NEW YORK UNIVERSITY
ROBERT F. WAGNER GRADUATE SCHOOL OF PUBLIC SERVICE

PADM-GP 2902: REGRESSION AND INTRODUCTION TO ECONOMETRICS
Spring 2012
Professor Leanna Stiefel

FACULTY INFORMATION
Email: leanna.stiefel@nyu.edu
Phone: (212) 998-5758
Office: Puck Building, 3098
Office Hours: Wednesdays, 4-5:30 p.m.

LECTURE
Mondays, 11:00 – 12:40, Silver 701

RECITATIONS (PADM-GP 290)
You are required to register for a recitation section, and you must attend that section when you choose to attend. You should attend the first two sessions, but attendance after that is optional although recommended. The recitation prior to the exam will serve as a review session. Other recitations will consist of working through textbook exercises, homework Q&A, and general Q&A. You are required to carefully read all recitation exercises prior to the recitation.

Section 1*: Thursdays 8:00-9:00a.m., 194 Mercer St. #305. Jeremy Cherson, jac912@nyu.edu
Section 2*: Tuesdays 8:35-9:35p.m., Silver 711. James Dunham, jd1938@nyu.edu
Section 3: Wednesdays 12:30-1:30p.m., 194 Mercer St. #304. Gundula Loffler, gkl220@nyu.edu

PREREQUISITE: CORE-GP 1011 or equivalent

COURSE DESCRIPTION
Multiple regression (econometrics) is the core statistical technique used by policy and finance analysts in their work. In this course, you will learn the theory and practice of econometric analysis. Specifically, you will learn how to evaluate whether regression coefficients are biased, whether standard errors (and thus t statistics) are valid, and whether regressions used in policy and finance studies support causal arguments.

In addition, employing one dataset, you will compute the statistics discussed in class using Stata, an econometric statistical package, and you will see how the results reflect the concepts discussed in class. Finally, with a group of your classmates, you will choose either to do a project that involves estimating your own regression model and applying the techniques we learn in class, or to present and discuss a published study from a list provided below.

* For the weeks of January 23 and January 30, recitation sections 1 and 2 will instead meet Friday and Saturday; students will sign-up for one of three timeslots. More information will be posted to Blackboard.
COURSE REQUIREMENTS AND GRADING

1. **(15%) Problem sets and computer exercises**
   Problem sets and computer exercises must be submitted at the beginning of the class for which they are listed in the course schedule below. You will earn a grade of “1” if the entire problem set is completed and another “1” if the entire computer exercise is completed. You will receive partial credit if only part of the problem set or computer exercise is completed. Solutions will be posted to Blackboard on Wednesday after they are due and graded assignments will be returned to your student mailbox by the next class. Please be sure to write your mailbox number on your submissions. Also, for Stata output, submit only the last “run” of the analysis.

2. **(45%) Exam**
   An in-class exam will be given during Class 10 (see Course Schedule below). You may bring a non-graphing calculator and two pages (single-sided) of notes.

3. **(25%) Article or (35%) Project**
   In groups of four to five, either:
   
   a) **Present and discuss a published article** during one of the last three classes (see more below as well as the list of articles). This option constitutes 25% of your grade.

   For this option, during Class 6, hand in a ranking of the articles from your first to your last choice for presentation/discussion. If you have no preference or if some preferences are tied, please let me know. I will put you in groups. Close to the time of your presentation, schedule a time to meet with me to go over the article in detail.

   b) **Conduct a regression analysis, write a paper, and present your results** (see more below as well as the list of databases). This option constitutes 35% of your grade.

   For this option, email me by Friday before class 2 with your rankings of preferred data sets and I will put you groups. Your group will contact me to meet the week of Class 4 to discuss the project, including at least one specification that will be estimated.

   **Note:** All group members will complete peer evaluations that will factor into grades.

4. **(15% of grade if presenting an article; 5% of grade if doing the regression project)**
   **Explanation of how statistics support causality (“causal paragraph”)**
   Write one page or less that describes in your own words the primary causal relationship an article aims to test, what the main “threats” to the causal interpretation are, and how the statistical methods are used to address the threats. Then, describe one of the dependent variables that the author(s) analyze and the coefficient on the independent variable that is the main focus of the study; fully interpret the meaning of the coefficient. Use your own words; do not quote or paraphrase the author(s).

   Students presenting articles will write paragraphs for three articles (including the one they present), while students doing the regression project will write a paragraph for one article.
Hypothetical Example of Paragraph on Causality

In “Does Head Start Work?” Sopsyc (200X) aims to determine if attending a Head Start program *causes* increases in academic performance and health outcomes in the four years following attendance. The main threat to a causal interpretation of the effect of Head Start attendance is that the comparison groups of children attending other day care options (including none) are likely to differ not only in observable but also in unobservable ways from Head Start attendees and these differences may be correlated with academic performance or health outcomes as well. In other words, omitted, unobserved characteristics may be correlated with day care options and with the outcomes, causing bias. For example, children in the various options may differ with respect to innate health or academic endowments and these may affect outcomes. The author addresses this concern using *family fixed effects*, for families where there are at least two children per family, with only one attending Head Start. Average differences between the two children within families provide the estimate of the effects. Family fixed effects control for differences in genetic endowments across children, unobserved differences in parental interest in education or health outcomes, and perhaps some unobserved nurture of children as well.

Dependent variable: Score on test of verbal ability, ranging for 0 to 500
Independent variable: Dummy for whether attends Head Start Program or not (not stays home or attends another child care option)
Coefficient on Head Start Dummy: 6.3 (standard error 3.1)
Interpretation: Students attending a Head Start Program achieve a verbal test score that is, on average, 6.3 points higher than children who do not, holding other factors constant. The coefficient is statistically significant at the 5% level.

More on Presenting or Discussing a Published Article

You will present one of the articles listed in the COURSE MATERIALS section of the syllabus below during one of the last three class sessions. Your group will meet with me (about 1.5 hours) before your presentation date to ensure you understand the article. You will use PowerPoint to present and explain the article’s content to the class.

More on Doing and Presenting a Regression Project

You will present the results of your project during one of the last three class sessions (using PowerPoint), and then you will submit a paper in which you professionally present your results Monday, May 14. Feedback from the presentation should be incorporated into your paper. Model how your paper is written on the articles assigned in class.

This is a good option if you want to work directly with additional data this semester. The databases available are listed on Blackboard, in the Assignments area, and their descriptions are in a file labeled “Description of Databases.” They are all “donated” from Wagner professors’ and doctoral students’ research projects.

The regression project will take more time than the article presentation, and for that reason project groups will write causal paragraphs for only one of the assigned articles. Students who choose to do the regression project should:
1. Have time before the exam to devote to specifying a model and running regressions.

2. Want a more intensive hands-on experience with analyzing, writing about, and presenting statistics.

3. Read chapter 11, “Running Your Own Regression Project,” in the course text.

The end result will be an 8-10 page paper, including two tables, organized by five sections (to mimic the articles we will read, but there will not be a literature review), and a presentation of your results during one of the last three class sessions.

**Note:** It is *highly* recommended that students considering additional coursework in statistics select the regression project option. Courses commonly taken after completing PADM-GP 2902 include PADM-GP 2875: Estimating Impacts in Policy Research, PADM-GP 3148: Applied Research Capstone, and RESCH-GE 2110: Applied Statistics (Large Databases in Education).

**Paper Outline**

I. **Introduction** – what is the goal of your statistical regression study? Why is it interesting; why do we care? (This does not have to be momentous – but you should explain why the results could be interesting or valuable to someone)

II. **Data** – describe your sources and discuss the descriptive statistics that are presented in Table 1.

III. **Model and Empirical Strategy** – what is your model (equation), and how does it achieve the goal of your analysis? Why are the specific variables used and measured as they are? Do you have any prior expectations about the signs of coefficients? How will you estimate this model? (Usually OLS with fixed effects.)

IV. **Results** – Discuss the results that are presented in Table 2.

V. **Conclusions** – what does your model say about your goal or issue? What is the next step in this research?

Appended at end of paper:

- Table 1 (with title): Descriptive statistics of all the variables in your model(s).
- Table 2 (with title): Results of your models, presented as in the papers we read.
- Final cleaned up Stata log file of your results.
COURSE MATERIALS


2. **Required:** Stata/IC 12, purchased and loaded onto your computer by the end of the first week of class.

   You should purchase this software through NYU’s Direct-ship GradPlan in order to obtain a student discount: [http://www.stata.com/order/new/edu/gradplans/gp-direct.html](http://www.stata.com/order/new/edu/gradplans/gp-direct.html).

   You should purchase Stata/IC 12 (not Small Stata). The least-cost option is a 6-month license, at $65. If you are planning to take Estimating Impacts or the research capstone, you may want to consider a one-year or perpetual license. Stata 12 is not platform-dependent and will run on either Windows or Mac operating systems.


4. **Required:** Computer Exercises and Data Set to download from Blackboard. See the course schedule below for when assignments are due in class. The exercises are in one folder in the Assignments area on Blackboard.

   Download newschools97034.dta, Class 3 Exercise 2012.do, Computer Exercises 2012.doc, and Class 1 Handout.doc from Blackboard, by the first week of class, saving them to a folder on your computer reserved for PADM-GP 2902 work. Then, watch the video on using Stata (under Assignments → Computer Exercises on Blackboard).

5. **Required:** Studies to read critically. Go to the class Blackboard site to obtain the articles. We will read some set of the following articles, depending on how many students want to do the article presentation.


BLACKBOARD
You will need to have access to the class Blackboard site, found under “Academics” on your NYU Home site (https://home.nyu.edu/) or at http://classes.nyu.edu/. All announcements and class related documents (problem sets, computer exercises, databases, solutions, optional exercises, occasional class notes, etc.) will be posted here. If you have not activated your NYU net account or have forgotten your password, you can activate or change your password at http://start.nyu.edu. You must activate your account in order to access course materials and announcements on Blackboard.

Once you have accessed Blackboard, please change your e-mail address to whatever you use most often by going to “Tools” and then “Personal Information.” Class announcements will be distributed via Blackboard’s e-mail list; thus it is important that you have an active e-mail account.

STATISTICAL SOFTWARE
We will use Stata/IC 12; no previous knowledge of Stata is necessary. In addition to learning Stata through the problem sets and computer exercises and in class, the Data Services Studio in Bobst (http://library.nyu.edu/dataservice/) offers short courses (tutorials) and on-site help with this package. Also, there is a short video produced by your professors that walks you through getting started (on Blackboard, under Assignments ➔ Computer Exercises).

Finally, recitations during the first two weeks of class will serve as Stata tutorials. *If you have a laptop, please bring it with the Stata installed.*

CLASS NOTES
Before each class, class notes will be available on Blackboard. *You should print them, bring them to class,* and use them to organize your notes.
SUMMARY OF COURSE GRADING AND DUE DATES
1. 15% Problem Sets and Computer Exercises
2. 45% Exam
3. 25% Article or 35% Regression Project
4. 15% or 5% Causal Paragraph

COURSE SCHEDULE (S= Studenmund text; B = Baum text)

<table>
<thead>
<tr>
<th>WEEK</th>
<th>Mon</th>
<th>READING</th>
<th>DUE IN CLASS</th>
<th>RECITATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1/23</td>
<td>S Chs. 1 &amp; 2 (34-38 &amp; 46-57)</td>
<td>See &quot;Class Topics&quot; below</td>
<td>Stata Tutorial I</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B Ch. 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2*</td>
<td>1/30</td>
<td>S Chs. 2 (39-45) &amp; 4</td>
<td>Problem Set Class 2</td>
<td>Stata Tutorial II</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B Ch. 2 (2.1.1-2.1.13; 2.2.1-2.2.4)</td>
<td></td>
<td>S Ch. 4, ex. 6 (a,b), 11 (a,b,c)</td>
</tr>
<tr>
<td>3</td>
<td>2/6</td>
<td>S Chs. 3 &amp; 5 (not including appendix)</td>
<td>Problem Set Class 3</td>
<td>Computer Exercise Class 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B (3.9.1; 4.3.1-4.3.4; 4.6 stop at 4.6.1)</td>
<td></td>
<td>S Ch. 5, ex. 8, 9, 12, 14 (a-d)</td>
</tr>
<tr>
<td>4**</td>
<td>2/13</td>
<td>S Chs. 6 &amp; 7 (207-213, 218-220, 223-226); Appendix Ch. 5</td>
<td>Problem Set Class 4</td>
<td>S Ch. 7, ex. 6, 12 (a,b,e)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2/27</td>
<td>S Ch. 7 (213-218, 220-223, 226-232)</td>
<td>Problem Set Class 5</td>
<td>Computer Exercise Class 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B (4.5.1-4.5.3; 5.2.2; 7.1 intro., 7.1.1, 7.1.2 until interactions)</td>
<td></td>
<td>S Ch. 7 ex. 7 (a), 14</td>
</tr>
<tr>
<td>6***</td>
<td>3/5</td>
<td>S Chs. 8 &amp; 9</td>
<td>Problem Set Class 6</td>
<td>Computer Exercise Class 6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B (5.2.9, 7.1.2, all of 7.2)</td>
<td></td>
<td>S Ch. 8 ex. 6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>S Ch. 9 ex. 11</td>
</tr>
<tr>
<td>7</td>
<td>3/19</td>
<td>S Ch. 10</td>
<td>Problem Set Class 7</td>
<td>Computer Exercise Class 7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B (4.3.7, 6.1.4, 6.3)</td>
<td></td>
<td>S Ch. 10 ex. 4 (a), 9(a,b), 11</td>
</tr>
<tr>
<td>8</td>
<td>3/26</td>
<td>S Ch. 16</td>
<td>Problem Set Class 8</td>
<td>Computer Exercise Class 8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B (6.2, 6.1.2, 6.1.3)</td>
<td></td>
<td>S Ch. 16 ex. 3</td>
</tr>
<tr>
<td>9</td>
<td>4/2</td>
<td>S Ch. 13</td>
<td>Problem Set Class 9</td>
<td>Computer Exercise Class 9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B Ch. 9 through 9.1.3</td>
<td></td>
<td>Exam Review</td>
</tr>
<tr>
<td>10</td>
<td>4/9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>4/16</td>
<td>S Ch. 14 (396-97) &amp; Table 11-2</td>
<td>Problem Set Class 11</td>
<td>Computer Exercise Class 11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B Ch. 10 through 10.1.3</td>
<td></td>
<td>S Ch. 13 ex. 3, 12 (a,b,c)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>S Ch. 14 ex. 14</td>
</tr>
<tr>
<td>12</td>
<td>4/23</td>
<td>Articles TBA (Presentations Week)</td>
<td>Explanation of Causality</td>
<td>None</td>
</tr>
<tr>
<td>13</td>
<td>4/30</td>
<td>Articles TBA (Presentations Week)</td>
<td>Explanation of Causality</td>
<td>None</td>
</tr>
<tr>
<td>14</td>
<td>5/7</td>
<td>Articles TBA (Presentations Week)</td>
<td>Explanation of Causality</td>
<td>None</td>
</tr>
<tr>
<td>15</td>
<td>5/14</td>
<td></td>
<td>Regression Project Papers</td>
<td>None</td>
</tr>
</tbody>
</table>

* email professor before class 2 if you want to conduct a regression analysis and write paper.
** schedule a meeting with professor to discuss regression analysis and at least one equation.
*** hand in article rankings during class if you choose this option.
CLASS TOPICS

Class 1: OLS Bivariate Regression Model with Error Term
Theoretical regression line; deterministic versus stochastic relationships; population versus sample regression line; error and residual; OLS estimators

Note: Purchase and install Stata/IC 12; download and save newschools97043.dta, Class 3 Exercise 2012.do, and Exercises Computer 2012.doc from Blackboard. Watch video on using Stata. Print and bring Class 1 Handout.doc to class.

Class 2: OLS Multiple Regression and Assumptions about Error Term
Reducing bias; interpretation of coefficients; BLUE assumptions

Class 3: Hypothesis Testing in Multiple Regression Context
Significance tests; confidence intervals; F test; \( R^2 \); Adjusted \( R^2 \); interpretation of computer output

Class 4: Functional Form Part I: Polynomials and Dummy Variables
Functional form; using qualitative data (dummies); joint tests of significance; curvilinear relationships

Class 5: Functional Form Part II: Interactions and Logarithmic Transformations
Interactions of dummies; interactions of continuous and dummy variables; continuous by continuous interactions; logarithmic transformations (percent change or elasticity transformations)

Class 6: Multicollinearity and Autocorrelation

Class 7: Heteroskedasticity

Class 8: Panel Data Estimation Part I
Pooled cross-sectional and time series data; panel data; fixed effects estimation.

Class 9: Panel Data Estimation Part II; Introduction to Linear Probability Models

Class 10: Exam

Class 11: Qualitative Dependent Variables; Simultaneous Equation Models

Class 12-14: Presentations

May 14: Papers Due in Professor Stiefel’s Mailbox (Puck, 3rd Floor) by 5 p.m.
STUDENT INFORMATION SHEET

Name_______________________________________________________________

Name of undergraduate institution and year of graduation:

Previous experience in statistics? When?

Previous professor for statistics if Wagner?__________________________________________

Program (circle one) Health      Public and Nonprofit      Urban Planning      International      Other

Expected specialization (circle) Finance Policy Management International Other

Fulltime            Part time

Brief description of work experience (if any):

Specific areas of interest in public policy, finance, planning or international:

What do you hope to learn or gain from taking this course?

Other information of interest, especially to help me get to know you faster.