University of Pittsburgh
Joseph M. Katz Graduate School of Business

Finance PhD Student Empirical Research Workshop

Summer 2023

Workshop Description

This workshop is an overview of computing concepts related to empirical research in financial economics. The workshop will introduce participants to best practices for conducting reproducible empirical research. Participants will establish basic programming skills, as a foundation for future research assistance and independent projects. The workshop will provide an overview of research resources used in empirical corporate finance research, as well as provide hands-on applications of data exploration, manipulation, and visualization.

Instructor: Madeline Marco Scanlon

- Office: Mervis Hall, 249
- Workshop Documents: OneDrive folder
- Contact: mms238@pitt.edu

Meeting Information

- Wednesday, 8/16: 5:00pm - 8:00pm (Mervis Hall, Conference Room 101)
- Saturday, 8/19: 12:00pm - 3:00pm (Mervis Hall, Conference Room 101)
- Monday, 8/21: 5:00pm - 8:00pm (Mervis Hall, Conference Room 101)
- Wednesday, 8/23: 5:00pm - 8:00pm (Mervis Hall, Conference Room 101)

Student Learning Objectives

This workshop is designed to assist participants in the following objectives:

- To access and apply commonly used data sources in empirical finance research.
- To practice and conduct reproducible research.
- To develop understanding of statistical software and packages used in empirical finance research.
- To establish skills associated with being a successful research assistant and financial economist.

At the end of the session, participants are responsible for independently completing the Replication Project (detailed in the Workshop Outline). The due date for this deliverable is Wednesday, August 30 by 8:00pm (EST).

Before the First Session

Statistical Software Download and Installation:

The following resources are a mix of licensed and open-source programming languages for statistical computing. Please read and follow the installation instructions available from the hyperlinks provided below. Stata is widely used in empirical economic research, while R can be used for more advanced statistical learning. SAS will be used as a way to access specific databases in WRDS.
If you run into any problems installing the software, please reach out to the Katz IT Help Desk in person (Mervis Hall, Room 109) or via phone at (412) 648-1601.

- **Stata**: A personal copy of Stata can be obtained through the MyPitt Software Download Service. Note: This is an annual license that will need to be renewed one year after installation.
- **R/RStudio**: In order to use R effectively, we need a suitable editing environment, i.e. RStudio[1]. A personal copy of R and RStudio can be obtained from the POSIT website. You must install R (left-hand side) before you install RStudio (right-hand side).
- **SAS**: A personal copy of SAS can be obtained through the MyPitt Software Download Service.

Request Access to Data Sources:

1. **Wharton Research Data Services (WRDS)**: WRDS provides instant access to important databases in the fields of finance, accounting, banking, economics, management, marketing and public policy. To create an account, visit the WRDS registration page.
   While we not have time to cover all of the WRDS offerings, we will explore:
   - COMPUSTAT: COMPUSTAT provides standardized North American and global financial statement (annual/quarterly) and market data for over 80,000 active and inactive publicly traded companies. It also provides aggregates, industry segments, banks, market prices, dividends, and earnings
     - Execucomp: Executive compensation data collected from company's annual proxy (DEF14A SEC form)
   - CRSP: CRSP provides comprehensive security price, return, and volume data for the NYSE, AMEX and NASDAQ stock markets. It also provides stock indices, beta-based and cap-based portfolios, treasury bond and risk-free rates, mutual funds, and real estate data.
   - I/B/E/S (Academic & Guidance): IBES provide both summary and individual analyst forecasts of company earnings, cash flows, and other important financial items, as well as buy-sell-hold recommendations.

   Please take a moment to locate/review the associated Overviews and Manuals provided in WRDS. These will be helpful references for all your data-related questions.
2. **Securities Data Company (SDC)**: SDC contains historical data on M&A, M&A events, M&A advisors, and new issues. Details related to accessing this data will be provided at a later date.
3. **Factiva**: To access Factiva, you must follow this link from a computer attached to the Pitt Oakland campus computer network. *The Pitt IT virtual lab WILL NOT work.*
   - Factiva contains a combination of global content, business search, and monitoring capabilities, functioning as a news aggregator and archive.

   Please take a moment to review the Factiva guidance on search parameters for strings here.

[1]For more context, RStudio is an integrated development environment (IDE) developed specifically for R programming. Although R can be run without RStudio, RStudio provides a more user-friendly experience with additional functionality.

Become Familiar with Coding Syntax:

There are numerous resources available online or in-print that will help you develop your statistical computing skills, however the best way to learn a new language is by working on practical examples. While this workshop will provide an introduction, developing a strategy to independently expand your coding skills will be an invaluable resource throughout your career. In this workshop, we will primarily focus on examples in Stata and R, while we will use SAS for extraction of data from particular WRDS providers.

Learn to Code Resources:

- **Stata**: Stata modules provided by the Institute for Digital Research and Education at UCLA
- **R**: R for Data Science; Modern Data Science with R
- **SAS**: SAS modules provided by the Advanced Research Computing: Statistical Methods and Data Analysis at UCLA

Quick Reference/Cheatsheets:

- **Stata**: https://www.stata.com/bookstore/stata-cheat-sheets/
- **R**: https://posit.co/resources/cheatsheets/?type=posit-cheatsheets/
  tidyfinance: https://www.tidy-finance.org/r/wrds-crsp-and-compustat.html
- **SAS**: https://www.sas.com/content/dam/SAS/documents/technical/certification/exam-content/base-programming-ref-sheet.pdf

Specific Questions/Forums:

- **Stata**: Statalist
- **R**: StackOverflow
- **SAS**: SAS Support Communities

Review Coding Style Guides:

As part of the publication process, journals require the submission of data and code used in analysis, along with sufficient documentation to permit replication. While the following suggestions should not be viewed as prescriptive, developing an easy to follow, sufficiently commented coding style makes collaboration easier.

- **Overall Resources:** Code and Data for the Social Sciences: A Practitioner's Guide
  In addition to providing generalized coding style guidance in the appendix, this is an invaluable resource for learning best practices in data architecture. Examples are written in pseudocode or Stata, but the general principles can apply to any language.
- **Stata**: The Stata Journal: Suggestions on Stata programming style
- **R**: The tidyverse Style Guide
- **SAS**: Guidelines for Coding of SAS Programs

Workshop Outline

Each session will consist of a lecture and a lab. The lecture component will include an overview of a programming language/syntax and an introduction to a new data source. The lab component will allow participants to complete an applied analysis using material from lecture.

Each lab will have an optional reading based on a published paper that uses the session's covered data providers. While the readings are optional, familiarizing yourself with the paper before class will provide context to the analysis and prepare participants for RA assignments.

Papers may be downloaded electronically from the University of Pittsburgh Library System using the citations in the References section at the end of the syllabus. A good habit to encourage early is regularly reviewing published papers from the top journals in finance and economics.

****A laptop with pre-installed software is required for every session.****

Session 1: Wednesday, 8/16

- Introduction to Stata Programming: Data Exploration, Manipulation, and Visualization
- Data Overview: Compustat (with SAS Extraction)/Execucomp
- Lab 1
- Optional Readings: Mitton (2022); Duchin et al. (2017)

Session 2: Saturday, 8/19

- Event Studies: A Basic Primer
- Data Overview: CRSP
- Lab 2
- Optional Readings: Miller (2023); MacKinlay (1997)

Session 3: Monday, 8/21

- Introduction to R Programming: Data Science in tidyverse
- Data Overview: SDC
- Lab 3
- Optional Readings: Harford (2005)

Session 4: Wednesday, 8/23

- Advanced Topics in Data Science: Text as Data, Fuzzy Matching, Web Scraping, and Accessing APIs
- Overview of Additional Computing Resources Available to Researchers
- Data Overview: Factiva and IBES (Academic and Guidance)
- Optional Readings: Ahern and Sosyura (2015)

Replication Project:

At the conclusion of the workshop, participants will be responsible for replicating a published paper in the participant's coding language of choice. While collaboration for discussion is encouraged, each participant must independently submit original code and analysis.

- Replication Paper: Kahle, K. and Stulz, R. M. (2021). Why are corporate payouts so high in the 2000s? *Journal of Financial Economics*, 142(3):1359–1380
- Due Date: Wednesday, August 30 8:00pm (EST)

Best Practices for Research Assistants

- RA assignments are excellent opportunities to engage in the research process under the guidance of faculty. Treat each project as a learning opportunity: You will only get out, what you put in to the experience.
- Develop skills that will make you a valuable potential coauthor: Be productive. Be efficient. Be detail-oriented. Be thorough. Be willing to learn unfamiliar techniques/processes.
- Take the initiative with RA assignments. As an active participant, you have greater opportunity to take advantage of mentorship and feedback available from faculty.
- Meet regularly with your faculty advisor and provide deliverables on schedule. Remember, early interactions with faculty will help later in the program when it comes to appointing a dissertation chair and selecting a committee.
- Be proactive in managing assigned responsibilities, in addition to your other coursework/projects. Keep an open line of communication with your faculty advisor, regarding approaching deadlines/conflicts. This can help manage expectations.

References

Ahern, K. R. and Sosyura, D. (2015). Rumor has it: Sensationalism in financial media. *The Review of Financial Studies*, 28(7):2050–2093.

Duchin, R., Gilbert, T., Harford, J., and Hrdlicka, C. (2017). Precautionary savings with risky assets: When cash is not cash. *The Journal of Finance*, 72(2):793–852.

Harford, J. (2005). What drives merger waves? *Journal of Financial Economics*, 77(3):529–560.

Kahle, K. and Stulz, R. M. (2021). Why are corporate payouts so high in the 2000s? *Journal of Financial Economics*, 142(3):1359–1380.

MacKinlay, A. C. (1997). Event studies in economics and finance. *Journal of Economic Literature*, 35(1):13–39.

Miller, D. L. (2023). An introductory guide to event study models. *Journal of Economic Perspectives*, 37(2):203–30.

Mitton, T. (2022). Methodological variation in empirical corporate finance. *The Review of Financial Studies*, 35(2):527–575.