FINA 6900-001: SPECIAL PROBLEMS DOCTORAL SEMINAR

FALL 2025

Instructor: Professor Amy Cyr-Jones (Cyr)

Email: amy.cyr@unt.edu

Class: Thursdays, 2-4:50pm

Classroom: BLB 285

Office hours: By appointment

Course Overview and Learning Objectives

This course is designed to equip first-year Ph.D. students with the fundamental skills required for empirical research in finance. The focus is on understanding key financial databases, learning how to access and process data, and developing the technical proficiency needed for research assistant (RA) work and independent projects.

Students will gain hands-on experience with proprietary financial datasets, such as CRSP and Compustat, while also exploring free sources of financial data. Additionally, the course will introduce best practices in data cleaning, merging, and statistical analysis using Stata, SAS, and Python.

By the end of the semester, students will be able to efficiently extract, manipulate, and analyze financial data, making them valuable collaborators in empirical research projects. By the end of this course, students will:

1. Develop a foundational understanding of financial research databases and their applications
2. Gain practical experience in accessing and extracting data from WRDS and other financial data sources
3. Learn best practices for cleaning, merging, and managing financial datasets
4. Build proficiency in statistical analysis using Stata, SAS, and Python
5. Understand how AI and machine learning tools can assist in empirical finance research
6. Learn to critically evaluate research methodologies and avoid common pitfalls such as p-hacking and data mining biases
7. Apply their knowledge by replicating key empirical finance studies

Class Format

This course meets once per week for 3 hours. The structure of each class follows a blended approach.

- **Lecture (first half)**: Each session begins with an in-depth lecture covering the week's topic, including theoretical background, database demonstrations, and empirical techniques.
- **Practical Application (second half)**: Students will work on coding exercises, database queries, and case studies to apply what they have learned. This portion will involve group work and live problem-solving. During this time, students can ask questions about the lecture content, assignments, or issues they are encountering in their research.

Course Materials and Class Preparation

There is no textbook for this course, but that does not mean that there is no material that students will be expected to review between class sessions. Papers and associated materials for each week will be available to students through Canvas. You are expected to complete all assigned materials before the start of each class, participate during discussions, and work through hands-on exercises during the sessions.

Grading

Grading Component	Percentage of Overall Grade
Assignments	25%
Exam 1	15%
Exam 2	20%
Final Project	40%
Total	**100%**

Final Project

The ability to replicate and critically analyze empirical research is essential for finance scholars. For your final assignment, you will conduct an empirical research replication project. You will select a published empirical finance paper and replicate its core findings using publicly available data. This process will allow you to gain hands-on experience in empirical research, improve your critical thinking skills, and enhance your ability to reproduce and extend academic studies. You will deliver a written research report summarizing the replication process, findings, challenges, and conclusions.

The final assignment is due **December 11, 2025 at 11:59pm**.

Extra Credit

There is no extra credit in this course and no individual extra credit opportunities will be granted.

Attendance

Research has shown that students who attend class are more likely to be successful. You should attend every class unless you have a university-excused absence such as active military service, a religious holy day, or an official university function as stated in the Student Attendance and Authorized Absences Policy (PDF) (https://policy.unt.edu/policy/06-039). If you cannot attend a class due to an emergency, please let me know.

ADA Accommodation Statement

The University of North Texas makes reasonable accommodations for students with disabilities. To request accommodations, you must first register with the Office of Disability Access (ODA) by completing an application for services and providing documentation to verify your eligibility each semester. Once your eligibility is confirmed, you may request your letter of accommodation. ODA will then email your faculty a letter of reasonable accommodation, initiating a private discussion about your specific needs in the course.

You can request accommodations at any time, but it's important to provide ODA notice to your faculty as early as possible in the semester to avoid delays in implementation. Keep in mind that you must obtain a new letter of accommodation for each semester and meet with each faculty member before accommodations can be implemented in each class. You are strongly encouraged to meet with faculty regarding your accommodations during office hours or by appointment. Faculty have the authority to ask you to discuss your letter during their designated office hours to protect your privacy. For more information and to access resources that can support your needs, refer to the Office of Disability Access website (https://studentaffairs.unt.edu/office-disability-access).

Academic Success Resources

UNT strives to offer a high-quality education in a supportive environment where you can learn, grow, and thrive. As a faculty member, I am committed to supporting you, and I want to remind you that UNT offers a range of mental health and wellness services to help maintain balance and well-being. Utilizing these resources is a proactive way to support your academic and personal success. To explore campus resources designed to support you, check out mental health services

(https://clear.unt.edu/student-support-services-policies), visit unt.edu/success, and explore unt.edu/wellness. To get all your enrollment and student financial-related questions answered, go to scrappysays.unt.edu

Academic Integrity

The University of North Texas promotes the integrity of learning and embraces the core values of trust and honesty. Academic integrity is based on educational principles and procedures that protect the rights of all participants in the educational process and validate the legitimacy of degrees awarded by the University. In the investigation and resolution of allegations of student academic dishonesty, the University's actions are intended to be corrective, educationally sound, fundamentally fair, and based on reliable evidence. Read the full Student Academic Integrity policy at https://policy.unt.edu/policy/06-003.

Generative AI Use

Throughout the semester, you will or may use specific Generative AI (GenAI) tools for certain assignments, with guidance on responsible use. These assignments help build ethical resilience and GenAI literacy, preparing you for careers in a GenAI-oriented workforce. In accordance with the UNT Honor Code, unauthorized use of GenAI tools is prohibited. Using GenAI content without proper credit or substituting your own work with GenAI undermines the learning process and violates academic integrity. If you're unsure whether something is allowed, please seek clarification.

Schedule

Session	Date	Topic	Assignments Due
1	8/21	Introduction to Financial Data and Research Methods	Class 1 Prep • Read Bloomfield, Nelson, Soltes (2016): Gathering Data for Archival, Field, Survey, and Experimental Accounting Research. Submit summary of research methods and an example of each type of data. • Read *Statistical Non-Significance in Empirical Economics* (Abadie, working paper) • Read *Presidential Address: The Scientific Outlook in Financial Economics* (Harvey, 2017)
2	8/28	AI in Research	Class 1 Assignments: • Come up with one empirical research question in your area that can be answered using CRSP/Compustat

			• Analyze 2 firms' data from CRSP and Compustat using Stata and SAS and submit descriptive statistics with code Class 2 Prep Work: • Read chosen finance paper and submit summary and key findings • Read GitHub page for *Hedging the AI Singularity* (Chen, working paper) and note key takeaways • Read *Prompt Engineering* (Boonstra) – Kaggle • Explore UNT's AI resource page
3	9/4	Extracting and Structuring Financial Data	Class 2 Assignments: • Finish and submit the AI-assisted coding and summarization of the Fama/French data • Write an example of both a good and bad prompt for web-scraping and submit output of both prompts with a write-up of the differences • Use AI engine to conduct a literature review on a finance or accounting question and submit the results along with a 1-page analysis of the results. Class 3 Prep Work: • Watch *Merging CRSP and Compustat Data* video on WRDS • Study *Using SAS Studio* on WRDS (slide deck) • Study *Getting Started* on WRDS (slide deck)
4	9/11	Merging Data	Class 3 Assignments: • Submit CRSP/Compustat analysis, visualizations, and write-up Class 4 Prep Work: • Study NYU *Merging Data Sets* graphic and Stata, SAS, and Python code for all merge cases (one-to-one, one-to-many, etc.) • Read *Data Quality Problems Troubling Business and Financial Researchers: A Literature Review and Synthetic Analysis* (Liu, 2020) • Read *SDC M&A User Guide* (WRDS) • Read *WRDS Overview of BoardEx*
5	9/18	Merging Bank Data	Class 4 Assignments: • Submit merge workflow, write-up, final dataset, descriptive statistics, and code (3 points) • **Use Stata:** Mert, Fahimeh, Xuan, Ishtiaq • **Use SAS:** Ajay, Nasim, Paul, Yu Class 5 Prep Work:

			• *To be assigned*
6	9/25	Merging ETF Global with CRSP/Compustat	Class 5 Assignments: • Final merged dataset of Bank Regulatory – Holding Companies Financial with CRSP/Compustat Merged – Bank Annual database • Financial highlights table • Financial highlights table (for combined banks not included in Bank Regulatory database) • **Use Stata:** Ajay, Nasim, Paul, Yu • **Use SAS:** Mert, Fahimeh, Xuan, Ishtiaq Class 6 Prep Work: • Read *Do ETFs Increase Volatility?* (Ben-David et al. 2018) – Introduction and Section I • Read WRDS *Overview of ETFG Research* (located on ETF Global page) • Submit a ½ page write-up of a research question involving ETFs (besides volatility) and how you would answer it using ETF Global data
7	10/2	Factor Models and Asset Pricing	Class 6 Assignments: • Merge ETF Global with Compustat financial variables (Ben-David, et al. 2018) • Estimate ETF ownership for each firm. Submit the estimates along with the code used. Class 7 Prep Work: • Read *Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk* (Sharpe 1964) – Introduction • Read *The Cross-Section of Expected Stock Returns* - Introduction and I. Preliminaries • Read *Common Risk Factors in the Returns on Stocks and Bonds* (Fama and French, 1992) • Read *WRDS Overview Fama-French Portfolios & 3 Factors*
8	10/9	**Exam 1**	Class 7 Assignments: • Construct and submit code and output for market, size, and value factors. Exam Prep: • Study Guide

9	10/16	Cross-Sectional and Time Series Analysis	Class 9 Prep Work: • Read *Time Series for Macroeconomics and Finance* (Cochrane, Chapter 2) • Read *Analyst Prep: Time-series Data vs. Cross-sectional Data* • Read *15.4 HAC Standard Errors* from *Introduction to Econometrics with R*
10	10/23	Working with External Financial Data Sources	Class 9 Assignments: • Estimate return predictability of asset growth following Cooper, Gulen and Schill (2008). Submit results, code, and a ½ page write-up of any roadblocks you encounter. Class 10 Prep Work: • Read *Measuring Economic Policy Uncertainty* (Baker, Bloom, and Davis 2016) • Read *Asset Pricing Explorations for Macroeconomics* (Cochrane and Hansen) – Introduction/Overview and Concluding Remarks
11	10/30	Scraping and Analyzing Text Data from EDGAR (pt 1)	Class 10 Assignments: • Merge Baker, Bloom, and Davis Uncertainty data with CRSP and analyze the impact of uncertainty on asset prices. Submit results and code along with a 1-page write-up of the results. Class 11 Prep Work: • Read *When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks* (Loughran and McDonald, 2011) • Read *Frequently Asked Questions About Form 13F* (sec.gov) – Introduction, Questions 1-3; 6-8; 56 • Read *Investor Bulletin: How to Reda a 10-K* (sec.gov)
12	11/6	Scraping and Analyzing Text Data from EDGAR (pt 2)	Class 11 Assignments: • Combine extracted voting results with CRSP/COMP merged data, AI Data, SEC 13F, and Jon Zytnick - Proxy Advisor Recommendation Data • Submit univariate and multivariate analysis of ownership and voting outcomes as a function of ownerships structure, the issue voted on, and firm characteristics. Class 12 Prep Work: • Read *Introduction to Web Scraping with R* (sscc.wisc.edu, 2025) • Read *Introduction to Natural Language Processing* (geeksforgeeks.org) • Read *What is Sentiment Analysis?* (geeksforgeeks.org)
13	11/13	Review/Catch Up	Class 12 Assignments:

			• Scrape 10-K filings from EDGAR for select firms, conduct basic NLP analysis, and submit findings and code used.
14	11/20	**Exam 2**	Exam Prep: • Study Guide
	11/27	Thanksgiving Break – No Class	
15	12/4	Final Assignment Questions/Work Time	
	12/11		**FINAL ASSIGNMENT DUE**

Schedule is subject to change*. Any revision to the course schedule will be communicated on Canvas.*

UNT is committed to the safety and security of the campus community and recognizes the value of maintaining university operations. However, in the event university operations are disrupted due to an emergency or other hazardous conditions threatening the safety, health, or welfare of the campus community, UNT may be required to cancel classes, suspend operations, or close campus in its entirety. Students will be notified by Eagle Alert if there is a campus closing that will impact a class.