Fall 2020
DRAFT Syllabus – POGO611
Advanced Data Analysis for Policy and Government

Instructor: Naoru Koizumi
[Office] Arlington Founders Hall 520
[Phone] 703-993-8380
[Email] nkoizumi@gmu.edu (for main communication)
[Office Hours] Mondays 1:00 – 3:00 pm (and by appointment)

Teaching Assistant: Abu-Bakkar Siddique
[Office] Arlington Founders Hall #526
[Email] asiddi@masonlive.gmu.edu
[Office Hours] Tuesdays 3:00 – 5:00 pm (and by appointment)

TABLE OF CONTENTS

· Course Description
· Course Time and Location
· Learning Outcome Measures
· Course Materials
· Course Software
· Course Topics
· Course Grading
· Lecture Schedule
· Homework and Exam Policies
· Project Information
· Student with Special Needs
· Journal: New Voices in Public Policy
· SPP Policy on Plagiarism

COURSE DESCRIPTION

PREREQUISITE: POGO511 or equivalent
This course is an introduction to advanced methods in statistical and regression analysis. It is
designed to give the student a set of skills that can be applied in the workplace or to prepare them
for more advanced courses in statistical analysis. The course begins with a brief review of the
Ordinary Least Square (OLS) methods and then moves on to the Maximum Likelihood (ML)
method and studies some Generalized Linear Models (GLMs) that use ML estimators. The
course then discusses the assumptions underlying the method of OLS (i.e., Gauss-Markov
Theorem) and studies several detection methods and remedies for the violation(s) of the GM
Theorem. The course also covers regression analysis with time series data and some forecasting techniques. The course stresses applications using an advanced statistical package, STATA.

COURSE TIME AND LOCATION

- **Time:** Mondays 7:20 – 10:00 pm
- **Location:** We meet on Zoom every week at the time specified above (It is expected that some of the classes will be held asynchronously (i.e., not at the time specified above, but via recorded video/s), but we will discuss this option during the first class.

LEARNING OUTCOME MEASURES

1. Knowledge and understanding
   - Students should be able to read and understand the journal articles that involves basic time-series/panel-data statistical analyses
   - Students should be able to critically assess policy arguments, comments, reports and other materials that use statistical analysis
   - Where relevant, students should be able to make policy remarks and draw policy implications/conclusions based on the findings of various policy studies that apply statistical techniques.

2. Develop quantitative and qualitative skills
   - Students should be able to conduct basic time-series/panel-data statistical analyses to analyze various policy issues.
   - Students should be able to interpret the STATA outputs of various statistical analyses covered under this course.
   - Students should acquire basic programming skills necessary to conduct basic time-series/panel statistical analysis in STATA.
   - Students should learn how to find, collect, organize and clean datasets necessary to analyze the topic of their interests.

3. Professional development and leadership
   - Students should learn working effectively in a team setting through homework assignment and course project.
   - Students should learn how to make professional presentations through a course project.
   - Students should learn how to seek assistance when needed and communicate effectively with classmates, TA and Professor.
   - Students should learn how to deliver assignments and outputs in a timely fashion.
COURSE MATERIALS

REQUIRED TEXTS:
- *Introductory Econometrics* by Wooldridge, Jeffrey (Thomson South-Western)
  (Currently Version 7, but any version is ok)

RECOMMENDED TEXTS:
- *Statistics with Stata - Updated for Version 12* by Hamilton (Duxbury Resource Center)
- *Applied Logistic Regression* by D. Hosmer and S. Lemeshow (Wiley InterScience Publication)
- Useful online statistical texts (to supplement the above texts)
  - [http://davidmlane.com/hyperstat](http://davidmlane.com/hyperstat)

(return to contents)

COURSE SOFTWARE

- STATA 14 (IC or SE) or above
  Students are recommended to purchase a 6-month (or longer) license of Stata, Version 14 or above, either IC or SE. If you plan to use a large dataset for your project, you may want to invest in SE rather than IC. For the product (IC vs. SE) comparison, see: [http://www.stata.com/products/which-stata-is-right-for-me/](http://www.stata.com/products/which-stata-is-right-for-me/)
- Software Price:
  **Stata/IC software**
  - $48.00/6 months
  - $94.00/one year
  - $225.00/perpetual
  **Stata/SE software**
  - $125.00/6 months
  - $179.00/one year
  - $425.00/perpetual
- How to Order:
  **Online:** http://www.stata.com/order
  **Email:** orders@stata.com
  **Phone:** 800-782-8272 (Monday-Friday, 8 to 5 Central Time)
  Delivery is via electronic download. You will receive download instructions when you place your order and then your License and Activation Key within one day.
If you do not want to purchase the software, you can always use the computers in the Arlington computer lab (3rd floor). Stata should be installed in all machines in the lab. Note that all exams and problem sets require you to use Stata.

You do NOT need to purchase STATA before the semester begins. We will discuss your options during the first class.

COURSE TOPICS

Regressions with Cross-Sectional Data
- OLS Estimators
- Goodness-of-Fit Tests
- Dummy Variables and Interaction terms
- Multicollinearity
- Prediction
- Maximum Likelihood Estimators (MLEs)
- Logit and Probit Regressions
- Other log-linear Regressions

OLS Assumptions and the Violations
- Gauss-Markov (GM) Conditions
- Tests for the GM Violations
- Weighted Least Squares (WLS) Estimation
- Generalized Least Squares (GLS)

Regressions with Time-Series and Panel Data
- Lagged Variables
- Fixed-Effects and Random-Effects Estimations
- Instrumental Variables (IV)/2 Stage Least Square (2SLS) Estimations
- Simultaneous Equation Modeling

COURSE GRADING

<table>
<thead>
<tr>
<th>Component</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 Problem Sets</td>
<td>30%</td>
</tr>
<tr>
<td>2 Exams</td>
<td>30% (Total)</td>
</tr>
<tr>
<td>- Mid Term</td>
<td>10%</td>
</tr>
<tr>
<td>- Final (Comprehensive)</td>
<td>20%</td>
</tr>
<tr>
<td>Project</td>
<td>30%</td>
</tr>
<tr>
<td>Class Participation</td>
<td>10%</td>
</tr>
<tr>
<td>Lecture #</td>
<td>Date</td>
</tr>
<tr>
<td>-----------</td>
<td>--------</td>
</tr>
<tr>
<td>1</td>
<td>Aug. 24</td>
</tr>
<tr>
<td>2</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>Sep. 7</td>
</tr>
<tr>
<td>3</td>
<td>14</td>
</tr>
<tr>
<td>4</td>
<td>21</td>
</tr>
<tr>
<td>5</td>
<td>28</td>
</tr>
<tr>
<td>6</td>
<td>Oct. 5</td>
</tr>
<tr>
<td>7</td>
<td>13 (Tuesday)</td>
</tr>
<tr>
<td>8</td>
<td>19</td>
</tr>
<tr>
<td>10</td>
<td>Nov. 2</td>
</tr>
</tbody>
</table>
Endogeneity Bias [W: Ch. 15]  
Instrumental Variable (IV) / 2 Stage Least Squares (2SLS)

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Sample Selection Bias / Heckman (Heckit) Model [W: Ch. 17] Simultaneous Equations Model [W: Ch. 16] Final Exam distributed</td>
<td>PS3 Due</td>
</tr>
<tr>
<td>11</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>16</td>
<td>Hierarchical Linear Model (HLM) Missing data Imputations Propensity Score Analysis</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>23</td>
<td>Student presentations</td>
<td>Final Exam Due</td>
</tr>
<tr>
<td>14</td>
<td>30</td>
<td>Student presentations</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dec. 7 Student presentations (Reading Period)</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Dec. 14</td>
<td>No lecture (Exam Period)</td>
<td>Project Paper Due</td>
</tr>
</tbody>
</table>

**HOMEWORK AND EXAM POLICIES**

Both homework (problem sets) and exams will be uploaded to the course blackboard site for download.

**HOMEWORK**  
The homework is collected at the end of class on the due date. Late submission will be penalized. Although it is permitted to work as a group, the final product must be original.

**EXAMS**  
Both exams are open book and open notes. However, no collaboration is allowed for the exams. Students are expected to solve the exams alone (without any assistance from others). Further, the final products must be original. Late submission will be penalized.

**PROJECT INFORMATION**

**1. PROJECT DESCRIPTION:**

Students are asked to conduct either his/her own or a team (2 students per group at max) project. Each group is expected to undertake a case study involving a statistical analysis of some data set. The only substantive requirement is that your analysis includes at least one example of advanced analysis you have learned in the class. Each group is expected to find a dataset used for
the project (from websites or elsewhere) and make a brief (about 15 minutes) presentation at the last class. The selected websites that contain useful datasets are listed in the following section.

The report is expected to be around 15-20 pages, double spaced including figures and tables. The report should state: (i) motivation of the problem and key question/s, (ii) descriptive analysis of your data, (iii) the model specification for your analysis and (iv) summary and discussion of your findings. All source datasets and material used in the analysis should be cited explicitly.

Each group is expected to submit an electronic form of the presentation (ppt format). Each group is also expected to submit both a hardcopy and electronic forms (docx format) of the report.

FOR STUDENTS WITH SPECIAL NEEDS

If you are a student with a disability and you need academic accommodations, please see me and contact the Disability Resource Center (DRC) at 993-2474. All academic accommodations must be arranged through the DRC.

JOURNAL: NEW VOICES IN PUBLIC POLICY

New Voices in Public Policy: I will consider nominating the very best papers in this course for publication in New Voices in Public Policy. New Voices is a student- and faculty-reviewed journal that shares the finest student work with the rest of the world.

SCHAR POLICY ON PLAGIARISM

The profession of scholarship and the intellectual life of a university as well as the field of public policy inquiry depend fundamentally on a foundation of trust. Thus any act of plagiarism strikes at the heart of the meaning of the university and the purpose of the School of Public Policy. It constitutes a serious breach of professional ethics and it is unacceptable.

Plagiarism is the use of another’s words or ideas presented as one’s own. It includes, among other things, the use of specific words, ideas, or frameworks that are the product of another’s work. Honesty and thoroughness in citing sources is essential to professional accountability and personal responsibility. Appropriate citation is necessary so that arguments, evidence, and claims can be critically examined.

Plagiarism is wrong because of the injustice it does to the person whose ideas are stolen. But it is also wrong because it constitutes lying to one’s professional colleagues. From a prudential perspective, it is shortsighted and self-defeating, and it can ruin a professional career.
The faculty of the School of Public Policy takes plagiarism seriously and has adopted a zero tolerance policy. Any plagiarized assignment will receive an automatic grade of “F.” This may lead to failure for the course, resulting in dismissal from the University. This dismissal will be noted on the student’s transcript. For foreign students who are on a university-sponsored visa (e.g., F-1, J-1, or J-2), dismissal also results in the revocation of their visa.

To help enforce the Schar school policy on plagiarism, all written work submitted in partial fulfillment of course or degree requirements must be available in electronic form so that it can be compared with electronic databases, as well as submitted to commercial services to which the School subscribes. Faculty may at any time submit student’s work without prior permission from the student. Individual instructors may require that written work be submitted in electronic as well as printed form. The Schar school policy on plagiarism is supplementary to the George Mason University Honor Code; it is not intended to replace it or substitute for it.