



**15.034 'Metrics for Managers:  
Big Data and Better Answers**

**Fall 2016**

**DRAFT** Course Syllabus

*Faculty:*

**Professor Joseph Doyle**  
E62-516 [jjdoyle@mit.edu](mailto:jjdoyle@mit.edu)

**Professor Roberto Rigobon**  
E62-515 [rigobon@mit.edu](mailto:rigobon@mit.edu)

*Teaching Assistant:*

**Mert Demirer** [mdemirer@mit.edu](mailto:mdemirer@mit.edu)



## Course Goals

This course enables students to transform data into knowledge to make better business decisions. Whether students anticipate producing or consuming data analytics, the course provides the tools to identify problems *and* a framework to solve them. That framework is **econometrics** (or ‘metrics as we like to call it): modern empirical methods used by economists.

How will you learn these tools? We will introduce the topics and then use in-class labs so students are **learning by doing**. To complement the in-class learning, we will also have short problem sets and group projects. At the end of the course, students will be able to evaluate the quality of evidence supported by data, and implement a whole toolkit aimed at providing credible answers to difficult questions.

Specifically, we will develop the foundations of econometrics and regression analysis. We offer some theory, but mostly practical examples including: forecasting using high-frequency financial data, developing a trading model, running experiments to test the effects of pricing, product quality, and marketing decisions, and discovering “natural” experiments that allow analysts to explore Big Data to go beyond correlations. One workhorse model we will introduce is instrumental-variable estimation.

**Big Data** is a hot topic these days, and we will explore the potential opportunities and pitfalls that arise with the flood of data currently being generated.

## Course Complements & Pre-Requisites

There are no pre-requisites for the course, as we will define terms as we go. The course is complementary to other data-analysis courses at the school, such as Analytics Lab, Statistical Consulting, Analytics Edge and Data Mining; we believe the more you see these concepts the better you will be at using them. What distinguishes this course is that we view these statistical concepts through the framework provided by economics and econometrics: a way of modeling the empirical problem to evaluate correlations and “get to causality”.

## Readings

We will rely on articles posted on Stellar, plus book chapters in the Course Reader available at study.net, including textbook treatments of the main topics. For those interested in a textbook, you will see from the readings that we are partial to Stock and Watson’s *Introduction to Econometrics*, Angrist & Pischke’s *Mastering Metrics* and Wooldridge’s *Introductory Econometrics*. Another useful text is Greene’s *Econometric Analysis*. We will have these textbooks on reserve at Dewey Library.



### Recitations/Office Hours

We will hold weekly recitations to aid you in learning how to use STATA and answer questions about the course and the projects. These are optional but we have found them to be appreciated by students. We aim to have them on Fridays. We will also offer office hours, including on Monday afternoons.

### Evaluation:

Grading will be determined as follows:

Class Participation	20%
Problem Sets	20%
<u>Team Projects</u>	
Data Exploration Project	10%
034 Hedge Fund Project	25%
034 Consulting Project	25%

Problem sets will be designed to be brief, but they will enable you to become more competent with statistical programming for the projects and internalize the lessons from the course.

For the team projects, students may join groups of up to five people. The first project is a data exploration exercise using cutting-edge data used to understand the extent to which management matters for firm performance.

The second project will use [Billion Prices Project](#) data to predict official measures of inflation and build a trading model using these predictions (the 034 Hedge Fund!).

The third project asks you to investigate whether core teams (study groups) matter for performance at Sloan and beyond. We will provide de-identified data including admissions criteria (e.g. GMAT scores), performance at Sloan (e.g. GPA), and exit survey results (including salary and industry). This is a unique dataset that you can analyze: you are the experts on how such study groups function. The answers can help shape core-team design (034 Consulting!).

We welcome suggestions to substitute the above projects with a question and data of your own choosing. The data would need to be accessible to the class and we would determine if it is appropriate for a semester-long course.



## Software

As part of the course you will learn to use R, which is a straightforward but powerful statistical software package. This is a valuable tool to go well beyond Excel in your analysis of data. This will be the basis for in-class labs as well as problem sets and projects.

R is free software and we will post how to download it (and what packages to include) prior to the start of the semester.



Course **DRAFT** Outline

*Sept 8*    *Thurs*    **Course Introduction: Forecasting, Causal Effects, and Models in the Age of Big Data**

Read: The Parable of Google Flu

*Sept 12*   *Mon*    **Recitation: STATA Primer**

*Sept 13*   *Tues*    **Regression Basics 1: Introduction**

Read: Stock & Watson p111-122

Due: Problem Set 1

*Sept 15*   *Thurs*   **Regression Basics 2: Key Assumptions & Interpretation**

Read: Stock & Watson p126-131; Wooldridge p83-88

Read: Stock & Watson p158-160

*Sept 19*   *Mon*    **Recitation**

*Sept 20*   *Tues*    **Regression Basics 3: Omitted Variable Bias**

Read: Wooldridge p88-92

Due: Problem Set 2

*Sept 22*   *Thurs*   **Regression Basics 4: Multivariate Regression**

Read: Wooldridge p69-81

Read: Stock & Watson p123-126; Wooldridge p80-84



- Sept 26** *Mon* **Recitation**
- Sept 27** *Tues* **Regression Basics 5: Statistical Inference**  
Read: Wooldridge p83-88, 123-131  
Due: Problem Set 3
- Sept 29** *Thurs* **Application: Randomized Experiments**  
Read: The Discipline of Business Experimentation, *HBR*  
SuperCrunchers p46-63  
Victory Lab p1-13
- Oct 3** *Mon* **Recitation**
- Oct 4** *Tues* **Application: Advertising Experiments**  
Review: Urban et al. "Morphing Banner Advertising"  
Price Stickiness and Customer Antagonism  
Due: Problem Set 4
- Oct 6** *Thurs* **Time Series Forecasting**  
Read: Signal and the Noise p337-352  
Anderson Chapter 6
- Oct 10** *Mon* **No Recitation**
- Oct 11** *Tues* **No Class: Holiday**
- Oct 13** *Thurs* **Big Data Tools 1: Forecasting**  
Watch: Machine Learning 1 Video on Stellar (Minute 31-end)
- Oct 18** *Tues* **Big Data Tools 2**  
Due: Problem Set 5
- Oct 20** *Thurs* **Big Data & Billion Prices**



- Oct 24 Mon SLOAN INNOVATION PERIOD: NO RECITATION**
- Oct 25 Tues SLOAN INNOVATION PERIOD: NO CLASS**
- Oct 27 Thurs SLOAN INNOVATION PERIOD: NO CLASS**
- Oct 31 Mon No Recitation**
- Oct 31 Mon Due: Write up for: Does Management Style Affect Profits?**
- Nov 1 Tues Group Project 1: Team Presentations**
- Nov 3 Thurs Introduction to Freakonomics**  
Read: *Freakonomics* (excerpt)
- Nov 7 Mon Recitation**
- Nov 8 Tues Panel Data: Fixed & Random Effects**  
Read: Wooldridge p484-491
- Nov 10 Thurs Natural Experiments 1: Difference-in-Differences**  
Read: Mostly Harmless Econometrics: Ch1-2  
Skim: Vertical Relationships and Competition in Retail Gasoline Markets
- Nov 14 Mon Recitation**
- Nov 15 Tues Natural Experiments 2: Finding Random Variation**  
Skim: Returns to Physician Human Capital
- Nov 17 Thurs Natural Experiments 3: Regression Discontinuity & Event Studies**  
Skim: After Midnight



*Nov 18* *Fri* **DUE: Group Project 2: Part 1**

*Nov 21* *Mon* **DUE: Group Project 2: Part 2**

*Nov 22* *Tues* **Group Project 2: Team Presentations**

*Nov 24* *Thurs* **No Class: Holiday**

*Nov 28* *Mon* **No Recitation**

*Nov 29* *Tues* **Instrumental Variables 1: Introduction**

Read: Wooldridge p512-521; 528-530

Instrumental Variables and the Search for Identification

Skim: Evidence of a Shift in the Short-Run Price Elasticity of Gasoline Demand

*Dec 1* *Thurs* **Applications: Measuring the Effects of Walmart**

Read: Walmart and Local Economic Development

*Dec 7* *Mon* **Recitation**

*Dec 6* *Tues* **Probit/Logit & Selection Models**

Read: Stock & Watson: Chapter 11, p383-396, 400-408

Due: Problem Set 6

*Dec 8* *Thurs* **Lessons**

*Dec 12* *Mon* **No Recitation**

*Dec 12* *Mon* **Due: Write up for Group Project 3**

*Dec 13* *Tues* **Group Project 3: Team Presentations**





### **STUDENT SUPPORT SERVICES**

If you are dealing with a personal or medical issue that is impacting your ability to attend class, complete work, or take an exam, please discuss this with Student Support Services (S<sup>3</sup>). The deans in S<sup>3</sup> will verify your situation, and then discuss with you how to address the missed work. Students will not be excused from coursework without verification from Student Support Services. You may consult with Student Support Services in 5-104 or at 617-253-4861. Also, S<sup>3</sup> has walk-in hours Monday-Friday 9:00 -10:00am.

### **STUDENT DISABILITY SERVICES**

MIT is committed to the principle of equal access. Students who need disability accommodations are encouraged to speak with Kathleen Monagle, Associate Dean, prior to or early in the semester so that accommodation requests can be evaluated and addressed in a timely fashion. Even if you are not planning to use accommodations, it is recommended that you meet with SDS staff to familiarize yourself with the services and resources of the office. You may also consult with Student Disability Services in 5-104 or at 617-253-1674. If you have already been approved for accommodations, please contact your professor early in the semester so that we can work together to get your accommodation logistics in place.