# Modern Methods for Applied Regression
# PPHA41430 - **DRAFT**

August 22, 2024

Instructor: Guillaume A. Pouliot
Teaching Assistant: Jake Nicoll
Lecture room and time: TBD
Section room and time: TBD

## Course Description

The course aims to take a mature and modern take on regression methods. The starting point is that standard reduced form methods such as ordinary least-squares (OLS) are almost always misspecified in practice. For instance, textbooks and research material often refer to OLS as estimating the conditional expectation, but this is only true if the conditional expectation is linear, which it almost never is in practice.

From that point, two avenues can be taken. A first avenue is to ask whether the estimator is still meaningful. In particular, we can ask whether the OLS still estimate, in some sense, a best approximation to the (nonlinear) conditional expectation, even though it cannot estimate the conditional expectation itself. We will see that this is sometimes the case, and the course will give a careful treatment of estimators with such a best approximation property.

A second avenue is to ask whether a more flexible estimator may in fact recover the desired estimand. This is where machine learning methods "come to the rescue". In the case of reduced form methods, recent development in double/debiased machine learning (DML) have allowed is to tackle such modeling issue using modern machine learning methods off the shelf. We will cover deep learning methods "from scratch" and use them to recover correct specification in reduced form estimation.

Regression analysis requires inference. We cover three fascinating topics in inference. First, we cover Markov chain Monte Carlo Methods (MCMC), treating an host of methods

using the Metropolis-Hastings algorithm as a unifying concept and algorithm. Second, we explore gains in robustness from using randomization-based inference methods –on observational data– to carry out inference with reduced form regression methods such as OLS. Finally, we consider approaches to producing confidence intervals which may be considered simultaneously without invalidating their coverage guarantees.

Finally, we cover reinforcement learning, and pay special attention to connections with econometric and statistical questions.

## Background and Goals

Background in linear algebra, probability, statistics and econometrics. Students who have successfully completed a first undergraduate course in econometrics are considered to have sufficient background.

The goal of the course is to provide students with a firm conceptual understanding of the reduced form regression methods and statistical inference, as well as practical experience with a selection of corresponding statistical and econometric methods.

## Grades

Students will be evaluated based on participation, weekly problem sets, and a final presentation.

- Participation: 15%
- Weekly problem set: 70%
- Final presentation: 15%

For the final presentation, students must collect or create a dataset and carry out and present a regression analysis of the data.

## Timeline

**Topics in estimation**

Week 1 Linear Regression for grown ups

Ordinary linear regression (OLS) is the workhorse of reduced form regression analysis. We study its meaning when the conditional expectation, the oft cited estimand of linear regression, it itself nonlinear and thus cannot be estimated (OLS).

readings: "Linear Regression for Grown Ups" handout; Mostly Harmless Econometrics, Chapter 3 up to and including 3.1.2.; Gary Chamberlain Lecture Notes, Lecture note 1

Week 2 Quantile regression

While OLS allows a characterization or approximation of the conditional expectation, quantile regression allows for the study of the entire conditional distribution. We study quantile regression and pay special attention to misspecification, inference, and computation.

readings: "Modern Introduction to Quantile Regression" handout

Week 3 Instrumental variables

**Topics at the intersection of estimation econometrics and Machine Learning**

Week 4 Backgrounder on machine learning

Week 5 Machine learning solutions for reduced form regression problems

**Topics in inference**

Week 6 Markov chain Monte Carlo

Week 7 Randomization-based inference

Week 8 Simultaneously valid inference

**Topics in inference**

Week 9 Reinforcement learning

**Student presentations**

Week 10 Student presentations

# References

Angrist, J.D. and Pischke, J.S., 2009. Mostly harmless econometrics: An empiricist's companion. Princeton university press.

Chamberlain, Gary. Lecture Notes.

https://github.com/paulgp/GaryChamberlainLectureNotes/tree/main/RawLectures