BEHAVIOR OF PRICE CONVERGENCE TO FUNDAMENTAL VALUES

ANALYSIS OF MACHINE LEARNING-BASED VALUE-TO-PRICE STRATEGY RETURNS OF UNDERVALUED STOCKS

STETTNER, PATRICK LUKAS (42668)

WESTERMANN, NICO (42690)

Master Thesis

Stockholm School of Economics

2025

Behavior of price convergence to fundamental values: Analysis of machine learning-based value-to-price strategy returns of undervalued stocks

Abstract:

This study investigates whether machine learning (ML)–based intrinsic value estimates enhance the performance of value-to-price (V/P) investment strategies relative to traditional theory-driven valuations. We replicate a Gradient Boosting Machine (GBM) model to generate monthly intrinsic values for U.S. equities and form high V/P quintile portfolios using both ML and Residual Income Model (RIM) valuations from 1984 to 2024. ML-based portfolios outperform RIM-based and market portfolios in cumulative and risk-adjusted terms, with equal-weighted alphas of 6.3% and value-weighted alphas of 2.5%, all for portfolios with a 3-year holding period. Monthly portfolio formation using ML estimates does not improve short-term returns below one year, indicating that superior performance is based on more accurate long-term valuation signals rather than increased portfolio turnover. Our results demonstrate that ML-based, peer-driven intrinsic value estimates provide robust and economically meaningful signals for value investing, offering an advantage over traditional RIM-based approaches.

Keywords:

Machine Learning, Intrinsic Value, Value Investing, Portfolio Strategy, Asset Pricing

Authors:

Stettner, Patrick Lukas (42668)

Westermann, Nico (42690)

Supervisor:

Magnus Dahlquist, Professor, Department of Finance

Master Thesis
Master Program in Finance
Stockholm School of Economics

Acknowledgment

We would like to sincerely thank our academic supervisor, Magnus Dahlquist, Handelsbanken Professor of Finance at the Stockholm School of Economics, for his guidance and expertise throughout this thesis. His thoughtful feedback and insightful suggestions greatly improved our analysis and strengthened the overall quality of our work.

Further, we are grateful to Sina Molavipor, Researcher at SEBx, for his valuable perspectives and constructive input, which clarified complex aspects and enriched our understanding of machine learning techniques and applications.

Contents

1. Introduction

One of the key questions in value investing is determining an accurate estimate of the real intrinsic value of a firm; a true value which is unobservable in itself. Several theory-based models have been developed over time, both from practitioner- and research-driven perspectives. (e.g., Pinto et Al. (2019)) Lately, the rise of machine-learning (ML) value estimates has created a new, more flexible, data- and peer-driven approach to intrinsic value estimation. Especially in the context of value investing, where one aims to identify stocks that, given the firm's underlying fundamentals, seem to be undervalued by the market, the comparison by peers can bring new insights, and more precise value estimates bear the potential to generate more accurate value-to-price (V/P) ratios and, in turn, higher returns when applied in a V/P investment strategy (Frankel & Lee, 1998).

Constructing V/P ratios from ML-derived intrinsic values and analyzing the resulting time series of realized returns allows us to assess whether more accurate, data-driven valuations translate into superior investment performance. ML value estimation models have typically been evaluated by comparing the V/P gap and price or earnings movement over a fixed horizon, such as one month or one year ahead like Chen et Al. (2022). We develop this one step further by implementing ML value estimates into the traditional V/P investing strategy, which previously relied on theory-based approaches such as residual income model (RIM) value estimates. (Ohlson, 1995) Furthermore, due to data limitations, the RIM usually provides only one value estimate per year, while some ML models can offer estimates on a higher temporal frequency.

In this study, we replicate a ML valuation approach by Geertsma and Lu (2023) using Gradient Boosting Machines, input these value estimates into a V/P investing strategy, and analyze the resulting portfolio returns. We benchmark these returns against those obtained using RIM-based value estimates. Additionally, we explore the opportunity of using monthly-formed portfolios, taking advantage of the higher frequency of ML-derived valuations.

2. Literature Background and Hypothesis

Foundations of Value Investing and Intrinsic Value

The concept of value investing rests on the notion that market prices occasionally fall below a firm's intrinsic value, implying that prices are expected to rise over time as they converge toward their true fundamental value. (Priel & Rokach, 2024) Such deviations can be viewed as temporary outliers in how the market prices fundamentals across firms. If one can accurately model how the market typically values firm characteristics, then deviations from that expected relationship may signal mispricing and potential investment opportunities. Understanding how prices converge toward intrinsic value first requires an examination of what constitutes intrinsic value and how it can be estimated.

Traditional Intrinsic Value Estimation models

Traditionally, a firm's intrinsic value is understood as the present value of all future discounted cash flows. (Williams, 1938) However, this measure is not directly observable, which has led researchers to develop various models to estimate it. Traditional approaches may include classic valuation models such as discounted cash flow. (Pinto, Robinson, & Stowe, 2019) These models depend on assumptions about future variables such as growth rates. To overcome this limitation, researchers have proposed refined valuation frameworks to enhance the precision of intrinsic value estimates.

Among these approaches, one of the widely used ones in the context of value investing research and V/P trading strategies is the Residual Income Model (RIM) proposed by Ohlson. (Ohlson, 1995) He uses current book value and adds cash flows not explained by the Capital Asset Pricing Model (CAPM) as residual income into the RIM.

Traditional valuation models such as the RIM are theory-driven, estimating what a firm's price should be based on accounting and finance principles. While they provide a structured framework, their reliance on fixed functional forms and linearity assumptions limits predictive power. In contrast, machine learning models can be data-driven and peer-based, learning how the market prices fundamentals on average across firms without the previously mentioned restriction. Deviations between model-implied and observed prices may thus reflect either model error or genuine mispricing, with the latter indicating potential excess-return opportunities. (Geertsma & Lu, 2023)

Machine Learning Valuation models

Machine learning (ML) methods have become increasingly relevant in financial valuation, being applied across diverse domains such as financial instruments (Beereddy, 2021), land valuation (Jafary, Shojaei, Rajabifard, & Ngo, 2024), and equity selection based on fundamentals analysis as described in Buczynski et Al. (2021), Nobre and Neves (2019), and Li (2024). Hanauer et Al. (2022) used models to pick stocks with relevance for a value investing strategy. Several studies show that ML-based valuation models outperform traditional methods when predicting returns derived from firm fundamentals. (e.g. Geertsma and Lu (2023)) A key advantage of ML models is the ability to relax linearity assumptions, an often-cited limitation of traditional valuation approaches. (Mignot, 2025), (Wolff & Echterling, 2023), (Kim & Ryu, 2015)

Multiple ML models have been used in literature for value and fundamentals investing: Huang et Al. (2021) employ Feed-forward Neural Networks, Random Forests, and Adaptive Neural Fuzzy Inference Systems to predict stocks' behavior based on fundamentals; Shen (2011) applies Artificial Neural Networks to implement a value strategy; and Hanauer et Al. (2022) use tree-based models to identify mispricing.

Geertsma & Lu (2023) employ Gradient Boosting Machines (GBM) to estimate the current intrinsic firm values based on fundamentals, they measure accuracy by seeing whether the price moves directionally towards their value estimate. Their GBM models deliver substantially more accurate valuations than traditional approaches to predict returns based on next month price movements towards their value estimate, with errors about 25–30% lower than traditional models. Further, the GBM has an out-of-sample R-squared (R^2) of 0.76-0.78, whereas the best traditional model attains a R^2 of 0.51. Thus, the GBM models have a better out-of-sample explanatory power by approximately 0.25 R^2 points. Abnormal returns stay significant, even after controlling for up to 13 different risk factors. (Geertsma & Lu, 2023)

Building on their findings, our study examines whether the ML-based valuation model of Geertsma and Lu (2023), deployed in the time frame from 1980 to 2019, maintains comparable valuation performance in the true out-of-sample time period from 2020 to

2024. For our first hypothesis **(H1)** we formally state our null hypothesis as the assumption that the model's estimation accuracy substantially deteriorates in the more recent market environment, and H1 as the alternative that the model remains robust and continues to deliver meaningful valuation across periods.

Value-to-Price trading strategy

The book-to-market (B/M) ratio has long been established as a key value factor in asset pricing, with Fama and French (1993) showing that portfolios sorted on B/M systematically capture differences in average stock returns, reflecting a persistent value premium and exposure to common risk factors such as size and book-to-market. Portfolio sorts as an empirical tool, however, date back earlier (e.g., Basu (1977); Banz, (1981)), providing a systematic way to study return patterns across firm characteristics.

Based on Ohlson's RIM framework, Frankel & Lee (1998) show that a long-short trading strategy based on sorting for Value-to-Price ratio (V/P) quintiles provides significant abnormal returns compared to the capital asset pricing model (CAPM) over a period of three years. They find that the V/P ratio is just as good as Book-to-Price strategy (B/P), where price is the equivalent of Fama and French's market price, in the first year, but over a longer time horizon of three years a V/P strategy creates returns up to twice the returns of a B/P strategy; both the outperformance of V/P over B/P (Gonçalves & Leonard, 2023) as well as the preference for longer holding periods within value investing (e.g., Huefner et Al. (2023)) are supported by later research.

While the initial question was raised whether those returns are due to mispricing or rather due to an unobserved risk factor, Ali et Al. (2003) as well as others (e.g., Bartram and Grinblatt (2018)) show that even after controlling for several risk factors a V/P strategy still carries abnormal returns.

However, despite this insight into return development over time, most recent studies with ML models focus on the predictive power of intrinsic value for returns or earnings at fixed horizons (e.g., one-year ahead (Chen, Cho, Dou, & Lev, 2022)), while few investigate how returns and valuation gaps evolve dynamically over time; a limitation also highlighted by Huefner et Al. (2023).

Convergence behavior

To understand convergence dynamics, it is important to distinguish between return behavior and valuation-gap changes reflected in the V/P ratio. Johnson & Xie (2004) examine extreme V/P stocks by shorting the lowest and longing the highest V/P quintiles in a buy-and-hold strategy to identify which stocks drive returns. Consistent with prior work, they find that returns accrue mainly after the first year, reflecting short-term momentum. The high V/P quintile outperforms the low one, likely due to transaction costs and short-sale constraints. Most excess returns arise from converging stocks, those moving from extreme to middle quintiles, driven primarily by price rather than value adjustments.

It appears that the behavior or extent of over-/undervaluation over time is less investigated than the returns of V/P strategies. Notably, Capozza & Israelsen (2010) conclude that within one year circa 15-30% of the difference between value estimate and stock price are removed.

Taken together, the literature reveals a gap that motivates this study. Although ML models have demonstrated superior intrinsic-value accuracy relative to traditional RIM approaches, they have rarely been used to construct V/P ratios or generate return time series, leaving open whether more accurate, data-driven estimates translate into stronger investment performance. To build a more holistic understanding of mispricing dynamics, we jointly analyze the evolution of returns and valuation ratios over time, testing whether ML-based V/P sorting yields significant abnormal returns consistent with gradual price convergence toward intrinsic value (**H2**). A secondary benefit of the ML framework is its higher valuation frequency, unlike RIM, creating the possibility of higher-frequency portfolio formation and allowing us to examine how the information content of ML-derived signals evolves over time. By comparing ML-based V/P portfolios with RIM benchmarks, we assess both the economic relevance of ML valuations and the potential value of more frequently updated signals.

3. Data and Methodology

3.1. Data Selection and Preparation

We use Wharton Research Data Services (WRDS) as the source for all price data as well as all data needed for the intrinsic value estimations through ML and the RIM. Our time series starts on Jan 1980 and includes all monthly price data and quarterly fundamentals data until Dec 2024.

For the ML Model, we follow the approach by Geertsma and Lu (2023). We retrieve return data from the Center for Research in Security Prices (CRSP) on a monthly basis, while we retrieve the quarterly fundamentals (see Appendix A.1 for a detailed list of all the variables in the ML Model) from Compustat. We include all equities listed on AMEX, NASDAQ, and NYSE where market equity value is available, and we exclude penny stocks by filtering for total assets, sales and book equity to be above the 10th decile every month. As Geertsma and Lu we avoid look-ahead bias by using a three-month lag to all fundamentals, indicating that investors only get access to fundamentals a quarter after the reporting date. To produce a monthly panel, if there is no new fundamental data we take the previous month's fundamentals as this is also the information that would be available to an investor. This approach creates a monthly panel dataset with fundamentals and pricing data of 13,658 individual firms forming 1,464,081 individual firm-months.

We source the data for the RIM from Compustat while analyst forecasts are obtained from I/B/E/S, following Frankel & Lee (1998). We include all equities listed on AMEX, NASDAQ, and NYSE for which book values, EPS forecasts, and prices are available, and we exclude firms with negative book values, extreme future ROEs (FROE > 100%), or prices below \$1. We align Compustat fiscal-year-end data for year *t*-1 with I/B/E/S EPS forecasts for year *t* to ensure that all inputs to the RIM are publicly available prior to portfolio formation. We include firms with fiscal-year-ends between June and December and use May forecasts, guaranteeing that one- and two-year-ahead EPS correspond to the correct fiscal year. Those filters are aligned with Frankel and Lee. As I/B/E/S data only start in 1980 and given our three-period RIM, we have a lack of data for the period of 1980-1983. Hence, to ensure sufficient data availability as well as model consistency, all return analysis and comparisons are done for the time frame 1984-2024.

3.2. Methodology

The ML valuation model provides intrinsic-value estimates at monthly intervals, which allows us to refresh the V/P ratio and form portfolios at a higher frequency than is possible with RIM, which updates intrinsic value only once per year. When comparing the two models, we use yearly portfolios with the same value estimation month for both models. Both the ML-based and RIM-based strategy are evaluated using monthly realized returns. We subsequently examine whether the ML-based model benefits from its ability to update valuations monthly. This setup enables us to test firstly whether the ML-based intrinsic values generate economically meaningful differences in portfolio performance compared to RIM-based ones, and secondly whether the additional flexibility of monthly ML updates provides incremental information beyond the traditional annual signal.

3.2.1. Intrinsic Value Estimation

Previous literature has indicated that ML models using fundamental values outperform traditional approaches. By estimating ML-based intrinsic values and benchmarking ML-value based V/P returns with RIM-value based V/P returns as traditional approach, we investigate whether a more sophisticated ML value also outperforms in the context of a V/P trading strategy.

Machine Learning Value Estimation

We estimate intrinsic values using the machine learning intrinsic value estimates with the same methodology as Geertsma and Lu (2023). They use a Gradient Boosting Machines (GBM) as a decision-tree based algorithm enables robust out-of-sample predictions without overfitting. The general rational behind this peer-based valuation approach is to build a common benchmark of the value of a firm based on its fundamentals and then based on a company's peers evaluate that specific entity's intrinsic value. Following this logic, every month the model is trained on the companies in the dataset and based on the training dataset it provides estimates for the three valuation multiples market-to-book (*m2b*), enterprise-value-to-assets (*v2a*) and enterprise-value-to-sales (*v2s*). Combining those multiples with a company's fundamentals in that month we get intrinsic value estimates. For the model training, the data is split into a training, validation, and testing dataset with a ratio 60:20:20. In every month a randomly selected group of companies is allocated in one of five blocks, where one of those blocks is the out-of-sample dataset.

Geertsma & Lu use metrics such as percentage valuation errors of the predicted value of equity (ε), RMSE, out-of-sample R-squared (r_{oos}^2), and the Pearson correlation between actual and predicted equity values ($\rho_{M,\bar{M}}$) to evaluate model performance. For details on precision metrics and the model construction see Appendix A.2 and A.3.

In addition to replicating the original model, we evaluate its performance in a true out-of-sample setting by extending the analysis beyond the original period (Jan 1980 – Dec 2019) to the subsequent period (Jan 2020 – Dec 2024). This approach tests whether the model's predictive accuracy is specific to the timeframe used in the original study or whether it maintains robust valuation performance in a completely new period. By applying the model to a period not used in the original paper, we can assess the temporal generalizability of the results and the extent to which the original findings reflect genuine predictive power rather than period-specific tuning.

Finally, we compare the value estimates, derived from the three log-version multiples, as well as the average of several estimates and check for the most precise model; we will use this one further on to produce the required value estimate.

Residual Income Value Estimation

As a traditional approach, we compute the intrinsic value of each firm using the RIM as described by Frankel & Lee (1998) based on Ohlson (1995). The RIM expresses firm value as sum of current book value and present value of expected future residual income:

$$V_0 = B_0 + \sum_{t=1}^{\infty} \frac{RI_t}{(1+r)^t}, \text{with } RI_t = NI_t - rB_{t-1} \tag{1}$$

where B_0 is current book value per share, NI_t is the forecasted net income per share, and r is the cost of equity.

For each fiscal year t, we start with the book value per share B_t from Compustat and forecast one- and two-year-ahead EPS (F_EPS1, F_EPS2) from I/B/E/S, along with the long-term growth rate (LTG). Forward-looking returns on equity $FROE_{t+1}$ and $FROE_{t+2}$

are derived from these forecasts, and the expected residual incomes for years $t + 1$ and $t + 2$ are computed as

$$RI_{t+1} = FROE_{t+1} \cdot B_t - r \cdot B_t \qquad RI_{t+2} = FROE_{t+2} \cdot B_{t+1} - r \cdot B_{t+1} \tag{2}$$

where r is the cost of equity and B_{t+1} is the book value updated for net income and dividends. The intrinsic value V_t is then obtained as:

$$V_t = B_t + \frac{RI_{t+1}}{1 + r} + \frac{RI_{t+2}}{(1 + r)^2} + \frac{RI_\infty}{(1 + r)^2} \tag{3}$$

with the terminal residual income RI_∞ calculated using *LTG*. When *LTG* is missing, we set long-term residual income growth to zero. Forward book values and residual incomes are checked for economic plausibility, and extreme values are trimmed to avoid distortion in the resulting valuations. Finally, value-to-price ratios (*V/P*) are computed for cross-sectional comparison, ensuring consistency between forecasted ROEs, book values, and market prices, and replicating the methodology and assumptions of Frankel & Lee. Appendix A.4 describes details on edge cases, including missing *LTG* and adjustments for negative or zero earnings.

The difference between model-implied and observed prices can reflect either (i) model misspecification error, or (ii) genuine market mispricing. While the former reflects a lack of explanatory power, the latter represents exploitable deviations from fundamental value. To disentangle these, we assess whether their residuals forecast future returns. Following Geertsema & Lu (2023), a model yielding return-predictive residuals suggests it captures fundamental value while market prices reflect temporary mispricing.

3.2.2. V/P Strategy and Return Analysis

We focus our empirical analysis on understanding how intrinsic value estimates derived from the ML and the RIM translate into realized stock returns. In particular, we examine whether portfolios formed on the basis of value-to-price (V/P) ratios yield significant abnormal returns relative to standard asset pricing benchmarks. Our analysis centers on four core dimensions: (i) portfolio sorting and weighting, (ii) cumulative return performance, (iii) risk-adjusted performance, and (iv) robustness to sorting frequency. Complementary aspects such as crisis behavior, industry patterns, and convergence timing are discussed as extensions.

Portfolio Sorting and Quintile Investing

We follow the established portfolio-sorting procedure of Frankel and Lee (1998). Each June, all firms are ranked by their estimated V/P ratios, computed either from ML-based or from the RIM-based intrinsic value estimates, and are then divided into quintiles. Portfolios are held for periods of minimum one year, but often also longer, and hence, overlap annually, following the convention in the literature.

Because our study focuses on undervalued stocks, we form a long-only portfolio of firms in the highest V/P quintile (i.e., those with the greatest estimated underpricing). Prior research has shown that the majority of the value-premium effect is concentrated in this segment, while short positions in the lowest V/P quintile contribute still significant but less additional returns, potentially due to transaction costs and short-sale frictions.

To assess the robustness of the results, we compute portfolio returns using both value-weighted (VW) and equal-weighted (EW) schemes. VW returns reflect the market capitalization-weighted performance of the portfolio, emphasizing larger firms, whereas EW returns give each firm an identical weight, thereby highlighting the behavior of smaller and mid-cap firms that may be more prone to mispricing. To mitigate distortions from illiquid or extremely small firms, we exclude the bottom decile of firms based on total assets, sales, and book equity each month, as described in the data section.

We monitor both cumulative and periodic portfolio returns over the several holding period (1, 2, and 3 years). Cumulative returns reflect the compounded performance over time, while periodic returns provide month-to-month variation used in risk-adjustment analyses.

Cumulative Return Analysis

We first evaluate the unconditional performance of ML- and RIM-based V/P portfolios by comparing cumulative returns for both VW and EW weighting schemes. This provides an initial view of how strongly estimated undervaluation translates into realized gains and whether the ML-based valuation approach outperforms the traditional RIM benchmark.

We benchmark cumulative returns against the average market return (proxied by the CRSP value-weighted market index) to assess relative performance. Furthermore, to understand the role of investment horizon, we initially compute cumulative returns for several holding periods (1, 2, and 3 years). Afterwards, if not specified otherwise, we use the three-year holding period returns for subsequent analyses to maintain focus and comparability with prior work.

This cumulative return analysis serves as the initial insights into the economic relevance of ML-based intrinsic value estimates.

Risk-Adjusted Performance

To assess whether the potentially observed excess returns represent compensation for risk or genuine mispricing, we perform multi-factor regressions using the Fama–French five-factor (FF5) (Fama & French, 2015) model in combination with the Momentum factor. Those factors include the market (market-minus-"risk-free rate": mktrf), size (small-minus-big: smb), value (high-minus-low: hml), profitability (robust-minus-weak: rmw), investment pattern (conservative-minus-aggressive: cma), and momentum (up-minus-down: umd). Monthly excess portfolio returns are regressed on the corresponding risk factors, and we estimate alphas using Newey–West adjusted standard errors (Newey & West (1987)) to correct for autocorrelation with a lag using Newey and West's (1994) automatic lag selection procedure to determine the maximum lag order of correlation.

A statistically and economically significant positive alpha would indicate that ML-based intrinsic value estimates capture persistent mispricing beyond traditional risk exposures.

Robustness to Sorting Frequency

As a robustness check, we evaluate whether the frequency of portfolio formation affects the performance of the ML-based strategy. Specifically, we compare abnormal returns obtained under annual (June-to-June) sorting, our baseline specification, with those based on monthly sorting using updated ML intrinsic value estimates. This comparison allows us to assess whether the ML-based V/P signal is stable over time or whether its predictive power decays quickly and requires frequent rebalancing.

If performance remains stable across frequencies, this suggests that the ML valuation signal is persistent and less sensitive to timing; conversely, a decline or increase under monthly sorting could imply short-lived or noise-driven predictive effects.

4. Results

4.1. Machine Learning Model Performance

4.1.1. Replication of Geertsma & Lu (1980-2019)

Table 1 reports the valuation performance of the ML models across the three base target multiples market-to-book (m2b), enterprise-value-to-assets (v2a), and enterprise-value-to-sales (v2s) in their log-transformed form. Following Geertsma and Lu (2023), we evaluate model precision using median and mean percentage valuation errors (|ε|), root mean squared errors (RMSE), out-of-sample R² for both the target variable and predicted equity value, and the Pearson correlation between actual and predicted values ($\rho_{m,\hat{m}}$).

TABLE 1

Model performance comparison for different ML models based on peer multiple (1980-2019 & 2020-2024)

Model	med(|ε|)	avg(|ε|)	med(ε)	avg(ε)	sd(ε)	rmse($\hat{y}$)	rmse($\hat{M}$)	$r^2_{oos}(\hat{y})$	$r^2_{oos}(\hat{M})$	$\rho_{m,\hat{m}}$
Time period 1980-2019 (Original)										
ln_m2b	0.2699	0.4372	0.2754	0.1578	1.0724	0.5230	9148748	0.5485	0.7577	0.8901
ln_v2a	0.2731	0.4470	0.2791	0.1652	1.1563	0.5228	8447700	0.7867	0.7934	0.9028
ln_v2s	0.2767	0.4584	0.2831	0.1720	1.2390	0.5349	8307617	0.7402	0.8002	0.9043
Time period 2020-2024 (Testing)										
ln_m2b	0.3521	0.6069	0.3626	0.2698	1.4042	0.6631	49673995	0.5887	0.6979	0.8356
ln_v2a	0.3535	0.6113	0.3650	0.2733	1.5803	0.6530	50638435	0.7427	0.6860	0.8288
ln_v2s	0.3591	0.6177	0.3721	0.2756	1.2989	0.6647	51104283	0.7115	0.6802	0.8249

Table 1: Performance metrics across three peer multiple specifications for 1980-2019 (original) and 2020-2024 (testing) periods. The ln_m2b model exhibits the lowest valuation errors across both periods, with median absolute error of 0.270 (original) and 0.352 (testing), consistently outperforming ln_v2a (0.273, 0.354) and ln_v2s (0.277, 0.359). While ln_v2a achieves marginally lower RMSE($\hat{M}$) and higher R², ln_m2b delivers superior valuation accuracy in minimizing absolute deviation between intrinsic and market values. All models demonstrate stable performance (R²_oos: 0.55-0.79) without material overfitting. Based on its consistently lowest valuation errors and robust correlation ($\rho_{m,\hat{m}}$ = 0.89, 0.84), the ln_m2b model is selected for return analyses.

The results confirm that the replicated ML models achieve valuation accuracy consistent with those reported by Geertsma and Lu (2023). Across all specifications, valuation errors remain within comparable bounds, and out-of-sample R² values range between 0.55 and 0.79, confirming stable performance and the absence of material overfitting Among the three different models, the *ln_m2b* model exhibits the lowest valuation errors and competitive predictive fit, with high correlation between predicted and actual values. We therefore adopt this model's intrinsic value estimates for subsequent return analyses.

4.1.2. True Out-of-Sample Performance (2020-2024)

Extending the evaluation window beyond the original period provides a test of temporal robustness and checks for application period robustness. Performance metrics in Table 1 show that, while overall predictive precision remains high, valuation errors increase in the 2020–2024 period. Nonetheless, the rank correlation between actual and predicted valuations remains above 0.82, suggesting that relative valuation accuracy is largely preserved.

Based on those findings we reject the null hypothesis of substantially deteriorating model performance, as the ML-based intrinsic value model maintains strong relative valuation accuracy ($\rho > 0.82$) in genuine out-of-sample context.

4.2. Return Analysis

4.2.1. Cumulative Portfolio Returns

To evaluate the economic relevance of the ML-based intrinsic value estimates, we examine cumulative returns of the high V/P quintile portfolios derived from both the ML and RIM valuation models. Both equal-weighted (EW) and value-weighted (VW) portfolios are formed annually in June and held thereafter.

Across both weighting schemes, the ML-based strategies display strong cumulative performance. (Figure 1) The EW ML portfolios consistently and substantially outperform the market, with cumulative returns peaking near 600%. This EW variant exhibits large fluctuations and pronounced spikes. The VW ML portfolios follow smoother trajectories but retain a substantial advantage over the market, peaking at around 200%. Moreover, in both schemes, the 2-year and 3-year holding periods consistently outperform the 1-year period, indicating a stable ML-derived value signal across investment horizons.

The RIM-based portfolios (Figure 2) present a more differentiated pattern. Under equal weighting, the 2- and 3-year RIM portfolios clearly surpass the market, achieving returns of approximately 200%, while the 1-year RIM EW portfolio delivers notably weaker performance but still achieves modest outperformance. For the VW portfolios, the 2- and 3-year horizons again perform similarly, though their outperformance over the market is marginal, suggesting that value weighting stabilizes overall performance but significantly dampens returns compared to the EW scheme.

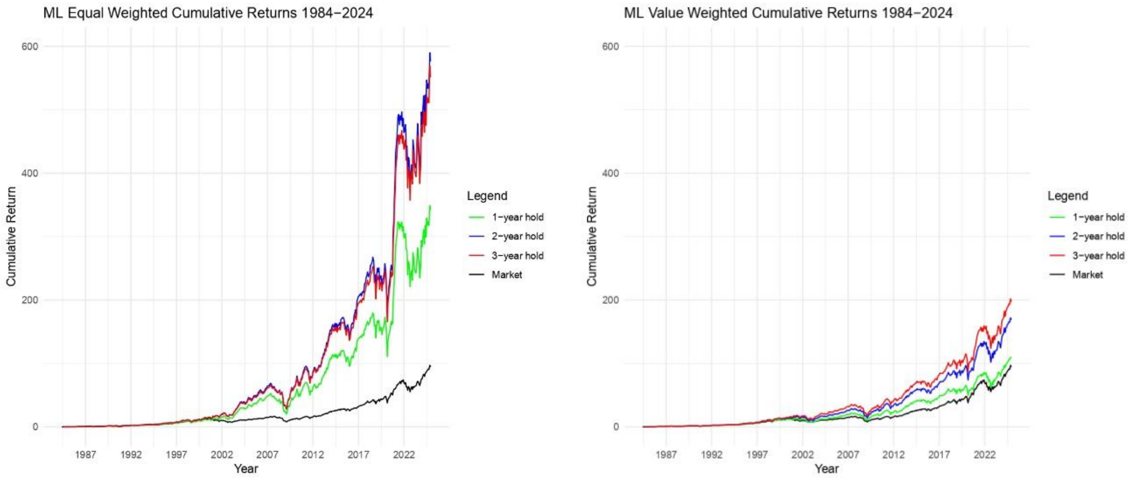

Figure 1: Cumulative returns of the ML-based high V/P portfolios using equal- and value-weighting. Across all holding periods (1-, 2-, and 3-year), ML portfolios substantially outperform the market, with EW returns showing greater volatility with visibly stronger returns over time; VW returns following smoother trajectories.

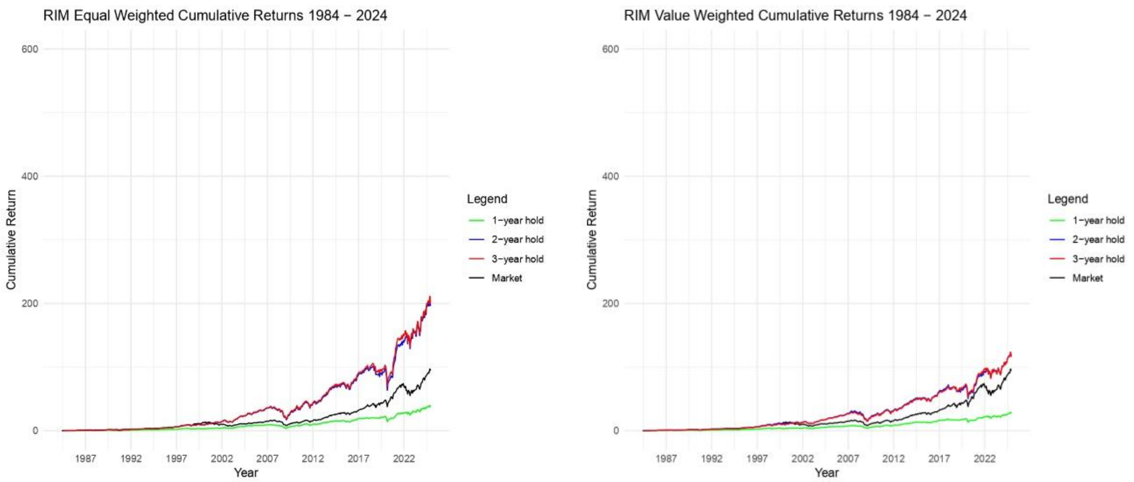

Figure 2: Cumulative returns of the RIM-based high V/P portfolios using equal- and value-weighting. Both equal-weighted and value-weighted 1-year returns underperform the market. 2- and 3- year holding period returns behave similarly and outperform the market. EW returns are visibly stronger over time.

Figure 3 provides a comparison between the ML and RIM strategies. In the EW setting, the ML portfolios generate considerably higher cumulative returns than both the RIM portfolios and the market, albeit with higher volatility. The RIM portfolios achieve the second-strongest performance, clearly outperforming the market. The VW results show the same ordering but exhibit smoother, lower return dynamics: the ML VW portfolios dominate, followed by the RIM VW portfolios, which marginally exceed the market.

Overall, these findings indicate that while both valuation approaches identify value opportunities, the ML-based intrinsic value estimates produce higher cumulative returns and greater scale of outperformance than the RIM approach, particularly in the equal-weighted setting. The RIM approach still yields economically meaningful performance, but its outperformance margin is less consistent across weighting schemes and horizons.

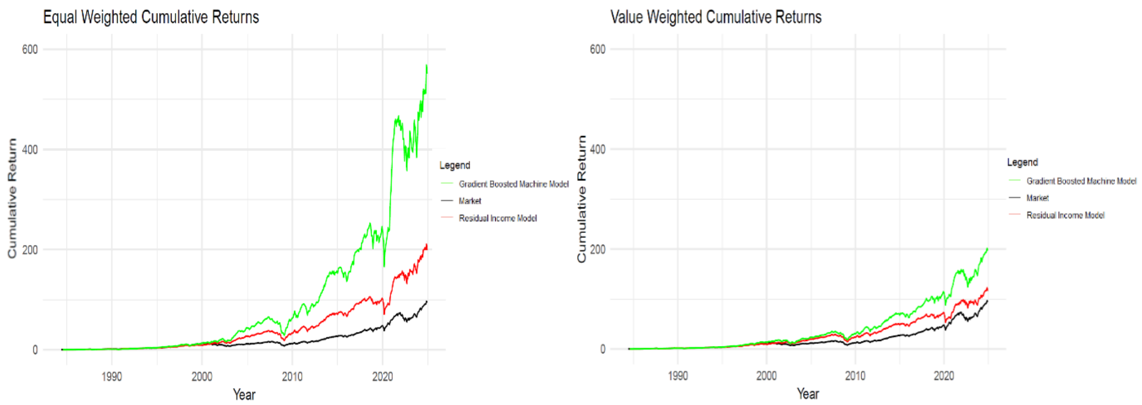

Figure 3: Comparison of cumulative returns for ML- and RIM-based portfolios against the market. In both weighting schemes, ML portfolios achieve the strongest performance, with more pronounced cumulative return for EW portfolios. RIM portfolios also outperform the market but remain below the ML strategies.

4.2.2. Risk-Adjusted Performance

To assess the returns of the ML-based approach accounting for several different risk factors, we estimate Fama-French five-factor (FF5) + Momentum regressions for each high V/P portfolio and report the intercepts (alphas). Standard errors are Newey–West adjusted to account for autocorrelation. To provide a complete picture of economic significance, Table 2 reports compound annual growth rates (CAGR), volatility, Sharpe ratios of cumulative returns, and FF5+Momentum alphas for the 3-year high V/P portfolios across the in-sample period (1984–2019), the out-of-sample period (2020–2024), and the full sample.

During 1984–2019, the ML-based portfolios exhibit solid performance: the EW portfolio delivers 16.0% CAGR, Sharpe 0.7, and alpha 6.1% (t = 6.38), while the VW portfolio achieves 13.7% CAGR, Sharpe 0.7, and alpha 2.8% (t = 3.56), indicating that the ML-derived V/P signal captures meaningful return premia under standard market conditions.

In 2020–2024, the EW portfolio continues to perform well with 22.8% CAGR, Sharpe 0.9, and alpha 8.0% (t = 2.74). The VW portfolio, however, delivers weaker results with 13.6% CAGR and an insignificant negative alpha of –1.4% (t = –0.70), suggesting that size-weighted exposures may behave differently in the post-COVID macroeconomic environment. This short and atypical window limits the ability to fully judge time-period out-of-sample return persistence, though the ML model's precision metrics remain stable.

Across the full sample, the EW portfolio delivers 16.9% CAGR and alpha 6.3% (t = 7.23), while the VW portfolio records 13.9% CAGR and alpha 2.5% (t = 3.37). Overall, the ML-based high V/P strategy maintains economically meaningful and broadly consistent performance, with the EW specification showing particularly robust results.

TABLE 2

Annualized Performance of ML-based High V/P Portfolios with 3-Year Buy-and-Hold Strategy

Period	Portfolio	CAGR (%)	Volatility (%)	Sharpe Ratio	Alpha (%)	t-stat
1984-2019	EW 3Y	16.0	18.6	0.7	6.1	6.38
	VW 3Y	13.7	16.5	0.7	2.8	3.56
	Market	11.2	15.4	0.6	–	–
2020-2024	EW 3Y	22.8	24.2	0.9	8.0	2.74
	VW 3Y	13.6	15.4	0.7	-1.4	-0.703
	Market	16.2	16.8	0.9	–	–
Full Period	EW 3Y	16.9	19.2	0.7	6.3	7.23
	VW 3Y	13.9	16.4	0.7	2.5	3.37
	Market	11.8	15.6	0.6	–	–

Table 2: Annualized performance of ML-based high V/P portfolios shows strong and persistent outperformance. Equal-weighted 3-year strategies deliver high CAGRs (16–23%) and sizable alphas (6–8%), while value-weighted portfolios earn moderate but positive abnormal returns except in 2020–2024.

Tables 3 and 4 compare ML- and RIM-based high V/P portfolios across holding periods (1984-2024) and show that ML-based portfolios deliver positive, statistically significant alphas. ML-based portfolios also exhibit larger factor exposures. RIM-based alphas are smaller, mostly significant for EW and generally insignificant for VW portfolios. The fact that our RIM model yields mostly insignificant alphas for VW portfolios is a contrast to prior studies, potentially due to the sample period but may be subject to future research.

Taken together, these results suggest that the ML-based valuation signal captures mispricing information not explained by conventional risk factors and generates stronger, more consistent abnormal returns than the RIM-based model, supporting **H2** that ML-based intrinsic value measures improve risk-adjusted performance.

TABLE 3

ML-based V/P portfolio abnormal FF5+MOM regression (monthly) for equal- and value-weighted portfolios

Variable	Equal-weighted portfolio returns			Value-weighted portfolio returns		
	1 Year	2 Years	3 Years	1 Year	2 Years	3 Years
mktrf	1.006***	0.965***	0.934***	0.992***	0.976***	0.974***
	(0.026)	(0.022)	(0.020)	(0.026)	(0.021)	(0.022)
smb	0.780***	0.767***	0.740***	0.035	0.061*	0.068**
	(0.053)	(0.047)	(0.043)	(0.038)	(0.035)	(0.034)
hml	0.203***	0.202***	0.194***	0.149***	0.169***	0.154***
	(0.045)	(0.040)	(0.038)	(0.044)	(0.034)	(0.036)
rmw	-0.089	-0.057	-0.026	0.001	0.004	0.010
	(0.056)	(0.051)	(0.049)	(0.044)	(0.049)	(0.052)
cma	0.160**	0.173***	0.174***	–0.002	0.069	0.161***
	(0.064)	(0.054)	(0.050)	(0.061)	(0.054)	(0.061)
umd	–0.349***	–0.268***	–0.225***	–0.241***	–0.207***	-0.169***
	(0.033)	(0.027)	(0.026)	(0.036)	(0.027)	(0.028)
Constant	0.005***	0.005***	0.005***	0.002*	0.002***	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 3: FF5+Momentum estimates for ML-based high V/P portfolios. The regressions show stable positive alphas across horizons; 0.005 for equal-weighted and 0.002 for value-weighted portfolios. Factor exposures are economically meaningful: market betas near 1.00, large SMB coefficients for equal-weighted portfolios (0.74–0.78), moderate HML loadings (≈0.15–0.20), and strongly negative UMD coefficients (–0.17 to –0.35).

TABLE 4

RIM-based V/P portfolio abnormal FF5+MOM regression (monthly) for equal- and value-weighted portfolios

Variable	Equal-weighted portfolio returns			Value-weighted portfolio returns		
	1 Year	2 Years	3 Years	1 Year	2 Years	3 Years
mktrf	0.881***	0.909***	0.899***	0.861***	0.921***	0.924***
	(0.039)	(0.020)	(0.017)	(0.037)	(0.020)	(0.019)
smb	0.362***	0.389***	0.384***	0.010	0.001	0.009
	(0.048)	(0.038)	(0.035)	(0.043)	(0.033)	(0.033)
hml	0.363***	0.368***	0.356***	0.363***	0.357***	0.350***
	(0.045)	(0.038)	(0.033)	(0.047)	(0.039)	(0.040)
rmw	0.193***	0.203***	0.197***	0.160***	0.172***	0.154***
	(0.054)	(0.051)	(0.046)	(0.045)	(0.035)	(0.035)
cma	0.095	0.107**	0.118***	0.098	0.130***	0.141***
	(0.073)	(0.053)	(0.045)	(0.068)	(0.049)	(0.048)
umd	–0.096***	–0.100***	–0.087***	0.003	-0.050**	-0.059***
	(0.036)	(0.025)	(0.023)	(0.029)	(0.022)	(0.020)
Constant	0.001	0.001*	0.001**	-0.0003	-0.0003	-0.0002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 4: FF5+Momentum estimates for RIM-based high V/P portfolios. Alphas are very small but mostly significant for equal-weighted portfolios (0.001) but generally insignificant for value-weighted portfolios. Market betas are below one (0.86–0.91), SMB loadings moderate for equal-weighted (0.36–0.39), HML exposures sizable (≈0.35–0.37), and UMD coefficients mildly negative for equal-weighted and near zero for value-weighted portfolios.

Our findings show that when ML-based portfolios are built once per year (in June) and held for 36 months, they produce stable positive alphas for both equal- and value-weighted portfolios (≈0.005 EW; ≈0.002 VW). Equal-weighted portfolios show large small-cap (SMB) and contrarian (negative UMD) exposures, while value-weighted portfolios exhibit smaller SMB, both having moderate value (HML) loadings.

4.2.3. Robustness to Sorting Frequency

One distinct advantage of the ML–based intrinsic value model is its ability to generate monthly valuation estimates, enabling more frequent portfolio formation and timely capture of mispricing. In contrast, the RIM updates valuations only once per year, limiting investment timing. To assess whether the ML-based V/P strategy's performance benefits from higher frequency, we compare annual and monthly portfolio formation schedules.

In the monthly formation approach, new high V/P portfolios are formed every month using the latest ML intrinsic value estimates, maintaining holding horizons of up to 36-month leading to a maximum of 36 overlapping portfolios. The results highlight weaker very short-term performance: the 1-month alpha is negative (–0.024 EW; –0.011 VW) and the 6-month alpha is insignificant for VW, while longer-horizon alphas remain stable (≈0.005–0.006 EW; 0.001–0.002 VW). The factor structure is preserved, with market betas near one, large SMB for EW, modest SMB for VW, moderate HML, consistently negative UMD, and mostly insignificant RMW with modestly positive CMA loadings. Table 5 and 6 show detailed factor loadings and significance levels.

Overall, both formation frequencies generate economically meaningful abnormal returns over the long term. However, monthly formation does not materially improve very short-term alphas, producing negative or insignificant alphas for horizons below 12 months, particularly for value-weighted portfolios. These findings indicate that the predictive power of the ML-based V/P signal is robust for longer horizons, as factor exposures and long-term performance remain broadly consistent across annual and monthly updates. At the same time, the results show that monthly formation and the derived enabling of holding periods lower than one year do not enhance abnormal returns. This is consistent with the momentum argument in literature explaining significant abnormal returns only after at least one year.

Table 5

ML-based V/P portfolio returns (monthly) – Equal-weighted portfolios with Newey-West standard errors

Variable	1 Month	6 Months	12 Months	24 Months	36 Months
mktrf	1.082***	1.034***	1.009***	0.970***	0.937***
	(0.034)	(0.027)	(0.026)	(0.026)	(0.026)
smb	0.821***	0.862***	0.855***	0.812***	0.772***
	(0.048)	(0.044)	(0.044)	(0.044)	(0.044)
hml	0.246***	0.166***	0.181***	0.188***	0.179***
	(0.046)	(0.041)	(0.040)	(0.040)	(0.040)
rmw	-0.067	-0.053	-0.040	-0.034	-0.009
	(0.085)	(0.059)	(0.054)	(0.054)	(0.054)
cma	-0..021	0.091	0.120**	0.160***	0.171***
	(0.066)	(0.060)	(0.060)	(0.060)	(0.060)
umd	-0.365***	-0.416***	-0.354***	-0.277***	-0.233***
	(0.0402)	(0.040)	(0.036)	(0.036)	(0.036)
Constant	-0.024***	0.003***	0.005***	0.006***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Table 5. FF5+Momentum estimates for monthly ML-based high V/P equal-weighted portfolios. The regressions show strong positive alphas across horizons, starting negative at –0.024 (1 month) and rising to 0.005–0.006 for 12–36 months. Factor exposures are meaningful: large SMB coefficients (0.77–0.86), moderate HML loadings (0.17–0.25), negative momentum exposures (UMD –0.23 to –0.42), and mostly insignificant RMW with positive CMA loadings.

Table 6

ML-based V/P portfolio returns (monthly) – Value-weighted portfolios with Newey-West standard errors

Variable	1 Month	6 Months	12 Months	24 Months	36 Months
mktrf	1.054***	1.003***	0.994***	0.987***	0.978***
	(0.027)	(0.024)	(0.022)	(0.022)	(0.022)
smb	0.121***	0.089**	0.081**	0.077**	0.079**
	(0.040)	(0.035)	(0.034)	(0.034)	(0.034)
hml	0.130**	0.129***	0.151***	0.164***	0.154**
	(0.044)	(0.038)	(0.036)	(0.036)	(0.036)
rmw	0.049	0.077*	0.085**	0.086**	0.078**
	(0.049)	(0.044)	(0.036)	(0.040)	(0.040)
cma	-0.048	-0.033	-0.006	0.068	0.135***
	(0.067)	(0.054)	(0.046)	(0.046)	(0.046)
umd	-0.198***	-0.245***	-0.231***	-0.211***	-0.194**
	(0.040)	(0.031)	(0.029)	(0.029)	(0.029)
Constant	-0.011***	0.0004	0.001*	0.002**	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Table 6: FF5+Momentum estimates for monthly ML-based high V/P value-weighted portfolios. Regressions show positive alphas across horizons, with a negative 1-month alpha, an insignificant 6-month alpha, and positive alphas thereafter (0.001–0.002 for 12–36 months). Factor exposures are meaningful: modest SMB (0.077–0.121), moderate HML (0.129–0.164), negative UMD (–0.194 to –0.245), and small to moderate positive RMW and CMA loadings.

5. Discussion and Research Outlook

With the findings of our analysis in mind we want to point out some aspects that benefit from further research. Even though our approach of using a pure long strategy is justifiable based on the literature, the comparison of this long-only strategy with a long-short strategy is a consequential extension of our study. Our long-only strategy seems especially adequate for mutual funds due to its simplicity and lower exposure to short-selling risks, while long-short strategies offer additional opportunities for hedging and higher returns, albeit with increased complexity, margin requirements, and potential regulatory constraints. Extensions relevant to hedge funds might therefore include analyzing the performance and risks of long-short strategies relative to the market.

Further, while previous research has found abnormal returns with a RIM-based V/P strategy, with our long-only strategy we do not find significant alphas with the RIM-based approach. Detailed analysis of this discrepancy may help to distinguish relevant reasons and validate and bring clarity into using a RIM-based approach as traditional baseline.

Our study closes the gap of ML value estimation and insights into V/P trading return over time. In an initial screening analysis, we find the majority of stocks drop out of the top V/P quintile within a few months. Combining insights into those dynamics with the monthly return framework enables an extension to our analysis that could include an inspection of a correlation between duration in the extreme quintile and abnormal returns. Subsequently, investigating company characteristics at investment of stocks that show a certain duration period may enable more precise identification of a sub-group with a certain duration in extreme V/P quintiles where stocks carry higher abnormal returns.

Additionally, we find that the average V/P per quintile over time as well as the excess returns above the market visually show indications that they are extremer around times of crises like the dotcom bubble (2000-2003) or the great recession (2007-2009). (Appendix A.5) This opens the opportunities for new angles to use V/P trading strategies to exploit extreme mispricing in sentiment-driven markets. This may include but is not limited to options such as signal-weighted portfolios where stocks with more extreme V/P ratios are weighted stronger or strategies where an investor only invests if the formed portfolios' quintile V/P averages pass a certain threshold.

A common discrepancy between the theory of portfolio formation and the practical implications of portfolio and factor trading is the recognition of transaction costs. In our study, we neglect transaction costs, which may become relevant when comparing different portfolio formation frequencies. For instance, forming portfolios once per year implies a single round of buying and selling per year, while forming portfolios monthly, even if each portfolio is held for several years, would result in twelve times the transaction volume over the same period. Based on common benchmarks in the U.S. equity market, transaction costs are often estimated around 25 basis points per trade (e.g., Domowitz (2002)), although more recent evidence suggests that actual institutional trading costs may be lower, typically around or below 10–20 bps (Frazzini et Al. (2018)). While this seems acceptable compared to our annualized overall alpha of 2.5% it still represents a non-negligible drag on net performance for higher-frequency portfolio formation. Transaction costs also relate to the discussion of multi-signal portfolios highlighted by DeMiguel et al. (2017), where overlapping signals may allow investors to reduce unnecessary trades. While our yearly long-only implementation is relatively conservative in turnover, exploring higher-frequency formation strategies could provide further insights into the trade-off between potential performance adjustments and transaction costs.

While the use of overlapping portfolios mechanically induces autocorrelation in event-time raw returns (Fama, (1998)), this does not invalidate our risk-adjusted results. Applying Newey–West standard errors corrects the induced autocorrelation in the regression residuals, ensuring statistically reliable risk-adjusted alpha estimates.

Beyond turnover and transaction costs, practical considerations may further affect the strategy's implementability. Even with a long-only buy-and-hold approach, investors may face constraints such as liquidity, capital allocation, and institutional restrictions, which can limit the ability to perfectly replicate the portfolios suggested by our model. Despite these constraints, observed abnormal returns appear economically meaningful relative to common transaction cost benchmarks, suggesting the strategy could still generate positive net gains in practice. At the same time, these limitations highlight that higher-frequency or more complex implementations would require careful balancing of potential performance improvements against operational costs and market frictions, offering a clear opportunity for future research.

6. Conclusion

Our study extends the V/P investment literature by replacing theory-driven models like the RIM with a data- and peer-driven ML model for estimating intrinsic values. We also extend ML fair value research by evaluating not only point-in-time prediction metrics but also time-series returns based on ML value estimates.

Before comparing returns, we replicate the ML approach by Geertsma and Lu (2023) and can not only confirm their model's precision but also show that it is not overfitted to a specific time period and performs well even in the timeframe after the modelling time period in their paper. We then analyze the returns from a V/P strategy.

We find that, not only do cumulative returns of an ML-based V/P formation outperform the RIM-based model and the market but also that risk-adjusted performance is notably stronger. For equal-weighted portfolios, annual ML formation delivers significant alphas of 6.3%, while value-weighted portfolios achieve smaller but positive alphas of 2.5%. Monthly formation, in contrast, produces negative or insignificant short-term returns for horizons below one year, suggesting that higher-frequency rebalancing does not yield higher significant alphas. RIM-based portfolios in our study show lower annualized returns and mostly insignificant alphas, particularly value-weighted portfolios. These results highlight that the superior performance of the ML-based V/P strategy arises primarily from more accurate long-term valuation signals rather than increased portfolio turnover.

We acknowledge that our study could benefit from several potential future extensions. While we deploy a long-only strategy, a long-short strategy may hold additional exploitation potential. Additionally, we neglect transaction costs and liquidity in our analysis; two factors that have been previously identified as influence factors in portfolio trading strategies. Another limiting factor may be that our RIM-based V/P alphas were insignificant, which deviates from previous research findings in literature. Lastly, we find indications that V/P portfolios may incorporate additional arbitrage opportunities when considering the spread of V/P signals. Extensions of our study could include portfolio weightings based on V/P signals, or looking at returns of portfolios formed in certain crisis periods like the dotcom-bubble (2000-2003) or the great recession (2007-2008).

References

Ali, A., Hwang, L.-S., & Trombley, M. A. (2003). Residual-Income-Based Valuation Predicts Future Stock Returns: Evidence on Mispricing vs. Risk Explanations. *The Accounting Review, 78*(2), 377-396. Retrieved from https://www.jstor.org/stable/3203258

Banz, R. (1981). The Relationship between Return and Market Value of Common Stocks. *Journal of Financial Economics, 9*(1), 3-18. Retrieved from https://doi.org/10.1016/0304-405X(81)90018-0

Bartram, S. M., & Grinblatt, M. (2018). Agnostic Fundamental Analysis Work. *Journal of Financial Economics, 128*(1), 125-147. Retrieved from https://doi.org/10.1016/j.jfineco.2016.11.008

Basu, S. (1977). Investment Performance of Common Stocks in Relation to Their Price-Earnings Ratios: A Test of the Efficient Market Hypothesis. *The Journal of Finance, 32*(3), 663-682. Retrieved from https://doi.org/10.1111/j.1540-6261.1977.tb01979.x

Beereddy, S. R. (2021). Fair Valuation for Financial Instruments using AI/ML. *Journal of Intelligent Systems and Applications in Engineering, 9*(4), 377-394. Retrieved from https://www.ijisae.org/index.php/IJISAE/article/view/7458

Buczynski, W., Cuzzokin, F., & Sahkian, B. (2021). A Review of Machine Learning Experiments in Equity Investment Decision-Making: Why Most Published Research Findings Do not Live Up to Their Promise in Real Life. *International Journal of Data Science and Analytics, 11*, 221-242. Retrieved from https://doi.org/10.1007/s41060-021-00245-5

Capozza, D. R., & Israelsen, R. D. (2010). How Quickly Do Equity Prices Converge to Intrinsic Value? *Journal of Investment Management, 8*(4), 1-13. Retrieved from https://dx.doi.org/10.2139/ssrn.1357841

Chen, X., Cho, Y. H., Dou, Y., & Lev, B. (2022). Predicting Future Earnings Changes Using Machine Learning and Detailed Financial Data. *Journal of Accounting Research, 60*(2), 467-515. Retrieved from https://doi.org/10.1111/1475-679X.12429

DeMiguel, V., Utrera, A. M., Nogales, F., & Uppal, R. (2017). A Transaction-Cost Perspective on the Multitude of Firm Characteristics. *Available at SSRN*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3066029

Domowitz, I. (2002). Liquidity, Transaction Costs, and Reintermediation in Electronic Markets. *Journal of Financial Services Research, 22*, 141-157. Retrieved from https://doi.org/10.1023/A:1016077023185

Fama, E. (1998). Market Efficiency, Long-Term Returns, and Behavioral Finance. *Journal of Financial Economics, 49*(3), 283-306. Retrieved from https://doi.org/10.1016/S0304-405X(98)00026-9

Fama, E. F., & French, K. R. (2015). A Five-Factor Asset Pricing Model. *Journal of Financial Economics, 116*(1), 1-22. Retrieved from https://doi.org/10.1016/j.jfineco.2014.10.010

Fama, E., & French, K. (1993). Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics, 33*(1), 3-56. Retrieved from https://doi.org/10.1016/0304-405X(93)90023-5

Frankel, R., & Lee, C. M. (1998). Accounting Valuation, Market Expectation, and Cross-Sectional Stock Returns. *Journal of Accounting and Economics, 25*(3), 283-319. Retrieved from https://doi.org/10.1016/S0165-4101(98)00026-3

Frazzini, A., Israel, R., & Moskowitz, T. (2018). Trading Costs. *Working Paper*. Retrieved from http://dx.doi.org/10.2139/ssrn.3229719

Geertsma, P., & Lu, H. (2023). Relative Valuation with Machine Learning. *Journal of Accounting Research, 61*(1), 329-376. Retrieved from https://doi.org/10.1111/1475-679X.12464

Gonçalves , A. S., & Leonard, G. (2023). The Fundamental-to-Market Ratio and the Value Premium Decline. *Journal of Financial Economics, 147*(2), 382-405. Retrieved from https://doi.org/10.1016/j.jfineco.2022.11.001

Hanauer, M. X., Kononova, M., & Rapp, M. (2022). Boosting Agnostic Fundamental Analysis: Using Machine Learning to Identify Mispricing in European Stock Markets. *Finance Research Letters, 40*, 102856. Retrieved from https://doi.org/10.1016/j.frl.2022.102856

Huang, S., Tan, H., Wang, X., & Yu, C. (2021). Valuation Uncertainty and Analysts' Use of DCF Models. *Accepted by Review of Accounting Studies*. Retrieved from https://dx.doi.org/10.2139/ssrn.3274391

Huang, Y., Capretz, L., & Ho, D. (2021). Machine Learning for Stock Prediction Based on Fundamental Analysis. *IEEE Symposium Series on Computational Intelligence (SSCI)*, (pp. 01-10). Orlando, FL, USA. Retrieved from https://doi.org/10.1109/SSCI50451.2021.9660134

Huefner, B., Rueenaufer, M., & Boesch, M. (2023). Value Investing via Bayesian inference. *Review of FInancial Economics, 41*(4), 465-492. Retrieved from https://doi.org/10.1002/rfe.1185

Jafary, P., Shojaei, D., Rajabifard, A., & Ngo, T. (2024). Automated Land Valuation Models: A Comparative Study of Four Machine Learning and Deep Learning Methods Based on a Comprehensive Range of Influential Factors. *Cities, 151*, 105115. Retrieved from https://doi.org/10.1016/j.cities.2024.105115

Johnson, W., & Xie, S. (2004). The Convergence of Stock Price to Fundamental Value. *S&P Global Market Intelligence - SSRN Publication*. Retrieved from http://dx.doi.org/10.2139/ssrn.590542

Kim, H., & Ryu, D. (2015). Measuring the Speed of Convergence of Stock Prices: A Nonparametric and Nonlinear Approach. *Economic Modelling, 51*, 227-241. Retrieved from https://doi.org/10.1016/j.econmod.2015.07.009

Li, S. (2024). Comparison of Machine Learning Models for Stock Prediction. *Advances in Economics, Management, and Political Sciences, 85*(1), 76-84. Retrieved from https://doi.org/10.54254/2754-1169/85/20240846

Mignot, S. (2025). Coevolution of Stock Prices and Their Perceived Fundamental Value. *Decision in Economics and Finance*. Retrieved from https://doi.org/10.1007/s10203-025-00517-w

Newey, W., & West, K. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica, 55*(3), 703-708. Retrieved from https://doi.org/10.2307/1913610

Newey, W., & West, K. (1994). Automatic Lag Selection in Covariance Matrix Estimation. *The Review of Economic Studies, 61*(4), 631-653. Retrieved from https://doi.org/10.2307/2297912

Nobre, J., & Neves, R. (2019). Combining Principal Component Analysis, Discrete Wavelet Transform and XGBoost to Trade in the Financial Markets. *Expert Systems with Applications, 125*, 181-194. Retrieved from https://doi.org/10.1016/j.eswa.2019.01.083

Ohlson, J. (1995). Earnings, Book Values, and Dividends in Equity Valuation. *Contemporary Accounting Research, 11*(2), 661-687. Retrieved from https://doi.org/10.1111/j.1911-3846.1995.tb00461.x

Pinto, J. E., Robinson, T. R., & Stowe, J. D. (2019). Equity Valuation: A Survey of Professional Practice. *Review of Financial Economics, 37*(2), 219-233. Retrieved from https://doi.org/10.1002/rfe.1040

Priel, R., & Rokach, L. (2024). Machine Learning-Based Stock Picking Using Value Investing and Quality Features. *Neural Copmuting and Applications, 36*, 11963-11986. Retrieved from https://doi.org/10.1007/s00521-024-09700-3

Shen, K.-Y. (2011). Implementing Value Investing Strategy by Artifical Neural Network. *Internationa Journal of Business & Information Technology, 1*(1), 12-22.

Williams, J. B. (1938). *The Theory of Investment Value.* Amsterdam: North-Holland Publishing Company.

Wolff, D., & Echterling, F. (2023). Stock Picking with Machine Learning. *Journal of Forecasting, 43*(1), 81-102. Retrieved from https://doi.org/10.1002/for.3021

Appendix

Appendix A.1: Datapoints and Variables Used in ML Model

Table A.1.1 – Retrieved Datapoints with Compusstat Code Used to Build Variables for ML Model

Compustat Code	**Description**
gvkey	Global Company Key
conm	Company Name
cik	CIK Number
exchg	Stock Exchange Code
gind	GIC Industries
gsubind	GIC Sub-Industries
sic	Standard Industry Classification Code
fqtr	Fiscal Quarter
acoq	Current Assets - Other - Total
actq	Current Assets - Total
aoq	Assets - Other - Total
apq	Account Payable/Creditors - Trade
atq	Assets - Total
ceqq	Common/Ordinary Equity - Total
cheq	Cash and Short-Term Investments
cogsq	Cost of Goods Sold
cshoq	Common Shares Outstanding
cstkeq	Common Stock Equivalents - Dollar Savings
dlcq	Debt in Current Liabilities
dlttq	Long-Term Debt - Total
doq	Discontinued Operations
dpq	Depreciation and Amortization - Total
dvpq	Dividends - Preferred/Preference
ibadjq	Income Before Extraordinary Items - Adjusted for Common Stock Equivalents
ibcomq	Income Before Extraordinary Items - Available for Common
ibmiiq	Income before Extraordinary Items and Noncontrolling Interests
ibq	Income Before Extraordinary Items
icaptq	Invested Capital - Total - Quarterly
invtq	Inventories - Total
lcoq	Current Liabilities - Other - Total

Table A.1.1 – Retrieved Datapoints with Compusstat Code Used to Build Variables for ML Model

Compustat Code	Description
lctq	Current Liabilities - Total
lltq	Long-Term Liabilities (Total)
loq	Liabilities - Other
ltq	Liabilities - Total
mibnq	Noncontrolling Interests - Nonredeemable - Balance Sheet
mibq	Noncontrolling Interest - Redeemable - Balance Sheet
miiq	Noncontrolling Interest - Income Account
niq	Net Income (Loss)
nopiq	Non-Operating Income (Expense) - Total
oiadpq	Operating Income After Depreciation - Quarterly
oibdpq	Operating Income Before Depreciation - Quarterly
piq	Pretax Income
ppentq	Property Plant and Equipment - Total (Net)
pstkq	Preferred/Preference Stock (Capital) - Total
pstkrq	Preferred/Preference Stock - Redeemable
rectq	Receivables - Total
revtq	Revenue - Total
saleq	Sales/Turnover (Net)
seqq	Stockholders Equity > Parent > Index Fundamental > Quarterly
spiq	Special Items
teqq	Stockholders Equity - Total
txdbq	Deferred Taxes - Balance Sheet
txditcq	Deferred Taxes and Investment Tax Credit
txpq	Income Taxes Payable
txtq	Income Taxes - Total
xidoq	Extraordinary Items and Discontinued Operations
xintq	Interest and Related Expense- Total
xoprq	Operating Expense- Total
xrdq	Research and Development Expense
xsgaq	Selling, General and Administrative Expenses
curcdq	ISO Currency Code (curcdq)
capxy	Capital Expenditures
cogsy	Cost of Goods Sold
cstkey	Common Stock Equivalents - Dollar Savings

Table A.1.1 – Retrieved Datapoints with Compusstat Code Used to Build Variables for ML Model

Compustat Code	Description
doy	Discontinued Operations
dpcy	Depreciation and Amortization - Statement of Cash Flows
dpy	Depreciation and Amortization - Total
dvpy	Dividends - Preferred/Preference
dvy	Cash Dividends
ibadjy	Income Before Extraordinary Items - Adjusted for Common Stock Equivalents
ibcomy	Income Before Extraordinary Items - Available for Common
ibcy	Income Before Extraordinary Items - Statement of Cash Flows
ibmiiy	Income before Extraordinary Items and Noncontrolling Interests
iby	Income Before Extraordinary Items
miiy	Noncontrolling Interest - Income Account
niy	Net Income (Loss)
nopiy	Non-Operating Income (Expense) - Total
oancfy	Operating Activities - Net Cash Flow
oiadpy	Operating Income After Depreciation - Year-to-Date
oibdpy	Operating Income Before Depreciation
piy	Pretax Income
revty	Revenue - Total
saley	Sales/Turnover (Net)
spiy	Special Items
txty	Income Taxes - Total
xidoy	Extraordinary Items and Discontinued Operations
xinty	Interest and Related Expense- Total
xopry	Operating Expense- Total
xrdy	Research and Development Expense
xsgay	Selling, General and Administrative Expenses
prccq	Price Close - Quarter

Table A.1.2 – Variable Construction following Geertsma & Lu (2023, pp. 367-369)	
Variable Name	**Calculation**
profitability_an	(sale - cogs - xint - xsga) / be
grossmargin_an	(sale - cogs) / sale
grossprofit_an	(sale - cogs) / at
roic_an	(ebit - nopi) / (ceq + lt - che)
ebitda2revenue_an	ebitda / sale
marginch_an	(ib / sale) - (l_ib / l_sale)
pchgm2pchsale_an	(((sale - cogs) - (l_sale - l_cogs)) / (l_sale - l_cogs)) - ((sale - l_sale) / l_sale)
deprn_an	dp / ppent
inventorygrowth_an	d_invt / l_invt
salesgrowth_an	(sale / l_sale) - 1
capexgrowth_an	d_capx / l_capx
assetgrowth_an	d_at / l_at
bookequitygrowth_an	(be / l_be) - 1
chlt_an	(lt / l_lt) - 1
chceq_an	(ceq - l_ceq) / l_ceq
chca_an	(d_act - d_che) / l_at
chcl_an	(d_lct - ifelse(is.na(d_dlc), 0, d_dlc)) / l_at
chnccl_an	d_lt - d_lct - ifelse(is.na(d_dltt), 0, d_dltt)
chfnl_an	ifelse(rowSums(!is.na(cbind(d_dltt, d_dlc, d_pstk))) >= 1 & !(sicn >= 6000 & sicn <= 6999), rowSums(cbind(d_dltt, d_dlc, d_pstk), na.rm = TRUE) / l_at, NA_real_)
chbe_an	(be - l_be) / l_at
investment_an	d_at / at
inventorychange_an	d_invt / (0.5 * l_at + 0.5 * at)
pchdeprn_an	((dp / ppent) - (l_dp / l_ppent)) / (l_dp / l_ppent)
pchsales2inv_an	((sale / invt) - (l_sale / l_invt)) / (l_sale / l_invt)
pchsale2pchinvt_an	((sale - l_sale) / l_sale) - ((invt - l_invt) / l_invt)
pchsale2pchrect_an	((sale - l_sale) / l_sale) - ((rect - l_rect) / l_rect)
pchsale2pchxsga_an	((sale - l_sale) / l_sale) - ((xsga - l_xsga) / l_xsga)

Table A.1.2 – Variable Construction following Geertsma & Lu (2023, pp. 367-369)	
Variable Name	**Calculation**
assetturnover_an	sale / at
assetturnover2_an	sale / ((at + l_at) / 2)
chcurrentratio_an	((act / lct) - (l_act / l_lct)) / (l_act / l_lct)
pchquickratio_an	((act - invt) / lct - (l_act - l_invt) / l_lct) / ((l_act - l_invt) / l_lct)
opleverage_an	(cogs + xsga) / at
liquid2assets_an	(che + 0.75 * (act - che) + 0.5 * (at - act)) / l_at
cash2assets_an	ifelse(!(sicn >= 6000 & sicn <= 6999) & !(sicn >= 4900 & sicn <= 4949), che / at, NA_real_)
sales2cash_an	sale / che
sales2rec_an	sale / rect
sales2inv_an	sale / invt
cf2debt_an	(ib + dp) / ((lt + l_lt) / 2)
debt2tang_an	(che + rect * 0.715 + invt * 0.547 + ppent * 0.535) / at

Appendix A.2: Detailed Machine-Learning Model Construction

The machine learning model for intrinsic value estimation follows the methodology of Geertsma & Lu (2023). We implemented data cleaning and return analysis in R, while using the Python code is adapted from the methodology by Geertsma & Lu for the ML model; the model pipeline is structured as follows:

Step 1: Data Preparation

All fundamental variables are lagged by three months to avoid look-ahead bias, and missing values are forward-filled to reflect the information actually available to investors. Outliers in firm fundamentals and targets are removed based on percentile thresholds or extreme values (e.g., negative book value). Firms are sorted into 5 cross-sectional blocks to allow block-wise out-of-sample testing.

Step 2: Feature Selection and Target Construction

The model predicts three valuation multiples: market-to-book (m2b), enterprise-value-to-assets (v2a), and enterprise-value-to-sales (v2s). Log-transformations of these multiples are also used to stabilize variance. Predictor variables are detailed in Appendix A.1.

Step 3: Dataset Splits

Data is split temporally into training, development, and testing sets (typically 60:20:20). Within each month, firms are assigned to one of five cross-sectional blocks; one block is held out as out-of-sample for evaluation. This approach ensures no overlap of firms across training and testing periods and prevents information leakage.

Step 4: Model Training

The LightGBM model is trained on the training set with development set evaluation to avoid overfitting. Categorical features are encoded as integers with missing values set to -1. Optional target outlier removal is applied via z-score thresholds. Missing input variables are automatically imputed by the package.

Step 5: Predictions and Performance Evaluation

Predictions are made for the train, dev, and test sets. Multiples are converted to predicted firm values and evaluated using a range of metrics: mean and median absolute error, mean and median absolute percentage error, mean and median absolute log percentage error, RMSE, R^2, and t-statistics from regressing actual on predicted values.

Appendix A.3: Machine-Learning Model Precision Analysis

The performance of the machine learning models is evaluated using a set of metrics consistent with Geertsma & Lu (2023) in chapter 4.1 (p. 337 ff.). These metrics capture both the accuracy and bias of predicted equity values across different target variables.

Percentage Valuation Error (ε)

For each firm-month observation, the percentage valuation error is defined as

$$\varepsilon = \frac{\widehat{M}}{M} - 1$$

where M is the actual equity value (CRSP price times shares outstanding) and $\widehat{M}$ is the predicted equity value. This metric allows comparison of predicted and actual equity values on a relative scale.

Root Mean Squared Error (RMSE)

Two RMSE metrics are reported:

- **Dollar RMSE:** evaluates the difference between predicted and actual dollar equity values.
- **Model RMSE:** evaluates the difference on the model target variable (e.g., log-multiples).

$$\mathrm{RMSE}(\hat{x}) = \sqrt{\frac{1}{N}\sum_{i}(x_i - \hat{x}_i)^2}$$

Out-of-Sample R-Squared (R^2_{oos})

The out-of-sample R^2 measures how well predictions perform relative to using the training sample mean as a predictor:

$$R^2_{oos}(\hat{p}) = 1 - \frac{\sum_i (p_i - \hat{p}_i)^2}{\sum_i (p_i - \bar{p})^2}$$

where p_i are actual values and $\bar{p}$ is the mean of the training sample.

Correlation and Bias Metrics

- **Pearson correlation ($\rho(M, \widehat{M})$)** quantifies the strength of the linear relationship between actual and predicted equity values.
- **Median Absolute Valuation Error (med(|ε|))** for robustness of predictions.
- **Median Valuation Error (med(ε))** evaluates systematic bias in over- or under-valuation.

These metrics are used to determine model performance and compare the value estimation models based on different multiple-approaches.

Appendix A.4: Detailed Residual Income Model Construction

This appendix details the implementation of the Residual Income Model (RIM) used to generate firm-level valuations, following Frankel and Lee (1998) and Ohlson (1995).

Data Sources

We combine four datasets: Compustat Fundamentals Annual (accounting data including book value, net income, dividends, total assets, fiscal year-end prices, and industry codes), I/B/E/S Summary Statistics (one- and two-year EPS forecasts and long-term growth, LTG), Fama-French factors (optional cost-of-equity estimation), and the IBES-CRSP linking table (maps I/B/E/S tickers to Compustat gvkeys via 8-character CUSIPs).

Sample Construction

The sample includes firms listed on major U.S. exchanges (NYSE, AMEX, NASDAQ; Compustat codes 11, 12, 14) with fiscal years ending June–December. Observations with prices ≤ \$1, missing analyst coverage, or non-positive book value are excluded. Only analyst forecasts announced through May are retained to match Frankel & Lee's statistical period. Firms must have both F_EPS1 and F_EPS2 available.

Cost of Equity

The baseline cost of equity is set at 12% for all firms. As a robustness check, an industry-adjusted Fama-French three-factor cost of equity can be computed. Zero rates are replaced with 0.0001 for numerical stability.

Linking Forecasts to Accounting Data

Compustat fundamentals dated fiscal year-end t are matched to I/B/E/S forecasts for fiscal year t+1 (so year t book value pairs with forecasts of year t+1 earnings). This ensures that forward EPS and LTG forecasts are aligned with historical fundamentals.

Forward ROE and Book Value Updates

Forward ROEs use average beginning-of-period book values. For year t,

$$\mathrm{FROE}_t = \frac{\mathrm{F\_EPS1}}{(B_{t-1} + B_{t-2})/2}, B_t = B_{t-1} \times [1 - (k) \times \mathrm{FROE}_t]$$

For year t+1:

$$\mathrm{FROE}_{t+1} = \frac{\mathrm{F\_EPS2}}{(B_t + B_{t-1})/2}, B_{t+1} = B_t \times [1 - (k) \times \mathrm{FROE}_{t+1}]$$

Book value is updated as B_t = B_{t-1} × [1 + (1 - k) × FROE_t], and similarly for subsequent years. Year t+2 earnings are EPS_{t+2} = F_EPS2 × (1 + LTG) or F_EPS2 if LTG is missing, with FROE_{t+2} computed similarly.

Dividend Payout Ratio

The payout ratio k is defined as dividends/net income if net income > 0, otherwise dividends/(0.06 × total assets). Values are clipped to [0,1].

Extreme Values

ROE and FROE are capped at ±100% and infinite values treated as missing. Value-to-price ratios are trimmed at the 1st and 99th percentiles.

Residual Income Valuation

The primary model V3_v uses three explicit forecast years plus a terminal residual income:

$$V = B_t + \frac{RI_t}{1+r} + \frac{RI_{t+1}}{(1+r)^2} + \frac{RI_\infty}{(1+r)^3 \cdot r}$$

with RI_t = (FROE_t − r) × base book value, and discounting applied appropriately. Alternative models use historical ROEs or shorter forecast horizons.

Edge Cases

Missing LTG is set to zero. Negative EPS forecasts are retained if resulting FROEs fall within [−1,1]. Observations with zero or negative book values are dropped. A minimum discount rate of 0.0001 is imposed to prevent numerical errors in terminal value calculations. Missing LTG is set to zero.

Appendix A.5: Further Descriptive Insights of the Analysis

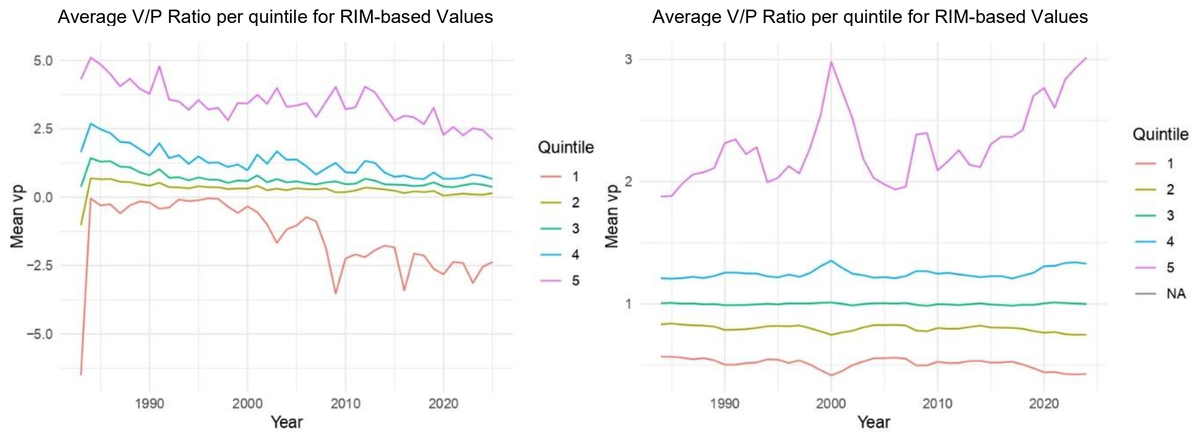

Figure A.5.1: Average V/P ratio in the five quintiles both for the RIM- and ML-based value estimates. We find that the range between the extreme quintiles for the RIM-based V/P ratios is larger with a spread of ca. 5.0, while the ML-based V/P ratios have a more fluctuating spread which only seldomly reaches a difference of up to 3.0. Additionally, it seems that the RIM-based V/P ratios decline over time, especially indicated by the corridor of the V/P ratios in the extreme quintiles. The shift from the top towards the negative quintile may influence the power of a long-only strategy based on a RIM. For the ML-based V/P ratios we find a more consistent pattern over time. Especially two things are worth noting: Firstly, even the low quintile average V/P ratio is rarely below zero which is a strong contrast to the RIM-based V/P ratios. Moreover, no declining trend is visible, but rather a larger spread around crisis times such as the dot.com bubble, the great recession, or towards Covid-19. More generally, the ML-based V/P ratios show a rising spread since 2010, mainly driven by the top V/P quintile.

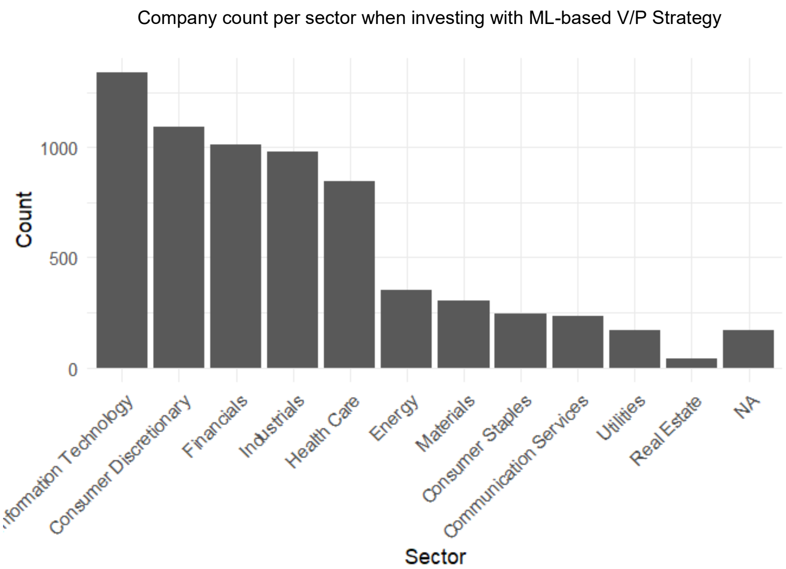

Figure A.5.2: Invested companies by sector when investing with ML-based V/P trading strategy. The figure shows that the highest count of invested companies is active within the sectors of Information Technology, Consumer Discretionary, Financials, Industrials, and Health Care.

Appendix A.6: Usage of Generative Pre-Trained Transformers

While we emphasize that the research, analysis, and drafting presented in this thesis were conducted through our own ideation, development, and execution, we acknowledge the limited use of Generative Pre-Trained Transformers (GPTs) during coding and drafting. We used OpenAI's ChatGPT and Anthropic's Claude, interchangeably, within the restricted scope described below.

Coding:
All model specifications, methodological choices, and coding logic were developed independently or drawn from established literature. GPTs were used only in later stages of coding for technical support tasks such as suggesting potential solutions for error handling, identifying syntax issues, or offering alternative ways to implement logic we had already designed.

Writing:
Similarly, the conceptual development and writing of the thesis were our own work. GPTs were used solely in the final stages of drafting to provide minor stylistic suggestions—for example, reducing repetition, improving clarity, or generating ideas for more concise phrasing. No substantive text, analysis, or argumentation was produced by GPTs without substantial verification and revision on our part.

Models used:

OpenAI. (2025). *ChatGPT* (Model 4o; Model 5.1) [Large language model].

URL: https://chat.openai.com/chat

Anthropic. (2025). *Claude* (Sonnet 4.5) [Large language model].

URL: https://claude.ai/