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

PADM-GP.2902.001 REGRESSION AND INTRODUCTION TO ECONOMETRICS
Fall 2011

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

STAFF INFORMATION
Tutor: Gundula Loffler
Email: gkl220@nyu.edu
Location: Aquino Conference Room
Hours: Fridays 11am to 1:30 pm

Tutor: Sara Lense
Email: sdlense@googlemail.com
Location: Aquino Conference Room
Hours: Wednesdays 5 to 6:30 pm and 8:45 pm to 9:45 pm

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 learn the theory and practice of econometric analysis. Specifically, you 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 compute the statistics discussed in class using Stata, an econometric statistical package, and you see how the results reflect the concepts discussed in class. Finally, with a group of your classmates, you 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 on the syllabus.
COURSE REQUIREMENTS AND GRADING

1. **(15%) Problems and computer exercises**

   Written and computer exercises are to be handed in during the class in which they are listed on the syllabus. You will be given a grade of “1” if the entire assignment is attempted and between a 0 and 0.5 if not. Answers will be posted on Blackboard following the class they are due and the assignments will be returned to your Puck mailbox, with comments. Please be sure to list your mailbox number on your papers. For Stata problems, please submit your last (most correct) run, not all the erroneous ones that might precede that one! These should not be more than about 10 pages long.

2. **(45%) Exam**

   An in-class exam will be given during Class 10 (November 21).

3. **(25%) Articles or (35%) Project**

   Note: there will be a peer assessment for the groups.

   **In a group of four or five students, 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 is worth 25% of your grade.

      For this option, during Class 6 (October 24), 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. NB: Some articles may be moved to other dates and/or not be presented at all depending on the number of students who choose to do the project option rather than the article option.

      Close to the time of your presentation, your group should schedule an hour and a half with me to go over the article in detail.

   b) **Do, present and write paper on a data and regression project** (see more below as well as the list of databases).  This is worth 35% of your grade.

      For this option, let me know by email by Friday before class 2 (week of September 19) with your rankings of preferred datasets and I will put you in a group. Then your group will contact me to meet during the week of class 4 (week of October 3) to show me at least one equation that will be estimated and discuss the project.
4. (15% of grade if presenting article; 5% of grade if doing data project) Explanations of how statistics support causality

Each week of presentations, for every article listed (once the final article list is established after class 6) for the students presenting articles (including the one that you present) and for one article of your choice for one week for students doing the project, write one page or less that describes in your own words the primary causal relationship the 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 authors analyze and the coefficient on the independent variable that is the main focus of the study and interpret in quantitative terms the meaning of the coefficient.

Hypothetical Example of Paragraph on Causality

In “Does Head Start Work?” Sopsyec (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, whereby 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 is 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 6.3 higher than children who do not. The coefficient is significant.
More on Presenting or Discussing a Published Article

A presentation of one of the articles listed in the COURSE MATERIALS section of the syllabus below is due one of the three last days of class (Some articles may be removed after class 6, depending on how many students sign up to do this option versus the project option.) Your group should meet with me before the class in which you present for about 1.5 hours to understand the article and you should use power point to present and explain the content to the class. I will help you to become a mini expert on this piece of work and this project will help you understand why we are learning all the techniques in the first part of the semester.

This is a good option if you want to understand statistics in action, and/or aim to be a policy analyst but do not necessarily want to be the person who does the computer work, and/or want to take your time becoming comfortable with data (and thus learn more over time, in future courses, such as PADM-GP.2875 or PADM-GP.3148 or RESCH-GE.2110).

More on Doing and Presenting a Data and Regression Project

A paper in which you professionally present the results of your project is due on Wednesday December 21; an in-class presentation of your results is due one the last three days of class. Feedback from the presentation should be incorporated into your paper. Model how your paper is written on the articles assigned in class.

If you want to work with additional data this semester, please sign up for the project. The databases available are listed on Blackboard, under assignments, and their descriptions are in a file labeled “Description of Databases” (under assignments). They are all “donated” from Wagner professors’ and doctoral students’ research projects. Look them over to try to decide which look like they might be an interesting basis for a study. If you can’t tell, but want to do the project, just let me know that (before class 2) and I will put you in a group.

The project will take more time than the (alternative) article presentation and for that reason project groups will write causal paragraphs on only one of the assigned articles (of your choice). Students who do the project should:

1. Have time before the midterm to devote to specifying a model and running regressions.

   - Want a more intense hands-on experience with analyzing, writing about and presenting statistics (in advance of taking other follow-on courses such as PADM-GP.2875: Estimating Impacts, or PADM-GP.3148: Applied Research Capstone, or RESCH-GE.2110: Applied Statistics, all three of which provide more hands-on work).
2. Read chapter 11 in STUDENMUND “Running Your Own Regression Project.”

The end result will be an 8-10 page paper, including two tables, organized into five sections (to mimic the articles we will read, but without a literature review), and a presentation of your results one of the last three days of class. You still will read the assigned articles and write one causal paragraph.

Paper Outline

Note: Discussion of the following two tables of results should be woven into the sections as indicated below, although the tables should be put in the back of the paper.

Table 1: Descriptive statistics of all the variables in your model(s)
Table 2: Results of your models, presented as in the papers we read

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 – just say something about why the results could be interesting or valuable to someone).

II. Data – describe your sources and discuss the descriptive statistics in Table 1

III. Model and Empirical Strategy – what is your model (equation) and how does it address 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 and fixed effects regression).

IV. Results – Discuss the results in Table 2

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


   This is available at the NYU bookstore, on Amazon and on other websites.

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

   You should purchase this software through Gradplan3 at NYU in order to obtain a student discount. Stata/Gradplan is at: [http://www.stata.com/order/new/edu/gradplans/gp-direct.html](http://www.stata.com/order/new/edu/gradplans/gp-direct.html)

   Student discounts are on the right-hand side. You should purchase Stata/IC 12 (**not** Small Stata). Least cost is a 6-month license, at $65 but if you are planning to take Estimating Impacts or the research capstone, you may want to consider a one or more-year license.

   Stata12 is not platform-dependent. It will run on either Windows or Mac operating systems.

3. **Computer Exercises and Data Set** to download from Blackboard. See syllabus for when these are due in class. The exercises are in one file under Assignments. The answers will be posted under the class when they are assigned (and due), after class.

   Download newschools97043.dta; Class 3 Exercise 2011.do; Exercises Computer 201.doc; and Class 1 Handouts.doc from Blackboard, *by the first week of class*, saving them to a folder on your computer reserved for P11.2902 work. **Watch the video** on using Stata (under Assignments, Computer Exercises on Blackboard).

4. **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 want to do this option versus the data project option) and you will pick one, with a group of your classmates, to present or discuss in front of the class, if this is your option. If you do the data project, you will read the articles that are presented but will not present any yourself.


**BLACKBOARD**

You need to have access to the class Blackboard site, found under “academics” on your NYU home location (https://home.nyu.edu/) or at http://classes.nyu.edu/. All announcements and class related documents (computer exercises, datasets, answers to exercises, 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.” Some class announcements may be distributed via Blackboard’s e-mail list, thus it is important that you have an active e-mail account.

If you do not already have Adobe Acrobat Reader installed on your computer, go to http://www.adobe.com/products/acrobat/ and follow the download instructions. You will need this (free) software in order to read and print some of the materials on Blackboard.
STATISTICAL SOFTWARE

We will use Stata/IC 12 software; no previous knowledge is necessary. In addition to learning Stata through the problem sets 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 and there is a short video produced by your professors that walks you through getting started (on Blackboard, Assignments, Computer Exercises).

Finally, there will be two extra sections of this class devoted to showing you how to use Stata and do the Stata homework for Class 3. They are on Wednesday September 21st, 8:45 to 10:45 pm, in Tisch Hall LC19 and Saturday, September 24th, 10 to 12 pm, in Tisch Hall LC19. Please plan to attend one of these and 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 these notes, bring them to class, and use them for taking notes. You will follow the class well only if have the class notes with you.
SUMMARY OF COURSE GRADING AND DUE DATES
NB: Monday, October 10 is an NYU holiday (Columbus Day).

1. 15% Problems and Computer Exercises  
2. 45% Exam  
3. 25% Studies or 35% Projects  
4. 15% or 5% Explanation of how statistics support causality

Key: S= Studenmund text

<table>
<thead>
<tr>
<th>CLASS</th>
<th>DUE DATE</th>
<th>READING</th>
<th>HOMEWORK</th>
<th>OTHER</th>
</tr>
</thead>
<tbody>
<tr>
<td>One</td>
<td>September 12</td>
<td>S Chs. 1 &amp; 2 (34-38 and 46-57)</td>
<td>Purchase and install Stata/IC 12 by Class Two. Download &amp; save newschools97043.dta; Class 3 Exercise 2011.do; Exercises Computer 201.doc; and Class 1 Handouts.doc from Blackboard. Watch video on getting started with Stata.</td>
<td></td>
</tr>
</tbody>
</table>
| Two   | September 19 | S Chs. 2 (39-45) & 4 | • S Ch. 1 ex. 9, 10 (a,b,c), 11  
  • S Ch. 2 ex. 6, 10, 13 (a,b,c) | Email Professor by Friday before class if you want to do Data Project instead of Article Presentation |
| Extra Section on use of Stata | Wednesday September 21 8:45 pm LC19 | | • Make sure you have done all assignments listed under HOMEWORK Class One  
  • If you have a laptop, bring it with Stata installed | To go over how to use Stata and homework for Class Three |
| Repeat Section on use of Stata | Saturday September 24 10am LC19 | | • Make sure you have done all assignments listed under HOMEWORK Class One  
  • If you have a laptop, bring it with Stata installed | To go over how to use Stata and homework for Class Three |
| Three | September 26 | S Chs. 3 & 5 (not including appendix) | • S Ch. 4, ex. 6 (a,b),11 (a,b,c)  
  • Computer Exercise Class 3 | |
<p>| Four  | October 3 | S Chs. 6 &amp; 7 (207 –213, 218 - 220, 223-226); Appendix Ch. 5 | • S Ch. 5, ex. 8, 9, 12, 14 (a-d) | Meet with Professor if doing Data Project; present ideas for one basic model (dv, iv’s) |</p>
<table>
<thead>
<tr>
<th>No Class</th>
<th>October 10</th>
<th>Columbus Day</th>
<th>NYU Holiday</th>
</tr>
</thead>
</table>
| Five     | October 17 | S Ch. 7 (213 – 218, 220-223, 226-232) | • S Ch. 7 ex. 6, 12 (a,b,e)  
• Computer Exercise Class 5 |
| Six      | October 24 | S Chs. 8 & 9 | • S Ch. 7 ex. 7 (a), 14  
• Computer Exercise Class 6 |
|          |            |              | Hand in article ranking for presentation if doing article presentation |
| Seven    | October 31 | S Ch. 10     | • S Ch. 8 ex. 6  
• S Ch. 9 ex. 11  
• Computer Exercise Class 7 |
| Eight    | November 7 | S Ch. 16     | • S Ch. 10 ex. 4 (a), 9(a,b), 11  
• Computer Exercise Class 8 |
| Nine     | November 14| S Ch. 13     | • S Ch. 16 ex. 3  
• Computer Exercise Class 9 |
| Ten      | November 21| Examination  |             |
| Eleven   | November 28| S Ch. 14, (396-97) & Table 11-2 | • S Ch. 13 ex. 3, 12 (a,b,c)  
• S Ch. 14 ex. 14  
• Computer Exercise Class 11 |
| Twelve   | December 5 | Articles to be announced (tba) | • Explanations of Causality  
Presentations articles and projects |
| Thirteen | December 12| Articles tba | • Explanations of Causality  
Presentations articles and projects |
| Fourteen* (Wed) | December 14 | Articles tba | • Explanations of Causality  
Presentations articles and projects |
| Project Papers | Wednesday December 21 | | Project papers due |
TOPICS AND ASSIGNMENTS

CLASS 1:  OLS BIVARIATE REGRESSION MODEL WITH ERROR TERM  
(September 12)

(Theoretical regression line; deterministic versus stochastic relationships;  
population versus sample regression line; error and residual; OLS estimators)

Reading:
-STUDENMUND: Chapters 1 and 2 (pages 34-38 and 46-57)

Other:
-Purchase and install Stata/IC 12 by Class Two (See Course Materials 2 section  
of syllabus.) Download and save newschools97043.dta, Class 3 Exercise 2011.do,  
and Exercises Computer 2011.doc from Blackboard. Watch video on using Stata.  
Print and bring Class 1 Handouts.doc to class.

CLASS 2:  OLS MULTIPLE REGRESSION AND ASSUMPTIONS ABOUT ERROR  
TERM  
(September 19)

Please let me know by the Friday before this class if you want to do the Data  
and Regression Project (instead of the article presentation)

(Motivation – reduction in bias and more plausible causality; interpretation of  
coefficients; BLUE; model assumptions)

Reading:
- STUDENMUND: Chapters 2 (pages 39-45) and 4

Exercises:
- STUDENMUND: Chapter 1, Exercises 9, 10 (a,b,c), 11 (you do not need  
datafile); Chapter 2, Exercises 6, 10, 13 (a,b, c).

SPECIAL SECTIONS: USING STATA, DOING COMPUTER HOMEWORK FOR  
CLASS THREE  
(Wed. September 21, 8:45 to 10:45 pm and repeat on Sat. September 24, 10am to 12 pm, both in Tisch Hall LC19)

If you have a laptop, bring it with Stata installed.
CLASS 3: HYPOTHESIS TESTING (t AND F) IN OLS MULTIPLE REGRESSION CONTEXT AND UNDERSTANDING STATA REGRESSION COMPUTER OUTPUT
(September 26)

(Significance tests; confidence intervals; F test; R^2; Adjusted R^2; Interpretation of regression computer output)

Reading:
- STUDENMUND: Chapters 3 and 5 (not including appendix)

Exercises:
- STUDENMUND: Chapter 4, Exercises 6 (a, b), 11 (a, b, c)

Computer Exercise Class 3: Introduction to Stata and the School Data Set; OLS Regression and Bias. NB: Watch the video on using Stata again if you have any trouble beginning this exercise. Also, please submit only the final version with the correct commands.

CLASS 4: FUNCTIONAL FORM PART I: POLYNOMIALS AND DUMMY VARIABLES, F TESTS (ANOTHER APPLICATION)
(October 3)

Please make an appointment to meet with me if you are doing the Data and Regression Project

(Flexibility of regression line; use of dummies; testing for significance of dummies as a group – F test again; curved lines with polynomials)

Reading:
- STUDENMUND: Chapters 6 and 7 (pages 207-213; 218-220; 223-226) and Appendix to Chapter 5

Exercises:
- STUDENMUND: Chapter 5, Exercises 8, 9, 12, 14 (a – d)
No Class on Columbus Day, October 10, NYU holiday

CLASS 5: FUNCTIONAL FORM PART II: INTERACTIONS AND LOGARITHMIC TRANSFORMATIONS (October 17)

(Interactions dummies, continuous, continuous and dummy; logarithmic or percentage change or elasticity transformations)

Reading:
- STUDENMUND: Chapter 7, pages 213-218, 220-223, 226-232

Exercises:
- STUDENMUND: Chapter 7, Exercises 6, 12 (a, b, e)

Computer Exercise Class 5: Polynomials, Dummy Variables and Joint F tests

CLASS 6: MULTICOLLINEARITY AND AUTOCORRELATION, PROBLEMS WITH STANDARD ERRORS (October 24)

Please hand in your preferences for articles to present or discuss if you are doing the article presentation (and not doing the data and regression project).

Reading:
- STUDENMUND: Chapters 8 and 9

Exercises:
- STUDENMUND: Chapter 7, Exercises 7 (a), 14

Computer Exercise Class 6: Ln transformations and Interactions
CLASS 7:  HETEROSKADASTICITY, PROBLEMS WITH STANDARD ERRORS  
(October 31)  

Reading:  
- STUDENMUND:  Chapter 10  

Exercises:  
- STUDENMUND:  Chapter 8, Exercises 6; Chapter 9, Exercises 11  

Computer Exercise Class 7: Multicollinearity and Autocorrelation  

CLASS 8:  FIXED EFFECTS REGRESSION WITH PANEL DATA PART I  
(November 7)  

(Pooled cross section and time series data; panel data; fixed effects regressions; random effects regressions; motivation – less omitted variable bias and better causality)  

Reading:  
- STUDENMUND:  Chapter 16  

Exercises:  
- STUDENMUND:  Chapter 10, Exercises 4 (a), 9 (a, b), 11  

Computer Exercise Class 8: Heteroskedasticity, White Test, Robust Standard Errors
CLASS 9: FIXED EFFECTS REGRESSION WITH PANEL DATA PART II AND INTRODUCTION TO LINEAR PROBABILITY MODELS
(November 14)

Reading:
- STUDENMUND: Chapter 13

Exercises:
- STUDENMUND: Chapter 16, Exercises 3

Computer Exercise Class 8: Fixed Effects

CLASS 10: EXAMINATION
(November 21)

You may bring one pages of notes (any font, any size paper, two-sided) and you should bring a calculator. All necessary statistical tables (t, F etc.) will be supplied.

CLASS 11: MORE ON QUALITATIVE DEPENDENT VARIABLES AND SIMULTANEOUS EQUATION MODELS
(November 28)

Reading:
- STUDENMUND: Chapter 14 and pages 396-97 (Table 11-2)

Exercises:
- STUDENMUND: Chapter 13, Exercises 3, 12 (a, b, c) ; Chapter 14, Exercise 14

Computer Exercise Class 11: Linear Probability, Probit and Logit
CLASS 12: PRESENTATIONS OF ARTICLES OR DATA PROJECTS  
(December 5)

These will be scheduled once students choose between article presentations and data projects.

Causal Paragraphs will be due the day an article is scheduled to be presented. The schedule will be posted after week 6.

**Homework:**

Hand in explanation of causality for this study.

CLASS 13: PRESENTATIONS OF ARTICLES OR DATA PROJECTS  
(December 12)

These will be scheduled once students choose between article presentations and data projects.

Causal Paragraphs will be due the day an article is scheduled to be presented. The schedule will be posted after week 6.

**Homework:**

Hand in explanation of causality for this study.

CLASS 14: PRESENTATIONS OF ARTICLES OR DATA PROJECTS  
(Wednesday December 14)

These will be scheduled once students choose between article presentations and data projects.

Causal Paragraphs will be due the day an article is scheduled to be presented. The schedule will be posted after week 6.

**Homework:**

Hand in explanation of causality for this study.

**WEDNESDAY December 21: Project Papers Due in my Puck mailbox by 5 p.m.**
GRADING CRITERIA

Grades will be assigned according to the following criteria:

A  **Excellent**: Exceptional work for a graduate student. Work at this level is unusually thorough, well reasoned, creative, methodologically sophisticated, and well written. Work is of exceptional, professional quality.

A-  **Very Good**: Very strong work for a graduate student. Work at this level shows signs of creativity, is thorough and well-reasoned, indicates strong understanding of appropriate methodological or analytical approaches, and meets professional standards.

B+  **Good**: Sound work for a graduate student; well-reasoned and thorough, methodologically sound. This is the graduate student grade that indicates the student has fully accomplished the basic objectives of the course.

B  **Adequate**: Competent work for a graduate student even though some weaknesses are evident. Demonstrates competency in the key course objectives but shows some indication that understanding of some important issues is less than complete. Methodological or analytical approaches used are adequate but student has not been thorough or has shown other weaknesses or limitations.

B-  **Borderline**: Weak work for a graduate student; meets the minimal expectations for a graduate student in the course. Understanding of salient issues is somewhat incomplete. Methodological or analytical work performed in the course is minimally adequate. Overall performance, if consistent in graduate courses, would not suffice to sustain graduate status in “good standing.”

C/-+  **Deficient**: Inadequate work for a graduate student; does not meet the minimal expectations for a graduate student in the course. Work is inadequately developed or flawed by numerous errors and misunderstanding of important issues. Methodological or analytical work performed is weak and fails to demonstrate knowledge or technical competence expected of graduate students.

F  **Fail**: Work fails to meet even minimal expectations for course credit for a graduate student. Performance has been consistently weak in methodology and understanding, with serious limits in many areas. Weaknesses or limits are pervasive.
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.