Data Science for Quantitative Finance

ECON 4984 /CMDA 4984

Virginia Tech, Spring 2025

Instructor:	Ali Habibnia	**Time:**	TR 5:00–6:15 PM
Email:	habibnia@vt.edu	**Location:**	DAV 325
Office:	Pamplin Hall 3116	**Prerequisites:**	See below
Office Hours:	Wednesday 3:30–4:30 PM	**Exam Date:**	May 10, 2025

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

Teaching Assistant: Jamshid Ardalankia ardalankia@vt.edu

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.gg/VUhEHpM4EG.

Course Overview

This course integrates data science techniques with quantitative finance, emphasizing financial modeling, algorithmic trading, portfolio construction, and risk management. Students will learn advanced statistical, econometric, and machine learning methods to analyze financial data, construct high-dimensional portfolios, and implement robust trading strategies in data-rich environments. The theoretical, computational, and mathematical underpinnings of these techniques will be examined thoroughly to provide a comprehensive understanding of modern quantitative finance.

Learning Objectives

Upon successful completion of the course, students will be able to:

- Implement the covered techniques in Python. (Hands-on labs may be combined within lecture or set separately for in-depth coding examples and problem-solving.)
- Understand the key concepts in quantitative finance and algorithmic trading, situating them within a unified theoretical framework.
- Apply statistical and machine learning methods to financial data for forecasting, analysis, and effective risk management.
- Build and optimize portfolios using high-dimensional estimation techniques, avoiding common pitfalls in portfolio applications.
- Develop and test algorithmic trading strategies, including pairs trading, under realistic market assumptions.
- Work proficiently with various sources of financial data and apply advanced estimation techniques in data-rich environments.

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.

Course Materials

Required Text/Materials:

- No official textbook is required; lectures, 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 Finance courses).

References:

- Linton, O. (2019). *Financial Econometrics: Models and Methods*. Cambridge: Cambridge University Press.
- Fan, Jianqing, and Yao, Qiwei. (2003). *The Elements of Financial Econometrics*. Wiley.
- James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical Learning*. Springer.
- Bodie, Z., Kane, A., and Marcus, A. J. (2014). *Investments*, 10th Edition. McGraw-Hill Irwin.
- Elton, E. J., Gruber, M. J., Brown, S. J., and Goetzmann, W. N. *Modern Portfolio Theory and Investment Analysis*. Wiley.
- Litterman, B., and the Quantitative Resources Group (GSAM). *Modern Investment Management*. Wiley.
- Bodie, Z., Kane, A., and Marcus, A. *Investments*. McGraw-Hill Irwin.

Tentative Weekly Topics

Week 1: Introduction to Quantitative Finance

- Overview of financial markets and instruments: equities, currencies, fixed income, derivatives, crypto, commodities.
- Financial mathematics vs. financial econometrics.
- Basics of algorithmic trading and data pipelines.
- Introduction to Python environment and tools for finance.

Week 2: Data Handling and Preprocessing

- Reading financial data from open sources and broker APIs.
- Data cleaning and web scraping for finance.
- Retrieving/visualizing historical and streaming market data.
- Discussion of alternative data (news, social media, satellite imagery) and data ethics.

Week 3: Review of Financial and Mathematical Foundations

- Key financial terms, market microstructure, and conventions.
- Linear algebra and probability review.
- Statistical properties of financial returns (stationarity, heavy tails, volatility clustering).

Week 4: Time Series Analysis: Asset Returns and Volatility

- Asset pricing models, efficiency, random walks.
- Mean-reversion (ARMA) and discrete-state models.
- Volatility modeling (ARCH, GARCH).
- Time series forecasting and applications to algorithmic trading.

Week 5: Probability Distributions in Finance

- Representations of a distribution (pdf, cdf, characteristic function).
- Normal and elliptical distributions.
- Mixture distributions and scenario-based modeling.
- Empirical distributions (MLE and historical simulation).

Week 6: Copulas & Dependence Modeling

- Definition, properties, and special classes of copulas.
- Modeling multivariate dependencies in financial data.
- Implementation and practical considerations (e.g., tail dependencies).
- Applications in risk management and pairs trading.

Week 7: Estimation & Dimensionality Reduction

- Shrinkage methods (LASSO, Elastic Net) and ridge regression.
- Principal component analysis (PCA) and factor analysis.

Week 8: Linear Factor Models & Extensions

- Regression-based factor models and principal component factor models.
- Application to asset returns and risk decomposition.
- Multifactor models for portfolio construction.

Week 9: Machine Learning Foundations in Finance

- Supervised vs. unsupervised learning paradigms.
- Feature engineering and cross-validation for financial data.
- Overfitting, regularization, and model selection.
- Practical pitfalls and best practices (look-ahead bias, data leakage).

Week 10: Advanced Machine Learning & Deep Learning

- Neural networks and deep architectures (MLP, CNN, RNN/LSTM).
- Reinforcement learning for algorithmic trading.
- Interpretability and eXplainable AI (LIME, SHAP).
- Autoencoders.

Week 11: Stochastic Calculus and Option Pricing (Intro)

- Efficiency and Lévy processes.
- Itô calculus fundamentals, Black-Scholes formula.
- Monte Carlo simulations and bootstrap methods.
- Applications to exotic derivatives and risk management.

Week 12: Portfolio Construction and Optimization

- Mean-variance optimization and convex programming.
- Benchmarking and risk-budgeting.
- High-dimensional portfolio simulation and robust estimation (shrinkage Ledoit-Wolf, factor models POET).
- Practical algorithmic solutions with Python.

Week 13: Big Data and HPC in Quant Finance

- Introduction to parallel and distributed computing approaches.
- Cloud computing for large-scale financial data analysis.
- GPU-accelerated libraries for ML and deep learning.

Week 14: Capstone Project

- Project presentations and peer feedback.
- Future directions in data science for finance.

Grading Policy and Assessments

Components and Weights:

- **Assignments/Homeworks (40%)**:
  - Weekly or bi-weekly problem sets involving theory, coding, and data analysis.
  - Some assignments might be based on DataCamp course completions.

  - Collaboration is encouraged, but each student must submit original work.
  - Extenuating circumstances for late submissions must be discussed in advance.

- **Final Project (40%)**:
  - A substantial project involving data analysis, modeling, or strategy 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.

- **Final Exam (20%)**:
  - May include theory-based questions, and problem-solving.

Letter Grade Scale: Grades follow a standard scale, e.g., [93–100: A], [90–93: A-], [87–90: B+], [83–87: B], etc. The instructor reserves the right to apply a final curve in the students' favor.

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.