ECON 4984: Big Data Economics

Virginia Tech, Spring 2019

Instructor:	Ali Habibnia	**Time:**	Tues/Thurs, 2 – 3:15pm
Email:	habibnia@vt.edu	**Location:**	Davidson Hall 201
Office:	Pamplin Hall 3055	**Prerequisites:**	ECON 2005 and 2006
Office Hours:	Thursday 4:00-5:00pm	**Final Exam:**	May 13, 2019 4:25pm

Course Pages: All course materials will be on Canvas.

Office Hours: Thursday 4:00-5:00pm, or by appointment, or post your questions in the page provided for this purpose on Canvas. **When emailing, please include "ECON 4984: Big Data" in your email header so I can keep track of responses more easily.

Teaching Assistant: He Jiang (jhe8@vt.edu), Tuesday 3:30-4:30pm @ Pamplin 3117.

Required Text: There is no textbook per se for this class; the materials will come from lecture notes and selected papers and book chapters. This will be discussed in class.

Optional Texts: This is a restricted list of interesting and useful books that will be touched during the course. You need to consult them occasionally.

- Trevor Hastie, Robert Tibshirani, Jerome Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*,Springer, 2017. (Available Online)
- Graham Elliott, and Allan Timmermann, *Economic Forecasting*, Princeton University Press, 2016.
- Shai Shalev-Shwartz, and Shai Ben-David, *Understanding Machine Learning From Theory to Algorithms*, Cambridge University Press, 2014. (Available Online)

Prerequisites: ECON 2005, ECON 2006. An undergraduate-level understanding of probability, statistics, linear algebra, and regression is assumed. Having taken econometrics or taking econometrics concurrently is strongly recommended.

Objectives: This course covers the theoretical, computational and statistical underpinnings of the big data analysis. The focus will be the econometric models and machine learning techniques to analyze the high-dimensional data sets a.k.a. Big Data and their implications in research focusing on interesting economic questions that arise from considering the rapid changes in data availability and computational technology. Big data econometric models provide a vehicle for modeling and analyzing complex phenomena and for incorporating rich sources of confounding information into economic models. The goal of this course is to give an applied, hands-on introduction to these methods. At the end of the course, students will be able to read and understand theoretical papers on the subject, to implement the techniques themselves in Python, and to apply the techniques to data used in economics and business.

Software: We will be using statistical software in this class. You are welcome to use the software of your choice, but class demonstrations will be in `Python`. I will provide a gentle introduction to programming using Python in the class. I will also show you how to use Google Colaboratory, Googles free cloud service for developers. With Colab, you can develop Python codes on Google servers that require no setup and runs entirely in the cloud. Please, bear in mind that all help with software in office hours will be limited to `Python`.

Tentative Course Outline:

- Python Crash Course; (NumPy, SciPy, pandas, matplotlib, scikit-learn, PyTorch)
- Fundamentals of Linear Algebra and Optimization
- Refresher on Regression Analysis; (Matrix Formulation, OLS, MLE, Logistic & Polynomial Regression)
- Big Data and the Curse of Dimensionality in Economics and Finance
- Regression with Many Regressors: Standard Approaches to Model Selection Algorithms
- Penalized Regression Methods: Lasso, Ridge, and Elastic Net
- Linear Dimensionality Reduction with an Emphasis on PCA
- Factor Models; Estimation and Inference
- Prediction with a Large Number of Covariates ("Big P")
- Analysis with Large Sample Sizes ("Big N")
- High-dimensional Methods and Inference on Structural and Treatment Effects
- A Brief Introduction to Bayesian Inference and Bayesian VARs
- Nonlinearity in Big Data Sets and Nonlinear Dimensionality Reduction
- Neural Networks and Deep Learning for Big Data Analysis
- Spark and Python for Big Data with PySpark

Grading Policy:

Assignments ..... bi-weekly - (25%)
Midterm ..... roughly half-way through - (25%)
Final Project ..... (20%)
Final Exam ..... cumulative - (30%)

Grading will nominally follow the typical breakdown on a total percentage scale, e.g., [93-100 A), [90-93 A-), [87-90 B+), [83-87 B), etc. All grades in Canvas will follow this scheme. However the instructor reserves the right to apply a final curve in the students favor.

Assignments

Class assignments are due at 5:00pm on the due date, and no late assignments will be accepted. Students are welcome to collaborate with one another, but are required to submit their own work as well as be able to reproduce it. All work must be shown and software must be used, when appropriate, with attached software output. If there is a truly extenuating circumstance requiring an extension, please email me in advance and let me know as soon as possible.

Exams

There will be two written exams. Midterm exam on Thursday March 7th (Tentative date), and Final Exam on May 13th. All in-class exams will be closed book and closed notes. Make-up exams will be offered for students who have well-documented emergencies approved by the instructor or reported in advance.

Final Project

The final project will be on a topic of your choice but relating to the lectures. You should manage to submit the first draft by the end of March to get my feedback on your projects. Deadline for the final version will be end of the semester.

Academic Honesty: Lack of knowledge of the academic honesty policy is not a reasonable explanation for a violation. The Undergraduate Honor Code pledge that each member of the university community agrees to abide by states:

As a Hokie, I will conduct myself with honor and integrity at all times. I will not lie, cheat, or steal, nor will I accept the actions of those who do. Students enrolled in this course are responsible for abiding by the Honor Code. A student who has doubts about how the Honor Code applies to any assignment is responsible for obtaining specific guidance from the course instructor before submitting the assignment for evaluation. Ignorance of the rules does not exclude any member of the University community from the requirements and expectations of the Honor Code.

For additional information about the Honor Code, please visit: https://www.honorsystem.vt.edu/

The Virginia Tech honor code pledge for assignments is as follows: I have neither given nor received unauthorized assistance on this assignment. I will not require you to paste that on your assignment, because that creates a logistical hassle when students forget. Nevertheless, that pledge is applied automatically. The honor code states that In the absence of a written honor pledge, the Honor Code still applies to an assignment.

Services for students with disabilities: Any student who feels that he or she may need an accommodation because of a disability (learning disability, attention deficit disorder, psychological, physical, etc.), please make an appointment to see me during office hours. Students with test-taking accommodations from SSD should contact me at least one week before each exam to make arrangements. You will need to take the exam at the SSD office. Once you take an exam, there is no remedy for a poor grade.

Students' Responsibility: You are strongly encouraged to complete the Student Perceptions of Teaching (SPOT) questionnaire. Constructive student feedback is important for enhancing the learning experience in this course. Changes to the class and instruction may result from suggestions that are shared with me. Comments about specific aspects of the course or instruction are most helpful.