BOSTON COLLEGE
Department of Economics

ECON2228.04: Introduction to Econometric Methods (Spring 2022)
O'Neill 257, T Th (9:00 – 10:15)

Christopher Maxwell
maxwellc@bc.edu

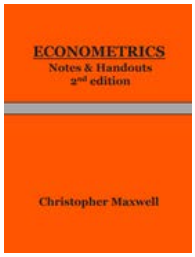
Maloney Hall, 337
Off. Hrs: by Zoom (tbd)

This is an introductory course in the use of econometric methods, with an emphasis on empirical applications. Our focus will be on learning ***to do*** econometrics, not just learning about econometrics or econometric theory.

While the course will cover the development of the formal tools of econometric analysis (simple and multiple regression analysis, estimation, inference, categorical variables, functional forms and so forth), we will also spend quite a bit of time on empirical methods (posing questions, building datasets, running regressions, looking at results, supplementing datasets, running more regressions, etc … until we can confidently say something about the questions at hand). As such, an important part of the course will be a set of empirical exercises and an empirical research project in which students will be building their own datasets and applying the various econometric methods developed in the course.

Prerequisites: An introductory statistics course such as EC 151. No exceptions. Students should also have an understanding of basic *Excel* (which will be used at times to assemble datasets and verify calculations) and basic calculus.[1] I also assume that you've had some previous exposure to *Stata*, the computer language that we will be using to run regressions (you should have seen Stata in EC151). (See far below for more about Stata.)

Course reference text: The course is built around a set of handouts, which are posted to Canvas. To simplify things for you, I have also self-published the handouts through Amazon. I have printed hardcopies for everyone (those copies are free to you and will be distributed in class). You can also order the book through Amazon. I sell the book at my cost, which is $8.86 (it's 372 8.5 x 11 pages), and will only make it available for you all for about two weeks or so. It should be available now.[2] I have also posted to Canvas the pdf for the text.



Some additional texts: There is no need to purchase any of these (most are available at O'Neill). I list them because it is sometimes useful to see different presentations of the material.

- Jeffrey M. Wooldridge, *Introductory Econometrics: A Modern Approach*, 7th ed., Cengage Learning, 2019.

[1] The course website and Canvas site contain a few links to online materials and tutorials.
[2] Let me know if you'd like a copy and are not able to purchase it through Amazon, and we can discuss shipping options.

- Angrist, Joshua D. and Jörn-Steffen Pischke, *Mastering 'Metrics: The Path from Cause to Effect*, 2014.
- Stock, James H. and Mark W. Watson, *Introduction to Econometrics*, 4th ed., Addison-Wesley, 2019.
- Studenmund, A.H., *Using Econometrics: A Practical Guide*, 7th ed., Pearson, 2017.
- Bailey, Michael A., *Real Econometrics: The Right Tools to Answer Important Questions*, 2nd ed., Oxford, 2019.

Grading (Exams: 75%; Research Project: 10%; Exercises/qFlips/Quizzes: 10%; Labs: 5%):

- **Exams (75%):** Three exams: Two mid-term exams and an optional final exam.

  Anticipated dates (these may change):

  - Mid Term Exam #1 on Half #1 material; Thursday, March 3rd : OLS Analytics and Assessment (this is the last class before Spring Break; plan accordingly)
  - Mid Term Exam #2 on Half #2 material; Thursday, May 5th (last class in the semester): OLS Estimation and Inference + Topics
  - Optional Final Exam: Thursday, May 12th, 12:30 PM

  For mid term exams, you are allowed one *cheat sheet* (8.5 x 11 or A-4) and the use of a calculator; for the final exam, you are allowed two *cheat sheets* and the use of a calculator.

  Exam weights and the optional final exam:

  - If you take the optional final exam: Each mid-term exam counts for 21% of your course grade, and the final exam counts for the remaining 33% (so the exam weights are: 21%-21%-33%).
  - If you decide not to take the optional final exam: Each mid-term exam counts for 37.5% of your course grade.

  You must commit to taking the final exam at the time you receive the exam. To allow you to make a fully informed decision about whether or not to take the (optional) final exam, conditional course grades, which assume that you are not taking the final exam, will be posted to Canvas as quickly as possible after the end of classes. At that time , I'll be happy to provide you with a sense of how final exam performance will impact your course grade.

  There are no make-up exams in this course. If you miss either mid-term exam, then you must take the final exam (exam weights will be adjusted proportionately). All exam grades are curved. While every exam is different, exam scores in general seem to average about 70% of total available points, which curves to about an 85 or so.

- **Research Project (10%):** Independent replication of an econometric analysis that has been published in an academic journal. More details below.
- **Exercises/qFlips/Quizzes (10%):** There will be four Exercises over the course of the semester. These are team projects, which you will typically have two weeks to complete. As well, there will be about ten quizzes or short online assignments (called *q(uick)Flips*), which

will typically follow each Unit. Course grades for Exercises/qFlips/Quizzes will be curved, after dropping the lowest score… it's OK to miss one; don't miss two! More details below.

- **Labs (5%):** Course-wide labs, focused on using Stata in empirical/econometric analysis. If you are not registered for a Lab, please do so… now! While I am always happy to tackle Lab questions, you should know that the Labs are completely independent of the 2228 courses.

Canvas: Historically, I did not use Canvas for the course, except to post course-related scores and grades. All other course-related materials were posted to my oh-so-spiffy (not!) course website: http://www.cmaxxsports.com/ec228 . (To access the website, just google *ec228*.) With the move to *OnlineU*, however, I migrated all course content from my website to Canvas. So while we will use the website a bit (for qFlips), almost all of the course content is now on the courses' Canvas website. Do not expect my ec228 website to be up-to-date or complete.

Accommodations: If you are a student with a documented disability seeking reasonable accommodations in this course, please contact Kathy Duggan (x2-8093; dugganka@bc.edu) at the Connors Family Learning Center regarding learning disabilities and ADHD, or staff in the Disability Services Office regarding all other types of disabilities, including temporary disabilities. Advance notice and appropriate documentation are required for accommodations.

Academic Integrity: You will be held to Boston College's standards of academic integrity. If you have any questions as to what that means, see BC's academic integrity policies webpage.[3]

Pass/Fail: It is perfectly fine, of course, to take the course Pass/Fail… but it is definitely not OK to do so and shirk on group projects/exercises. That is not fair to your teammates… and they will come to hate you! Accordingly: If you are taking the course Pass/Fail, please let me know at the start of the semester, and I will monitor goings-on and make adjustments if necessary.

The Research Project (10% of your course grade): This is an applied/empirical project, which will kick off with team assignments at around the time of the first mid-term exam. In the past, there have been two phases to the project: Phase I – Independent Replication and Phase II - Improvement. However, this semester I'm dropping Phase II, and so the Project is now focused solely on independently replicating an existing published piece of econometric analysis… of your choosing. [4]

1. **Team assignments:** The Research Project kicks off with team assignments around the time of the first mid-term exam. I will assign the teams, which will likely have three members each.
   - ***Kickoff/Team assignments:*** Project teams assigned around Mon. March 14th
2. **Topic selection:** Topics should showcase interesting econometric analysis, and need not be restricted to topics in Economics. It's important to get an early start, as empirical research is always slow going!. To help you in that regard, I'll ask teams to email me a one paragraph

[3] https://www.bc.edu/content/bc-web/academics/sites/university-catalog/policies-procedures.html#academic_integrity_policies

[4] Published here means published in an academic journal (so no unpublished senior theses, web blogs, or the like).

description of their topic/paper of interest, within two weeks of team assignments. I will compile and circulate those blurbs and we'll discuss them in class at some point.

- ***Topic selection:*** Due around Weds. March 30th , about two weeks after *Team assignments*)

3. **Independent replication:** Independently replicate both the summary statistics of interest presented in the paper (to show that you have indeed replicated the construction of the dataset) as well as at least one set of regression results of interest.

   Your end-of-semester deliverable will be a *PowerPoint* presentations (or the equivalent), which should be concise and to the point; ***Shorter is always better***. I will say more about the format of the deliverable when teams are assigned.

   Your PowerPoint presentation should discuss your data sources in detail and how your dataset was constructed. Credit will reflect in part the level of difficulty.[5] In some cases you may be able to obtain data from the original authors, which obviously greatly simplifies (trivializes) the replication phase. But that is not *independent replication*. You can do that if you want, but your grade will suffer mightily (since building datasets is hard work, and copying and pasting is not). If you do work with the authors' data and/or programs, be sure to give them full credit for such. Not doing so is plagiarism.

   At the end of the semester teammates will assess their own and each other's performance using the posted *Peer Evaluation* form (see posting to Canvas). Students' grades will reflect both their individual performance as well as the quality of the final team submission.

   - ***Independent Replication:*** Due around Fri., April 29th ; the next to last week of classes, and about three class-weeks after *Topic selection*.

Leave plenty of time for *Replication*. You'll find this far more challenging and time consuming than you could ever imagine. And yes, ***independent replication means independent replication!***

Shirkers take notice... I repeat: *Peer evaluation* forms will be distributed at the end of the semester, so that team members can assess each other's performance… as well as their own.

Empirical work is slow going. Be sure to leave yourself enough time to complete the assignment to your satisfaction.

Exercises/qFlips/Quizzes (10% of your course grade; grades on Exercises/qFlips/Quizzes are curved):

***Exercises*:** There will be four or so empirical exercises, which together with *qFlips* and Quizzes (see following) count towards 10% of your course grade. Exercises will be team assignments, usually with two students per team, and are graded on a 10 point scale. I will assign teams, which will differ from Exercise to Exercise. Answers are submitted online (one submission set per team; as well, one set of snapshots uploaded to Canvas).

[5] If you want a sense of *degree of difficulty*, just ask.

In some cases, the Exercises are designed to give you practice with the techniques and tools we have developed in the course… other times, they are designed to introduce you to new material, which we have not yet covered in the course. These will take some time to complete, so please do not leave them until the last moment. A good rule of thumb is that Exercises will take about a week to complete… so budget your time accordingly (no sympathy for teams getting a late start).

***qFlips*:** We will have about a half dozen custom tailored *quickFlips* this semester. They will typically be self-paced *online* assignments, and are designed to give you some rudimentary practice with concepts and applications that we are covering in the course. Some of these will be team assignments; I will assign teams, which will typically have two members.

***Quizzes*:** There will also be three or so short Canvas quizzes over the course of the semester. These will be multiple choice exams and typically feature about eight questions. Quizzes are to be completed individually (though they are open book; open notes; open lifelines; open etc etc).

Each qFlip and Quiz is graded on a two point scale ($1/5^{th}$ the value of an Exercise). I anticipate close to perfect scores on these as they are primarily designed to reinforce the learning of the Unit material (you have an unlimited opportunity to revise/update your qFlip answers).

Grading: I will drop your lowest qFlip/Quiz score, just in case you inadvertently miss a deadline; but do not miss two deadlines! These scores (max: two points each) are then added to your Exercise scores (max: ten points each), and the total is scaled up to a maximum of 100.

Course Topics (I have posted (see Canvas) course notes/handouts for each Unit… and slideshows and videos too! As well, you may want to consult the Wooldridge text.)

Introduction

Unit 1 – Introduction & Getting Started: Estimating the relationship between x and y; causality v. correlation; data types; economic v. statistical significance; robust analysis (how many regressions did you run?); art v. science; sample statistics (sample means, variances, standard deviations, covariances, and correlations); standardizing data; OLS as minimization of SSRs (FOCs and SOCs)

Simple Linear Regression (OLS/SLR) Models

Unit 2 – OLS/SLR analytics (single explanatory variable): *In the beginning* (SLR.1: the data generation process); residuals and sum squared residuals (SSR); OLS, FOCs and SOCs, Sample Regression Function (SRF), predictions and residuals; economic significance/meaningfulness (elasticity and *beta* regressions)

Unit 3 – OLS/SLR assessment: Sum Squared Explained (SSE) and Sum Squared Total (SST); SST = SSE + SSR (w/ constant term in the model); Goodness of Fit (GOF) metrics - Coefficient of Determination (R^2), Mean Squared Error (MSE) and Root MSE (RMSE); comparing SLR models using GOF metrics

Boston College
econ2228 S22: Econometric Methods

Multiple Linear Regression (MLR) Models

Unit 4 – OLS/MLR analytics I (adding, and subtracting, explanatory variables): Comparing SLR and MLR analytics; interpreting coefficients I – *ceteris paribus* (partial effects and the SRF); interpreting coefficients II – partial correlations (*WhatsLeft* and W*hatsNew)*; an overview of omitted variable bias (endogeneity)

Unit 5 – OLSMLR assessment: Comparing SLR and MLR assessment (GOF metrics); shortcomings of R^2; adjusted R^2; comparing MLR models using GOF metrics

Unit 6 – OLS/MLR analytics II: The *collinearity* regression; multicollinearity, R^2_j's and Variance Inflation Factors (VIFs); Omitted variable bias/impact (endogeneity); simple v. partial correlations

Mid-Term Exam #1 (about here)

***Way-Too-Fast* Review of Statistics**

Unit 7 – Review of Estimation and Inference: Our focus will be on estimation of the population mean; LUEs (Linear Unbiased Estimators); BLUEs (Best Linear Unbiased Estimators); point and interval estimators; standard errors, t statistics, p-values; confidence levels; critical values; confidence intervals; hypothesis testing; significance levels

Estimation and Inference in Regression Analysis

Unit 8 – SLR Estimation: Gauss-Markov assumptions (SLR.1 – SLR.5); Population Regression Function (PRF); conditional means; means, variances, standard deviations and standard errors of OLS estimators (intercepts and slopes); unbiasedness (OLS coefficients; MSE); LUEs; homoskedasticity; *BLUE*: The Gauss-Markov Theorem

Unit 8a – Heteroskedasticity: Issues (OLS standard errors no longer correct; LUE but not BLUE); White-corrected standard errors (*robust* inference); working towards BLUE (weighted least squares… but where do those weights come from?)

Unit 9 – SLR Inference: Add SLR.6 to the mix; normally distributed errors; variances, standard deviations and standard errors; t statistics; t-tests (Null hypotheses); p values; confidence intervals; hypothesis tests; economic v. statistical significance (elasticities v. p-values); *Convergence I* (t stats and R^2)

Unit 10 – MLR Estimation and Inference: Compare to SLR; *What's new? … Not much!*; now MLR.1-MLR.5; n-k-1; multicollinearity, standard errors and Variance Inflation Factors (VIFs); MLR.6; heteroskedasticity and *robust* standard errors

Topics

Unit 11 – Dummy Variables and Fixed Effects: Dummies revisited; on the RHS and on the LHS; uses on the RHS (slope and intercept dummies); quieting the endogeneity critics (fixed effects); Examples (sovereign debt ratings, gender bias in wages, and death penalty deterrence)

Unit 12 – F-tests and Convergence: Extension of t-tests to more complicated null hypotheses; testing linear restrictions with the F(q, n-k-1) distribution; *Convergence II*: connects Goodness-

of-Fit metrics and inference stats (t stats and incremental R^2, SSR and SSE); reported F stats (for the regression) and associated p values; relation to adjusted R^2; *Babies and bathwater*; Chow tests; machine learning

Mid-Term Exam #2 (about here)

Further Topics

Unit 13 – Linear Probability Models (LPMs) and Functional Forms: Dummies on the LHS; Linear Probability Models (LPMs); exploring functional forms (quantile dummies; linear splines; logarithms and exponentials; polynomials; cubic splines, and fixed effects)

Unit 14 – Further Topics I: *Differences-in-Differences* (*Deflategate*; NBA Referee Own-Race Bias); *Regression Discontinuity Designs* (Highway Fatalities & Daylight Savings Time); *Instrumental Variables* (The Oregon Health Insurance Experiment (Medicaid)); Maximum Likelihood Estimation (MLE); limited dependent variables; logit and probit models; censored and truncated regression models

Unit 15 – Further Topics II: OLS asymptotics (large sample properties; consistency; convergence in distribution); misspecified models; proxy variables; missing data; outliers; non-random samples; forecasting and *prediction* intervals;

Stata @ Boston College

There are a large number of statistical software packages that you can use to do econometric analysis. We will use Stata, one of the more popular packages and the package that receives the most support at Boston College. If time permits, we may also work a bit in SAS, R and/or Python.

I will be providing more details as the semester develops, but for now: Stata is available to BC students through the "application server", known as the *apps server*.. In the past you've needed Citrix Receiver installed on your computer to access the apps server (and if you were not connected to the BC network, you also needed to use VPN to access the apps server). But those days are over. You can now use the "Light Version" of Citrix Receiver to directly access the apps server (no need to install the Citrix Receiver or VPN.) To learn more, go to: https://www.bc.edu/content/bc-web/offices/its/support/software-hardware/apps.html .

Alternatively, and to avoid traffic jams with Citrix and the apps server, you may want to purchase a six-month *Stata/BE* license for $48. For details, go to: https://www.stata.com/order/new/edu/profplus/student-pricing/

We will devote some time to learning how to use Stata to build datasets and run regressions. You will discover that building datasets is long, hard, tedious and unrewarding work… and running regressions is relatively quick and easy… and a lot more fun!

Stata Resources

As the semester progresses, you may find the following resources of interest:

- Encountering Stata questions/issues/features?… just *Google* it (always include "UCLA").

And here are a few sites that might be helpful (the pdfs are posted to the course website):

- http://fmwww.bc.edu/GStat/docs/StataIntro.pdf
- https://stats.idre.ucla.edu/stata/modules/
- http://dss.princeton.edu/training/StataTutorial.pdf
- https://stats.idre.ucla.edu/stata/
- … and don't forget YouTube: https://www.youtube.com/results?search_query=stata

Examples and datasets (**bcuse** may be helpful here… I'll explain in class):

- http://fmwww.bc.edu/gstat/examples/wooldridge/wooldridge.html
- http://fmwww.bc.edu/ec-p/data/wooldridge/datasets.list.html
- http://fmwww.bc.edu/ec-p/data/ecfindata.php (link down at the moment)
- https://stats.idre.ucla.edu/other/dae/
- https://stats.idre.ucla.edu/other/annotatedoutput/

Also: Ben Lambert's *full course in econometrics* videos are terrific and come with a *British accent*!

- https://www.youtube.com/user/SpartacanUsuals/playlists