Testing ML Forecasts Under Transaction Costs

Bernard Shen - Senior Thesis Presentation (11/14)

The "Paper" vs. "Practice" Gap in ML Asset Pricing

The "Paper" Story: Machine Learning models (Neural Networks, LASSO, etc.) show powerful statistical performance in forecasting returns (e.g., Gu, Kelly, Xiu 2018), generating high "paper" Sharpe Ratios.

The EMH Puzzle: If these signals are real, why haven't they been arbitraged away? Are these signals rich enough to trade on, or is the "alpha" an illusion?

Possible Sources of the Gap:

- **Overfitting:** Models are finding spurious patterns.
- **Strategy Decay:** Signals were real but are now known and have decayed.
- **Adapting Markets:** New behaviors make old signals obsolete.
- **Friction (Execution Loss):** Signals are real, but implementation costs (market impact, fees) are higher than the signal's value.

The Research Question

Can Machine Learning models with superior statistical prediction metrics (like R^2 or MSE) maintain their performance advantage when subjected to realistic transaction costs during portfolio implementation?

In other words, does a 'smarter' model's edge survive market friction?

Where This Thesis Fits

Pillar 1: ML Prediction (The Signal)

- **Gu, Kelly & Xiu (2018):** Use their methods (LASSO, NNs) to generate return forecasts.

Pillar 2: Optimal Execution (The Strategy)

- **Almgren & Chriss (2000):** Provides a pragmatic, risk-averse framework for how to execute a large portfolio rebalance over time (e.g., decaying trade schedule).

Pillar 3: Market Impact (The Cost)

- **The "Square-Root Impact Law"** (e.g., Farmer et al. 2013, Torre & Ferrari 1999, Gatheral 2016): An empirically-verified model for how much trading costs.

My Contribution (The Bridge): I formally connect the ML signals from GKS with the costs of the Square-Root Law, using a realistic execution strategy from Almgren-Chriss.

The Research Plan (Experimental Design)

Step 1

Train ML Models (LASSO, NN) on a rolling window of historical data.

Step 2

Generate Forecasts for all stocks in the universe for the next month.

Step 3

Form Portfolios (e.g., Long top quintile, Short bottom quintile).

Step 4

Calculate Required Trades (Turnover) to rebalance from last month's portfolio.

Step 5

Simulate Execution: Apply Almgren-Chriss schedule to trades (Step 4) and apply Square-Root Law cost model.

Step 6

Compare **Gross Sharpe Ratio** (from Step 3) vs. **Net Sharpe Ratio** (from Step 5).

Methodology: Models & Methods

Prediction Models:

- *Baseline:* **LASSO Regression.** A linear model that enforces sparsity. Good for finding a few "durable" predictors.
- *Advanced:* **Neural Network (NN).** A non-linear, complex model that can capture "non-durable" or complex interaction effects.

Comparison Method:

- We will compare the validity of the models by simulating their performance degradation under different "treatments."
- **Control:** Zero transaction costs (the "paper" portfolio).
- **Scaled Treatment:** Realistic transaction costs, tested at different AUM levels ($10M, $100M, $1B) to see how performance scales with portfolio size.
- **True Benchmark:** Compare performance against 1/N and market portfolios.

Data & Resources

Data Sources:

- **CRSP:** Monthly stock returns, prices, volume, market cap.
- **Compustat:** Quarterly & Annual firm fundamentals (the "features" for the ML models).

Universe & Features:

- *Universe:* S&P 500 stocks.
- *Features:* A small, proven set of monthly-lagged features (e.g., momentum, size, value) consistent with the GKS paper.

Computational Resources:

- **Python:** Scikit-learn, Keras/TensorFlow for model building.
- **Midway3 Cluster:** This simulation is computationally intensive and requires a high-performance cluster for the rolling-window training of the neural networks.

Preliminary Results: Costs Matter

Proof-of-Concept simulation for LASSO at $1B AUM.

The "Paper" Story (Gross Performance):

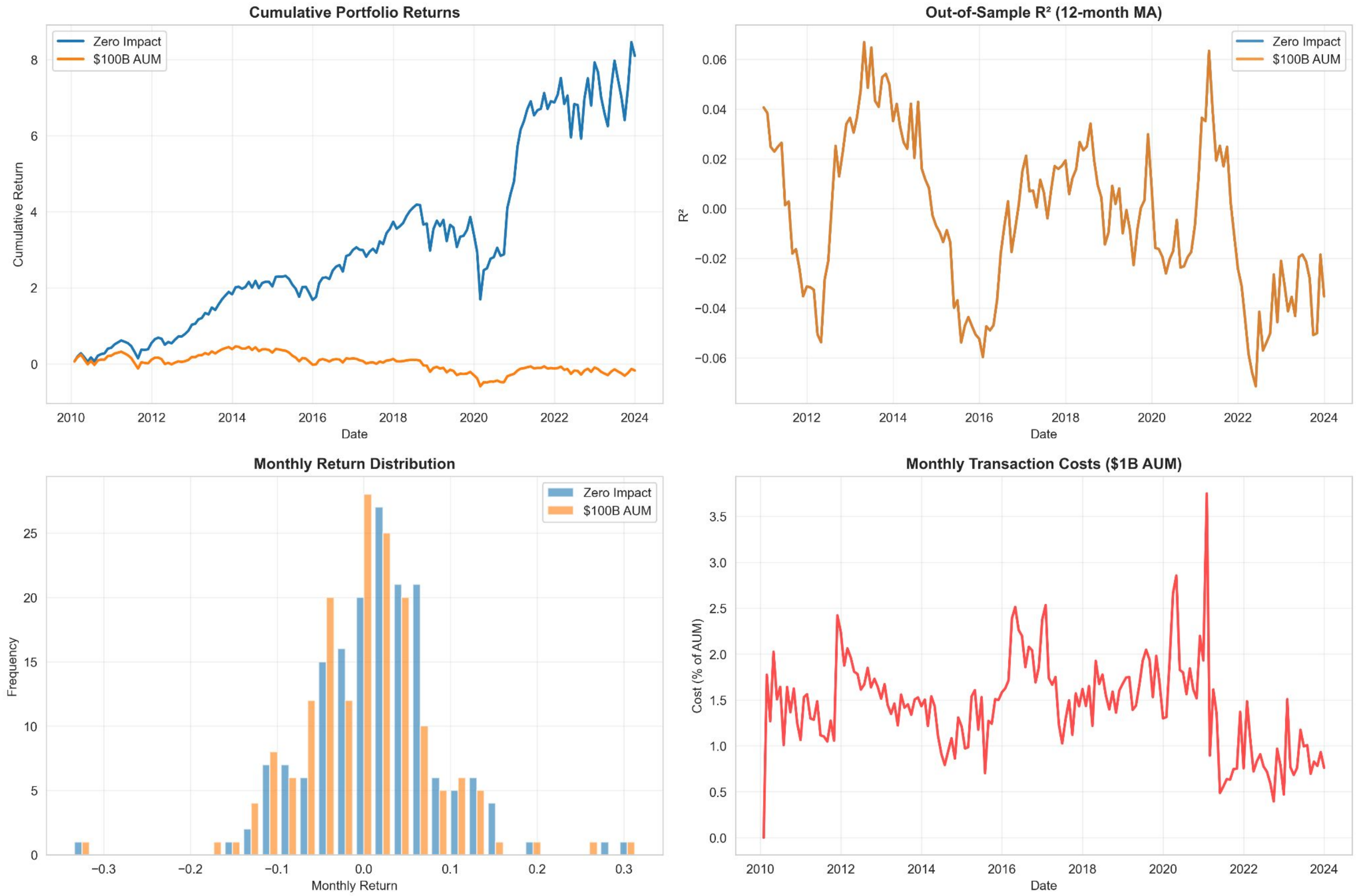
- Gross Sharpe Ratio: **0.730** (A profitable, investable strategy on paper)

The "Practice" Reality (Net Performance):

- Mean Monthly Turnover: **46.5%**
- Resulting Mean Monthly Transaction Cost: **1.44%**
- Net Sharpe Ratio: **0.082**

The Bottom Line:

- Performance Degradation: **-0.65 (88.9%)**
- The signal was consumed by trading costs.


Cumulative Portfolio Returns
Zero Impact
$100B AUM
Cumulative Return
Date
Out-of-Sample R² (12-month MA)
Zero Impact
$100B AUM
R²
Date
Monthly Return Distribution
Zero Impact
$100B AUM
Frequency
Monthly Return
Monthly Transaction Costs ($1B AUM)
Cost (% of AUM)
Date

Conclusion & Next Steps

The Bottom Line:

- Transaction costs are a critical, non-trivial dimension of model evaluation.
- The practical value of complex ML models may be far lower than their statistical value suggests.
- The "best" model is likely a tradeoff between prediction accuracy and implementation cost.

Next Steps:

- Complete data analysis pipeline on Midway3.
- Run full 30-year backtest simulation.
- Analyze performance degradation curves and test for statistical significance.

Write final thesis.