(Adv) Big Data Economics

ECON 5314G

Virginia Tech, Spring 2026

Instructor:	Ali Habibnia	**Time:**	TR 5:00–6:15 PM
Email:	habibnia@vt.edu	**Location:**	PAM 3008
Office:	Pamplin Hall 3116	**Prerequisites:**	Please see details below
Office Hours:	TR 12:30–1:30 PM	**Final Exam:**	May 9, 2026

Course Pages: All course materials will be provided on Canvas & my GitHub.

Office Hours: Tuesday and Thursday 12:30-1:30 pm, or by appointment, or post your questions on the page provided for this purpose on Canvas. When emailing, include "Adv Big Data Class" in your email header so I can track responses more easily.

Teaching Assistant: Jamshid Ardalankia (ardalankia@vt.edu), TBC @ Pamplin 3117.

Class Discord Server: An informal Discord server has been created for sharing interesting links, brainstorming final project ideas, and discussing technical or conceptual questions.
Join here: https://discord.com/invite/RPJbwT3u.

Required Text/Materials:

- No official textbook is required; lectures, selected papers, codes, slides, and instructor-provided notes form the core material.
- Students will have full access to DataCamp's online platform for six months (Python, ML, and Stat courses).

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

- Fan, J., Li, R., Zhang, C.-H., and Zou *Statistical Foundations of Data Science*, CRC Press, 2020
- 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)
- Gilbert Strang, *Linear Algebra and Learning from Data*, Wellesley Cambridge Press, 2019.

DataCamp Academic Group: In addition to our regular coursework, I have created an academic group on DataCamp, an interactive online learning platform specializing in data science and programming. This is a supplementary resource to enhance our lecture materials. DataCamp offers coding challenges, tutorials, and projects that are extremely beneficial for learners of all levels, especially in machine learning and programming. Given the diverse backgrounds and skill levels in coding and machine learning among our students, we will provide access to the entire DataCamp platform for 6 months. This will allow you to

explore and practice topics of your interest or those recommended for the final project. Occasionally, I may assign tasks based on DataCamp content, typically involving completing a course that requires about 3-4 hours.
Please use the following invite link to join our academic group: DataCamp Group Invite. It's important to register using your real name to ensure that only enrolled students gain access.

Prerequisites: ECON 3254 Analysis of Economic Data or ECON 4304 Econometric Methods or CMDA 3654 Intro Data Analytics & Visualization or STAT 3006 Statistical Methods. 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 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, implement the techniques themselves in Python, and 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, Google's 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 the software during office hours will be limited to `Python`. Students get free access for an entire semester to the DataCamp platform. DataCamp is an online learning platform that focuses on teaching students the comprehensive skills they need to become successful data scientists.

Tentative Course Outline:

- Preliminaries
  - Overview of Big Data and Big Data Visualization
  - Python Programming & Libraries; (NumPy, SciPy, pandas, matplotlib, scikit-learn, PyTorch)
  - Fundamentals of Linear Algebra and Optimisation for Machine Learning in Python
  - Refresher on Regression Analysis; (Matrix Formulation, OLS, MLE, Logistic & Polynomial)
  - Curse of Dimensionality
- Model Selection and Feature Extraction
  - 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
  - Economic forecasting in a Big Data environment
  - Estimation of Large Covariance and Precision Matrices
  - Feature Selection from an Information-Theoretic Perspective
  - A Brief Introduction to Bayesian Inference and Bayesian VARs

- Deep learning in Big Data Analytics
  - Nonlinearity in Big Data Sets and Nonlinear Dimensionality Reduction
  - Neural Networks and Deep Learning Autoencoders
  - Double Machine Learning for Treatment and Causal Inference

Grading Policy

Assignments	bi-weekly - (20%)
Midterm	roughly halfway through - (20%)
Final Project	(40%)
Final Exam	2nd Half - (20%)

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 12:00 pm 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 exams. The midterm and final exams. Make-up exams will be offered for students who have well-documented emergencies approved by the instructor or reported in advance.

Final Project

- A substantial project involving data analysis, modeling, or method development related to course content.
- An initial proposal or draft is due mid-semester (end of March). Final submission due by semester end.
- Students may work in small teams, but individual contributions must be clearly documented.
- Projects must include reproducible code, documented assumptions, and a clear evaluation protocol.
- Projects should go beyond replication and demonstrate independent analysis. Acceptable projects include model extensions, alternative assumptions, comparative empirical studies, or original applications to economic or financial data.

Policies

Academic Honesty: Lack of knowledge of the academic honesty policy is not a reasonable explanation for a violation. The Undergraduate Honor Code pledge 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 are responsible for abiding by the Honor Code. A student unsure of how the Honor Code applies to any assignment must obtain guidance from the course instructor before submission. Ignorance of the rules does not exclude any member of the University 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."* The instructor will not require you to paste that into each assignment. Nevertheless, the pledge applies automatically.

Services for Students with Disabilities: Any student who needs accommodation because of a disability (learning disability, ADD, psychological, physical, etc.) should make an appointment to see the instructor during office hours. Students with test-taking accommodations from SSD should contact the instructor at least one week before each exam to make arrangements. Once you take an exam, there is no remedy for a poor grade.

Students' Responsibility: All students are strongly encouraged to complete the Student Perceptions of Teaching (SPOT) questionnaire. Constructive feedback is vital for enhancing the learning experience in this course. Comments about specific aspects of the class or instruction are especially helpful.