Do U.S. Mutual Funds Have Stock-Selection and Market-Timing Skill?

Yiting Wang, Jiayu Wang, Yuqi Liu, Zhengning Li, Chencheng Liu

February 2026

1 Introduction

Active mutual funds charge fees in exchange for promised expertise: (i) *stock selection* (picking securities that outperform after risk adjustment) and (ii) *market timing* (dynamically adjusting market exposure to benefit from changing market conditions). Whether these skills exist in the data—and if so, for which funds and under what conditions—matters for households, institutions, and regulators because trillions of dollars are allocated based on perceived manager ability.

In this project, we will evaluate U.S. mutual fund performance using daily returns and fund characteristics from WRDS/CRSP, and test whether managers exhibit statistically and economically meaningful stock-selection and/or market-timing ability after accounting for factor exposures, fees, and multiple-testing issues.

2 Idea and Big Question

Core idea: Use the CRSP Survivor-Bias-Free U.S. Mutual Fund Database (daily returns + fund attributes) to estimate each fund's risk-adjusted alpha and timing coefficients, then distinguish skill from luck using modern statistical controls (e.g., false discovery rate, bootstrap), and study whether any skill is *persistent* and *predictable*.

Big questions we want to answer:

- **Selection:** Do U.S. equity mutual funds deliver positive risk-adjusted alpha net of expenses? What fraction of funds (if any) are truly skilled vs. just lucky?
- **Timing:** Do funds exhibit market-timing ability (nonlinear payoffs / state-dependent betas) at daily frequency?
- **Persistence and predictability:** If skill exists, does it persist out-of-sample? Which observable fund characteristics (fees, size, style, turnover proxies, etc.) are associated with future skill?

3 Why This Is Important

- **Investor welfare:** If most alpha is explained by luck, investors should prefer low-cost index funds; if a small subset is skilled, better screening methods could improve net returns.
- **Market efficiency and competition:** Persistent stock-selection skill implies informational advantages and limits to arbitrage; timing ability speaks to how managers process macro/market signals.
- **Practical deliverable:** A reproducible "skill map" of the mutual fund universe that separates factor betas, alpha, and timing, and quantifies uncertainty in a way that can be communicated to non-technical stakeholders.

4 Relation to Our Course (230J)

This project most directly builds on three core themes from 230J: (i) modern performance measurement in asset management, (ii) factor analysis as the benchmark for defining "alpha," and (iii) active return prediction with strict out-of-sample validation.

- **Asset management and modern performance measurement.** Following *Lecture – Asset Management* and the *Example of Modern Performance Measurement*, we evaluate mutual funds by decomposing returns into systematic exposures and residual performance, and interpret results in a manager-selection and monitoring context.
- **Factor models as the baseline for skill.** Consistent with *Lecture – Factor Analysis*, we treat multi-factor risk premia as the primary explanation of fund returns and define stock-selection ability as alpha after controlling for factor exposures.

Financial innovation and data science applications:

- **Financial innovation:** Create an interpretable "skill scorecard" (factor exposures, selection alpha, and timing metrics) to support practical manager screening and ongoing performance monitoring.
- **Data science:** Build a reproducible WRDS/CRSP pipeline for daily panel data, address common fund-data issues (e.g., share-class duplication and fund entry/exit), estimate models at scale, and evaluate results out-of-sample.

5 Data and Empirical Design

Primary data (WRDS/CRSP mutual funds): We will use the CRSP Survivor-Bias-Free U.S. Mutual Fund Database via WRDS. The `daily_returns` table provides daily total returns keyed by `crsp_fundno` and `caldt`, with `dret` available starting in the late 1990s (daily series begin in 1998).[1] We will focus on U.S. domestic equity funds using CRSP style/objective codes and filter out index funds and non-equity categories where appropriate.

Supplementary data:

- **Risk factors:** Daily factor returns (e.g., market excess return, size, value, momentum; optionally profitability and investment) and risk-free rate from standard academic sources (e.g., Fama–French factor data).

[1] CRSP Survivor-Bias-Free U.S. Mutual Fund Database Guide (WRDS/CRSP). The database also documents known biases such as duplicated histories across share classes that must be handled carefully. See: `https://wrds-www.wharton.upenn.edu/documents/410/CRSP_MFDB_Guide.pdf`.

- **Fund characteristics:** Fees/expense-related variables, fund size (TNA), age, and style classifications from CRSP mutual fund tables; optionally holdings (available for many funds beginning in the 2000s) to connect outcomes to portfolio choices.[2]

6 Methods: Separating Stock Selection vs. Timing

(A) Stock-selection ability (alpha)

For each fund i, estimate risk-adjusted alpha using daily regressions:

$$r_{i,t} - r_{f,t} = \alpha_i + \beta_i^\top \mathbf{f}_t + \varepsilon_{i,t},$$

where $\mathbf{f}_t$ includes CAPM and multi-factor specifications (FF3, Carhart 4-factor, and optionally augmented with sector/industry tilt, quality, low-volatility, and liquidity factors). We will:

- estimate alphas in rolling windows (e.g., 1–3 years of daily data) and evaluate out-of-sample performance;
- test **persistence** via sorting funds on past alpha and measuring future alpha;
- correct inference for **multiple testing** across thousands of funds using false discovery rate (FDR) and bootstrap-based "skill vs. luck" decomposition.

(B) Market-timing ability

We will use canonical timing tests and modern extensions:

- **Treynor–Mazuy:** add a quadratic term in market excess return to capture convexity:

$$r_{i,t} - r_{f,t} = \alpha_i + \beta_i(r_{m,t} - r_{f,t}) + \gamma_i(r_{m,t} - r_{f,t})^2 + \varepsilon_{i,t}.$$

- **Henriksson–Merton:** allow different betas in up vs. down markets (option-like timing).
- **State-dependent betas (data science extension):** model time-varying betas as a function of lagged state variables (volatility proxies, term spread, liquidity indicators). We will evaluate whether predicted beta shifts improve realized performance without overfitting (strict time-series cross-validation).

(C) Financial innovation and data science applications

- **Financial innovation:** construct a transparent "Skill Scorecard" that reports (i) factor exposures, (ii) alpha with uncertainty bands, (iii) timing coefficients, and (iv) an FDR-adjusted probability of true skill.
- **Data science:** build a reproducible WRDS extraction + cleaning pipeline; implement scalable estimation across funds; use robust validation (walk-forward evaluation, nested CV for hyperparameters, and leakage checks).

7 Related Literature (Hierarchical Structure)

Performance measurement and "alpha" benchmarks. Early work evaluates manager performance using CAPM-based alpha (Jensen, 1968). Modern practice defines alpha relative to multi-factor benchmarks, including the Fama–French 3-factor model (Fama and French, 1993), the Carhart 4-factor model with momentum (Carhart, 1997), and expanded factor sets such as the Fama–French 5-factor model (Fama and French, 2015).

Selection vs. timing decomposition. Beyond average alpha, market-timing ability is commonly tested using convexity/nonlinearity in market exposure, including the Treynor–Mazuy specification (Treynor and Mazuy, 1966) and the Henriksson–Merton regime-based timing framework (Henriksson and Merton, 1981; Henriksson, 1984).

Skill vs. luck and multiple testing. Because thousands of funds are evaluated simultaneously, statistical significance can be driven by chance. A central approach is to control false discoveries when interpreting estimated alphas; Barras, Scaillet, and Wermers (2010) propose an FDR-based decomposition of funds into skilled, unskilled, and zero-alpha groups.

Economic equilibrium, persistence, and implementation frictions. In competitive markets, capital flows may arbitrage away expected net alpha; Berk and Green (2004) formalize this equilibrium logic linking flows and performance. Empirically, performance can be decomposed into stock-picking, style, transaction costs, and expenses (Wermers, 2000), highlighting the importance of measuring returns net of implementable frictions.

Practitioner evidence and industry benchmarking. Industry scorecards such as S&P SPIVA regularly compare active funds against passive benchmarks, complementing academic tests by emphasizing net-of-fee outcomes and benchmark choice (SPIVA reports). Commercial evaluation frameworks (e.g., Morningstar-style ratings) are widely used but can be sensitive to methodology, motivating transparent and reproducible academic measures.

[2] CRSP product overview notes holdings coverage beginning 2003 and availability of daily/monthly returns and NAVs. See: `https://www.crsp.org/research/crsp-survivor-bias-free-us-mutual-funds/`.