Group project insturctions

Proposal due date: 9am, November 5th, 2012

Final due date: 9am, December 7th, 2012

Group assessment due date: 9am, December 10th, 2012

Assignment: This project requires you to pick an appropriate data set, propose a creative research question, and test this question using OLS regression analysis techniques. You should refer to chapters 16-18 of the Keller book in completing this project. You will work in groups of about four, and responsibilities should be shared as evenly as possible across all members of the group. You will be graded on this project in two ways. One, I will grade the proposal and final project that your group turns in. This grade will apply to all group members. Two, each group member will have a chance to give me feedback on each of her peer's effort. Only I will see these, and based on this input, I will assign an individual grade to each group member.

I will leave the topics of your papers up to you. The Internet is replete with economics data. For examples, see:

http://www.lib.umich.edu/govdocs/stecon.html

http://www.econdata.net/

http://usa.ipums.org/usa/

http://www.bls.gov/nls/

Your paper need not be about economics specifically. Papers concerning questions from other fields such as finance or marketing are fine. There is abundant data on political elections and sports, and good topics could emerge from these areas. You may be surprised what level of data on seemingly obscure topics you can find with a little googling. I impose only two limits on topics. One, you may not, under any circumstances, use "survey data" that your group collected itself. Any data you use should come from a reliable source, and you should be prepared to show me your raw data. Two, part of the grade for your final project will be based on the creativity and usefulness of your research question and data source. That is to say, a paper that does something really interesting, but struggles with data and estimation issues will probably get a higher grade than a well-executed project that is directly taken from an example from class or the book. Similarly, a paper that demonstrates the group has internalized the lessons of chapters 16-18 will receive a higher grade, all else equal.

What you need to do: There are three due dates associated with this project to keep in mind.

- **Monday, November 5, 2012, 9am:** Your group jointly submits a proposal, briefly outlining your topic, what model you plan to estimate, and what data you plan to use. The proposal should contain a brief description of the data, as well as a link to the data or a description of how you came by it.

- **Friday, December 7, 2012, 9am:** Your group jointly submits the final version of your paper

- **Monday, December 10, 2012, 9am:** You individually email me with "grades" for the other members of your group. I recommend you use the following system, though if you'd like to use a finer metric or give me verbal feedback, please do so:

    - **check plus:** group member went above and beyond what could reasonably be expected in terms of effort and creative input, and was instrumental to the project's success.

– **check:** group member's effort was entirely satisfactory, on par with the effort and creative input required to produce a good project.

– **check minus:** group member's effort or input was lacking. Group member missed agreed-upon meetings or otherwise shirked. Group member struggled to add to the project creatively.

Grade: Your grade will be based on the following:

- 10%: My assessment of your group's proposal

- 60%: My assessment of your group's final project

- 20%: Your peers' assessment of your input into the group project

- 10%: You will get all of these points if and only if you send me an email containing an assessment of your peers by Monday, December 10, 2012, at 9am

Suggested paper structure: Unless you have a compelling reason for using a different structure, your paper should roughly follow the following outline:

1. **Introduction** Explain your project, define the key independent variables, and denote data sources. The introduction is very important — it is a chance to engage the reader immediately. This section should include some motivation and background on the topic you have chosen as a group. Parts of your proposal can serve as a rough draft of the introduction section. This section should implicitly answer the question, why do we care?

   This section should quickly inform the reader on the research question, why the question is important, and how your group plans to go about answering or investigating the question. The introduction should not go into the level of detail of the remaining sections, but the introduction should convey the general idea of the research problem and your approach.

2. **Data** This section should describe the data you are using and how it is able to help you answer the research question. What type of data are you using? Does your data source have any weaknesses? what are they, and how can you reasonably get around them? What variables are you trying to measure using your data?

   Identify which independent variables are of main interest to your research. Also, explain why other variables are included in the regression model - are they meant to control for other important factors? Describe any dummy variables and why they are included.

   The goal in this section is to give the reader an accurate description of the data you are using, as well as the variables within the overall data set that stand out as being the most important. You will go into more detail regarding why you use certain variables in the regression equation section.

3. **Data Analysis** Enter your data into Excel or another statistical software program. Analyze the data to check for mistakes and report summary descriptive statistics like mean, standard deviation, minimum, and maximum. Describe your findings and observations. The point of this section is to make sure there is nothing weird going on in your dataset. If you have income or age as an independent variable, make sure none of them take on negative values. If you have people whose age is 600, you

know there is a problem. Identify any strange observations and determine if they are miscoded or should be kept in the data set.

4. **Model & Regression Equation** State your "model" as a regression equation, and explain why your model addresses the research question. Here, you specify the model in terms of the population parameters (β's). This model may not be the same model that you actually run in the data analysis section (i.e. if you lack good measurements for certain variables), but it is the model that you believe best explains your research problem. This section should justify why you include certain variables. You don't want to run a "kitchen sink" regression since that can lead to other problems. Give some intuitive or theoretical reason why you have included the variables you chose. An example would be if you are interested in how a social life influences someone's grade. The theoretical model may look like:

$$GPA = \beta_0 + \beta_1 * StudyHours + \beta_2 * \#ofPintNights + \beta_3 * Sports + \beta_4 * SororityOrFraternity + \epsilon$$

In this section you should also discuss how you think each one of these variables will influence the dependent variable. For example, you would expect β_1 to be positive since studying more should increase your GPA, while you believe β_2 will be negative since spending more time at Pazzo's may distract you from studying.

5. **Regression Results** Estimate regressions using Excel or similar software, and evaluate and interpret the results. You can report more than one regression if this adds to your overall research agenda. Are your coefficients significantly different than zero? What fraction of the overall variation have you explained? Are the signs and magnitudes of your estimates consistent with what you expected? In this section, your job is to answer the question "what does it all mean?". A reader should come away knowing what you think your results say about your research question.

6. **Regression Extensions, Interpretations, & Conclusions** Consider extensions to the regression equation, and summarize your most interesting findings. Overall, a conclusion section should tie the entire paper together. This is an opportunity to synthesize your results, highlighting the relationships that you want to emphasize. Also, you can point out strengths and limitations to your research design. One question you should answer in this section is "if I had the ideal data set, what would I have done differently." What are the implications of your findings? Finally, suggest extensions to the research for future work.