FALL 2010
NEW YORK UNIVERSITY
ROBERT F. WAGNER GRADUATE SCHOOL OF PUBLIC SERVICE

P11. 2902 REGRESSION AND INTRODUCTION TO ECONOMETRICS

Times and Place:
Mondays 4:55 to 6:35 pm
25 West 4th Street, room C-20

Prerequisite:
P11.1011 or equivalent

Tutors (all at Puck)
Name/email: Michelle Wong (mwong@nyu.edu)
Time/place: 6:45-8:45 pm Mondays, Jersey Conf.
(exceptions will be emailed)

Name/email: Douglas Coulter
(douglas.coulter@nyu.edu)
Time/place: Thursdays 3:00 -5:00 pm, Jersey Conf.
(exceptions will be emailed)

COURSE DESCRIPTION

In this course, you will learn how to use econometric methods, which are essential to the work of policy and finance analysts. Specifically you will learn how to estimate a regression equation and 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 computer 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 presenting and discussing a published study from a list provided on the syllabus.

FUTURE COURSEWORK AND BECOMING PROFICIENT WITH STATISTICAL DATA ANALYSIS

If you wish to become completely comfortable with hands-on work with statistics, data sets and statistical computer packages, then the following four courses at Wagner are highly recommended: P11.1011, P11.2902, P11.2875 and P11.3148 (research capstone). In addition E10.2110 Applied Statistics: Using Large Databases in Education Research can be taken at the same time (or after) P11.2902 and it will teach more Stata and use of large education databases.
That said, P11.2902 has considerable computer work assigned, and if you wish a more data-intensive experience in this course, you choose to do the data and regressions project (see below).

**COURSE REQUIREMENTS AND GRADING**

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

   Written and computer exercises must be handed in *during the class in which they are listed on the syllabus*. They will be graded 0, 1 or 2. Answers will be posted on Blackboard following the class they are due and the assignments will be returned to your Puck mailbox. Please be sure to list your mailbox number on your papers.

2. **(45%) Exam**

   An in-class exam will be given during Class 10 (*November 22*).

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

   Note that last semester exactly half the class chose each option.

   **In a group of four, five or six students, either:**

   - **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 25)**, 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 will schedule an hour and a half with me to go over the article in detail.

     - **Do and present 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 20)** 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 4)** 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 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 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 will be 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 will meet with me before the class in which you present for about 1.5 hours to understand the article and you will 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 P11.2875 or P11.3148 or E10.2110).

More on Doing and Presenting a Data and Regression Project

A paper will be due on the last day of class describing your project; an in-class presentation of your results will be due one the last three days of 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’ 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 and help you develop the project.

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.
2. Want a more intense hands-on experience with analyzing, writing about and presenting statistics (in advance of taking other follow-on courses such as P11.2875: Estimating Impacts, or P11.3148: Applied Research Capstone, or E10.2110: Applied Statistics, all three of which provide more hands-on work).
3. Read chapter 11 in STUDENMUND “Running Your Own Regression Project.”
The end result will be a paper, with two tables and 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 will still read the assigned articles (and write one causal paragraph).

Paper Outline (paper around eight to ten pages, with two tables)

Note: Discussion of the following two tables of results would be woven into the sections as indicated below, although the tables will be appended to 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


2. 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 all articles but will but will not present any.


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.

**BLACKBOARD**

You will 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 software; no previous knowledge is necessary and a student version is combined with your textbook. 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 the UCLA ATS website (http://www.ats.ucla.edu/stat/stata/default.htm) has useful online modules on Stata.

**CLASS NOTES**

Before each class, I will email you when class notes are available on Blackboard. You should print these notes, bring them to class, and use them for taking notes. You will only follow the class well if have the class notes with you.
SUMMARY OF COURSE GRADING AND DUE DATES

NB: Monday, October 11 is an NYU holiday (Columbus Day). Wednesday, December 15 runs on a Monday schedule and will be our last day of class.

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 13</td>
<td>S Chs. 1 &amp; 2 (34-38 and 46-57)</td>
<td>• S Ch. 1 ex. 9, 10 (a,b,c), 11</td>
<td>Email Professor by Friday before class if you want to do Data Project instead of Article Presentation</td>
</tr>
<tr>
<td>Two</td>
<td>September 20</td>
<td>S Chs. 2 (39-45) &amp; 4</td>
<td>• S Ch. 2 ex. 6, 10, 13 (a,b,c)</td>
<td></td>
</tr>
<tr>
<td>Three</td>
<td>September 27</td>
<td>S Chs. 3 &amp; 5 (not including appendix)</td>
<td>• S Ch. 4, ex. 6 (a,b), 11 (a,b,c)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Computer Exercise Class 3</td>
<td></td>
</tr>
<tr>
<td>Four</td>
<td>October 4</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>
<td>Meet with Professor if doing Data Project</td>
</tr>
<tr>
<td>No Class</td>
<td>October 11</td>
<td>Columbus Day</td>
<td>NYU Holiday</td>
<td></td>
</tr>
<tr>
<td>Five</td>
<td>October 18</td>
<td>S Ch. 7 (213 – 218, 220-223, 226-232)</td>
<td>• S Ch. 7 ex. 6, 12 (a,b,e)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Computer Exercise Class 5</td>
<td></td>
</tr>
<tr>
<td>Six</td>
<td>October 25</td>
<td>S Chs. 8 &amp; 9</td>
<td>• S Ch. 7 ex. 7 (a), 14</td>
<td>Hand in article ranking for presentation if doing article presentation</td>
</tr>
<tr>
<td>Seven</td>
<td>November 1</td>
<td>S Ch. 10</td>
<td>• S Ch. 8 ex. 6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• S Ch. 9 ex. 11</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Computer Exercise Class 7</td>
<td></td>
</tr>
<tr>
<td>Eight</td>
<td>November 8</td>
<td>S Ch. 16</td>
<td>• S Ch. 10 ex. 4 (a), 9(a,b), 11</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Computer Exercise Class 8</td>
<td></td>
</tr>
<tr>
<td>Nine</td>
<td>November 15</td>
<td>S Ch. 13</td>
<td>• S Ch. 16 ex. 3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Computer Exercise Class 9</td>
<td></td>
</tr>
<tr>
<td>Ten</td>
<td>November 22</td>
<td>Examination</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Date</td>
<td>Topic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>------------</td>
<td>------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eleven</td>
<td>November 29</td>
<td>S Ch. 14, (396-97) &amp; Table 11-2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- S Ch. 13 ex. 3, 12 (a,b,c)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- S Ch. 14 ex. 14</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Computer Exercise Class 11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Twelve</td>
<td>December 6</td>
<td>Articles to be announced (tba)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Explanations of Causality</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Presentations articles</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thirteen</td>
<td>December 13</td>
<td>Articles tba</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Explanations of Causality</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Presentations articles</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fourteen</td>
<td>December 15</td>
<td>Articles tba</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Wednesday)</td>
<td>- Explanations of Causality</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Presentations articles and projects</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Project papers due</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TOPICS AND ASSIGNMENTS

CLASS 1: OLS BIVARIATE REGRESSION MODEL WITH ERROR TERM
(September 13)

(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)

CLASS 2: OLS MULTIPLE REGRESSION AND ASSUMPTIONS ABOUT ERROR TERM
(September 20)

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).
CLASS 3: HYPOTHESIS TESTING (t AND F) IN OLS MULTIPLE REGRESSION
CONTEXT AND UNDERSTANDING STATA REGRESSION COMPUTER
OUTPUT
(September 27)
(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

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

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 11, NYU holiday

CLASS 5:  FUNCTIONAL FORM PART II: INTERACTIONS AND LOGARITHMIC TRANSFORMATIONS
(October 18)

(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 25)

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  
(November 1)

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 8)

(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 and Other Tests, Robust Standard Errors
CLASS 9:  FIXED EFFECTS REGRESSION WITH PANEL DATA PART II AND INTRODUCTION TO LINEAR PROBABILITY MODELS  
(November 15)

Reading:  
- STUDENMUND: Chapter 13

Exercises:  
- STUDENMUND: Chapter 16, Exercises 3

Computer Exercise Class 8:  Fixed Effects

CLASS 10:  EXAMINATION  
(November 22)

You may bring two pages of notes (any font, any size paper, one 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 29)

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
(November 6)

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 GROUP PROJECTS
(December 13)

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 GROUP PROJECTS
(Wednesday, December 15)

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.

Project Papers will be due this class

Homework:
Hand in explanation of causality for this study.

Project Papers Due
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      Other

Brief description of work experience (if any):

Specific areas of interest in public policy or finance:

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.