

Testing ML Return Forecasts Under Transaction Costs

Undergraduate Senior Thesis Research Proposal

1. Research Question

Can ML models with superior statistical prediction metrics (R^2 , MSE) maintain their advantage when subjected to realistic transaction costs during portfolio implementation?

2. Proposed Methodology

This project trains ML models to forecast stock returns, constructs portfolios from these forecasts, and measures performance degradation under simulated market impact.

Models: LASSO and neural network will be trained on a small set of monthly lagged features with proven predictive ability¹ for S&P 500 stocks, following a rolling window approach.

Portfolio Construction: a dollar-neutral long/short strategy, rank-weighted by predicted returns (position sizes scaled proportionally to the rank).

Market Impact: Square-root impact model with parameters from empirical studies, assuming multi-day execution consistent with the Almgren-Chriss framework.² The parameters will be calibrated to reflect execution slippage across a fragmented market, including non-displayed liquidity venues.

Design: Each model's portfolio decisions are simulated under two scenarios: (1) zero market impact (control), and (2) realistic transaction costs (treatment). Performance degradation is measured by comparing gross vs. net Sharpe ratios across methods, testing whether models with superior prediction accuracy maintain advantages after costs.³ Different AUM (\$10M, \$100M, \$1B) will be tested to measure how degradation scales with portfolio size. Benchmarks include 1/N and market portfolios.

3. Background

My background in computer science and economics provides a strong foundation for this research. I have extensive experience with Python, its data science libraries, and the execution of large-scale jobs on the Midway3 cluster. My research in applied microeconomics, supported by coursework in econometrics and financial econometrics, has provided me with the framework to execute empirical projects with methodological soundness.

4. Request

I am seeking a thesis advisor to supervise this research project and request access to the **CRSP** database and the **Midway3** cluster for data retrieval and model experimentation.

Prepared by: Bernard Shen

Date: November 23, 2025

¹Shihao Gu, Bryan Kelly, and Dacheng Xiu, "Empirical Asset Pricing via Machine Learning," *Review of Financial Studies* 33, no. 5 (2020): 2223–2273.

²Robert Almgren and Neil Chriss, "Optimal Execution of Portfolio Transactions" (December, 2000).

³Bryan Kelly, Semyon Malamud, and Kangying Zhou, "The Virtue of Complexity in Return Prediction," *Journal of Finance* 79, no. 1 (2024): 459–503.