submission created by your peers. You will provide candid but constructive feedback on their work. While your review will not affect your peers’ grade directly, you will help them improve their final project in the same way that you would want your peers to help you. Scrutinizing your peers’ work (which will be based on a different project than your own) may also help you realize ways that you could make your own work clearer to the target audience. The milestones and their deadlines are in the table below and on the course outline.

|  |  |
| --- | --- |
| Milestone | Due Date |
| 1. Group and topic selection | Lesson 8 |
| 2. Initial Analysis | Lesson 21 |
| Peer review | Lesson 22 |
| 3. Rough Draft | Lesson 30 |
| Peer review | Lesson 31 |

**Final Deliverables**

You will submit a final report and deliver a final presentation as a group. The final report will document all the work your group did and include an appendix with all your calculation and software outputs in sufficient technical detail to enable your results to be reproduced. The group final presentation will provide an overview of your work to an audience that is familiar with technical methods but not with your project.

**Final Report (200 points)**

The body of the final report should include sections that are aimed more at executive management and then others aimed more at technical researchers. The final report should include the following major sections:

**Executive Summary** – A complete but concise description of your work to entice potential readers into reading the remainder of the report. It should answer the following questions:

• Why should the reader care about the problem and results?

• What problem is the report going to try to solve?

• How did you go about solving the problem?

• What are the results?

• What are the implications?

The executive summary should be aimed at a layperson who is familiar with the topic you are studying but not necessarily with econometric terminology. It is meant for executives at the firm sponsoring the research.

**Motivation and Dataset Overview** – Why did you pick this problem? What is the problem you are actually trying to solve and why did it interest you. A clear problem statement is key because it will focus all of your data wrangling and analytical efforts.

Here is also a good place to provide a little background information about your topic. Be sure to include at least two outside sources (with proper citation) to help provide context for your problem and possible avenues of analysis.

The dataset overview should provide a short description of the dataset you used to conduct the research. At a minimum, you should address:

• Where the data came from – a specific source should be cited

• What type of data it is (cross-sectional, time-series, panel)?

• How was the data collected?

• Provide a list or table of variables used in the analysis and how they are defined (don't forget units).

• What are some basic descriptive statistics (e.g., number of observations, number of variables, what each observation represents, summary statistics if appropriate, density/sparsity of the data)?

The dataset overview should make it clear to the reader where the data came from and allow the reader to get the data in order to re-create your analysis. In total, the motivation and dataset overview section should set the stage for the rest of the report, giving the reader a clear picture of what the problem is, why the problem was important to you, and the data you will use to address the problem.

**Data Processing** – A detailed account of all the data manipulation you conducted in order to get the data into a proper format for analysis. The procedure you describe should be completely reproducible by another analyst with similar skills. This section should address such topics as:

• How you handled missing values (or evidence that there were no missing values).

• How you defined any new variables.

• Which observations you used in your analysis.

The level of data processing required will depend heavily on the topic and the data. In the past, this has been one of the most challenging aspects of the project; many groups spent the bulk of their time on this so-called "data wrangling" effort. So, **plan accordingly** and record everything you do so you can recreate it and write it up later. In the end, you should have a tidy data set that the reader can reproduce based on the steps you described.

**Analysis** – The analysis section of the final report should be the biggest section of the report. Here is where you describe all the analytical work you did with your tidy data set to come to some kind of resolution to your original problem statement. Things to address in this section include:

• The model (or models) you created and the evolutionary process used to create it (them).

• The econometric (or other mathematical) methods you used in your model and why.

• Any assumptions you made and justification for their use.

• Any tests you did to validate the assumptions.

This section can get pretty technical, and that is OK. You should provide all the information required to reproduce the research you conducted; so this means including Stata code as well as any other programming algorithms (e.g., **R** code) that you used. Be sure not to just copy and paste lines and lines of code, neglecting to explain the econometric theory and reasoning behind your coding efforts. You should use sound reasoning and judicious inclusion of code snippets to explain how you applied theory to create your models and ultimately use your data to answer the problem at hand. Use the code to tell the story of your analytical thought process and how you used a tool to do the rigorous mathematics. By the end of this section, the reader should be convinced that you did a thorough job examining the data and used your technical expertise to address the original problem statement. This is where your credibility as an analyst will be solidified.

**Results and Key Findings** – This section should describe the results of your analysis. Here is where graphics and tables will come in real handy. Your results should expand upon what you alluded to in the executive summary. Include an assessment of your confidence in the findings, using statistics relevant to the question at hand. Don't just paste Stata’s raw text output, be sure to format it into a Word-friendly table so that the reader is drawn to your key findings and not distracted by tons of minutia.

**Conclusions, Implications, and Recommendations for Further Work** – Draw a conclusion about the work you did and how you answered the original problem statement. Describe any implications your work has for the field you researched. Lastly, provide some recommendations for further work; if you had more time, money, resources, what else would you do to better address your problem.

There is no specific format or set page length. Include the major sections above and address everything in them as a minimum, but add sections and issues you deem necessary to craft your report. The bottom line is, you need to convince the reader that you:

• understand his problem

• collected and processed data pertinent to the problem

• conducted a thorough analysis of the data using sound econometric and mathematical techniques  
 • drew appropriate conclusions based on the results of your analysis

If you do this in 5 pages or 20 pages, as long as it's thorough, neat, and well written, you will get a good grade. The final report should contain any tables and figures necessary to support your story. Don't include data in the prose and then reproduce it in a table, that is just simply redundant. Graphical representations of your data or the results of your modeling efforts will greatly enhance your analysis. But, including every graphic that Stata prints out without referencing it in your report is simply distracting and will not enhance the quality of your analysis. Also, when copying tables from Stata (or other mathematical programs), be sure the formatting is clean and neat. Nothing takes away from the professionalism of a project more than sloppy cut-and-paste work. It may be best to just take the data and re-format it in a Word table to make it presentable. Watch those font sizes too, don't make your work so small that it is impossible to read.

**Final Presentation (50 points)**

Your group will deliver a presentation about your topic and analysis. The audience for the presentation is your instructor and your peers, all who have a good understanding of econometric methods but may not be familiar with your topic or dataset. All members of your group should participate in some way, and all should be prepared to answer questions from the class or instructor. The length of the presentation should be 15 to 20 minutes, not including questions. Your group should aim to communicate the following:

• What is the problem, what motivated you to analyze it, and what is the goal of your analysis?

• What data was available, and how did you process it to conduct your analysis?

• What econometric/mathematical methods did you use to analyze your problem?

• What were your key findings?

• What were the limitations of your analysis and/or the data?

• What would you recommend to the sponsor based on your findings?

• What else would you research/examine given more time/resources/skills?

Audience members are expected to follow their peers’ presentations closely and ask relevant questions after the allotted time. All cadets will also provide written feedback to their peers. Details and a link to the feedback form will be provided by email in advance of the presentations.

**Documentation**

In addition to documenting any help you received on the assignment, you should include appropriate references for any ideas or direct quotes that you include in your writing. Appropriate references requires both citations (brief references to an outside source included in your text) and complete references (all information needed to find the original source, listed at the end of your document). These citations and references should be in a standard format such as the APA style. Submissions without appropriate documentation are subject to the usual 5 point penalty *in addition to* deductions for missing information as specified in the grading rubric.

**Late Work**

The instructors expect that all students will submit all final project deliverables on time. Consistent with penalties for an unexcused absence for a GR (FOI 36-173), any component of the final project that is not delivered on time will incur a penalty of 25 percent of the total points for that component. Contact your instructor immediately if you have reason to believe that you may miss the final deadline.

**Grading**

Separate grading rubrics for each milestone and the final report and presentation address the specifics about your grade on this project. See them for further details.

Regarding the milestones, they basically address the following:

• **Milestone 1** is the project proposal and analysis plan. It ensures you have a group, a problem statement, a viable dataset, and a plan for how you plan to address your problem.

• **Milestone 2** is your initial analysis. At this point, you are half way through the semester and should be half done with the project. You should be pretty much done with your data wrangling and have an initial model developed. From this initial model you should be able to draw some preliminary results and explain what they mean. You should also have a good plan of action for the remainder of the semester that will refine the work you've done.

• **Milestone 3** is your rough draft. At this point, you should have a solid 80% solution to your problem. The data should be tidy and all issues in it fully addressed. You should have a firm model build with some solid results. The rough draft should mirror the format you plan to use in the final report, and you should be thinking about how to present all your material. From this point on, you should be refining your work, not building whole-sale new models or analyzing new data sets.