UNIVERSITY OF MANNHEIM
Business School

CHAIR OF FINANCIAL MARKETS AND FINANCIAL INSTITUTIONS

Master's Theses HWS 2025

Topic S1: Generative AI and Firm Values

Classification: Empirical Topic (Python or Stata)

Advisor: Sabrina Yufang Sun

AI has the potential to revolutionize many industries. This is, however, only possible if AI technologies are successfully adopted. Empirically, firms are not always successful in their AI adoption. According to a 2024 Gallup survey, 80% of the digital transformation initiatives fail. Given the importance of a successful digital transformation, it is important to understand whether and how capital market participants incorporate this factor into their firm valuations.

This thesis will take the perspective of large institutional investors – mutual funds, hedge funds, and pension funds, and investigate how these large institutional investors differentiate winning from losing firms in times of AI transformation. Specifically, the student will

(1) Empirically investigate the stock market reactions of large institutional investors during the launch of a major large language model (LLM)
(2) Empirically investigate the mechanisms behind the institutional investor reactions.
(3) Conduct a thorough review of the relevant literature.

Data and Requirements:

The student working on this topic will work closely with the advisor. Datasets from institutional investors and the stock markets are both provided by the advisor.

The student will use statistical software such as Stata or Python for empirical analyses, hence prior experience with Python or Stata is an advantage, but not a must. Stata and Python beginners are welcome. The advisor will provide some sample code for the student to start with. Also, a basic understanding of econometric foundation is a plus (but not a must).

Introductory Literature:

Eisfeldt, A. L., & Schubert, G. (2024). *AI and Finance* (No. w33076). National Bureau of Economic Research.

Eisfeldt, A. L., Schubert, G., & Zhang, M. B. (2023). *Generative AI and firm values* (No. w31222). National Bureau of Economic Research.

Disclaimer from Advisor:

My co-author and I are currently working on a closely related topic.

Topic S2: The Impact of Large Corporations on the Rules of Capital Markets

Classification: Empirical Topic

Advisor: Sabrina Yufang Sun

Classification: Empirical Topic (ChatGPT + Python)

In today's world, corporate power doesn't just shape markets through its transactions —it shapes the rules of the game. Leveraging ChatGPT as an analysis tool, this thesis explores how major U.S. corporations seek to influence financial regulation through the Securities and Exchange Commission (SEC), the key watchdog of Wall Street. From comment letters to behind-the-scenes meetings, companies use a range of tactics to push their interests. Understanding these strategies offers a fascinating look into the intersection of politics, business, and law. For students interested in finance, policy, or corporate strategy, this topic opens a window into how real-world decisions are made—and who truly holds the power to make them.

Specifically, the student will

(1) Use ChatGPT to analyze the communication records between large US corporations to the SEC.
(2) Use Python scripts to analyze the impact of these communications in capital markets.

Data and Requirements:

The project provides a great opportunity for students to learn Python as a beginner and to acquire or upgrade their skills of using large language models like ChatGPT on finance-related textual data. All textual data necessary for the project will be provided by the advisor.

ChatGPT and Python beginners are welcome. Prior experience with ChatGPT and Python are a plus, but not a must. In case of no prior experience, the student is expected to follow instructions closely and be highly motivated to learn new concepts and tools.

Introductory Literature:

Bertrand, M., Bombardini, M., Fisman, R., Hackinen, B., & Trebbi, F. (2021). Hall of mirrors: Corporate philanthropy and strategic advocacy. *The Quarterly Journal of Economics, 136*(4), 2413-2465.

Bertrand, M., Bombardini, M., Fisman, R., & Trebbi, F. (2020). Tax-exempt lobbying: Corporate philanthropy as a tool for political influence. *American Economic Review, 110*(7), 2065-2102.

DeHaan, E., Kedia, S., Koh, K., & Rajgopal, S. (2015). The revolving door and the SEC's enforcement outcomes: Initial evidence from civil litigation. *Journal of Accounting and Economics*, *60*(2-3), 65-96.

Disclaimer from Advisor:

I am currently working on this topic as part of my dissertation.

Topic S3: Non-Passive Securities Lending by Passive Investors

Classification: Empirical topic

Advisor: Annabelle Bröstl

Passively managed funds, such as index funds, are often viewed as purely mechanical investors whose investment decisions are determined by their benchmark index composition. However, these funds also make other, less visible decisions that can be influenced by conflicts of interest within their fund family. One such decision is securities lending, where funds lend shares to market participants such as hedge funds, thereby generating additional income for their investors.

The paper "Non-Passive Securities Lending by Passive Investors" (Gogar, Haushalter & Pisciotta, 2023) shows that passive funds' lending decisions are shaped not only by market fees and borrowing demand, but also by internal and external costs for their fund families. Internal costs arise when active funds within the same family hold large positions in the same security and lending is perceived to potentially harm its price. External costs occur when the fund family has business relationships (e.g., pension plan management) with the issuing company, creating incentives to avoid enabling short-selling that could strain those relationships. The authors find that such conflicts significantly reduce lending activity, particularly when funds use in-house lending agents who retain a large share of the lending revenues.

The goal of this thesis is twofold. First, the student should replicate the core empirical findings of Gogar et al. (2023). Second, the student should extend the analysis by investigating whether bank-affiliated mutual fund families adjust their securities lending behavior for the stocks of their parent bank—potentially due to reputational concerns or to avoid adverse price effects.

Requirements:

The empirical work requires access to mutual fund holdings and securities lending data (e.g., CRSP, Morningstar and SEC EDGAR). The databases are readily accessible for affiliates of the University of Mannheim. The candidate should feel comfortable working with large datasets using a statistical software program (such as STATA, R, or Python).

Introductory Literature:

Gogar, A., Haushalter, D., & Pisciotta, K. (2023). Non-Passive Securities Lending by Passive Investors. *Unpublished Working Paper.*

Golez, B., & Marin, J. M. (2015). Price support by bank-affiliated mutual funds. *Journal of Financial Economics*, *115*(3), 614-638.

Topic S4: Selecting Mutual Funds with Machine Learning and Fund Characteristics

Classification: Empirical topic

Advisor: Yue Wu

Classic mutual fund literature has shown that average active funds earn negative alpha net of expenses (Jensen, 1968; Gruber, 1996). Though some studies have justified the existence of a subset of managers that outperform their benchmarks (Fama and French, 2010; Berk and Van BinsBergen, 2015), the impersistent performance makes it difficult to identify outperforming funds ex ante.

DeMiguel et al. (2023) apply machine-learning methods to show that exploiting nonlinearities and interactions in the relation between fund characteristics and performance can help to construct tradable long-only portfolios of mutual funds that earn significant out-of-sample alphas net of all costs. The paper focuses on identifying tradable portfolios of funds by using exclusively past data to construct long-only portfolios, which has both academic and practical implications. By comparing the out-of-sample and net-of-costs performance of the portfolios of funds, the paper finds that gradient boosting and random forests can select long-only portfolios of funds that earn statistically significant alphas relative to the FF5 model augmented with momentum. Meanwhile, the equally weighted and asset-weighted portfolios earn negative annual net alphas, consistent with existing evidence that the average active fund underperforms passive benchmarks after costs.

This project aims to help students understand mutual fund characteristics, performances, and the application of machine learning methods on mutual fund investing. The main working tasks are:

1. providing a literature review about mutual fund performance evaluations;
2. replicating the main findings of DeMiguel et al. (2023);
1. extending the model by including the most recent data, and extending the model with the most recent available data.
3. Optionally, you could either use an alternative factor model to re-estimate the alphas in all fund portfolio models (see Table 3 and IA.3), or apply neural networks to see if neural network is a better machine learning method for predicting mutual fund performances (see Table IA.7).

Requirements:

The empirical work requires the use of large databases, i.e., CRSP and Compustat. The databases are readily accessible for affiliates of the University of Mannheim. The candidate should feel comfortable in the use of a statistical software program (such as STATA) and econometric methods. The authors provide their R code at https://data.mendeley.com/datasets/rpgb99m5zy/3: the student can either translate the code into Python code or directly use the R code, but Python is preferred for the machine learning part of the thesis. Subscribing to STATA is not mandatory if the student is comfortable with coding regressions in Python or other programming languages. Prior knowledge in Python and machine learning is a plus but not required.

Introductory Literature:

Berk, J.B. and Van Binsbergen, J.H., 2015. Measuring skill in the mutual fund industry. *Journal of Financial Economics*, *118*(1), pp.1-20.

DeMiguel, V., Gil-Bazo, J., Nogales, F.J. and Santos, A.A., 2023. Machine learning and fund characteristics help to select mutual funds with positive alpha. *Journal of Financial Economics*, *150*(3), p.103737.

Fama, E.F. and French, K.R., 2010. Luck versus skill in the cross-section of mutual fund returns. *The Journal of Finance*, *65*(5), pp.1915-1947.

Gruber, M.J., 1996. Another puzzle: The growth in actively managed mutual funds. *The Journal of Finance*, *51*(3), pp.783-810.

Jensen, M.C., 1968. The performance of mutual funds in the period 1945-1964. *The Journal of finance*, *23*(2), pp.389-416.

Topic S5: Mutual Fund Competition, Managerial Skill, and Alpha Persistence

Classification: Empirical topic

Advisor: Yue Wu

Can mutual fund managers generate positive alpha? Can they do so persistently? For over 70 years, as early as Jensen (1968) and represented by Carhart (1997) and Fama and French (2010), a rich literature has shown that neither funds in aggregate nor individual funds perform better than what would be expected by random chance.

Since the classic diseconomies of scale hypothesis has been continually challenged, Hoberg et al. (2018) propose an alternative microfoundation for understanding mutual fund performance: when competition is high in a given style market, funds are less likely to generate sustained alpha. To identify the competitors of the mutual fund, the paper innovatively creates a 3D space with regard to size, book-to-market ratio, and momentum, and calculates the spatial distance between each fund pair. The outperformance of a given mutual fund is thus defined as "customized peer alpha" (CPA)", and the paper shows that high CPA predicts high future alpha and that the relation is stronger in low-competition markets.

The project aims to help students understand how to measure mutual fund performance and the intuition behind the non-persistence of the performance. The working tasks are:

2. providing a literature review on mutual fund competition and performance;
3. replicating the main results of Hoberg et al. (2018), and extending the model with the most recent available data;
4. Including another 1 or 2 asset pricing factors to upgrade the 3D fund space into a multi-dimensional space. Can you still identify the competitors of the funds? Can you reach similar conclusions to the ones in Hoberg et al. (2018)? Alternatively, you can replace one of the existing dimensions in the 3D fund space with a new dimension and compare the results.

Requirements:

The empirical work requires the use of large databases, i.e., CRSP and Thomson Reuters. The databases are readily accessible for affiliates of the University of Mannheim. The candidate should feel comfortable in the use of a statistical software program (such as STATA or Python) and econometric methods. Subscribing to STATA is not mandatory if the student is comfortable with coding regressions in Python or other programming languages.

Introductory Literature:

Carhart, M.M., 1997. On persistence in mutual fund performance. *The Journal of Finance*, *52*(1), pp.57-82.

Fama, E.F. and French, K.R., 2010. Luck versus skill in the cross-section of mutual fund returns. *The Journal of Finance*, *65*(5), pp.1915-1947.

Hoberg, G., Kumar, N. and Prabhala, N., 2018. Mutual fund competition, managerial skill, and alpha persistence. *The Review of Financial Studies*, *31*(5), pp.1896-1929.

Jensen, M.C., 1968. The performance of mutual funds in the period 1945-1964. *The Journal of finance*, *23*(2), pp.389-416.

* Reading should focus on the paper required to be replicated and empirical part of the other papers. Theoretical models of the listed literature are not mandatory readings.

Topic S6: Category-Framing in Segment Reporting

Classification: Empirical topic

Advisor: Paul Seidel

Berger and Hann (2003) document that the adoption of SFAS 131 fundamentally altered the way U.S. firms report their business segments. By requiring companies to disclose segment information, the standard led to a measurable increase in the number of segments reported and provided investors and analysts with more disaggregated and potentially more informative data. Their study shows that this change enhanced the usefulness of segment disclosures, improving transparency and contributing to a closer alignment between external reporting and internal decision-making structures. This thesis will begin by replicating the central findings of Berger and Hann (2003), focusing on changes in the number and composition of reportable segments around the introduction of SFAS 131 and testing whether segment data indeed became more detailed and informative. The student will reconstruct the empirical setting and verify the robustness of the original results. After establishing this foundation, the project may then extend the analysis in more flexible directions (will be decided jointly by supervisor and student). Possible avenues include exploring how alternative segment-presentation choices—such as the degree of aggregation, the visibility of major expense categories, or the size of residual "other segment items"—affect the way investors perceive firm performance.

Specific tasks:

1. **Replicate the main findings of Berger and Hann (2003)** by reconstructing a pre-/post-SFAS 131 setting, measuring changes in the number and composition of reportable segments, and verifying whether the empirical patterns can be reproduced with standard market and analyst proxies.
2. **Explore one or more archival extensions (optional)** that descriptively compare segment presentation choices across firms and periods, or that benchmark simple market/analyst reactions around notable disclosure changes. The precise angle can be selected to match feasible samples.
3. **Design a small pilot survey (optional)** to gauge whether alternative segment-presentation frames influence perceived performance or basic allocation and valuation judgments.

Requirements and Data:

The student should be comfortable with using statistical software such as Stata or Python for empirical analysis and have a solid econometric foundation. The replication relies on standard financial databases—accessible through the university's subscriptions. The extension may draw on publicly available regulatory filings or involves own data collection.

Introductory Literature:

Berger, P.G. and Hann, R., 2003. The impact of SFAS No. 131 on information and monitoring. *Journal of Accounting Research*, 41(2), pp.163-223.

Tversky, A. and Kahneman, D., 1981. The framing of decisions and the psychology of choice. *Science*, 211(4481), pp.453-458.

Topic S7: Analyst herding and learning

Classification: Empirical topic

Advisor: Paul Seidel

Welch (2000) documents herding and dispersion in analysts' earnings forecasts and links these patterns to forecast accuracy and market responses. This thesis starts from that foundation and, in a second step, explores whether the *structure* of analyst communication helps explain when consensus is accurate and how quickly prices incorporate earnings news. The project sits at the intersection of the analyst literature (e.g., Welch; Hong & Kubik), social learning and information aggregation (DeGroot; Golub & Jackson), investor attention (Da, Engelberg & Gao), and network-views of information diffusion (Ozsoylev et al.).

Specific tasks:

1. **Replication (Welch, 2000).** Reproduce the core herding/dispersion and forecast-accuracy results for U.S. equities using contemporary data and standard event-time methods; document construction choices.

2. **Extension (directional, scoped during the project).** Explore network-related mechanisms around analyst communication and price efficiency—for example, relating periods of higher communication intensity or more concentrated influence to consensus accuracy and post-announcement price incorporation—subject to feasibility and data access. The precise specification will be finalized jointly together with student and supervisor after the replication.

Requirements:

The student should be comfortable working in Python with WRDS data, handling large panels. Familiarity with event studies and basic network concepts is helpful but not mandatory.

Data:

Replication relies on I/B/E/S (analyst EPS forecasts, consensus/dispersion, actuals) and CRSP (daily returns), with optional Compustat fundamentals and Fama–French factors for robustness. For the extension, publicly observable interaction data (e.g., news quotes/hyperlinks or analyst mentions) may be incorporated to form simple communication measures, subject to access and scope.

Introductory Literature:

Welch, I. (2000). Herding among security analysts. *Journal of Financial Economics, 58*(3), 369–396.

Hong, H., & Kubik, J. D. (2003). Analyzing the analysts: Career concerns and biased earnings forecasts. *Journal of Finance, 58*(1), 313–351.

Ozsoylev, H. N., Walden, J., Yavuz, M. D., & Yildiz, M. (2014). Investor networks and information diffusion. *Review of Financial Studies, 27*(5), 1323–1366.