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

P11.2902 REGRESSION AND INTRODUCTION TO ECONOMETRICS
Spring 2011
Professor Daniel L. Smith

FACULTY INFORMATION
Email: daniel.smith@nyu.edu
Phone: (212) 998-7443
Office: Puck Building, 3046
Office Hours: Wednesdays, 3:30-5 p.m. or by appointment

STAFF INFORMATION
Tutor Michelle Wong
Email mwong@nyu.edu
Location Puck Building, Adjunct Office 3047
Hours Tuesdays 2:30-4:30 p.m.

Tutor Jennifer Hustwitt
Email jh2699@nyu.edu
Location See Blackboard
Hours Thursdays 6:30-8:30 p.m.

Tutor Douglas Coulter
Email douglas.coulter@nyu.edu
Location Puck Building, Aquino (formerly Lafayette) Conference Room (3066)
Hours Fridays 1:30-3:30 p.m.

PREREQUISITE
P11.1011 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 on the syllabus.
COURSE REQUIREMENTS AND GRADING

1. (15%) Problems and computer exercises
Written and computer exercises must be submitted at the beginning of the class for which they are listed on the syllabus. You will be given a grade of “1” if the entire assignment is completed, and between a .5 and 0 if not. Answers will be posted on Blackboard following the class an assignment is due, and graded assignments, with comments, will be returned to your Puck mailbox. Please be sure to list your mailbox number on your submissions.

2. (45%) Exam
An in-class exam will be given during Class 10.

3. (25%) Article or (35%) Project
In a group of preferably four to five students (six maximum), 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, submit 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, your group will schedule an hour and a half with me to go over the article in detail. Some articles may be moved to other dates and/or may not be presented, depending on the number of students who choose to do the project option rather than the article option.

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, let me know via email by Friday before Class 2 (week of January 31) 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 February 14) to discuss the project, including at least one specification that will be estimated.

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; interpret in quantitative terms the meaning of the coefficient.

Students who choose the article presentation option will do this for every article listed (once the final article list is established after Class 6), including the article they will present, each week of presentations. Students doing the regression project will do this for one article of their choice for one week.
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 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.

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 will be due one of the three last days of class. Some articles may be removed after Class 6, depending on how many students choose to do this option versus the project option. Your group will meet with me (about 1.5 hours) before the class in which you present to understand the article, and you will use PowerPoint to present and explain its content to the class.

More on Doing and Presenting a Regression Project
A paper in which you professionally present the results of your project will be due Friday, May 13; an in-class presentation of your results will be due one of 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.

This is a good option if you want to work 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. If you want to do the regression project but are not sure which dataset you would like to work with, let me know before Class 2, and I will put you in a group.
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 (of your choice). Students who do the regression project should:

1. Have time before the examination 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 one of the last three days of class. You still will read the assigned articles and write one causal paragraph.

**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 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 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: Descriptive statistics of all the variables in your model(s)

Table 2: Results of your models, presented as in the papers we read
COURSE MATERIALS


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

2. Stata/IC 11, 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 11 (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 11 is not platform-dependent and will run on either Windows or Mac operating systems.

3. Computer Exercises and Data to download from Blackboard. See the syllabus for when assignments are due in class. The exercises are in one folder in the Assignments area on Blackboard. The answers will be posted after class on the due dates.

Download newschools97034.dta, Class 3 Exercise 2011.do, Exercises Computer 2011.doc, and Class 1 Handout.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 students want to do this option versus the project option). If you do the project, you will read all articles but will but will not present any.


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 (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.

STATISTICAL SOFTWARE
We will use Stata/IC 11; 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 February 2, 8:30 to 10 pm, in Tisch Hall LC-9 and Saturday, February 5, 10 am to 12 pm, in Tisch Hall LC-9. 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 to organize your notes.
SUMMARY OF COURSE GRADING AND DUE DATES

NB: Monday, February 21 is an NYU holiday (Presidents’ Day). The week of March 14-18 is Spring Break.

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

COURSE SCHEDULE
Key: S = Studenmund text

<table>
<thead>
<tr>
<th>CLASS</th>
<th>Monday</th>
<th>Tuesday</th>
<th>READING</th>
<th>HOMEWORK</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1/24</td>
<td>1/25</td>
<td>S Chs. 1 &amp; 2 (34-38 and 46-57)</td>
<td>See &quot;Topics and Assignments&quot;.</td>
</tr>
<tr>
<td>2</td>
<td>1/31</td>
<td>2/1</td>
<td>S Chs. 2 (39-45) &amp; 4</td>
<td>S Ch. 1 ex. 9, 10 (a,b,c), 11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>S Ch. 2 ex. 6, 10, 13 (a,b,c)</td>
</tr>
<tr>
<td>3</td>
<td>2/7</td>
<td>2/8</td>
<td>S Chs. 3 &amp; 5 (not including appendix)</td>
<td>S Ch. 4, ex. 6 (a,b), 11 (a,b,c)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Computer Exercise Class 3</td>
</tr>
<tr>
<td>4</td>
<td>2/14</td>
<td>2/15</td>
<td>S Chs. 6 &amp; 7 (207-213, 218-220, 223-226); Appendix Ch. 5</td>
<td>S Ch. 5, ex. 8, 9, 12, 14 (a-d)</td>
</tr>
<tr>
<td>5</td>
<td>2/28</td>
<td>2/22</td>
<td>S Ch. 7 (213-218, 220-223, 226-232)</td>
<td>S Ch. 7 ex. 6, 12 (a,b,e)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Computer Exercise Class 5</td>
</tr>
<tr>
<td>6</td>
<td>3/7</td>
<td>3/1</td>
<td>S Chs. 8 &amp; 9</td>
<td>S Ch. 7 ex. 7 (a), 14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Computer Exercise Class 6</td>
</tr>
<tr>
<td>7</td>
<td>3/21</td>
<td>3/8</td>
<td>S Ch. 10</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></td>
<td></td>
<td></td>
<td></td>
<td>Computer Exercise Class 7</td>
</tr>
<tr>
<td>8</td>
<td>3/28</td>
<td>3/22</td>
<td>S Ch. 16</td>
<td>S Ch. 10 ex. 4 (a), 9(a,b), 11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Computer Exercise Class 8</td>
</tr>
<tr>
<td>9</td>
<td>4/4</td>
<td>3/29</td>
<td>S Ch. 13</td>
<td>S Ch. 16 ex. 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Computer Exercise Class 9</td>
</tr>
<tr>
<td>10</td>
<td>4/11</td>
<td>4/5</td>
<td></td>
<td>EXAMINATION</td>
</tr>
<tr>
<td>11</td>
<td>4/18</td>
<td>4/12</td>
<td>S Ch. 14 (396-97) &amp; Table 11-2</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></td>
<td></td>
<td></td>
<td></td>
<td>Computer Exercise Class 11</td>
</tr>
<tr>
<td>12</td>
<td>4/25</td>
<td>4/19</td>
<td>Articles TBA (Presentations Week)</td>
<td>Explanation of Causality</td>
</tr>
<tr>
<td>13</td>
<td>5/2</td>
<td>4/26</td>
<td>Articles TBA (Presentations Week)</td>
<td>Explanation of Causality</td>
</tr>
<tr>
<td>14</td>
<td>5/9</td>
<td>5/3</td>
<td>Articles TBA (Presentations Week)</td>
<td>Explanation of Causality</td>
</tr>
<tr>
<td>Finals Week</td>
<td>5/13</td>
<td>5/13</td>
<td></td>
<td>Project Papers Due</td>
</tr>
</tbody>
</table>
TOPICS AND ASSIGNMENTS

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)

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

Other:
- Purchase and install Stata/IC 11 (See Course Materials section of syllabus.) ---
- Print and bring Class 1 Handouts.doc to class.

CLASS 2: OLS MULTIPLE REGRESSION AND ASSUMPTIONS ABOUT ERROR TERM

Please let me know by the Friday before this class if you want to do the 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).

CLASS 3: HYPOTHESIS TESTING (t AND F) IN OLS MULTIPLE REGRESSION CONTEXT AND UNDERSTANDING STATA REGRESSION COMPUTER OUTPUT

(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.

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

Please make an appointment to meet with me if you are doing the 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)

CLASS 5: FUNCTIONAL FORM PART II: INTERACTIONS AND LOGARITHMIC TRANSFORMATIONS

(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

Please submit your preferences for articles to present or discuss if you are doing the article presentation (and not doing the 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

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

(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

Reading:
- STUDENMUND: Chapter 13

Exercises:
- STUDENMUND: Chapter 16, Exercises 3

Computer Exercise Class 8: Fixed Effects

CLASS 10: EXAMINATION

You may bring two pages of notes (any font, 8.5x11” or smaller, one sided) and you should bring a calculator. All necessary statistical tables will be supplied.
CLASS 11: MORE ON QUALITATIVE DEPENDENT VARIABLES AND SIMULTANEOUS EQUATION MODELS

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

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

Homework:
Submit explanation of causality for this study.

CLASS 13: PRESENTATIONS OF ARTICLES OR DATA PROJECTS
These will be scheduled once students choose between article presentations and regression projects.

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

Homework:
Submit explanation of causality for this study.

CLASS 14: PRESENTATIONS OF ARTICLES OR DATA PROJECTS
These will be scheduled once students choose between article presentations and regression projects.

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

Homework:
Submit explanation of causality for this study.

FRIDAY May 13: Project Papers Due in my Puck mailbox (3rd Floor) 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.