Contents

1. Introduction

1.1 Research objective and motivation

This thesis is deliberately a purely empirical prediction and portfolio-implementation exercise. It does not derive expected returns from a structural asset-pricing theory, and it does not claim that the 318 predictors are theory-based, priced, or non-diversifiable risk factors. CAPM and FF5-plus-momentum are used as benchmark definitions for residual returns and as portfolio risk controls, not as a theoretical foundation for the full information set. The empirical difficulty is that expected returns are not observed directly: realized returns contain a large unpredictable component, while the available conditioning information is high-dimensional, correlated, and potentially nonlinear. Machine learning is used here as a flexible forecasting technology, but, without a supporting structural theory, that flexibility can fit sample-specific relationships as easily as persistent economic structure (Gu, Kelly, and Xiu 2020; Kozak, Nagel, and Santosh 2020).

This thesis examines two linked questions. The first is statistical: can a high-dimensional, point-in-time information set predict the cross-sectional ordering of subsequent U.S. equity returns and benchmark-adjusted returns? The second is economic: can that estimated ordering be translated into a feasible long-short portfolio whose returns remain positive after transaction costs, stock borrowing, financing, risk controls, and capacity constraints? Treating these as separate questions is essential. A model may rank securities correctly while providing insufficient information about payoff magnitudes, and a statistically informative signal may be uneconomic once turnover, liquidity, and covariance are taken into account.

The information set extends beyond the firm characteristics conventionally used in cross-sectional asset-pricing studies. It combines trailing returns and market state, price trends, volatility, liquidity and trading frictions, accounting fundamentals, institutional ownership and filing activity, option-implied information, analyst forecasts and revisions, and firm-related news intensity and tone. All predictors are aligned to

the monthly signal date. These variables are treated as candidate predictive covariates rather than risk factors. The objective is not to claim that a particular database or single feature explains expected returns; it is to test whether nonlinear models can identify cross-sectional regularities in this heterogeneous information set and whether those regularities survive separate economic evaluation.

The empirical analysis considers both total-payoff and model-relative targets. CAPM and the Fama-French five-factor model augmented with momentum are used to define alternative residual-return benchmarks (Sharpe 1964; Fama and French 2015; Carhart 1997). A positive prediction result for these residuals means only that the information set helps order subsequent returns not captured by the specified benchmark in the observed sample. It does not identify whether the residual reflects mispricing, omitted risk, time-varying exposures, model misspecification, or a pattern that will disappear outside the sample. It is therefore not, by itself, sufficient evidence that the score identifies a durable trading opportunity rather than noise. The portfolio analysis provides a separate historical implementation test, but it cannot resolve this economic-identification problem. Accordingly, the thesis does not attempt to establish that CAPM or FF5-plus-momentum either holds or fails universally.

1.2 Empirical hypotheses

For benchmark $b \in \{CAPM, FF5+M\}$, let $\alpha^{b}_{i,t+1}$ denote stock i's next-holding-period factor-adjusted return and let $\hat{s}^{b}_{i,t}$ denote a score estimated from information available at the monthly signal date. The first hypothesis concerns cross-sectional model-relative return predictability and is expressed through the monthly Spearman rank correlation:

$$H^{b}_{0}: E\left[\operatorname{corr}_{S}\left(\hat{s}^{b}_{i,t}, \alpha^{b}_{i,t+1}\right)\right] = 0,$$

against the one-sided alternative

$$H^{b}_{1}: E\left[\operatorname{corr}_{S}\left(\hat{s}^{b}_{i,t}, \alpha^{b}_{i,t+1}\right)\right] > 0.$$

The second hypothesis concerns directional payoff prediction. Let $\pi^{L}_{i,t+1}$ and $\pi^{S}_{i,t+1}$ denote the realized long- and short-side payoffs after the side-specific return

transformation. Separate scores are estimated because long and short positions have asymmetric payoff distributions and implementation costs:

$$H_0^d: E\left[\mathrm{corr}_S\left(\hat{s}_{i,t}^d, \pi_{i,t+1}^d\right)\right] = 0, H_1^d: E\left[\mathrm{corr}_S\left(\hat{s}_{i,t}^d, \pi_{i,t+1}^d\right)\right] > 0,$$

for $d \in \{L, S\}$. The associated economic test is whether the fixed rankings, after causal calibration into expected-payoff units, produce positive net excess return within a prespecified feasible set. This portfolio test is not a substitute for the predictive hypotheses; it evaluates the economic content of the estimates after implementation frictions.

1.3 Empirical design

The empirical analysis proceeds in two stages. The predictive stage compares XGBoost, multilayer perceptrons, Elastic Net, the provider-supplied LSEG StarMine Combined Alpha Model rank, and prespecified ensembles using walk-forward estimation. Model selection is based on a validation period, and predictive performance is measured independently of portfolio formation using monthly rank information coefficients and heteroskedasticity- and autocorrelation-consistent inference. Full, direct-CAM-free, and correlation-reduced information views distinguish the incremental contribution of the provider signal from that of the broader predictor panel.

The portfolio stage holds the selected rankings fixed. Because ordinal scores do not supply the cardinal expected returns required for allocation, each ranking is mapped into expected-payoff units using only outcomes that would have been observable at the decision date. These estimates enter a constrained optimizer that accounts for spreads, market impact, financing, stock-specific borrowing, covariance, factor and industry exposures, concentration, turnover, and liquidity. This structure follows the distinction in the literature between statistical return prediction and net-of-cost implementability (DeMiguel et al. 2020; Jensen et al. 2026).

Predictive-model and portfolio-configuration choices are made before the January 2024-August 2025 historical evaluation. A separate exposure experiment applies a

common multiplier to the fixed portfolio weights; it neither retrains the predictive models nor changes the underlying security selection.

1.4 Contribution

The contribution is empirical and computational rather than theoretical. First, the thesis constructs a point-in-time monthly panel that integrates conventional market and accounting variables with option-implied measures, ownership and filing information, analyst expectations, and timestamped firm-related news. Second, it compares linear and nonlinear ranking methods across total-payoff and factor-adjusted targets, while allowing long- and short-side prediction to differ. Third, it evaluates the full chain from statistical ranking to cardinal payoff calibration and constrained allocation. This design separates measured predictive content from the effects of calibration, portfolio constraints, trading costs, and exposure; it does not identify a new risk factor or structural mechanism.

1.5 Main findings

The predictive results indicate that the information set contains statistically measurable cross-sectional information. During January 2024-August 2025, the selected models produce mean monthly rank information coefficients of 0.058 for CAPM-adjusted returns, 0.048 for FF5-plus-momentum-adjusted returns, 0.067 for long-side net payoff, and 0.057 for short-side net payoff. All four estimates are positive, although validation evidence is substantially weaker for the short-side model.

The portfolio results show that the predictive rankings retain economic content after the stated implementation costs. The validation-selected 1.5% position-limit specification produces a 14.67% annualized net excess return, 6.32% realized volatility, a 2.320 Sharpe ratio, and a 1.34% maximum drawdown in the 20-month evaluation period. Its total-return CAGR of 21.34% is approximately equal to SPY's 21.33%, while its realized volatility and drawdown are lower. Most of the estimated

return originates from the long positions; the short sleeve has a negative average standalone contribution.

A separate exposure sensitivity applies a validation-selected multiplier of 2.05 to the same security-level weights. It raises evaluation-period CAGR to 39.17% with a 2.259 Sharpe ratio, but peak gross exposure reaches 2.62 and one-way participation reaches 24.4% of average daily volume at a $100 million capital base. This result illustrates the effect of scaling the same predictions; it is not evidence of additional predictive content or of implementability at that capital level.

2. Related literature and empirical motivation

2.1 From linear asset pricing to machine learning

Classical portfolio theory separates beliefs about expected return from the covariance structure that determines risk (Markowitz 1952). CAPM then connects expected return to market exposure (Sharpe 1964). Multi-factor models expand the relevant risk-adjustment set, including profitability and investment in the Fama-French five-factor model (Fama and French 2015) and momentum in the Carhart extension (Carhart 1997).

When an empirical asset-pricing model begins from theory-motivated risk factors, machine learning can relax functional-form restrictions while retaining an economic interpretation. That statement does not apply directly to the present design, which begins from a broad, heterogeneous information set rather than a closed set of theory-derived risk factors. Here machine learning is used to approximate a potentially nonlinear predictive mapping; estimated feature weights, splits, and interactions are not interpreted as prices of risk or as evidence that the underlying variables represent non-diversifiable shocks. Gu, Kelly, and Xiu compare linear, tree-based, and neural methods on a large U.S. equity panel and trace much of the nonlinear gain to interactions among a relatively compact set of momentum, liquidity, and volatility signals (Gu, Kelly, and Xiu 2020). Their finding that shallow neural networks often outperform deeper alternatives is especially relevant in a low-signal

monthly panel. Deep-learning approaches can learn nonlinear pricing structures and latent factors when the architecture and economic restrictions are designed carefully (L. Chen, Pelger, and Zhu 2024). Tree boosting is attractive for tabular financial data because it captures thresholds and interactions while providing regularization and scalable fitting (T. Chen and Guestrin 2016).

The literature does not imply that greater model complexity necessarily improves predictive performance. Financial panels contain weak signals, correlated predictors, time-varying coverage, and unstable relationships. With no structural restriction connecting most inputs to non-diversifiable risk, the central empirical problem is to distinguish reproducible structure from noise. Freyberger, Neuhierl, and Weber find that many familiar characteristics lack incremental predictive content once nonlinearities and joint selection are considered (Freyberger, Neuhierl, and Weber 2020). Kozak, Nagel, and Santosh similarly show why shrinkage and low-dimensional structure matter in a high-dimensional cross-section (Kozak, Nagel, and Santosh 2020). Regularization, validation, and linear benchmarks can reduce the opportunity to fit noise, but they cannot establish that a discovered relation is structurally stable. Elastic Net is therefore retained throughout this thesis as the principal linear comparator.

2.2 Characteristics, expectations, and risk

Stock characteristics may proxy for expected return, exposure to common shocks, or both. Kelly, Pruitt, and Su emphasize the relationship between characteristics and conditional covariances (Kelly, Pruitt, and Su 2019). This distinction matters for portfolio construction. A stock with a high forecast need not receive a large weight if it adds concentrated factor or idiosyncratic risk. Conversely, a security with modest standalone expected return can still improve a portfolio through covariance.

In this thesis, a feature and a risk factor are not interchangeable concepts. A risk-factor interpretation requires an economic mechanism and exposure to risk that cannot be eliminated through diversification. Most of the 318 inputs are firm-level

characteristics, transformations, coverage indicators, or provider signals for which no such claim is made. The CAPM and FF5-plus-momentum factors retain their conventional role in benchmark adjustment and portfolio risk control, while the remaining variables are evaluated only for incremental predictive content.

The thesis therefore keeps three objects separate:

- **ranking signal:** the model's cross-sectional ordering;
- **expected payoff:** a causal calibration of realized payoff conditional on rank; and
- **risk:** a point-in-time factor covariance and security-specific variance model.

These objects answer different empirical questions. Rank IC measures whether the estimated ordering is informative. The calibration stage examines whether differences in rank correspond to differences in subsequently realized payoffs. The portfolio stage then evaluates whether the calibrated payoffs remain economically relevant after covariance, concentration and implementation costs. Reporting these stages separately prevents a portfolio return from being interpreted as direct evidence about every underlying prediction model.

2.3 Analyst information

Analyst research contains information beyond a single earnings forecast. Recommendation ranks, changes, dispersion, revision breadth, target prices, and information age may reflect both public-information processing and informed expectations. Analyst reports have documented information content for returns and revisions (Asquith, Mikhail, and Au 2005).

The LSEG block in this thesis contains 97 analyst-related sources. One important component is the StarMine Combined Alpha Model rank, abbreviated CAM. CAM is a proprietary LSEG model output rather than a model estimated in this thesis. To avoid double counting, the model comparison evaluates a full information view, a direct-CAM-free view, and a correlation-reduced view. Prespecified ensembles can then add the direct CAM rank back to a learned CAM-free model.

2.4 Financial news

News can convey information about individual firms, related firms, industries, and aggregate market conditions. Media tone has predictive relationships with market activity and sentiment (Tetlock 2007), and firm-specific language contains information about subsequent earnings and returns (Tetlock, Saar-Tsechansky, and Macskassy 2008). Finance-specific dictionaries improve textual measurement relative to generic positive/negative word lists (Loughran and McDonald 2011).

The news design in this thesis is deliberately compact. Alpaca's Benzinga archive is stored in monthly point-in-time snapshots. Only stories published and last updated by the signal-close cutoff are eligible. Articles are linked directly to a stock when the article symbols identify it, and indirectly through approved company-family or SIC2 relationships. Text is summarized into stock-related counts, recency, novelty, breadth, and finance-specific sentiment features over 1-, 7-, and 30-day windows. No stock enters or exits the investable universe merely because it has or lacks news.

This design does not assign a uniform positive or negative interpretation to the absence of an article. Unrestricted article text is not supplied directly to the predictive models because doing so would materially change the dimensionality and stability of the historical text-processing design.

2.5 Missing data

Missingness is pervasive in asset-pricing panels. It often reflects coverage regimes, firm characteristics, and provider history rather than random noise. Recent work shows that simple imputation choices can materially affect inference (Freyberger, Neuhierl, and Weber 2025). The present thesis uses its missing-data treatment as a predictive encoding, not as a claim of inferential equivalence to a formal missing-data estimator.

Continuous variables are clipped cross-sectionally at the monthly 1st and 99th percentiles, converted to centered within-month ranks, and assigned the neutral value zero when missing. A separate missingness indicator allows every model to

distinguish an observed median-rank value from an unavailable observation. Binary inputs likewise receive a neutral value and an availability indicator. Categorical inputs use training-period vocabularies with explicit unknown states. Complete-case filtering is rejected because it would make the universe small, unstable, and selected on alternative-data availability.

2.6 Trading costs and implementability

Backtests that ignore implementation frictions can substantially overstate attainable returns. Trading cost depends on spreads, trade size, liquidity, and market impact; institutional evidence documents both scale and heterogeneity in these costs (Frazzini, Israel, and Moskowitz 2018). DeMiguel and coauthors show that transaction costs can change which combinations of characteristics matter for a portfolio because trades generated by different signals may offset one another (DeMiguel et al. 2020). More recent work argues that machine-learning portfolios should be judged by the implementable net-of-cost frontier rather than by frictionless prediction alone (Jensen et al. 2026). Short positions add borrow cost and availability constraints. Leverage and financing create further asymmetry.

The final optimizer subtracts transaction, financing, and stock-specific borrow costs directly from its objective. Its primary model includes spread plus square-root impact, a 0.5% annual financing rate, and stock-level borrow. A stress evaluation doubles spread and impact, raises financing to 1.5%, and uses stressed borrow rates. The same weights are evaluated under stress; weights are not re-optimized after seeing realized outcomes.

2.7 Position of the thesis in the literature

The empirical structure follows the comparative logic of the asset-pricing literature. Gu, Kelly, and Xiu proceed from methodology to a common U.S. equity panel, compare model families on statistical prediction, and then examine economic portfolio implications and variable importance (Gu, Kelly, and Xiu 2020). This thesis adopts that sequence and adds an explicit implementation stage:

Empirical question	Evidence used here	Why it is separate
Is information available before the trade?	Point-in-time timestamps, label-availability dates, and delayed execution	Establishes chronology before model quality is discussed
Do models rank stocks?	Monthly Spearman IC, HAC inference, and label-shuffle controls	Tests predictive ordering independently of portfolio choices
Which information and model family help?	Full, CAM-free, and correlation-reduced views across XGBoost, MLP, Elastic Net, direct CAM, and ensembles	Distinguishes nonlinear value from analyst information and redundancy
Can ranks become economic payoffs?	Causal side-specific rank calibration	Avoids treating an ordinal rank as a cardinal return forecast
Does the signal retain economic value?	Net-return optimization with costs, covariance, turnover, capacity, and exposure limits	Evaluates economic significance rather than a frictionless spread
Does exposure change the result?	Capacity-constrained 1.00-times reference and validation-selected 2.05-times specification	Separates security selection from the return, risk, and capacity effects of scaling

The design imposes distinct evidentiary requirements. A positive rank IC does not establish implementability, while a positive historical portfolio return does not prove that every component model independently forecasts returns. Conclusions are therefore stated at the level directly supported by each test.

3. Data and sample construction

3.1 Data sources and variable construction

The empirical analysis distinguishes externally supplied observations from variables constructed for this thesis. Wharton Research Data Services (WRDS) is the access platform through which the principal U.S. market, accounting and derivatives databases are obtained; it is not treated as a separate economic data source. The main sources and transformations are summarized below.

Database or provider	Information obtained	Variables or objects constructed in this thesis
CRSP, accessed through WRDS	Security identifiers, prices, returns, shares, volume and delisting information	Point-in-time universe, total shareholder returns, momentum, reversal, volatility, beta, liquidity and execution-period outcomes
Compustat, accessed through WRDS	Dated financial-statement observations	Valuation, profitability, investment, growth and leverage measures with reporting lags
OptionMetrics, accessed through WRDS	Option and implied-volatility observations	Implied-volatility level, skew, term-structure and option-activity measures
Institutional-ownership and SEC filing sources	Dated ownership observations, filing events and issuer information	Ownership changes, filing-event indicators and point-in-time identity variables
London Stock Exchange Group (LSEG) Data & Analytics	Analyst estimates, recommendations, target prices and provider-calculated CAM ranks	Point-in-time forecast revisions, dispersion, breadth and staleness measures; CAM itself remains a provider-supplied input
Benzinga news distributed through Alpaca	Timestamped articles, text and associated symbols	Direct, company-family and industry-related news intensity, sentiment, novelty, recency and breadth measures
Kenneth French Data Library	Market, size, value, profitability, investment and momentum factor returns	CAPM and FF5-plus-momentum residual-return targets and factor-adjusted evaluation measures

Accordingly, the external sources provide the underlying observations and, in the case of CAM, one direct analyst ranking. The thesis constructs the monthly point-in-time panel, trailing and cross-sectional transformations, missing-data encoding, prediction targets, model estimates, forecast combinations, rank-to-payoff calibration and portfolio weights. This distinction is maintained throughout the empirical analysis.

3.2 Universe and trading schedule

The research universe follows the project’s point-in-time U.S.-traded common-stock rules. At each monthly signal date, one eligible security line is selected per issuer

under the established PERMNO/PERMCO procedure. The long-eligible screen requires a price of at least $5, market capitalization of at least $300 million, 60-day median dollar volume of at least $5 million, a positive-volume fraction of at least 90%, and at least 126 usable daily returns. Eligibility is not conditioned on present-day survival, present-day ticker availability, analyst coverage, news coverage, or future shortability.

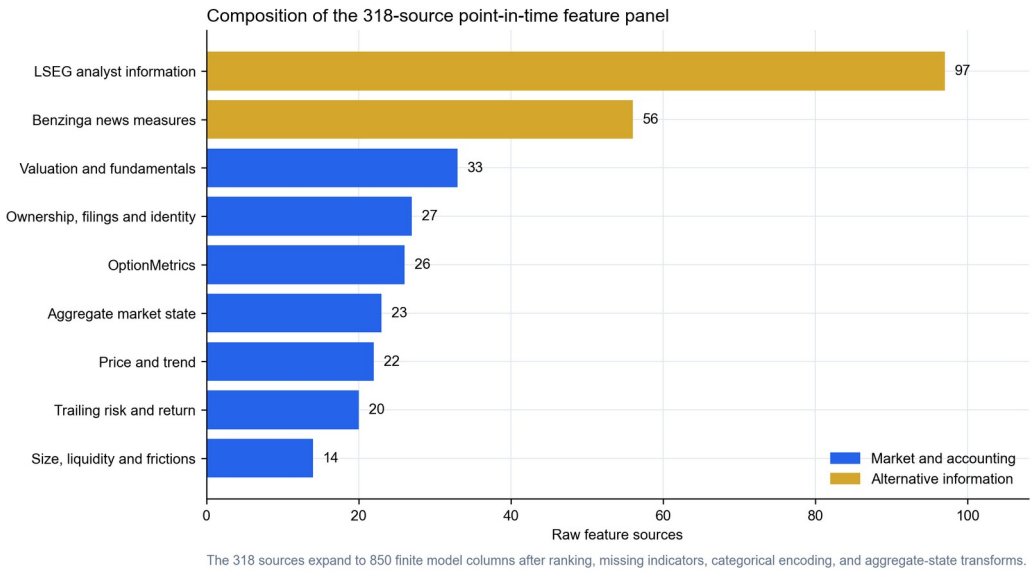
Signals are formed monthly at the signal close. Execution occurs at the authenticated close of the first trading day after the signal, and the position exits at the next rebalance close. Returns are total shareholder returns and are delisting-aware. This schedule prevents same-close formation and execution assumptions.

The final panel covers January 2015 through August 2025, comprising 128 monthly signal dates and 278,897 stock-month observations. All monthly cross-sections are constructed directly from the dated source observations; the December 2019 cross-section contains 2,051 securities.

3.3 Feature families

The 318 raw feature sources are grouped as follows:

Feature family	Raw sources	Examples
Price and trend	22	multi-horizon momentum, reversal, moving-average location
Trailing risk and return	20	volatility, downside risk, beta, idiosyncratic variation
Size, liquidity, and frictions	14	market cap, dollar volume, spreads, turnover, price
Valuation and fundamentals	33	book-to-market, profitability, investment, leverage, growth
Ownership, filings, events, and identity	27	institutional ownership, filings, exchange, ADR and incorporation flags
OptionMetrics	26	implied volatility, skew, term structure, option activity
Aggregate market state	23	market and macro regime variables known at signal time
LSEG analyst information	97	estimates, revisions, dispersion, target prices, CAM ranks
Benzinga news measures	56	direct/related article intensity, sentiment, novelty, recency, breadth
Total	318	



Composition of the point-in-time feature panel. Source: author's calculations from the source-variable registry.

Fifteen engineered variables are constructed only from contemporaneous or trailing inputs. Four planned segment variables are excluded because their source table contains zero observations. FF12 industry is retained for diagnostics only; it is never used to form, neutralize, gate, or optimize the portfolio. PERMNO, PERMCO, ticker, issuer name, outcome status, and eligibility flags are not model inputs.

The raw sources expand to 850 finite model columns after ranks, missing indicators, training-period categorical vocabularies, and aggregate-state transforms. There are no complete-case stock-months across all raw fields; the neutral-plus-indicator encoding is therefore essential to preserve a broad and historically consistent universe.

3.4 Benzinga news data

The Benzinga sample distributed through Alpaca contains approximately 1.95 million articles and covers all 128 monthly signal dates. An article enters the information set only when both its creation timestamp and its latest-update timestamp precede the signal close. Direct symbol links are preferred; approved company-family and SIC2 links capture economically related news. The encoding distinguishes an absence of stock-specific news from an absence of authenticated provider coverage.

3.5 LSEG analyst data

The LSEG analyst panel covers all 128 monthly signal dates. The underlying observations include forecast levels, recommendations, target prices, and their publication dates. The variables constructed in the thesis measure revisions, breadth, dispersion, and staleness. The StarMine CAM rank is retained separately as a provider-generated signal.

3.6 Prediction targets

Four targets are formed for each stock and holding month:

1. **CAPM alpha rank:** the centered cross-sectional rank of next-period CAPM residual performance.
2. **FF5-plus-momentum alpha rank:** the centered cross-sectional rank of next-period performance after the five Fama-French factors and momentum.
3. **Long net-opportunity rank:** the rank of long-side next-period payoff relative to cash and relevant frictions.
4. **Short net-opportunity rank:** the rank of the payoff from being short, so a high score denotes a larger expected short-side payoff.

The target transformation is cross-sectional and monthly. If $r_{i,t+1}$ is the next holding-period outcome and F_t is the empirical cross-sectional distribution, the centered rank target is

$$y_{i,t+1}^{rank} = F_t(r_{i,t+1}) - \frac{1}{2}.$$

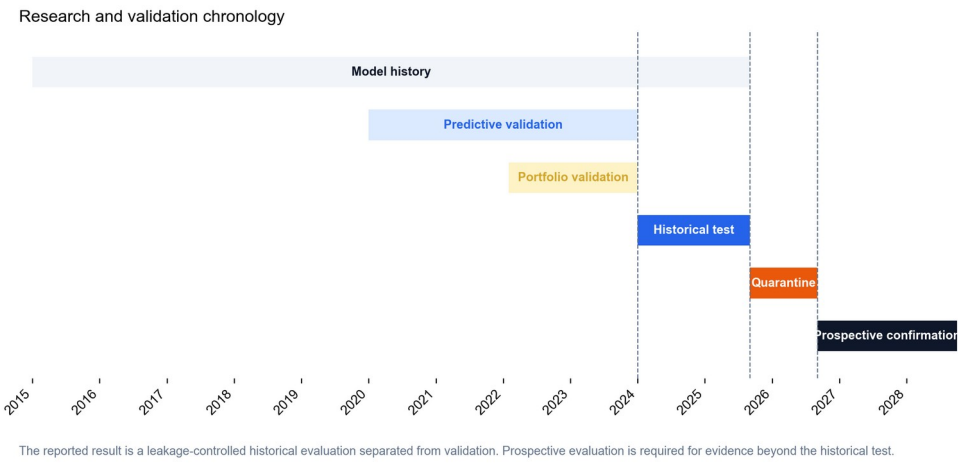
This target asks the model to order stocks rather than forecast an exact percentage return. The ordering is evaluated with monthly Spearman rank correlation.

3.7 Chronology

The common information panel begins in January 2015 because the news archive begins in 2015. The empirical chronology is:

- **Model history:** January 2015 onward, subject to each target's causal refit rule.

- **Predictive-model validation:** January 2020-December 2023, 48 months.
- **Portfolio-configuration validation:** February 2022-December 2023, 23 months.
- **Historical test:** January 2024-August 2025, 20 months.
- **Quarantine:** September 2025-August 2026.
- **Earliest prospective confirmation:** September 2026.



Research and validation chronology. Source: author's construction.

A row enters a fit only when its outcome availability date is no later than the current forecast signal date. Of 278,897 stock-month rows, 777 lack a realized outcome/availability date. They do not enter labeled training; among observed labels, no timing violation is found.

4. Predictive methodology

4.1 Information views

The predictive analysis compares three information views for every model family:

- **Full:** all encoded features, including direct CAM fields.
- **CAM-free:** removes the five direct CAM sources so the learned model can be combined with CAM without mechanically double counting it.
- **Correlation-reduced:** prunes source features with absolute rank correlation above 0.90 using only data available through December 2019. Ties are

resolved by observed coverage, registry order, and feature name; outcomes are never used.

Every view is reported. The historical test cannot select a feature view after outcomes are observed.

4.2 XGBoost

XGBoost captures nonlinear thresholds and interactions in tabular data. The validation grid varies depth, learning rate, estimator count, subsampling, and regularization. Continuous XGBoost predicts the centered rank directly; the pairwise version learns the within-month ordering through a ranking loss.

4.3 Neural network

The neural model is a feed-forward multilayer perceptron. Validation compares shallow two- and three-layer architectures with dropout, weight decay, and gradient clipping. Continuous networks predict ranks; pairwise networks learn which stock in a within-month pair should rank higher. Final neural forecasts average three fixed seeds to reduce initialization noise.

4.4 Elastic Net

Elastic Net provides the linear benchmark. Its objective combines squared error with L_1 and L_2 penalties:

$$\min_{\beta} \frac{1}{2n} \| y - X\beta \|_2^2 + \lambda \left(\rho \| \beta \|_1 + \frac{1-\rho}{2} \| \beta \|_2^2 \right).$$

Elastic Net tests whether the broad panel's useful content can be captured with sparse or shrunk linear combinations. Its penalty strength and mix of L_1 and L_2 regularization are chosen on the same causal validation scheme as the nonlinear models.

4.5 Direct CAM and ensembles

Direct CAM is an analyst-ranking measure supplied by LSEG, not a model estimated in this thesis. Prespecified ensembles combine estimated forecasts with one another

or with CAM. Both equal-rank and trailing-performance weights are evaluated; the validation-selected specifications use equal-rank combinations.

4.6 Walk-forward refitting and tuning

Refit schedules are quarterly or annual, with expanding histories for the alpha tasks and a rolling 60-month alternative for the payoff tasks. Hyperparameters are retuned annually or biennially using a trailing 12-month holdout. The selection metric is equal-month mean Spearman IC. Historical-test outcomes cannot change the view, objective, hyperparameters, ensemble, or schedule.

4.7 Predictive evaluation

For month t, the rank information coefficient is

$$IC_t = \mathrm{corr}_{Spearman}(\hat{s}_{i,t}, y_{i,t+1}).$$

The principal statistic is the equal-month mean, $\bar{IC} = T^{-1}\sum_t IC_t$, with a six-lag heteroskedasticity-and-autocorrelation-consistent standard error. Supporting measures include the positive-IC fraction, early/recent mean IC, rolling minimum mean IC, tail payoff spread, decile monotonicity, size/liquidity stratification, rank turnover, seed stability, and a label-shuffle control.

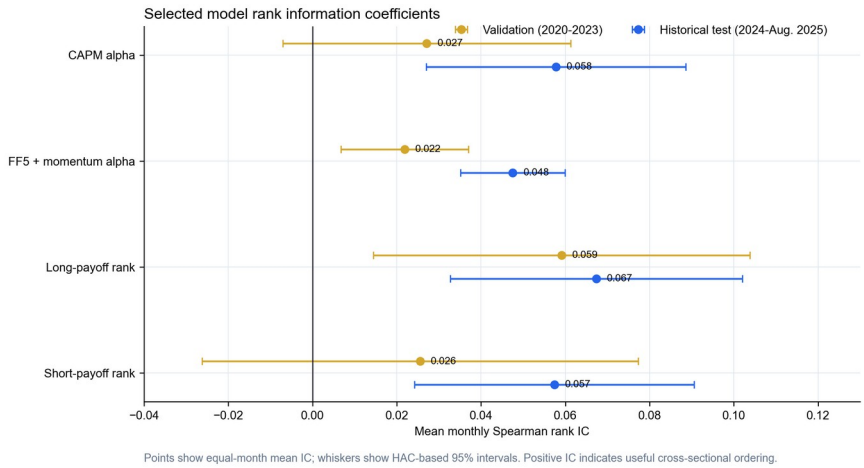
IC measures ordering, not exact-return accuracy. An IC of 0.06 means that, on average, the predicted and realized cross-sectional ranks have a positive Spearman correlation of six percent. That can be economically useful across many securities even when individual monthly returns remain difficult to forecast.

5. Predictive results

5.1 Selected models

Target	Selected model	Objective	Test IC	Test t	Val. IC	Val. t
CAPM alpha	CAM-free MLP + direct CAM	Continuous rank	0.0578	3.68	0.0271	1.55
FF5 + momentum alpha	CAM-free MLP + direct CAM	Pairwise	0.0475	7.51	0.0219	2.83
Long net payoff	CAM-free MLP + direct CAM	Continuous rank	0.0674	3.82	0.0591	2.59

Target	Selected model	Objective	Test IC	Test t	Val. IC	Val. t
Short net payoff	Correlation-reduced XGBoost + MLP	Pairwise	0.0574	3.39	0.0256	0.97



Selected model rank information coefficients. Source: author's calculations.

Three patterns are important.

First, the validation-selected CAPM, FF5-plus-momentum and long-payoff specifications combine a CAM-free neural network with direct CAM. The analyst score contains predictive information, while the neural model contributes information beyond that score. Excluding CAM from the neural input before combining the two forecasts avoids mechanical double counting.

Second, the short-payoff problem differs. Its selected model is a pairwise ensemble of correlation-reduced XGBoost and MLP without direct CAM. This supports the original hypothesis that the long and short sides need not use the same predictors or model.

Third, XGBoost alone produces lower validation performance in the alpha tasks, even after testing deeper trees and stronger regularization. The short-side ensemble is the relevant exception. Broad feature inclusion does not guarantee that boosting will disregard all uninformative variables: finite samples, correlated splits, changing coverage and noisy one-month outcomes can reduce out-of-sample performance. The outcome-blind correlation-reduced view performs better for the selected short-side ensemble.

5.2 Statistical interpretation

All four selected models have positive historical-test IC estimates and HAC t-statistics above 3.3. For the two asset-pricing targets, the CAPM-alpha estimate is 0.0578 with a t-statistic of 3.68, and the FF5-plus-momentum-alpha estimate is 0.0475 with a t-statistic of 7.51. Within the fixed historical test, these estimates reject the null hypothesis of zero average rank predictability. The earlier validation evidence is less uniform: FF5-plus-momentum and long payoff have validation t-statistics above 2.5, while the CAPM and short-payoff t-statistics are 1.55 and 0.97, respectively.

The hypothesis test concerns cross-sectional ordering, not exact percentage-return prediction. It also remains conditional on the factor specification. A predictable residual may represent omitted risk, factor-model misspecification or mispricing; the present design does not identify which interpretation is correct. The ordering is sufficient to define eligible sets, but portfolio formation additionally requires a mapping from ranks to expected-payoff units.

6. Portfolio-formation methodology

6.1 Portfolio universe and alpha sign restriction

The top 20% of the fixed long- and short-payoff rankings define the eligible tails, with a desired minimum of 40 securities on each side. If a security qualifies for both, it is assigned to the side with the larger calibrated expected payoff. The FF5-plus-momentum alpha estimate then imposes the following sign restrictions:

- a new long must have strictly positive causal calibrated FF5-plus-momentum alpha;
- a new short must have strictly negative causal calibrated FF5-plus-momentum alpha.

The gate does not determine the weight. It prevents the return selector from buying a stock with negative expected factor-adjusted performance or shorting one with positive expected factor-adjusted performance.

6.2 Rank-to-payoff calibration

Ranks are ordinal, so they cannot be inserted into an optimizer as percentage returns. For each side, a monotonic rank-to-payoff curve is estimated using only outcomes available before the signal. Rank bins are finer in the tails, months receive equal weight, at least 24 completed months are required, and isotonic regression preserves monotonicity.

Let $q_{i,t}$ be the percentile rank and $g_t(q)$ the causal side-specific isotonic fit. The expected payoff supplied to the optimizer is

$$\hat{\mu}_{i,t} = g_t(q_{i,t}).$$

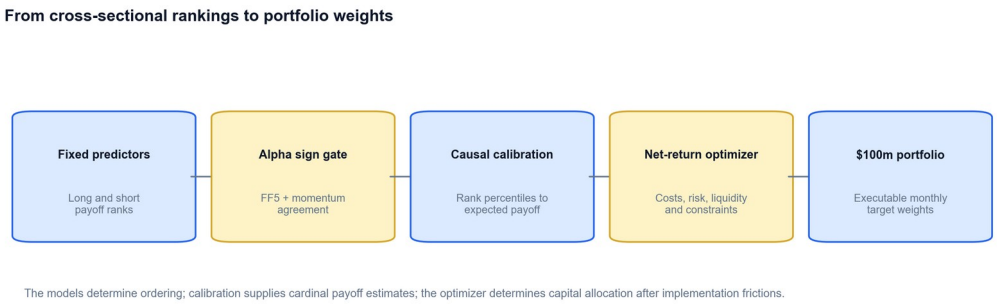
Long payoff is next-period total return minus cash; short payoff is its negative. Costs are charged once in optimization and accounting rather than being duplicated in the calibration target.

6.3 Objective

The portfolio objective maximizes expected monthly stock excess return net of implementation costs:

$$\max_{w} w^{\top}\hat{\mu}_t - C_{trade}(w, w_{t-1}) - C_{borrow}(w) - C_{finance}(w).$$

Risk is controlled through constraints rather than a fixed mean-variance penalty. The resulting objective evaluates expected net return subject to a prespecified feasible set, rather than combining return and risk through an estimated preference parameter.



From rankings to portfolio weights. Source: author's construction.

6.4 Portfolio constraints

The 1.5% specification retains every baseline rule except the maximum position weight:

Constraint	Fixed value
Capital base	$100 million
Maximum gross exposure	1.50
Minimum long gross	0.50
Minimum short gross	0.30
Net exposure interval	-0.10 to 0.30
Maximum absolute position weight	1.50%
One-way monthly turnover	50%
One-way ADV participation	10%
Absolute market and other factor exposure	10%
SIC2 net absolute exposure	5%
SIC2 gross exposure	30%
Forecast-volatility ceiling	10% annualized
Cash	Allowed

The 30% short floor is selected on validation. Removing it changes the entire feasible portfolio rather than only deleting short positions, and the free-short specification performs worse in validation. Inherited holdings that leave the eligible set may be reduced but not increased or reversed.

6.5 Risk model

The risk model combines a point-in-time six-factor covariance matrix with sector-shrunk security-specific variance. It covers market, size, value, profitability, investment, and momentum. The optimizer limits factor and industry exposure but does not impose exact market neutrality; FF12 is not used.

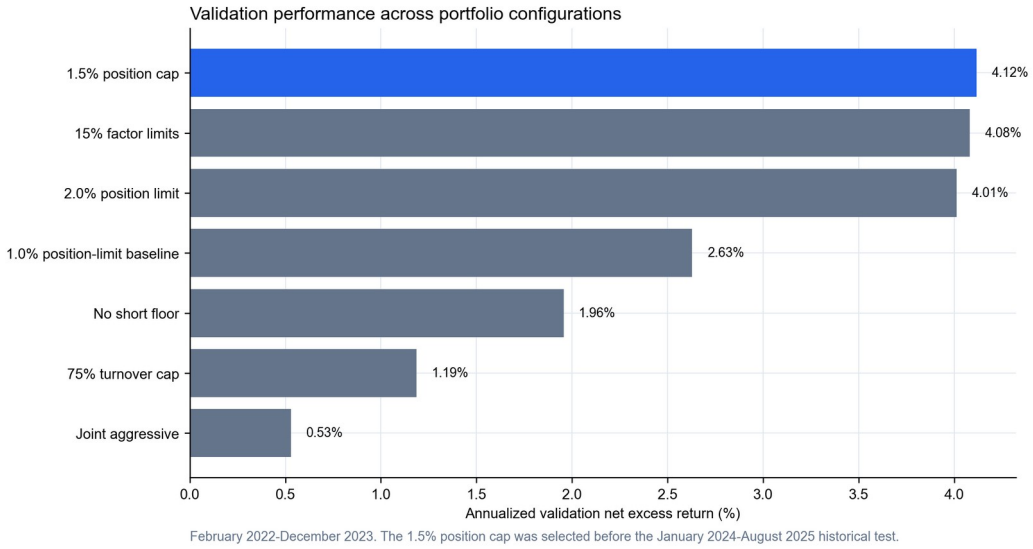
6.6 Cost model

Primary costs include quoted spreads, square-root market impact, stock-specific borrow, and 0.50% annual financing. The stress model doubles spread and impact, applies stressed borrow, and raises financing to 1.50%. Capacity is evaluated at

$100 million. Because the 10% ADV limit binds, the same weights should not be scaled to a larger capital base without a new capacity study.

6.7 Portfolio-configuration selection

Twenty-five prespecified portfolio configurations vary position concentration, side floors, turnover, liquidity, exposure, factor, industry, and volatility limits. Selection is based on annualized net excess return from February 2022 through December 2023. Observations from the subsequent evaluation period do not enter this choice.



Validation performance across representative portfolio configurations. Source: author's calculations.

The 1.5% position limit produces the highest validation net excess return, 4.12%, with a 0.831 Sharpe ratio. The 2.0% limit produces a slightly lower estimate; removing the short floor produces 1.96%, and the broader constraint relaxations also produce lower validation returns. The selected difference is therefore greater concentration in the highest predicted ranks rather than a general relaxation of the feasible set.

6.8 Validation-selected 2.05-times exposure specification

The portfolio-configuration study fixes the security-level weights. A separate validation exercise evaluates the effect of increasing exposure without retraining a predictive model or rerunning the optimizer. Each exposure specification applies a constant multiplier k:

$$w_{i,t}^{H} = k\, w_{i,t}.$$

Spread, impact, financing, borrow, turnover, and NAV are recomputed for every multiplier rather than scaled mechanically. The selection rule maximizes validation CAGR subject to a Sharpe-ratio floor of 0.75 and positive stressed-cost excess return. Using February 2022-December 2023 data only, the rule selects $k = 2.05$; the multiplier is then held fixed for the historical evaluation.

This exercise measures sensitivity to exposure rather than defining an implementable allocation rule. Peak gross exposure reaches 2.76 in validation and 2.62 in the historical evaluation, and the resulting liquidity demands exceed those imposed on the capacity-constrained reference portfolio.

7. Economic results

7.1 Historical-test performance

The 1.5% position-cap portfolio is evaluated from January 2024 through August 2025. It produces:

Metric	1.0% baseline	1.5% specification	Difference
Annualized net excess return	9.94%	14.67%	+4.73 pp
Stressed-cost net excess return	8.33%	12.96%	+4.63 pp
Total-return CAGR	15.85%	21.34%	+5.49 pp
Realized volatility	5.77%	6.32%	+0.55 pp
Net Sharpe ratio	1.723	2.320	+0.597
Stressed-cost Sharpe	1.449	2.054	+0.606
Maximum drawdown	-1.96%	-1.34%	+0.62 pp
Positive months	65%	75%	+10 pp
Net excess return without best two months	5.83%	10.21%	+4.38 pp

The portfolio has positive net excess return in 15 of 20 months. Performance remains positive after excluding the best two months and under the standard stressed-cost model.

7.2 Inference

The 1.5% specification’s mean monthly net excess return has a HAC t-statistic of 3.487. Relative to the 1.0% reference, the annualized difference is 4.726 percentage

points, with a paired HAC t-statistic of 5.135 and a three-month block-bootstrap 95% interval of +2.984 to +6.524 percentage points. These estimates support a difference from the fixed reference within the historical comparison, but they do not establish persistence in future samples.

7.3 Benchmark comparison

The portfolio and SPY end at almost identical wealth: total-return CAGR is 21.339% versus 21.330%. Portfolio volatility is 6.31% versus 10.98%, maximum drawdown magnitudes are 1.34% and 6.89%, respectively, and return correlation with SPY is 0.26. The active arithmetic return is slightly negative. The evidence therefore indicates similar realized wealth accumulation with lower measured risk, rather than positive active performance relative to SPY.

7.4 Exposure and capacity

The portfolio averages 101.3% gross and 27.9% net long exposure, with about 60 longs and 62 shorts. One-way turnover averages 50%, peak ADV participation reaches the 10% limit, and maximum gross exposure is 1.28. Factor, industry, and accounting checks reconcile within economically immaterial solver residue.

7.5 Long and short sleeves

The long sleeve contributes +16.28% annualized before portfolio-level cost. The short sleeve contributes -0.54%, and primary costs subtract 1.07%, leaving 14.67% net excess return.

The short positions do not produce a positive standalone contribution in this sample. They remain part of the joint allocation because covariance, factor exposure, industry concentration and downside protection are evaluated simultaneously. The weak validation t-statistic and negative average contribution imply that the evidence is more persuasive for the long-side ranking and combined portfolio than for independent short-side alpha.

7.6 Why the 1.5% cap helps

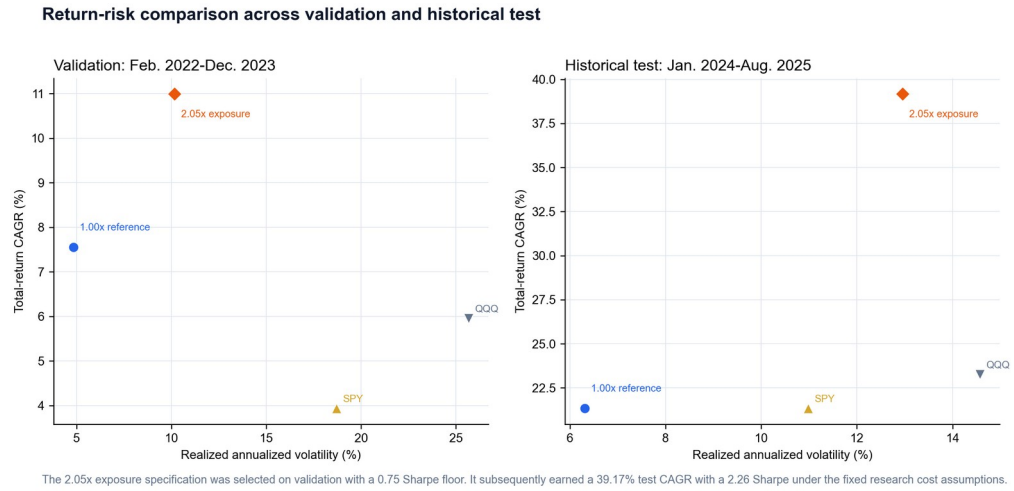
The 1.5% cap does not operate through materially greater leverage: mean gross exposure rises only from 0.972 to 1.013 and annual cost increases by seven basis points. Instead, the optimizer holds fewer and larger positions among the highest predicted ranks. Long contribution rises from 14.13% to 16.28%, the magnitude of the short contribution changes from -3.19% to -0.54%, and the 1.5% specification outperforms the 1.0% reference in 17 of 20 months.

7.7 Validation-selected scaled-exposure specification

The 2.05-times specification addresses the sensitivity of the fixed security-level decisions to greater exposure, subject to a minimum validation Sharpe ratio and positive stressed-cost excess return. The table reports the unscaled reference, the scaled specification and both benchmarks under the same research cost assumptions.

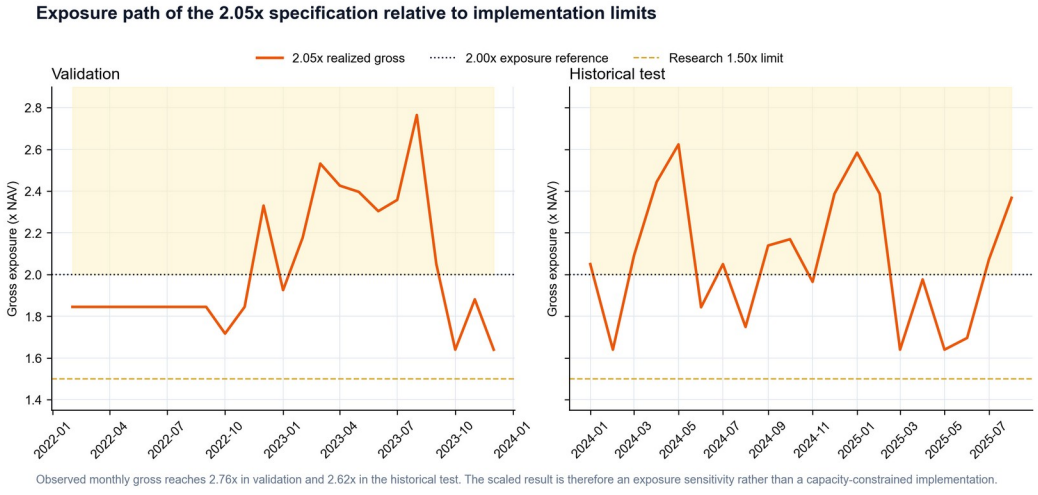
Period	Implementation	CAGR	Volatility	Sharpe	Max. DD	Ending $100
Validation	1.00x reference	7.55%	4.96%	0.831	-2.77%	$114.97
Validation	2.05x exposure specification	10.99%	10.17%	0.752	-6.16%	$122.13
Validation	SPY	3.93%	18.71%	-	-18.37%	$107.67
Validation	QQQ	5.95%	25.67%	-	-27.11%	$111.72
Historical test	1.00x reference	21.34%	6.32%	2.320	-1.34%	$138.04
Historical test	2.05x exposure specification	39.17%	12.96%	2.259	-3.76%	$173.49
Historical test	SPY	21.33%	10.98%	-	-6.89%	$138.02
Historical test	QQQ	23.26%	14.57%	-	-8.63%	$141.69

The scaled specification has a higher realized CAGR than SPY and QQQ in both periods. In validation, its Sharpe ratio is 0.752 and its stressed-cost Sharpe ratio is 0.266, compared with 0.831 for the unscaled reference. In the historical test, it produces a 39.17% CAGR, 29.27% annualized net excess return and a 1.895 stressed-cost Sharpe ratio. The unscaled reference has the higher historical-test Sharpe ratio, 2.320 versus 2.259.



Return-risk comparison across validation and historical evaluation. The capacity-constrained 1.00-times reference, validation-selected 2.05-times exposure specification, SPY and QQQ are shown separately under the same cost assumptions. Source: author's calculations.

The gain comes from scaling, not from different stocks. Mean gross exposure rises from 1.01 to 2.08, peak gross reaches 2.62, turnover reaches 103.8%, and annual cost rises from 1.07% to 2.99%. At $100 million, peak ADV participation reaches 24.4% and the largest position reaches 3.53%.



Monthly gross exposure of the validation-selected 2.05-times specification. The path is shown relative to two-times gross exposure and the 1.50-times limit of the capacity-constrained reference. Source: author's calculations.

The 2.05-times result is therefore interpreted as an exposure sensitivity, while the 1.00-times portfolio remains the capacity-constrained reference. The scaled specification is not supported as a $100 million implementation because its realized exposure and trading intensity exceed the reference capacity limits.

8. Robustness and alternative approaches

8.1 Rank-weighting baselines

Equal weight, linear tail, Gaussian tail, and squared-tail weighting are included as transparent baselines. They show whether ranking quality has value without a cardinal optimizer. These baselines are useful, but they cannot adapt weights to stock-specific cost, covariance, borrow, stale positions, or factor constraints as directly as the expected-net-return optimizer.

8.2 Alpha-gate alternatives

Return-only, CAPM-gated, FF5-plus-momentum-gated and both-alpha-gated portfolios are compared in the initial portfolio study. The validation-selected specification uses the FF5-plus-momentum gate. CAPM remains part of the predictive evidence but is not used in the fixed portfolio specification.

8.3 Free-short and hedge alternatives

Removing the short floor changes the full feasible set, including long allocations, net exposure, covariance, financing and turnover. In validation, the free-short specification produces 1.96% annualized net excess return, compared with 4.12% for the 1.5% specification. Explicit SPY-hedge specifications reduce measured risk but also reduce validation return. The results do not establish a general proposition about mandatory short exposure; they document the role of the short floor within this fixed empirical specification.

8.4 Direct portfolio-weight learning

A direct decision-learning robustness test fits neural policies that learn portfolio weights from all 850 encoded features and from a CAM-free diagnostic view. The full-feature gated policy reaches a historical Sharpe ratio of 0.934, and the CAM-free gated diagnostic reaches 1.004. Both remain below the fixed rank-calibration-optimizer portfolio. Direct decision learning is therefore an interesting future research direction, but it does not outperform the principal specification in this sample.

8.5 Alternative predictive and optimization specifications

The additional robustness study evaluates confidence-adjusted payoffs, opportunity-breadth scaling, specialist mixtures, multitask and multihorizon networks, decision-aware masks, and latent residual risk. Only the three-factor latent-risk specification satisfies the validation criterion. In the historical evaluation it produces a 14.55% annualized net excess return and a 2.282 Sharpe ratio, compared with 14.67% and 2.320 for the principal specification. The paired annualized return difference is -0.115 percentage point, with a 95% block-bootstrap interval of -0.447% to +0.204%. The principal monthly return path is also reproduced exactly from the fixed predictive inputs and portfolio rules. None of the alternatives outperforms it under the stated validation criterion.

8.6 Stress costs

Under doubled spread and impact, higher financing, and stressed borrow, annualized net excess return remains 12.96% and the Sharpe ratio remains 2.054. Stressed annual cost is 2.77%. This robustness is economically important because turnover is at its monthly cap and the portfolio reaches its ADV limit.

8.7 Intramonth intervention and survivor reoptimization

The principal portfolio is rebalanced monthly. A separate post-diagnostic analysis asks whether an exceptional price movement or a daily model signal should trigger an earlier reduction or exit. Each specification starts from the same monthly weights, translated to no more than 1.80 times gross exposure, and applies the same financing and trading-cost assumptions. A signal is executed at the following trading-session close. Intramonth rules may reduce or close an existing holding but cannot introduce a new security or re-enter it before the next monthly rebalance.

The first analysis compares take-profit, stop-loss, trailing-stop, safety, XGBoost, Elastic Net, and ensemble rules. The validation-selected 15% stop-loss performs poorly in the historical evaluation. A five-day depth-2 XGBoost signal raises return but also volatility, leaving its Sharpe ratio below that of the monthly benchmark. A

second analysis evaluates compact three-seed MLP models and prespecified MLP-XGBoost ensembles. The selected five-day ensemble has the highest Sharpe ratio among the daily variants. A final specification retains those warning signals and, after each executed batch, reoptimizes only across the surviving positions. It cannot add a security, re-enter an exited name, reverse a position, or restore more gross exposure than the batch removed; cash remains feasible.

January 2024-August 2025 policy	Ann. net excess return	Volatility	Sharpe	Max. drawdown	Paired HAC t vs monthly
Monthly 1.80x exposure benchmark	21.63%	11.03%	1.961	-8.21%	-
Five-day XGBoost signal	23.00%	12.03%	1.913	-6.58%	0.62
Five-day MLP-XGBoost signal; cash after exit	25.18%	11.81%	2.132	-6.47%	0.96
MLP-XGBoost signal plus survivor reoptimization	25.67%	12.14%	2.114	-6.40%	0.92

Survivor reoptimization restores 41.7% of the gross exposure released by the warning signals. Its annualized return is 0.49 percentage point above that of the cash-after-exit ensemble, but its Sharpe ratio is slightly lower and the paired HAC t-statistic of the difference is 0.64. None of the daily specifications demonstrates a statistically supported improvement over the monthly specification. They are therefore reported as post-diagnostic evidence rather than as part of the principal result.

9. Limitations

9.1 Historical validation boundary

The empirical design uses point-in-time timestamps, admits labels only after their availability dates, and places execution after signal formation. The portfolio specification is selected using February 2022-December 2023 data before evaluation from January 2024 through August 2025. Chronology, target construction, and portfolio accounting checks identify no look-ahead or target leakage in the reported

estimates. Nevertheless, this evidence remains historical. The intramonth analyses were designed after earlier outcomes had been examined and are therefore reported only as post-diagnostic robustness exercises.

9.2 Theory-free specification and structure-versus-noise uncertainty

The broad information set is selected for empirical coverage rather than derived from a structural model of expected returns. This makes it possible to search for nonlinear predictive regularities, but it sharply limits economic interpretation. Many inputs are correlated transformations, coverage indicators, or proprietary signals; a flexible model may therefore exploit sample-specific correlations, provider regimes, omitted exposures, or other unstable relationships. Walk-forward estimation, regularization, validation-period model choice, label-shuffle controls, correlation-reduced views, and linear benchmarks reduce this risk, but they cannot determine whether the fitted relation is durable structure or noise.

A positive rank IC shows historical ordering, and a positive net backtest shows that the fixed rule was economically successful under the stated sample and implementation assumptions. Neither result identifies the underlying mechanism or proves that the signal represents a tradable opportunity that will persist. This limitation is especially important because the final evaluation contains only 20 months. A credible distinction between durable opportunity and noise requires prospective evidence from the unchanged specification.

9.3 Short test window

Twenty monthly observations cannot establish regime durability. The test includes positive and negative equity months but remains dominated by a strong market, so even a statistically positive Sharpe estimate is uncertain.

9.4 Weak short-side validation

The selected short-payoff model has a validation IC of 0.0256 but a HAC t-statistic of only 0.97. Its historical test is stronger, yet the average final short-sleeve contribution

is negative. The thesis therefore cannot claim that the short selector independently produces reliable short alpha.

9.5 Capacity

The capacity-constrained reference reaches its 10% ADV limit at $100 million. The 2.05-times specification reaches 24.4% of ADV and 103.8% one-way turnover, so it is an exposure sensitivity rather than evidence of unconstrained capacity. The common panel begins only in 2015, and alternative-data coverage changes across firms and time. Any prospective application would require renewed capacity estimates and independent verification of the portfolio constraints under the intended capital base.

9.6 Benchmark comparison and external validity

The capacity-constrained portfolio matches rather than exceeds SPY's realized CAGR. The 2.05-times specification has higher realized CAGRs than SPY and QQQ, but its 20-month active-return bootstrap intervals include zero; superiority to the benchmarks is therefore not statistically established. Moreover, the scaled result depends on leverage, financing, and liquidity assumptions that cannot be validated by a historical simulation alone. These estimates should therefore be interpreted as evidence from the stated sample and implementation model, not as a forecast of attainable future returns.

10. Conclusion

This thesis conducts a purely empirical test of whether machine-learning methods can extract cross-sectional predictive information from a high-dimensional, point-in-time panel of U.S. equities. It does not propose a structural asset-pricing theory or interpret the broad predictor panel as a set of priced risk factors. The selected specifications produce positive evaluation-period rank information coefficients between 0.048 and 0.067 for CAPM-adjusted returns, FF5-plus-momentum-adjusted returns, and directional net payoffs. For the two model-relative alpha targets, the

corresponding HAC t-statistics are 3.68 and 7.51. Within the stated benchmark definitions and sample, these estimates reject the narrow null hypothesis of zero cross-sectional rank predictability.

This result should not be interpreted as a comprehensive rejection of CAPM or FF5-plus-momentum. The estimated alphas are residuals relative to particular factor specifications, and predictability may reflect omitted risks, model misspecification, changing factor exposures, genuine mispricing, or an unstable sample-specific relation. The evidence establishes that the selected information set predicts the ordering of model-relative residual returns in the historical sample. It does not identify the economic mechanism and, by itself, cannot determine whether the rankings represent durable opportunities or noise. Distinguishing among these explanations would require theory-guided asset-pricing tests and prospective evidence beyond the scope of this dissertation.

The economic analysis evaluates whether the estimated ordering remains relevant after implementation costs. The fixed ranks are converted into expected-payoff estimates using only previously observed outcomes and supplied to an optimizer subject to explicit cost, covariance, liquidity, exposure, turnover, and capacity restrictions. The 1.5% position cap selected in the validation period permits greater concentration in the highest predicted ranks without changing the predictive models.

In the January 2024-August 2025 historical test, the unscaled specification produces a 14.67% annualized net excess return, 6.32% volatility, a 2.320 Sharpe ratio and a maximum drawdown magnitude of 1.34%. Its total-return CAGR is approximately equal to SPY’s, while its realized volatility and drawdown are lower. The scaled 2.05-times specification raises historical-test CAGR to 39.17% and retains a 2.259 Sharpe ratio, but it requires peak gross exposure of 2.62 times NAV and reaches 24.4% of average daily volume at the $100 million scale. The scaled result therefore describes sensitivity to exposure rather than additional predictive content.

The long positions account for most of the estimated portfolio return. The short positions contribute negatively on average, and the short-payoff model has weak

validation inference. The evidence for independent short-side alpha is therefore substantially weaker than the evidence for the long-side ranking and the combined portfolio. Post-diagnostic daily intervention exercises do not satisfy the criterion for replacing the monthly specification.

The findings support conditional cross-sectional return predictability and positive historical net-of-cost performance in the stated sample. They do not establish a universal failure of the factor benchmarks, identify a new non-diversifiable risk factor, or prove persistent real-time alpha. The absence of a structural theory, the high dimensionality of the information set, and the 20-month evaluation window make the distinction between durable structure and noise particularly difficult. Capacity constraints also bind, and prospective execution has not yet been observed. External validity and practical tradability must therefore be assessed using the unchanged specification on subsequently realized data.

References

Asquith, Paul, Michael B. Mikhail, and Andrea S. Au. 2005. ‘Information Content of Equity Analyst Reports.’ *Journal of Financial Economics* 75 (2): 245–82. https://doi.org/10.1016/j.jfineco.2004.01.002.

Carhart, Mark M. 1997. ‘On Persistence in Mutual Fund Performance.’ *The Journal of Finance* 52 (1): 57–82. https://doi.org/10.1111/j.1540-6261.1997.tb03808.x.

Chen, Luyang, Markus Pelger, and Jason Zhu. 2024. ‘Deep Learning in Asset Pricing.’ *Management Science* 70 (2): 714–50. https://doi.org/10.1287/mnsc.2023.4695.

Chen, Tianqi, and Carlos Guestrin. 2016. ‘XGBoost: A Scalable Tree Boosting System.’ In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–94. https://doi.org/10.1145/2939672.2939785.

DeMiguel, Victor, Alberto Martín-Utrera, Francisco J. Nogales, and Raman Uppal. 2020. ‘A Transaction-Cost Perspective on the Multitude of Firm Characteristics.’ *The Review of Financial Studies* 33 (5): 2180–2222. https://doi.org/10.1093/rfs/hhz085.

Fama, Eugene F., and Kenneth R. French. 2015. ‘A Five-Factor Asset Pricing Model.’ *Journal of Financial Economics* 116 (1): 1–22. https://doi.org/10.1016/j.jfineco.2014.10.010.

Frazzini, Andrea, Ronen Israel, and Tobias J. Moskowitz. 2018. ‘Trading Costs.’ SSRN. https://doi.org/10.2139/ssrn.3229719.

Freyberger, Joachim, Andreas Neuhierl, and Michael Weber. 2020. ‘Dissecting Characteristics Nonparametrically.’ *The Review of Financial Studies* 33 (5): 2326–77. https://doi.org/10.1093/rfs/hhz123.

———. 2025. ‘Missing Data in Asset Pricing Panels.’ *The Review of Financial Studies* 38 (3): 760–802. https://doi.org/10.1093/rfs/hhae003.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu. 2020. ‘Empirical Asset Pricing via Machine Learning.’ *The Review of Financial Studies* 33 (5): 2223–73. https://doi.org/10.1093/rfs/hhaa009.

Jensen, Theis Ingerslev, Bryan Kelly, Semyon Malamud, and Lasse Heje Pedersen. 2026. ‘Machine Learning and the Implementable Efficient Frontier.’ *The Review of Financial Studies*. https://doi.org/10.1093/rfs/hhag022.

Kelly, Bryan, Seth Pruitt, and Yinan Su. 2019. ‘Characteristics Are Covariances: A Unified Model of Risk and Return.’ *Journal of Financial Economics* 134 (3): 501–24. https://doi.org/10.1016/j.jfineco.2019.05.001.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh. 2020. ‘Shrinking the Cross-Section.’ *Journal of Financial Economics* 135 (2): 271–92. https://doi.org/10.1016/j.jfineco.2019.06.008.

Loughran, Tim, and Bill McDonald. 2011. ‘When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks.’ *The Journal of Finance* 66 (1): 35–65. https://doi.org/10.1111/j.1540-6261.2010.01625.x.

Markowitz, Harry. 1952. ‘Portfolio Selection.’ *The Journal of Finance* 7 (1): 77–91. https://doi.org/10.1111/j.1540-6261.1952.tb01525.x.

Sharpe, William F. 1964. ‘Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk.’ *The Journal of Finance* 19 (3): 425–42. https://doi.org/10.1111/j.1540-6261.1964.tb02865.x.

Tetlock, Paul C. 2007. ‘Giving Content to Investor Sentiment: The Role of Media in the Stock Market.’ *The Journal of Finance* 62 (3): 1139–68. https://doi.org/10.1111/j.1540-6261.2007.01232.x.

Tetlock, Paul C., Maytal Saar-Tsechansky, and Sofus Macskassy. 2008. ‘More Than Words: Quantifying Language to Measure Firms’ Fundamentals.’ *The Journal of Finance* 63 (3): 1437–67. https://doi.org/10.1111/j.1540-6261.2008.01362.x.