This course provides an opportunity for students to [i] do sustained work on an empirical project; and [ii] practice in downloading and organizing data; and [iii] develop the ability to present complex regression results through visualization techniques, to audiences that are not well versed in the technicalities of empirical research.

These exercises are especially useful for your development as economists as it is now rare for our students to write master’s theses. This is true whether you expect to complete your graduate studies with this master’s degree or if you expect to pursue further graduate training. The success of the course depends critically on your input to the course, both in conducting and reporting the progress on your project and in providing constructive criticism to your colleague(s).

A. Empirical Project {300 points}

1. **Select a topic.** The topic you select should be susceptible to analysis using econometric techniques learned in E570 and/or E581. In some cases, it may be possible to use a more advanced technique that was not covered in any of those courses but, if so, please discuss with me prior to embarking on that attempt. There are several options:
   a. Continue on a topic you addressed in one of the earlier classes;  
      [This is the option most typically exercised.]
   b. Build on research you are doing for a research project, employer or faculty member;
   c. Select a new topic from one suggested by a faculty member who will also provide a reasonably clean data set.
   d. Select a new topic and identify your own data set. Example—OECD data set for cross country data over time.

A topic and data set should definitely have been identified at the latest by the third meeting.  
**[Jan. 24]** Students should be prepared to indicate what their topic is and what question or questions they hope to answer. No formal literature review is expected at this time.

2. **Perform a literature search.** What theoretical models/methods have been used to analyze the question? What are the theoretical implications for the question? Are they clear cut or are they ambiguous? What empirical studies are there relating to your question? What data set(s) have been used? What empirical technique/models...
have been used? What are the empirical findings so far? What gaps are there in the findings? What gap will you fill? Or will you provide additional evidence that will either confirm or contradict existing findings? Tools: EconLit, SSCI, classic articles and SSCI.

A written literature review should be turned in for a grade by the week. Each student should also be prepared to present the basic findings of their literature review to the class, preferably in the form of a Power Point (or equivalent) presentation. [Feb.] {25 pts.}

3. Analyze data. You should run the appropriate regressions using the Stata regression package.

Initial data analysis should be concluded by the week. Students should be prepared to do a Power Point presentation (preferred) on the nature of their data, how it will enable them to answer the question they are asking, any shortcomings of the data, anticipated problems, etc. Variable definitions and descriptive statistics should be presented. This could also include initial correlation matrix analysis, scatterplots, etc. No formal regression analysis needs to be presented at this time. The material to be presented in class should be shared with the instructor and the discussant no later than the Monday preceding the class. There will be a set of presenters and discussants the week [Feb.] with the roles reversed for the following week [Feb.]. Presenters may take up to 30 minutes. The discussants should restrict themselves to no more than 10 minutes. {25 pts.}

March 11-16 SPRING BREAK

Initial regression analysis should be concluded by the week. Students should be prepared to present a Power Point (or equivalent) presentation of their initial findings with their base technique. Students should indicate shortcomings, if any and indicate whether other functional forms will be performed, whether diagnostic tests need to be undertaken. How will the student improve the analysis in future weeks. Again, same format as above. Students are expected to describe the basic techniques they are using at the level of Wooldridge or Kennedy...1 presenters and 1 discussant in the week [March]...with roles reversed the next meeting [March] {50 pts.}

week—Review of issues like heterogeneity, endogeneity and causality. Students work on projects and consult with instructor on progress. [March]

week [April] Students should be prepared to present updated and improved project results at this point. Again student discussants will be assigned on a rotating basis. One set of students will present in week 11 [April] with three discussants and the roles reversed the following week [April]. {50 pts.}

and week—Final polished Power Point (or equivalent) presentation of results with discussant, instructor and general comments. One set of presenters and discussants on [April] with the roles reversed the following week [April]. {50 pts.}

week – Final written paper turned in by May. {100 pts.}
Papers are typically about twenty pages in length, in addition to relevant tables of results...but that is just a guideline.

Your final Stata do file and your data set should also be turned in at the same time. You should be sure that your do file follows good practice as suggested by Long’s chapter 2.

B. Organization and Interpretation of Data Using Stata {150 pts.}

i. Downloading and Organizing Data {75 pts.}...January 10 – Jan. 31

This process should be governed by the material in Long’s *Regression Models for Categorical Dependent Variables Using Stata*, chapters 2 and 3 [just the part on missing data].

[You were asked to buy this book for E581 so it is not a new purchase.]

Students will download and organize data for the following:

[i] PSID—single wave [2007]—to replicate Mroz specification for labor supply of married women. Download equivalent variables to those he downloaded and described in Tables III and V. Get to the point where you can run the 2sls specification on \textit{wife’s work hours} =
**Grading:** Grading will be on my usual scale:

<table>
<thead>
<tr>
<th>Percentage</th>
<th>Letter</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>97-100</td>
<td>A+</td>
<td>291-300</td>
</tr>
<tr>
<td>93-96</td>
<td>A</td>
<td>278-290</td>
</tr>
<tr>
<td>90-92</td>
<td>A-</td>
<td>270-277</td>
</tr>
<tr>
<td>87-89</td>
<td>B+</td>
<td>261-269</td>
</tr>
<tr>
<td>83-86</td>
<td>B</td>
<td>248-260</td>
</tr>
<tr>
<td>80-82</td>
<td>B-</td>
<td>240-247</td>
</tr>
</tbody>
</table>

and so on.

As with all graduate classes, any grade below a B is an indicator that the student did not demonstrate the minimum competency expected of graduate students. The assignment weights for grading are given in the schedule above in **bold** letters inside the `{}`. The Empirical Project –is worth 300 points in all and the Data Organization and Visual Interpretation Assignments are worth 150 points overall—a total of 450 points for the course. Hence one needs 405 total points to make it to the bottom of the A- range.