

University of Mannheim · Business School · Chair of
International Finance · 68161 Mannheim

Prof. Dr. Stefan Ruenzi
L9, 1-2
68161 Mannheim
Germany

<http://intfin.bwl.uni-mannheim.de>

Topic R1: Capturing lottery features through aggregation

Advisor: Alexandros Pierides

Topic R2: Luck versus Skill in Mutual Fund Performance

Advisor: Alexandros Pierides

Topic R3: Asset Pricing During the Day and Night

Advisor: Qi Zhang

Topic R4: The Profitability of Stock Return Anomalies in Modern Markets

Advisor: Qi Zhang

Topic R5: Non-Standard AI Errors

Advisor: Paul Seidel

Topic R6: Navigating through the fog

Advisor: Paul Seidel

A general remark:

All topics focus on empirical analyses (as most seminar topics or master's thesis offered in the finance area). The empirical work requires obtaining data and the use of a statistical software program (such as STATA or Python) to process and analyze the data. Data can be retrieved from databases that are typically readily accessible for affiliates of the University of Mannheim or is provided by the advisors. You should be motivated to acquire necessary skills for databases, software, and econometric methods during the writing phase.

Topic R1: Capturing lottery features through aggregation

Classification: Empirical topic

Advisor: Alexandros Pierides

A large literature in behavioral finance argues that investors exhibit a preference for lottery-like stocks, namely securities with a small probability of generating extremely large payoffs. This preference has been linked to several prominent asset pricing anomalies, including the MAX effect, the skewness premium, and the underperformance of speculative stocks. Empirical research in this area relies heavily on proxies intended to capture lottery-like characteristics, such as maximum daily returns, expected skewness, or idiosyncratic skewness. However, the large number of competing measures raises an important question regarding which proxies most effectively capture the lottery feature embedded in stock returns.

In a recent paper, Jiang et al. (2025) examine how different lottery proxies relate to one another and whether aggregating them improves the measurement of lottery characteristics. The authors show that commonly used proxies capture distinct dimensions of lottery-like behavior and therefore contain complementary information. Motivated by this observation, they construct an aggregate lottery measure that combines multiple existing proxies into a unified indicator. Their results demonstrate that the aggregated measure exhibits substantially stronger predictive power for future stock returns than individual lottery proxies alone. The findings suggest that lottery preferences are multidimensional and that combining information across measures provides a more robust characterization of speculative demand in financial markets.

The aim of this thesis is twofold. First, the student is expected to replicate the main empirical findings of Jiang et al. (2025). Second, the student may extend the analysis by examining settings in which lottery preferences are expected to be particularly pronounced. For example, the predictive power of the aggregated lottery measure may be stronger during periods of elevated investor sentiment or among stocks that are predominantly held by retail investors.

Requirements: The empirical work requires the use of large databases, e.g., CRSP/Compustat. These databases are readily accessible to affiliates of the University of Mannheim. The candidate should feel comfortable with the use of a statistical software program (such as STATA) and econometric methods.

Introductory Literature:

- Bali, T. G., Cakici, N., and Whitelaw, R. F. (2011). Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of financial economics*, 99(2):427–446.
- Barberis, N. and Huang, M. (2008). Stocks as lotteries: The implications of probability weighting for security prices. *American Economic Review*, 98(5):2066–2100.
- Jiang, L., Zhou, G., and Zhu, Y. (2025). Which proxy: Capturing lottery feature through aggregation. *Financial Management*, 54(2):331–362.

Topic R2: Luck versus Skill in Mutual Fund Performance

Classification: Empirical topic

Advisor: Alexandros Pierides

A central question in asset management is whether mutual fund performance reflects managerial skill or merely luck. While some funds consistently outperform benchmark models, it is difficult to determine whether such performance arises from genuine investment ability or from random variation in returns. This distinction is important not only for investors allocating capital across funds, but also for understanding market efficiency and the value added by active portfolio management. As the mutual fund industry has grown substantially over recent decades, identifying whether skilled managers exist and whether they can be detected empirically has become a major topic in empirical finance.

In a highly influential paper, Fama and French (2010) examine whether mutual fund returns provide evidence of managerial skill after accounting for common risk factors and sampling variation. Using bootstrap techniques and a large sample of U.S. equity mutual funds, the authors compare the observed distribution of fund alphas to the distribution that would arise purely by chance. Their results show little evidence that the average actively managed mutual fund outperforms benchmark models after fees and expenses. However, they also find that a small group of managers appears to possess genuine skill, particularly among top-performing funds. The findings suggest that while most mutual fund performance can be explained by luck and costs, exceptional managerial ability may still exist in a limited subset of funds.

A related strand of the literature addresses the multiple testing problem that arises when evaluating thousands of mutual funds simultaneously. Barras, Scaillet, and Wermers (2010) apply false discovery rate methods to distinguish genuinely skilled managers from funds that appear successful purely due to chance. Their findings suggest that only a small fraction of mutual funds exhibit true positive skill, while many apparently successful funds can be explained by luck. This approach complements the bootstrap methodology of Fama and French (2010) and has become an important tool in the mutual fund performance literature.

The aim of this thesis is threefold. First, the student is expected to replicate the main empirical findings of Fama and French (2010). Second, the student may investigate whether evidence of managerial skill has changed over time, particularly following the rapid growth of passive investing and ETFs. Increased competition from passive investment products may have altered the prevalence of skilled active managers in the mutual fund industry. Third, the student may examine whether evidence of skill varies across fund characteristics. For example, skilled managers may be more likely to be found among funds with high active share, where managers have greater freedom to deviate from benchmark portfolios and generate alpha.

Requirements: The empirical work requires the use of large databases, e.g., CRSP/Compustat. These databases are readily accessible to affiliates of the University of Mannheim. The candidate should feel comfortable with the use of a statistical software program (such as STATA) and econometric methods.

Introductory Literature:

- Barras, L., Scaillet, O., and Wermers, R. (2010). False discoveries in mutual fund performance: Measuring luck in estimated alphas. *The journal of finance*, 65(1):179–216.

- Berk, J. B. and Green, R. C. (2004). Mutual fund flows and performance in rational markets. *Journal of political economy*, 112(6):1269–1295.
- 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):1915–1947.

Topic R3: Asset Pricing During the Day and Night

Classification: Empirical topic

Advisor: Qi Zhang

Traditional asset pricing theory predicts that assets with higher systematic risk should earn higher expected returns. In the Capital Asset Pricing Model (CAPM), this relationship is captured through market beta, which measures the sensitivity of an asset's return to movements in the overall market. However, Black et al. (1972) demonstrate that the security market line (SML) for US stocks is too flat relative to the CAPM prediction, which shows weak relation between returns and beta.

In a recent study, Hendershott et al. (2020) revisit the risk-return relationship by separating stock returns into overnight (close-to-open) and intraday (open-to-close) components. They show that the relationship between beta and returns differs dramatically depending on whether the market is open or closed. Specifically, high-beta stocks earn significantly higher returns overnight, while they underperform low-beta stocks during the trading day. As a result, the traditional close-to-close beta-return relation appears weak because positive overnight and negative intraday risk premia offset each other.

The aim of this thesis is as follows: first, the student is expected to replicate the main empirical findings of Hendershott et al. (2020). In particular, the student should construct beta-sorted portfolios and examine whether the relation between beta and returns differs between overnight and intraday trading periods. Second, the student should extend the analysis to also cover the most recent years. Third, the student should further explore whether the day-night effect varies across firm characteristics or periods of high market volatility.

Requirements: The empirical work requires the use of large databases, e.g., Compustat/CRSP. The databases are readily accessible to affiliates of the University of Mannheim. The candidate should feel comfortable with the use of a statistical software program (such as STATA) and econometric methods.

Introductory Literature:

- Black, F., Jensen, M. C., & Scholes, M. (1972). The capital asset pricing model: Some empirical tests.
- Frazzini, A., & Pedersen, L. H. (2014). Betting against beta. *Journal of Financial Economics*, 111(1), 1–25.
- Hendershott, T., Livdan, D., & Rösch, D. (2020). Asset pricing: A tale of night and day. *Journal of Financial Economics*, 138(3), 635–662.
- Lou, D., Polk, C., & Skouras, S. (2019). A tug of war: Overnight versus intraday expected returns. *Journal of Financial Economics*, 134(1), 192–213.

Topic R4: The Profitability of Stock Return Anomalies in Modern Markets

Classification: Empirical topic

Advisor: Qi Zhang

A large empirical asset pricing literature documents hundreds of stock return anomalies that appear to predict future returns. These anomalies are typically based on firm characteristics such as valuation ratios, profitability, investment, momentum, or trading activity. While many of these anomalies historically generated economically significant abnormal returns, an important open question is whether anomaly profits remain economically meaningful after accounting for trading frictions, publication effects, and modern market efficiency improvements.

In a recent study, Chen and Velikov (2023) explore the profitability of more than 200 stock return anomalies after incorporating transaction costs, post-publication decay, and changes in market structure. Using CRSP and Compustat data, they show that the majority of anomaly profits become economically small once implementation costs and data-mining concerns are considered. Their findings suggest that many traditional anomaly strategies may no longer provide substantial abnormal returns in modern financial markets.

The aim of this thesis is as follows: first, the student is expected to replicate the main empirical findings of Chen and Velikov (2023). In particular, the student should examine whether anomaly returns remain significant after accounting for transaction costs and post-publication effects. Second, the student should extend the analysis to also cover the most recent years. Third, the student should further investigate whether anomaly profitability differs across firm size groups or liquidity conditions.

Requirements: The empirical work requires the use of large databases, e.g., CRSP/Compustat. These databases are readily accessible to affiliates of the University of Mannheim. The candidate should feel comfortable with the use of a statistical software program (such as STATA or Python) and econometric methods.

Introductory Literature:

- Chen, A. Y., & Velikov, M. (2023). Zeroing in on the expected returns of anomalies. *Journal of Financial and Quantitative Analysis*, 58(3), 968–1004.
- Hou, K., Xue, C., & Zhang, L. (2020). Replicating anomalies. *Review of Financial Studies*, 33(5), 2019–2133.
- Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1), 1–22.
- McLean, R. D., & Pontiff, J. (2016). Does academic research destroy stock return predictability? *Journal of Finance*, 71(1), 5–32.

Topic R5: Non-Standard AI Errors

Classification: Empirical topic

Advisor: Paul Seidel

Recent advances in artificial intelligence have made it increasingly plausible that large parts of empirical research can be delegated to AI agents. Coding agents can collect data, write analysis scripts, make methodological choices, interpret results, and generate research reports. This creates substantial opportunities for productivity in empirical finance, but it also raises a fundamental question: if several AI agents receive the same data and the same research question, do they arrive at the same empirical answer? If not, AI-based research may be subject to an additional source of uncertainty that is not captured by conventional standard errors.

This question is closely related to the literature on nonstandard errors. Menkveld et al. (2024) argue that empirical results vary not only because samples are noisy, but also because researchers make different analytical choices when testing the same hypothesis. They document this by letting 164 research teams test the same hypotheses on the same data and show that researcher-to-researcher variation can be sizeable. In a recent paper, Gao and Xiao (2026) transfer this idea to AI agents. They deploy 150 autonomous Claude Code agents to independently test six hypotheses about market quality trends in NYSE TAQ data for SPY from 2015 to 2024. Their results show that AI agents also exhibit sizeable nonstandard errors, driven by differences in measurement choices, empirical specifications, and model-specific styles.

The aim of this thesis is twofold. First, the student is expected to replicate the main empirical findings of Gao and Xiao (2026). In particular, the student should examine whether independent AI agents produce different estimates when given the same data and research task, and whether this dispersion can be interpreted as AI-based nonstandard error. Second, the student may extend the analysis by studying which design choices account for most of the dispersion, whether different AI models or prompting protocols lead to systematically different empirical results, or whether structured guidance reduces nonstandard errors in AI-generated research.

Requirements: The empirical work requires the use of databases, e.g., CRSP. The databases are widely accessible to affiliates of the University of Mannheim. The candidate should feel comfortable with the use of a statistical software program (such as Python) and econometric methods. Moreover, the student should be familiar with or interested in recent AI tools.

Introductory Literature:

- Gao, R., and Xiao, S. C. (2026). "Nonstandard Errors in AI Agents." arXiv preprint arXiv:2603.16744.
- Menkveld, A. J., Dreber, A., Holzmeister, F., Huber, J., Johannesson, M., Kirchler, M., Razen, M., Weitzel, U., Declerck, F., and Moinas, S. (2024). "Nonstandard Errors." *Journal of Finance*, 79(3), 2339–2390.
- Brogaard, J., Hendershott, T., and Riordan, R. (2014). "High-Frequency Trading and Price Discovery." *Review of Financial Studies*, 27(8), 2267–2306.

Topic R6: Navigating through the fog

Classification: Empirical topic

Advisor: Paul Seidel

A large literature in accounting and finance studies whether corporate disclosures are difficult to process. Traditional measures such as the Fog Index, document length, or file size are designed to capture the readability of financial reports for human readers. These measures have been linked to analyst behavior, investor attention, and the speed with which information is incorporated into prices. However, modern capital markets increasingly rely on machine-based information processing. Analysts, data vendors, institutional investors, and researchers increasingly use large language models and AI agents to extract structured information from corporate disclosures. This raises a new question: disclosure quality may no longer depend only on whether humans can read a filing, but also on whether AI systems can reliably extract, structure, and cite economically relevant facts from it.

Recent research on financial disclosure has started to examine machine-readable reporting formats and machine access to corporate filings. For example, prior work studies how firms adapt their language when machines are listening, whether machine access to filings improves market efficiency, and how machine-readable reporting formats affect investors. However, these studies mostly focus on whether information is available in a machine-readable format or whether machines access the information. They do not directly measure whether an AI system can actually retrieve, reconcile, and operationalize the relevant economic content of a disclosure. This thesis therefore focuses on AI Extractability: the extent to which a corporate filing allows an AI agent to extract key accounting numbers, narrative risk information, segment data, liquidity constraints, or forward-looking statements in a structured and verifiable way.

The aim of this thesis is twofold. First, the student is expected to construct a simple AI Extractability measure for a sample of corporate filings. This measure may be based on standardized extraction tasks in which an AI model is asked to retrieve selected accounting facts or narrative information from 10-K or 10-Q filings and return them in a structured format, including source references. The quality of these outputs can be evaluated using dimensions such as numerical accuracy, citation validity, consistency across prompts, and agreement with XBRL or Compustat data. Second, the student should examine whether firms with more AI-extractable disclosures experience faster capital-market learning. Possible outcomes include post-filing drift, analyst forecast revisions, abnormal trading volume, bid-ask spreads, or return volatility around filing dates. The student may further extend the analysis by comparing AI Extractability to traditional readability measures such as Fog, document length, or file size.

Requirements: The empirical work requires the use of databases, e.g., CRSP. The databases are widely accessible to affiliates of the University of Mannheim. The candidate should feel comfortable with the use of a statistical software program (such as Python) and econometric methods. Moreover, the student should be familiar with or interested in recent AI tools.

Introductory Literature:

- Li, F. (2008). "Annual Report Readability, Current Earnings, and Earnings Persistence." *Journal of Accounting and Economics*, 45(2–3), 221–247.

- Loughran, T., and McDonald, B. (2014). “Measuring Readability in Financial Disclosures.” *Journal of Finance*, 69(4), 1643–1671.
- Cao, S., Jiang, W., Yang, B., and Zhang, A. L. (2023). “How to Talk When a Machine Is Listening: Corporate Disclosure in the Age of AI.” *Review of Financial Studies*, 36(9), 3603–3642.
- Barbopoulos, L., Dai, R., Putniņš, T. J., and Saunders, A. (2024). “Market Efficiency When Machines Access Information.” *Journal of Financial Economics*, 154, 103776.