84702 – Regression Analysis for Political Science II

Instructor: Daniel Hansen
Semester: Spring 2021
Email: dshansen@andrew.cmu.edu
Class time: 2:20pm to 3:40pm, Tuesdays/Thursdays
Location: Doherty Hall 2105
Office Hours: 3:50 to 4:50 each T/TH; or
                by appointment; or
                by ProMinute Office Hours app

Course Description and objectives:

This course is the second of the quantitative methods sequence for graduate students in the Institute for Politics and Strategy. Statistical analyses are ubiquitous even in everyday life – covering elections (infamously), sporting events to measure team and player performance, survey results, in addition to informing important policy debates such as minimum wage effects or climate change. Whether your career is bound for the private sector, government, or academia, it is likely you will benefit tremendously with a deep grounding in various statistical methods. This course will expose you to a broad range of models suitable for analysis in diverse situations. It will extensively cover time series and panel data analysis, modeling nonlinear and interaction effects, survival analysis and ordinal outcomes (especially helpful for analyzing survey results), and very importantly – how to identify, model and mitigate endogeneity problems. This broad exposure will empower you to understand and critique a wide variety of studies.

The learning objectives for this course are thus, explicitly, to: (a) give you an intuitive understanding of a broad range of statistical models; (b) bolster your ability to critique scholarly analysis; (c) enable you to competently execute your own research; and lastly (d) enhance your comfort in working with data in the R statistical environment. The emphasis for the course is to teach these econometric techniques at an intuitive level and leave you able to estimate these models on different data in the future on your own initiative.

Texts and Resources:

The only required text which you must acquire is **"Introductory Econometrics: A Modern Approach" by Jeffrey Wooldridge**. 7th or 6th edition is fine. Note in the reading list, this text is referenced simply with a "W".

I will also upload readings from other texts, including: **"A Guide to Econometrics" 6E by Peter Kennedy.** In the reading list, these chapters are referenced with a "K". Other assorted readings will come from other various texts and I will also supplement with academic articles employing some of the methods we are learning.

The course will rely on R and R Studio. I encourage everyone at the beginning of the semester to ensure their version of these programs are updated, as this can sometimes cause glitches. Of course these are freely available to download. A useful R code resource is the "R cookbook", available here:

Course Structure, Assignments, and Grading

Course Grade Composition	
Attendance & Participation	6%
Exams (10% each)	30%
HW Sets (6% each)	18%
Article Presentations	6%
Project Presentation	10%
Project Paper	30%
Total	100%

The grading scale for this course will be as follows:

A = 90 to 100%
B = 80 to 89%
C = 70 to 79%
D = 60 to 69%
R = 59% and below

For Masters students (which is everyone), the +/- system will be used. A (-) sign will be attached from 90 to 92.9%; 93 and above is an A. For all lower grades, the scale is: (+) is earned at 87% to 89.9%, a "straight" grade is between 83% and 86.9%, and a (-) is earned between 80% and 82.9%. This pattern repeats until a D is earned between 60% and 66.9%. Below a 60 results in an E/F.

Attendance & Participation

This course will make an effort to engage in conversation about what we are learning, as much as possible. This requires you to have diligently and scrupulously done the readings, be ready to offer answers, or simply be open about confusions you have. While it should be easy for students to earn this part of the grade, I may deduct if you seem persistently unprepared. If I am concerned about this I will contact you privately. Note as well, it is perfectly ok to have done your best with the readings and simply have some confusion – you won't be deducted for giving incorrect answers in class discussion! If you show active participation or a willingness to ask questions, you should be fine here. You will have two excused absences automatically, but anything beyond this may result in deductions. Communicate with me proactively if you perceive there are personal issues preventing consistent attendance or engagement, and I am happy to work with you on a specific basis to resolve any challenges you have.

Article Presentations

For each student in the class, I will assign a day to present an academic article which employs some of the regression techniques we will be learning about in the class. Note that **ALL students** are required to have read the article, **NOT** just the presenter for the day. As much as possible, I am aligning the article with the topic of the day. These presentations will come roughly at or slightly past midway through the class session, after I have done some opening lecture and discussion about the subject at hand. I fully expect that you won't have a "full" grasp on the material in the reading I have assigned, and that is OK. However, I expect you to have carefully read the assigned textbook chapters for that day as well, which should help you to understand the methods used in the academic article.

The point of this presentation is to show a real world example of political science research employing these techniques. In this way, the textbook chapter/s will map cleanly onto a cogent (hopefully interesting) example. It will also help you to see how scholars discuss these models within their actual research – you can see how they motivate it, discuss strengths and weaknesses, etc. Sometimes, you may even notice they made questionable choices!

In **giving your presentation, you will**: (a) have a simple, clear introduction – what is the author arguing? Why is it interesting? (b) briefly discuss the authors theory. (c) be explicit and detailed about the research design, what model/s they used and why was that the appropriate choice (especially as it relates to what we've learned in class), and SHOW their empirical results (screenshots of tables/figures are good). Discuss and interpret them. (d) make a brief effort to critique the operationalization. It should be roughly 15 minutes in length, and be ready for your classmates to ask questions! I expect you to answer any questions they raise.

You will be **graded** on a 4-category scale.

100% = It's so good, how can I complain?
90% = Good, but some errors; missed some obvious aspects of the empirical design & results
50% = Apparent lack of preparation, very confused about the paper, sloppy presentation style
10% = "I would guess the student spent 10 minutes on this"

Homework Sets

There will be 3 homework sets, worth 6% each. They will cover roughly the previous 4 days of lecture, wherein I will give you computer exercises which engage with the material covered in class. The problems will allow you to use the code we've already examined to execute and provide answers.

Exams

There will also be 3 exams. However, unlike the homework sets, these will not be computer exercises but will instead be a series of conceptual questions relating to the textbook material. These are meant to assess your intuitive understanding of how the models work and specific features related to them. The exams are not meant to be tricky, but to check your baseline comprehension. If you do the readings and pay active attention in class, you should be fine. The exams will be posted

to Canvas, and you may take the exam anytime during the allotted 24 hour day. You may take it during normal class hours if you desire, or you may choose a different time of the day.

Research & Data Project

This will be broken into component parts – a presentation (10%) and an actual research paper (30%). The paper should be roughly 5 to 7 pages in length, without considering output tables and figures which you will have to include in the paper (1.5 spacing). You will first **make a proposal** to me of your project. You will tell me: the theory/hypothesis you want to test, which what variables (it's ok if you don't know how to find them – start conceptual, and I can guide you to appropriate data sources). You will **deliver the proposal to me on March 18th**. You may choose one of two options:

(X) Conduct an "**augmented replication**" analysis. This means you will identify an article of interest from a prominent social science journal, locate the replication materials for the associated paper, and conduct a basic replication analysis of some of the results from the authors' paper. After replication, you will then conduct a basic sensitivity analysis of some kind. THEN, you will "augment" it by adding your own twist to the research. It could mean you test something within the authors data – perhaps an interactive hypothesis they didn't explore, or a nonlinearity, or perhaps you felt they should have used a different regression model. Or, you may need to merge in a new variable or two to test your extra hypothesis/es.

(Z) Propose your **own (semi)-original idea**, and test it. If you choose this path, I would want you to find one or two papers related to it, just to give your paper a baseline, and from their you can build a coherent paper and research design. If you go this route, I will help you out by doing more to ensure you have all of the data needed (it need not be a large dataset, just some basic measures to allow you to accomplish your statistical objective).

<u>Paper</u>

The paper must, whether option X or Z, demonstrate some sensitivity and stability analysis. It must provide a basic theoretical discussion, before then executing a careful empirical design. It should have a basic Intro (what the paper is examining, briefly and concisely preview your findings), and a basic literature review. For the literature review, you may use some of the original paper's literature discussion provided you place it within your own voice and interpretation. Offer criticisms of the main study you are relying on. Then, briefly sketch out your theory and expectations. All of this first part should be concise – give enough of this to establish intellectual momentum. But the whole point of this paper is to showcase your statistical modeling, and that's where the main essence of the paper should be.

So after the intro, lit review, and theory, move towards a careful and meticulous statistical analysis. Explain how your variables conceptually map onto your theory/innovation (why they are a good fit to test what you are testing), discuss the data being used. Evaluate other researcher's choices – what model did they use and was it an appropriate choice? Do you need to consider a different model? Assess how sensitive your results are to alternative specifications (model, variable inclusion, or other modeling choices). If there is an endogeneity problem or threats to inference, then explain this problem (even if you cannot "solve" it). Acknowledge it and move forward (I will be impressed

though if anyone accuses an author of ignoring endogeneity and then actually estimates a correction!). Then of course, present very clear results, with nice tables and figures. The discussion of the results should convey that you understand the interpretation, what the effects mean, the magnitude of the effects, and how well it matches theoretical expectations. It will **be due May 10th**.

Presentation

Your presentation should be roughly 15 to 17 minutes. It should be very clear and easy to follow. Have a nice and clear introduction: don't launch into messy details - set the scene for us. What are you testing in your paper, what's the motivation (why is it interesting/important), and in a sentence tell us your results. Then, move towards a brief review of the what the existing studies say on the issue, and briefly outline the theory. Then walk us through your statistical analysis. At the end you will receive questions from your audience. Be aware that you **may** need to incorporate some minor changes before your final draft is due, based on comments in the Q&A or advice I may give you.

Learning during the era of COVID-19

We are clearly in highly unusual times, and you should know that I understand and am very sympathetic to the challenges we all are facing. Be proactive and conscientious regarding your personal approach to these difficulties. Develop a routine which maximizes your ability to learn and function in this new format. Talk with friends, consult with CMU resources, and be proactive in reaching out to me so I understand your personal circumstances. I am devoted to ensuring your ability to succeed.

I encourage **everyone** to take public health guidelines seriously, use all precautions and socially-distance as best as you can. I ask each of you to remember that neglecting these guidelines affects not just you, but others. Your fellow community members should not have to bear the consequences of your negligence. We suffer and succeed as a community, so let us all show how great our community really is.

Given the in-person plus remote format for the course, every class will require all students to log in to the Zoom link, whether in person or not. This will also enable us to use the break-out rooms feature for the course. I encourage everyone to turn on their video as it helps provide for the best class environment and social experience. Each class will be start with everyone's microphone muted. If a remote student wishes to ask a question, please use the raise hand feature on Zoom or briefly message in the chat window. I will do my best to notice you and will ask you to unmute yourself to ask a question or participate, and as a fail-safe I will assign a graduate student to help me in monitoring for remote questions.

General policies

- I have a zero-tolerance policy for cases of academic dishonesty. You need to make a choice beforehand that you will not engage in such disreputable conduct. Violations of the university policy on plagiarism or other cases of academic dishonesty will not be tolerated and can lead to serious action which may follow you far beyond your years in this university. Don't go there.

- Technology: I am a big believer in the use of computers, laptops and tablets for coursework. However, there is rampant abuse of this technology. People use it for social media purposes and many other activities during class. This is extremely disrespectful to your fellow classmates and can negatively influence your participation grade if I observe it during class.
- There will be no make-up exams unless there are severe cases of extenuating circumstances. In such cases documentation will be required for arrangements to be made.

As a student, you may experience a range of challenges that can interfere with learning, such as strained relationships, increased anxiety, substance use, feeling down, difficulty concentrating and/or lack of motivation. These **mental health concerns** *or stressful events may diminish your academic performance and/or reduce your ability to participate in daily activities. CMU services are available, and treatment does work. You can learn more about confidential mental health services available on campus at: http://www.cmu.edu/counseling/. Support is always available (24/7) from Counseling and Psychological Services: 412-268-2922.*

Take care of yourself. *Do your best to maintain a healthy lifestyle this semester by eating well, exercising, avoiding drugs and alcohol, getting enough sleep, and taking some time to relax. This will help you achieve your goals and cope with stress.*

All of us benefit from support during times of struggle. You are not alone. There are many helpful resources available on campus, and an important part of the college experience is learning how to ask for help. Asking for support sooner rather than later is often helpful.

If you or anyone you know experiences any academic stress, difficult life events, or feelings of anxiety or depression, we strongly encourage you to seek support. Counseling and Psychological Services (CaPS) is here to help. Call 412-268-2922 and visit their website at http://www.cmu.edu/counseling/. Consider reaching out to a trusted friend, faculty member, or family member for help in getting connected to support services.

If you or someone you know is feeling suicidal or is in danger of self-harm, call for help immediately, day or night:

CaPS: 412-268-2922
Re:solve Crisis Network: 888-796-8226
If the situation is life threatening, call the police:
    On campus: CMU Police: 412-268-2323
    Off campus: 911
If you have questions about this or your coursework, please let me know.

Readings and Course Schedule

Date	Subject	Reading 1	Reading 2	Presentation	Assignments
2-Feb	Class Introduction				
4-Feb	Recapping Regression				
9-Feb	Discussion on review/readings	K2 and 3			
11-Feb	Transformations, conceptual issues	W6	McElreath 5.3	Bodea & Ye (inter)	
16-Feb	Binary analysis	W7		Houle	
18-Feb	Heteroskedasticity	W8			
23-Feb	*Break Day*				
25-Feb	Time Series I	W10 (334 to 356)			Give HW
2-Mar	Time Series II	K19	W11 (select sections)	Masaaki	
4-Mar	**Exam I**				
9-Mar	Panel Data I	W13		Gray	**HW 1 Due Wed 10th**
11-Mar	Panel Data II	W14			
16-Mar	Nonlinear	K16		Bodea & Houle	Give HW
18-Mar	Survival	Rabe-Hasketh 14		Svolik	Proposal Due
23-Mar	**Exam II**				
25-Mar	*Group Day I*				
30-Mar	Endog I (IV models)	W15 (495 to 510)	K9 (exclude 9.3.4)	Acemoglu & Robinson	**HW 2 Due**
1-Apr	Endog II (Simult)	W16		Bernard	
6-Apr	*ISA Conference - No class*				
8-Apr	Endog III (selection)	K17.3	Von Stein	DeMerrit	Give HW
13-Apr	*Group Day II*	No reading; Show up happy			
15-Apr	*Carnival - No class*				
20-Apr	Bayesian Modeling	K14		Karreth	**HW 3 Due**
22-Apr	Kennedy's 10 Commandments	K22		EU paper	
27-Apr	**Exam III**				
29-Apr	*Presentations Day I*				
4-May	*Presentations Day II*				
6-May	*Presentations Day III*				
10-May					**Paper due**