DISSERTATION PROPOSAL

Mingjun Sun

"Machine Learning in Asset Pricing"

Friday, November 21, 2025
4:00pm
Tepper 5219

Chapter 1

Cross-sectional information such as firm characteristics has long been central to explaining differences in expected returns, whereas the time-series information has received comparatively less attention. Therefore, I study the informational content of cross-sectional and time-series predictors in a high-dimensional factor framework. I empirically analyze four high-dimensional factor models, incorporating nonlinear transformations of firm characteristics and time-series macroeconomic predictors. I find that time-series predictors add little value beyond cross-sectional information, while incorporating nonlinearities of firm characteristics enhances the pricing performance of out-of-sample mean-variance efficient portfolios. Furthermore, both cross-sectional and time-series information can be effectively summarized by a low-dimensional stochastic discount factor comprising dominant principal components of factors. Theoretically, I demonstrate that in the absence of near-arbitrage opportunities, dominant principal components of factors must capture the majority of cross-sectional information.

Chapter 2

Motivated by evidence that portfolio holdings contain information about asset prices and about investors' preferences over individual stocks, I develop a demand-based approach to identify economically linked firms. Using U.S. actively managed mutual fund data, I build time-varying demand-based links based on correlations in the dynamic portfolio weight changes and also the latent demand estimates derived from the asset demand system in Koijen and Yogo (2019). Demand-based links differ from standard industry groupings and exhibit strong explanatory power for cross-sectional comovement in firm fundamentals. A firm's returns can be explained by those of its demand-based linked firms. Building on this insight, I develop a cross-stock reversal strategy that exploits these demand-revealed connections. I propose to develop a model in which investment managers' demand reveals inter-firm links. This chapter highlights the role of institutional demand in firm links and asset pricing.

Chaper 3

The IPO firm's capital structure decisions, along with its short- and long-term IPO returns, are crucial for investors' decision-making. Therefore, I study the extent to which an IPO firm's capital structure and returns are explained by text-based similar public firms. I construct an IPO firm's set of public comparables by identifying text-based connected stocks. Specifically, I measure cosine similarity between OpenAI's text embeddings of business descriptions in the IPO firm's S-1 (or F-1) filing and those in public companies' 10-K filings. Using this business network, I propose to study the following research questions: (i) whether the leverage levels of connected stocks explain an IPO firm's capital structure at listing; (ii) whether the pre-IPO returns of connected stocks explain first-day and long-run returns in the cross-section; and (iii) whether the idiosyncratic volatility of connected stocks explain IPO returns.

Proposed Committee: Burton Hollifield (Chair), Bryan Routledge, Patrick Blonien, Miguel Oliveira