York University
Winter 2021 - Remote delivery
AP/Econ 4210 3.0 M: Econometrics

– Syllabus –
Subject to change

Official description

The objective of this course is to provide students with a number of tools for conducting and understanding empirical research in economics. The focus of the course will be on empirical methods for cross-sectional and panel data. The main topic of the course is the linear regression model, its estimation and inference. We cover least squares and maximum likelihood estimators. Other topics include heteroskedasticity, endogeneity, instrumental variable estimation, and simultaneous equations, difference-in-differences. In addition to analytic exercises, students will receive practical questions requiring handling and analyzing data using the statistical software package `R`. Prerequisite: AP/ECON 3500 3.00 or 3210 3.00. Course credit exclusions: GL/ECON 4260 3.00, SC/MATH 3330 3.00. Note: This course can be taken along side Econ 4140.

Note: Econ 3500 is **not** required for this course, provided you've taken Econ 3210. It is highly recommended you have at least **one** of Econ 3210 or Econ 3500.

What we're gonna do

This is an intermediate-level course in econometrics for students at the York (Keele campus). The goal is to equip you with a modern approach to data analysis and econometrics, focusing on the use of data to answer causal questions. You will learn about different empirical techniques that economists use: random assignment, linear regression, difference-in-differences, instrumental variables and regression discontinuity design. The workhouse of the course is the linear regression model, its estimation and inference.

You will learn about applications of these techniques. In particular, you will get hands on experience and gain familiarity with `R` and `Rmarkdown` to get data, manipulate data, perform data visualization, estimate econometric models, and present and communicate results.

Learning Outcomes

By the end of this course you should be able to:

1. Understand the notion of causality, its importance in empirical work, and differentiate causality from correlation or association.
2. Interpret and comment on tables of estimated coefficients from a wide range of econometric models, in various formats.
3. Use your understanding of the methods to assess the validity and quality of empirical studies, including the ability to judge whether a method may or may not work in a specific research context.
4. Identify the appropriate empirical strategy to answer causal questions using data: random assignment, regressions, instrumental variables, difference-in-differences, and panel designs, and their associated regression specification.
5. Clearly articulate each method's requirements/assumptions, typical use, and limitations, and know how to interpret their quantitative results.

6. Apply these methods to actual data-sets, using the `R` programming language and output and communicate the results in `Rmarkdown`.

Lectures and tutorials

This course will be delivered in a 'flipped' style delivery. This means that readings and videos will be posted prior to class, and students are expected to be familiar with the material *before* the lecture. Lectures will be roughly one-hour long followed by tutorials. See the Course schedule for more details.

Lectures

Before each lecture, you are expected to do the required readings. Readings will usually be taken from the textbook, but will occasionally come from additional sources, typically via reading/video that I will create and post.

Tutorials

In the tutorials, we deal with the practical application of the concepts and methods used in class. This will involve `R` and `Rstudio`. You must have a working computer, a current installation of `R` and `Rstudio` and an internet connection.

Lectures and Office Hours

	Day	Time	Duration (min)	Place
Lecture:	T	16:00	~ 180	Zoom
Office Hours (Ben):	T	15:00	~ 60	Zoom

Office Hours

I will be available for office hours. During these times students are encouraged to ~~drop by~~ schedule a zoom session. If you cannot attend the scheduled office hours, I will be available by appointment as well. I will provide additional office hours before the midterm and final exam.

E-mail

E-mail is a good way to contact me. I will attempt to answer reasonable questions via email as quickly as possible. Please include "Econ 4210" in the subject line to facilitate this process. Your name (as it is written on Moodle) and student ID are required in every email if you expect to get a response. My answering-your-email-vibe depends on the quality of your email. In particular, emails should be straight forward and ask specific questions that I can answer either directly or by referring to the appropriate material. Good emails are concise, specific and have an obvious solution. You should think of an email as a last resort: the answer to your question should not be something that you could look up yourself. If you have an issue with `R`, you should Google it first or check out slack exchange before contacting me. If you have a question about an administrative issue, you should check the course schedule or syllabus or ask a friend before me.

Course Work

Homework

Done online via MyLabs. During the course of the term, there will be ten (10) homework assessments. The homework will be done individually. Only the best eight (8) homework assessments will count toward your final grade. A student may write all ten (10) homework assessments regardless of previous performance. However, only the best eight (8) will count toward the final grade.

- **Missed homework will get a grade of zero, and will automatically count toward your lowest 2 grades. Thus, there is no need to contact me if you miss a homework. *There will be no extensions, no exceptions.***
- Homework will consist of answering numeric, multiple choice, and true-false type questions on MyLabs. Some questions will involve data and I recommend you use `R` to answer these questions.
- The homework schedule will be posted well in advance. There will be at least 6 days to complete each homework. Students will have two-attempts, and there is no time limit on each attempt.

Empirical Assignments

During the course of the term, there will be six (6) assignments to be done in `R` and `R markdown`. Each assignment will be worth 6 percent of your final grade. This work can be done in groups up to four, but each student must upload their own assignment. No work will be accepted after the due date. Students are encouraged to study the suggested answers to each assignment, regardless of their performances on the assignment. If a student finds a problem with the grading of a assignment, she/he should immediately talk to the me. The deadline for a regrading request is **one week** from the day the assignment grade is posted to e-class, regardless of when the student actually receives it.

Assignment due dates and instructions are posted well in advance, and it is your responsibility to ensure adequate time to complete the work and deal with any issues, *including technical issues*. **Failure to submit an assignment on time will result in a grade of zero**. Assignments are considered submitted by the time all the files have been uploaded in the correct format to e-class according to the assignment instructions. **Assignments cannot be submitted late. There will be no exceptions.** Make sure to allow ample time for submission before the deadline; excuses such as: *the website is slow*, *I only submitted one minute late*, *I have had a stomach bug on the last day before deadline*, *I forgot to upload one of the files*, etc. are not valid excuses. These rules are there to limit unwarranted individual requests, which take up valuable time that I could spend improving the course content.

- Empirical assignments will consist of longer-form empirical questions and must be completed in `Rstudio` and submitted as an `HTML` document constructed in `Rmarkdown`. Assignments may involve downloading and loading data, installing and loading packages, writing code, constructing (nice) tables and figures, interpreting and communicating statistical output.
- The assignment schedule will be posted well in advance. There will be at least 14 days to complete an assignment.

Quizzes

During the course of the term, there will be four (4) quizzes. The quizzes will be done individually. Only the best three (3) quizzes will count toward your final grade. Each quiz will be approximately 45 minutes to an hour long. A student may write all four (4) quizzes regardless of previous performance. However, only the best three (3) will count toward the final grade. **Missed quizzes will get a grade of zero, and will automatically count toward your lowest grade. Thus, there is no need to contact me if you miss a quiz.** There will be no extensions, no exceptions.

- **Missed quizzes will get a grade of zero, and will automatically count toward your lowest grade. Thus, there is no need to contact me if you miss a quiz.**
- Quizzes will consist of answering numeric, multiple choice, and true-false type questions on MyLabs. Some questions will involve data and I recommend you use `R` to answer these questions.
- The Quiz schedule will be posted well in advance. There will be at least 3 days to complete each quiz. Students will have one attempt per quiz. Each quiz has a 45 minute time limit.

Final exam

A final exam, scheduled by the registrar, will be worth 30 percent of the final grade. I will discuss the details at the appropriate time. I will post a practice final exam so that students are fully aware of the style of questions that will appear on the final exam. I will provide solutions to the practice final exam. You do not

have to ask for them, I promise I will post them. Emailing me will not speed up this process. The final exam is cumulative.

Grade

The final grade in the course will be determined as follow:

Component	Symbol	Calculation/grading	Weight toward grade	Individual weight
Homework (8/10)	**HW**	MyLabs	16%	2%
Assignments (6)	**A**	Me/TA	30%	5%
Quizzes (3/4)	**Q**	MyLabs	24%	8%
Final Exam	**FE**	Me	30%	-

The grading scheme for the course conforms to the 9-point grading system used in undergraduate programs at York (e.g., A+ = 9, A = 8, B+ = 7, C+ = 5, etc.). Assignments and tests will bear either a letter grade designation or a corresponding number grade (e.g. A+ = 90 to 100, A = 80 to 90, B+ = 75 to 79, etc.).

How to be successful in this course.

Eighty percent of success is showing up – **Woody Allen**

The most straightforward way to success (ie, a good grade and perhaps a letter of reference for a grad program) is to show up, pay attention, participate, do the homework (they are relatively easy, low-stress, grade inflators), and not leave all of your studying to the day before the a quiz/final. I want you to succeed. I am willing to help if you struggle. But there has to be a certain level of commitment on your end, too.

Learning how to use new software/programming language can be a frustrating, non-linear process. The only way to learn is by doing. There will be many failures along the way. . . but you will get there.
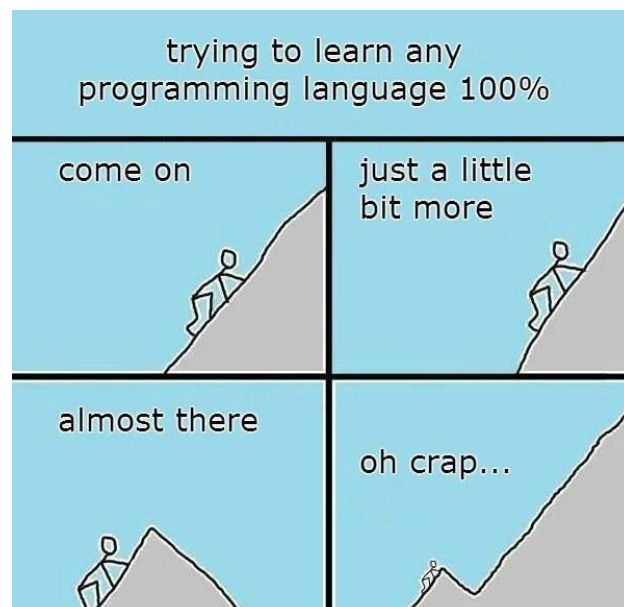

Figure 1: Learning Curve

Why is there so much course work?

Because I want everyone to do well. It helps to have consistent practice and feedback. The homework component is not overly demanding or time consuming. The empirical assignments are meant to emphasize key concepts and build your skill-set through practice and feedback. Since each new concept builds on the previous ones, its important to keep up with the material and have a chance to demonstrate your mastery of each concept on a routine basis. This course outline attempts to establish a feedback loop: Students complete a task and are evaluated, receive feedback, and can adjust accordingly. The feedback loop is a process to check and confirm the understanding of specific, manageable, focused concepts at regular intervals to build capacity and depth over time. The feedback is important for me, also, so that I can address areas of weakness. And it works.

Reading

The following textbook is required for the course:

Stock and Watson, *Introduction to Econometrics, 4th Edition.* Pearson. With MyLabs.

Topics to be Covered

- Introduction to `R`, `Rstudio` and `R markdown`
- Basic data analysis/programming/data visualization
- Review of Basic Statistical Concepts
- Review of Regression and Inference
- Multiple Regression Analysis: Additional topics and issues
- Instrumental Variables Estimation
- Introduction to Panel Estimation/ difference-in-differences
- Limited Dependent Variables
- Program Evaluation/potential outcomes
- Introduction to Machine Learning/Prediction

Important Information

All students are expected to familiarize themselves with the following information, available on the Senate Committee on Curriculum & Academic Standards webpage:

- York's Academic Honesty Policy and Procedures/Academic Integrity Website
- Ethics Review Process for research involving human participants
- Course requirement accommodation for students with disabilities, including physical, medical, systemic, learning and psychiatric disabilities
- Student Conduct Standards
- Religious Observance Accommodation

The departments guidelines for deferred standing

Can be found at this link

Why Econ 4210?

1. Answering questions using data is a big deal:

In both the public and private sectors, decision/policy makers are increasingly relying on data to make informed, data-based decisions. As an example, there are some characteristics of several leading tech firms' job postings:

1. Netflix job posting keywords: Causal Inference; Experimental Design; Advertising Effectiveness,
2. Facebook's Data Science Team features RCTs and causal inference,

3. Amazon offers a reduced form/causal/program evaluation *track*,
4. Google seeks experience in experimental design and causal inference.

From a random Deliveroo job ad (source) looking for an economist:

> Experimenting at this scale presents some unique challenges and we're investing heavily in building a world-class platform for designing, deploying, and analyzing product experiments. We're looking for experts in statistical inference and estimation to join our growing team of data scientists and help us develop innovative statistical solutions for industrial-scale experimentation.

And its not just the private sector – policy evaluation is an important part of government decision making at all levels. **Answering causal questions using data is what this course is about.**

2. You plan on getting a graduate degree

This course, along with the other core economics courses, are essential for success in MA Economics programs. This course, in particular, is a strong signal of a students potential in an MA Economics program and, thus, highly useful in making admittance decisions. If you are planning on an MBA instead, this course is still a great fit, a strong signal of potential, and will help set you apart from other candidates a bit. In Master's of Public Policy and Master's of Public Administration programs, this course will help you get a leg up on their statistical and program evaluation courses.

3. You want to stand out

There are just under 2500 economics majors at York (Keele) and only about 25 students take this course per year (roughly, just over 1 percent). You want to be in this **1 percent.**

Why `R`?

1. Its free

`R` is a free, open-source programming language for statistical computing. All of our work in this class will be done using `R`. You will need regular, reliable access to a computer running an up-to-date version of `R`. Download here

`RStudio` is a free, open-source, industry leading, `R` programming environment. It contains a built-in code editor, many features to make working with `R` easier, and works the same way across different operating systems. Use of `RStudio` is required for the labs, and strongly recommended in general. Download here

`R markdown` provides an authoring framework for data science. You can use a single `R Markdown` file to both:

1. save and execute code,
2. generate high quality reports that can be shared with an audience.

2. You'd like a job at some point, and `R` can help.

`R` is used in nearly every industry, particularly in fields that require data analytics. Many large companies use `R` (source):
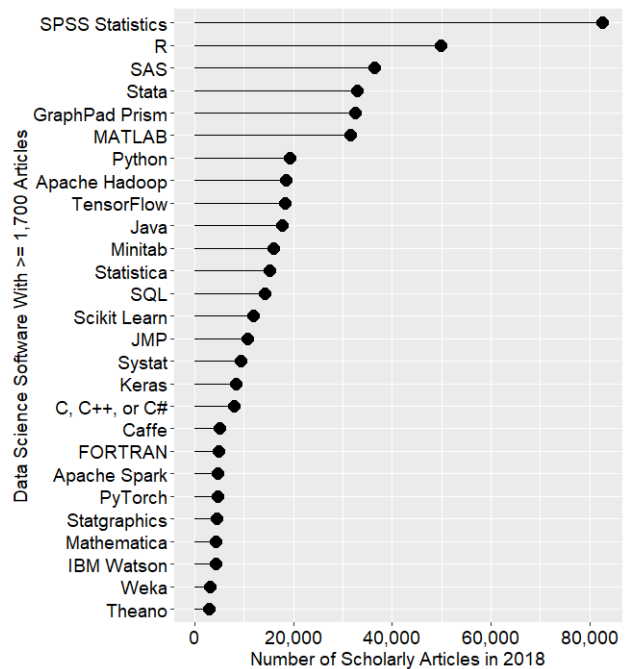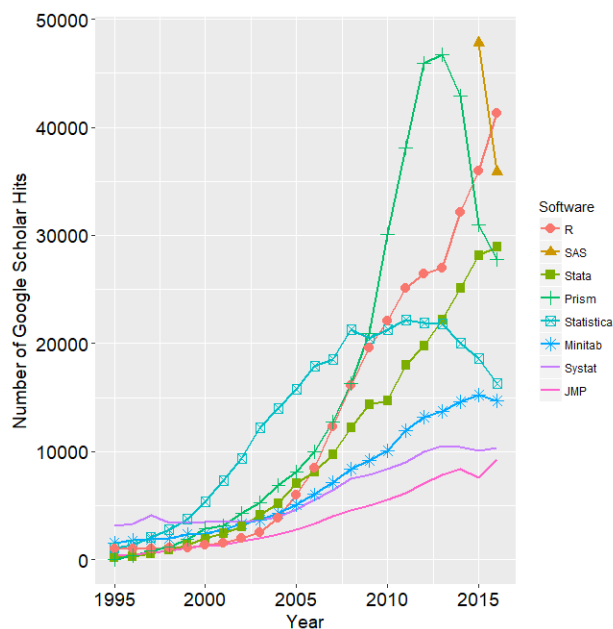
At facebook, `R` is used for data analytics and visualization, among other things (source):
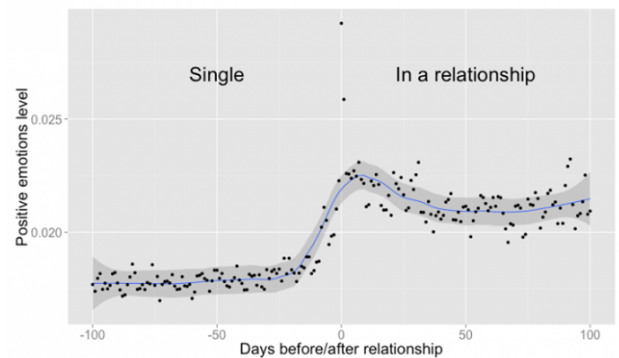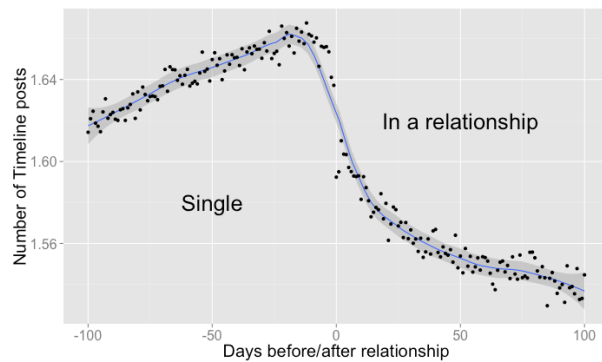
3. `R` is popular and growing in cutting-edge research

`R`'s popularity is growing in scholarly research (source).

Don't take it from me, here are some other sources:

1. Fast Comapny,
2. Turn a hobby into a career,
3. 8 Reasons,
4. Because these guys use `R`, and they do great data visualization.