How To Bet On Winners (and Losers)

Christian Brownlees* André B.M. Souza*

July 15, 2025

Abstract

We cast the construction of long-short portfolios as a statistical decision problem in which the investor seeks to buy the top-performing stocks (the "winners") and sell the worst-performing ones (the "losers") on the basis of stock characteristics. We derive the optimal portfolio selection rule implied by a loss function that accounts for different types of misclassification errors in portfolio construction. This approach leads to a return classification problem and the optimal rule buys or sells stocks based on their probabilities of being winners or losers, conditional on the stock characteristics. When returns are generated by an additive regression model and misclassification costs satisfy a symmetry condition, the optimal rule simplifies to the conventional sorting procedure based on expected returns. An empirical application using U.S. stock data shows that portfolios constructed using the optimal rule achieve higher Sharpe ratios compared to those built using conventional methods. Our results demonstrate that predictive signals in the cross-section of stock returns go beyond expected returns, and that properly optimized portfolio selection rules based on these signals can generate substantial economic value for investors.

JEL: G10, G11, G12, G17, C10, C14, C58.

Keywords: Machine Learning, Return Prediction, Asset Allocation, Classification, Cross-Section of Returns.

---

* Universitat Pompeu Fabra and Barcelona School of Economics. e-mail: `christian.brownlees@upf.edu`, ESADE Business School. e-mail: `andre.souza@esade.edu`. Replication files for this paper can be found in https://github.com/andrebmsouza/hbw. We thank seminar participants at the University of Glasgow, University of Zurich, BI Norwegian Business School, CREST-ENSAE, Universitat Pompeu Fabra, Toulouse School of Economics, Fundação Getulio Vargas and Erasmus University Rotterdam. We would also like to thank Andrew Patton for a detailed discussion of the paper. Christian Brownlees acknowledges financial support from the Spanish Ministry of Science and Technology (Grant MTM2012-37195); the Ayudas Fundación BBVA Proyectos de Investigación Cientìfica en Matemáticas 2021; the Spanish Ministry of Economy and Competitiveness through the Severo Ochoa Programme for Centres of Excellence in R&D (SEV-2011-0075), and André B.M Souza acknowledges financial support from the MINECO project PID2021-126300NB-I00.

1 Introduction

One of the classic objectives of empirical asset pricing is to detect predictive signals for the cross-section of stock returns. Several recent studies have approached this as a *return prediction* task and relied on regression techniques to estimate expected returns on the basis of stock characteristics (Lewellen, 2015; Zaffaroni and Zhou, 2024). Among the most promising developments in this area is the application of machine learning, which has shown considerable effectiveness in capturing complex relationships between firm-specific characteristics and expected returns (Freyberger *et al.*, 2020; Leippold *et al.*, 2022; Bali *et al.*, 2023; Chen *et al.*, 2023). An important finding from this body of work is that long-short portfolios obtained by buying the stocks with the highest expected returns and selling the stocks with the lowest expected returns generate large economic gains to investors (Kelly *et al.*, 2020).

In this work we frame the challenge of detecting predictive signals for the cross-section of stock returns as a *return classification* task and we propose a novel method to construct long-short portfolios based on return class probabilities. Our approach is grounded in statistical decision theory. We consider an investor that seeks to construct a portfolio by buying the top-performing stocks (the "winners") and selling the worst-performing ones (the "losers") on the basis of stock characteristics. We derive the optimal portfolio selection rule implied by an appropriate loss function that accounts for different types of misclassification errors in portfolio construction. The optimal rule buys or sells stocks based on their probabilities of being winners or losers, conditional on the stock characteristics. An empirical application using U.S. stock data shows that portfolios constructed with our optimal rules outperform those built using conventional techniques.

A growing body of research frames the detection of cross-sectional return predictability as a classification problem (Rapach *et al.*, 2024; He *et al.*, 2024; Makridakis *et al.*, 2024). This approach is appealing as binning returns into distinct classes expands the search for predictive signals beyond expected returns to the entire return distribution. This is of interest, for instance, in the presence of characteristics that are predictive of higher order moments of the distribution of returns, such as volatility and skewness. An important yet often overlooked issue in this literature concerns the mapping of return classification outputs into long-short portfolios, a key step in the assessment of the economic value of predictive signals. We contribute to this literature by providing optimal rules to map return class probabilities into long-short portfolios.

Our framework considers an investor that seeks to construct a long-short portfolio on the basis of stock characteristics. The investor is interested in choosing a portfolio selection rule—a function that maps stock characteristics into a buy, sell or no-trade decision for each stock in the cross-section of stocks. Consistent with the empirical asset pricing literature, the investors' objective is to determine which positions to take on each stock, rather than how much capital to allocate to each position. We introduce a loss function to rank competing portfolio selection rules. The proposed loss function is analogous to those used in the ranking and selection literature (e.g. Gu and Koenker, 2023) and incorporates tuning parameters to control the relative costs associated with different types of misclassification errors in portfolio construction. The tuning parameters may be interpreted as measures of the investor's degree of risk aversion, allowing the framework to be tailored to different risk preferences.

The optimal portfolio selection rule is obtained as the rule that minimizes the expected loss. The optimal rule buys stocks that have a sufficiently high probability of being winners and sells those with a sufficiently high probability of being losers, with the tuning parameters controlling the shape of the buy and sell regions. A key feature of our framework is that the optimal rule does not rely on parametric assumptions about the distribution of returns and does not impose a specific functional form linking characteristics to returns. Instead, it is solely determined by the probability that each stock becomes a winner or a loser, given its characteristics. In the special case in which returns are generated by an additive regression model with symmetric errors, the optimal portfolio selection rule simplifies to a function of each stock's expected return, volatility and the distribution of the standardized errors. In this setting, the optimal rule buys (sells) a stock when its expected return exceeds (falls below) a threshold that increases (decreases) with the stock's volatility. If, in addition, the misclassification costs satisfy a symmetry condition, the optimal portfolio selection rule corresponds to the conventional strategy based on sorting stocks according to expected returns.

We put our framework to the test in a sample of U.S. stocks from 1957 to 2021. Our dataset consists of equity return data for more than 7500 unique stocks over 756 months. Additionally, for each stock we consider a set of 94 characteristics (Kelly *et al.*, 2020). We construct long-short portfolios on the basis of different versions of the optimal portfolio selection rule that stem from our framework. We consider the optimal selection rule based on classification as well as the optimal selection rule based on regression, which arises when returns are assumed to be generated by an additive regression model. We refer to the portfolios constructed using the optimal rules as the optimal portfolios. We benchmark the optimal portfolios with portfolios constructed using conventional rules from the literature. For classification, we consider as a benchmark most-likely-class selection, that is, the rule that buys a stock if its most likely class is the winner class and sells it if its most likely class is the loser class (Rapach *et al.*, 2024). For regression, we consider as a benchmark the selection rule based on expected return sorts, the standard tool in the literature. We refer to the resulting portfolios as the benchmark portfolios. The implementation of the optimal rules requires estimates of the conditional probability of being a winner or a loser (in the classification case) or the expected return and conditional volatility of each stock (in the regression case). We estimate these quantities using linear methods (logistic and linear regression) as well as machine learning techniques (extreme gradient boosting for classification and regression). The optimal portfolios require the selection of appropriate values for the tuning parameters associated with the loss function. We determine these tuning parameters in a data-driven way by maximizing the Sharpe ratio of the resulting portfolios on a validation sample. We label portfolios based on classification with linear methods as LIN-C and those with machine learning techniques as ML-C. Similarly, portfolios based on regression with linear methods and machine learning techniques are labeled LIN-R and ML-R, respectively.

A number of results emerge from our analysis. First, optimal portfolios consistently outperform their benchmark counterparts. Second, optimal portfolios constructed using classification methods outperform those based on regression with either standard sorting or optimal selection rules. Third, portfolios built using machine learning techniques outperform those based on linear methods. Overall, optimal portfolios based on classification with machine learning achieve the highest expected returns and Sharpe ratios. Among the benchmark portfolios, the highest average monthly returns are 3.61% (ML-R) for equal-weighted portfolios and 1.68% (ML-R) for value-weighted portfolios. In comparison, the best performing optimal portfolios have average monthly returns of 6.12% (ML-C) and 4.00% (ML-C) for equal-weighted and value-weighted portfolios, respectively. The highest Sharpe ratio among benchmark portfolios is 2.88 (ML-C) for equal-weighted and 0.94 (LIN-R) for value-weighted portfolios. In contrast, the optimal portfolios deliver Sharpe ratios of 2.92 (ML-C) and 1.24 (ML-C) for equal-weighted and value-weighted portfolios, respectively.

We incorporate transaction costs in our analysis to provide a realistic assessment of the performance of the portfolio selection rules, following the methodology introduced in Ledoit and Wolf (2025). Once transactions costs are taken into account, none of the benchmark portfolios convincingly outperform the market. In contrast, optimal portfolios based on classification with machine learning techniques generate sizable gains to investors. The value-weighted benchmark ML-C portfolio achieves a Sharpe ratio of 0.48—the highest among all benchmark portfolios, though only marginally above the market Sharpe ratio of 0.46 over the same period. In contrast, the optimal ML-C portfolios have Sharpe ratios that are about 30% higher than the market Sharpe ratio, using equal weights, and about 80% higher, using value weights.

We examine whether the Fama-French 5-factor model (Fama and French, 1995) augmented with the momentum, short-term-reversal and long-term-reversal factors (FF5+) explains the excess returns generated by the portfolios. Gross of transaction costs, optimal portfolios based on classification generate positive and significant risk-adjusted returns, whereas their benchmark counterparts do not. In contrast, both benchmark and optimal portfolios based on regression generate positive and significant risk-adjusted returns. Risk adjusted returns of optimal portfolios are systematically larger than those of their corresponding benchmark portfolios. Moreover, the share of variation of portfolio returns explained by the FF5+ model is substantially smaller for the optimal portfolios than for the benchmark portfolios. Net of transaction costs, only the optimal ML-C portfolio yields positive and significant risk-adjusted returns.

We then study how the portfolios vary along the characteristics of their constituent stocks. The most pronounced differences between long and short positions emerge along characteristics such as short-term reversal, 12-2 months momentum, change in 6-month momentum, sales-to-price ratio, and industry momentum. Stocks in the long leg of the portfolios tend to have more extreme rankings of short-term reversal, size, idiosyncratic volatility, bid-ask spreads and return volatility. In contrast, stocks in the short leg of the portfolios tend to have more extreme rankings of maximum daily returns over the past month, return volatility, bid-ask spreads, idiosyncratic volatility and 12-2 months momentum. Optimal portfolios tend to select stocks with more extreme characteristic values than benchmark portfolios. This pattern also holds when comparing classification-based portfolios with their regression-based counterparts, and machine learning portfolios with portfolios constructed using linear methods.

Our results show that portfolios based on classification methods and optimal portfolio selection rules outperform portfolios based on regression, be it using optimal or benchmark portfolio selection rules. This suggests that considering the impact of characteristics on the distribution of returns—beyond their effect on expected returns—may lead to better portfolios. We also find that machine learning methods outperform linear methods, and the performance gains are substantially larger when optimal portfolio selection rules are employed. We attribute the performance of optimal portfolios based on classification to the strong discriminatory power achieved by these methods. In fact, for the same level of false positive rates (defined as buying a non-winner or as selling a non-loser), classification methods achieve twice as large true positive rates (defined as buying a winner or as selling a loser) relative to regression methods.

Our paper relates to several strands of the literature. First, our paper relates to the empirical finance literature which has long employed characteristic-sorted portfolios to construct investment strategies and to identify pricing anomalies (Jegadeesh and Titman, 1993; Fama and French, 1992, 1993). Second, our paper relates to the growing literature that constructs long-short portfolios using expected returns (Lewellen, 2015; Kelly *et al.*, 2020, to name a few) and return class probabilities (Rapach *et al.*, 2024; He *et al.*, 2024; Makridakis *et al.*, 2024; Del Viva *et al.*, 2024). Third, our paper relates to the literature that investigates the theoretical properties of portfolio sorts (Cattaneo *et al.*, 2020; Daniel *et al.*, 2020; Cattaneo *et al.*, 2023), and to the literature on mean-variance optimal characteristic-based portfolios (Brandt *et al.*, 2009; Ledoit *et al.*, 2019; Olmo and McGee, 2022). Finally, our work is related to the literature on optimal ranking and selection, which includes, among others Gu and Koenker (2023); Hirano and Porter (2009); Andrews *et al.* (2024); Gelman and Price (1999).

The remainder of this paper is organized as follows. Section 2 introduces the methodological framework. Section 3 details the empirical study. Section 4 contains our conclusions.

2 Methodology

2.1 How to Bet on Winners and Losers

One of the primary objectives of empirical asset pricing is to detect signals that are predictive for the cross-section of stock returns. A popular approach in the literature is to frame the detection of predictive signals as a return prediction problem (Lewellen, 2015; Freyberger *et al.*, 2020; Cattaneo *et al.*, 2020). Let R_{it} denote the return of stock i for $i = 1, \ldots, n$. Then, interest lies in learning expected returns conditional on past information, which are assumed to be determined by a set of stock characteristics, that is

$$\mathbb{E}_t(R_{i\,t+1}) = \mathbb{E}(R_{i\,t+1}|\boldsymbol{X}_{i\,t}) \ ,$$

where $\boldsymbol{X}_{i\,t} = (X_{1,i\,t}, \ldots, X_{p,i\,t})'$ denotes the p-dimensional vector of characteristics (Kelly *et al.*, 2020). Rather than approaching the detection of predictive signals as a return prediction problem, a growing literature casts this as a return classification problem (Rapach *et al.*, 2024; He *et al.*, 2024; Han, 2022). Let $Y_{i\,t}$ denote the return class of stock i for $i = 1, \ldots, n$, that is, the random variable taking values in $\{1, \ldots, K\}$ and such that

$$\begin{aligned}
Y_{i\,t+1} = 1 \quad &\Longleftrightarrow \quad \{\infty < R_{i\,t+1} \leq c_1\} \\
Y_{i\,t+1} = k \quad &\Longleftrightarrow \quad \{c_{k-1} < R_{i\,t+1} \leq c_k\} \text{ for } k = 2, \ldots, K-1, \\
Y_{i\,t+1} = K \quad &\Longleftrightarrow \quad \{c_{K-1} < R_{i\,t+1} < \infty\} \ ,
\end{aligned}$$

for some appropriate choice of class thresholds $-\infty < c_1 < c_2 < \ldots < c_{K-1} < \infty$. Then, interest lies in learning class probabilities conditional on past information, which are assumed to be determined by a set of stock characteristics, that is

$$\mathbb{P}_t(Y_{i\,t+1} = k) = \mathbb{P}(Y_{i\,t+1} = k|\boldsymbol{X}_{i\,t}) \text{ for } k \in 1, \ldots, K \ .$$

Framing the detection of cross-sectional predictive signals as a classification problem broadens the search for predictive signals beyond expected returns to the entire return distribution. This is appealing, for instance, in the presence of characteristics that are predictive of higher order moments, such as volatility and skewness.

Arguably, the usefulness of a predictive signal lies in its ability to generate economic value to investors. The conventional strategy to assess the economic value of predictive signals is based on the performance of long-short portfolios constructed from such signals. Several procedures may be used to map predictive signals into long-short portfolios. In the case of return prediction, the standard strategy consists in buying (selling) a stock if its expected return is in the top (bottom) decile of the cross-section of expected returns (Kelly *et al.*, 2020). By contrast, it is unclear how to map class probabilities—of which there are K—into long-short portfolios. One proposal, put forth in Rapach *et al.* (2024), is to buy a stock if its most likely class is the winner class and sell it if its most likely class is the loser class. This strategy is known to be the optimal classification rule when classes are unordered (see, e.g. Hastie *et al.*, 2009), in which case wrongly labelling a loser stock as a winner carries the same cost as labelling a non-winner as a winner. In the portfolio selection setting these errors carry different costs for investors, so they should also carry different misclassification costs.

In this work we propose novel long-short portfolio selection rules based on return class probabilities. We cast the construction of a long-short portfolio as a decision theoretic problem and introduce the optimal portfolio selection rule implied by an appropriate loss function. We introduce the portfolio selection rule $\delta(\boldsymbol{X}_{i\,t}, \ldots, \boldsymbol{X}_{n\,t})$ as a function that maps stock characteristics into a buy (1), sell (-1) or no-trade (0) decision for each stock, that is

$$\delta : \mathbb{R}^{p\times n} \rightarrow \{-1, 0, 1\}^n \ .$$

We use the shorthand notation $\delta_{i\,t}$ to denote the i-th component of $\delta(\boldsymbol{X}_{i\,t}, \ldots, \boldsymbol{X}_{n\,t})$. We assume the investor is concerned with minimizing the sum of the misclassification losses for the stocks in the top class (the winners) and the bottom class (the losers) and introduce a loss function to rank competing portfolio selection rules. We define our loss function for a generic portfolio selection rule δ as

$$\mathcal{L}_{\lambda_W,\lambda_L}(\delta) = \mathcal{L}^W_{\lambda_W}(\delta) + \mathcal{L}^L_{\lambda_L}(\delta) \ , \tag{1}$$

where

$$\mathcal{L}^W_{\lambda_W}(\delta) = \sum_{i=1}^{n} \underbrace{\mathbb{1}(\{Y_{i\,t+1} = K\} \cap \{\delta_{i\,t} \neq 1\})}_{\text{False Negative Error}} + \lambda_W \underbrace{\mathbb{1}(\{Y_{i\,t+1} \neq K\} \cap \{\delta_{i\,t} = 1\})}_{\text{False Positive Error}}, \text{ and}$$
$$\mathcal{L}^L_{\lambda_L}(\delta) = \sum_{i=1}^{n} \underbrace{\mathbb{1}(\{Y_{i\,t+1} = 1\} \cap \{\delta_{i\,t} \neq -1\})}_{\text{False Negative Error}} + \lambda_L \underbrace{\mathbb{1}(\{Y_{i\,t+1} \neq 1\} \cap \{\delta_{i\,t} = -1\})}_{\text{False Positive Error}} \ ,$$

with $\lambda_W \geq 0$ and $\lambda_L \geq 0$ tuning parameters measuring the relative costs of selecting, respectively, a false winner and false loser. These tuning parameters may be interpreted as risk aversion parameters in a sense that will be made clear below. We emphasize that the proposed loss takes into account the ordering of the classes. In fact, the contribution of the i-th stock to the loss is given by

		Loser		Neutral		Winner
		$Y_{i\,t+1}=1$	$Y_{i\,t+1}=2$	$\dots$	$Y_{i\,t+1}=K-1$	$Y_{i\,t+1}=K$
Sell	$\delta_{i\,t}=-1$	0	λ_L	$\dots$	λ_L	$1+\lambda_L$
No-trade	$\delta_{i\,t}=0$	1	0	$\dots$	0	1
Buy	$\delta_{i\,t}=1$	$1+\lambda_W$	λ_W	$\dots$	λ_W	0

The loss function depends on the tuning parameters λ_W and λ_L, and is akin to the loss function used in the literature on ranking and selection (Gu and Koenker, 2023). The tuning parameters may be set a priori (e.g. one may set $\lambda_W = \lambda_L = 1$) or optimized according to a data-driven criterion.

Given a loss function, the optimal portfolio selection rule may be defined as the rule that minimizes the expectation of the loss in (1), that is

$$\delta^* \in \arg\min_{\delta} \mathbb{E}_t(\mathcal{L}_{\lambda_W,\lambda_L}(\delta)) \ . \tag{2}$$

The following proposition characterizes the solution to this optimization problem. We refer to the optimal rule established in the proposition below as the optimal portfolio selection rule for classification.

Proposition 1 (Optimal Portfolio Selection Rule for Classification)**.** *The optimal portfolio selection rule δ^* is such that*

$$\delta^*_{i\,t} = \begin{cases} 1 & \text{if } p^W_{i\,t} \geq \max\left(\frac{\lambda_W}{1+\lambda_W}, \frac{\lambda_W-\lambda_L}{1+\lambda_W} + \frac{1+\lambda_L}{1+\lambda_W}p^L_{i\,t}\right) \\ -1 & \text{if } p^L_{i\,t} \geq \max\left(\frac{\lambda_L}{1+\lambda_L}, \frac{\lambda_L-\lambda_W}{1+\lambda_L} + \frac{1+\lambda_W}{1+\lambda_L}p^W_{i\,t}\right) \\ 0 & \text{otherwise} \end{cases} , \tag{3}$$

where $p^L_{i\,t} = \mathbb{P}(Y_{i\,t+1} = 1|\boldsymbol{X}_{i\,t})$ and $p^W_{i\,t} = \mathbb{P}(Y_{i\,t+1} = K|\boldsymbol{X}_{i\,t})$.

In words, the optimal portfolio selection rule δ^* buys (sells) stocks that have a sufficiently high probability of becoming winners (losers), and does not trade on stocks that do not have a sufficiently high probability of being winners or losers. Figure 1 plots the selection region implied by equation (3) for $\lambda_W = \lambda_L = 1$ (left panel) and for $\lambda_W = 1/3$ and $\lambda_L = 1/2$ (right panel). As can be seen in Figure 1, the parameters λ_W, λ_L control the shape of the selection region. We remark that the selection region depends on λ_W, λ_L and the ratio between the two quantities. In particular, increasing λ_W while keeping λ_L fixed reduces the "buy" region. Conversely, increasing λ_L while keeping λ_W fixed reduces the "sell" region. Setting $\lambda_W < \lambda_L$ increases the area of the "buy" region relative to the "sell" region. We emphasize that the optimal selection rule does not depend on specific distributional assumptions on the returns and, importantly, it does not depend on the choice (or number) of intermediate thresholds $c_2, \ldots, c_{K-2}$.

[FIGURE 1 ABOUT HERE]

It is interesting to explore the optimal portfolio selection rule under the additional assumption that stock returns are generated by an additive regression model as in Kelly *et al.* (2020). Specifically, we assume that

$$R_{i\,t+1} = \mathbb{E}(R_{i\,t+1}|\boldsymbol{X}_{i\,t}) + \epsilon_{i\,t+1} \ , \tag{4}$$

where $\epsilon_{i\,t+1}$ is an error term with variance $\mathrm{Var}(R_{i\,t+1}|\boldsymbol{X}_{i\,t})$. We further assume that the standardized residuals $Z_{i\,t+1} = \epsilon_{i\,t+1}/\sqrt{\mathrm{Var}(R_{i\,t+1}|\boldsymbol{X}_{i\,t})}$ have distribution F_Z, which is symmetric and admits an inverse. The following proposition characterizes the optimal portfolio selection rule for the additive regression model given in (4). We refer to the optimal rule established in the proposition below as the optimal portfolio selection rule for regression.

Proposition 2 (Optimal Portfolio Selection Rule for Regression)**.** *If returns are generated by the additive regression model given in* (4)*, the optimal portfolio selection rule δ^* is such that*

$$\delta^*_{i\,t} = \begin{cases} 1 & \textit{if } F_Z\left(\frac{\mu_{i\,t} - c_{K-1}}{\sqrt{\sigma^2_{i\,t}}}\right) \geq \max\left(\frac{\lambda_W}{1+\lambda_W}, \frac{\lambda_W - \lambda_L}{1+\lambda_W} + \frac{1+\lambda_L}{1+\lambda_W} F_Z\left(\frac{c_1 - \mu_{i\,t}}{\sqrt{\sigma^2_{i\,t}}}\right)\right) \\ -1 & \textit{if } F_Z\left(\frac{c_1 - \mu_{i\,t}}{\sqrt{\sigma^2_{i\,t}}}\right) \geq \max\left(\frac{\lambda_L}{1+\lambda_L}, \frac{\lambda_L - \lambda_W}{1+\lambda_L} + \frac{1+\lambda_W}{1+\lambda_L} F_Z\left(\frac{\mu_{i\,t} - c_{K-1}}{\sqrt{\sigma^2_{i\,t}}}\right)\right) \\ 0 & \textit{otherwise} \end{cases} \ , \tag{5}$$

where $\mu_{i\,t} = \mathbb{E}(R_{i\,t+1}|X_{i\,t})$ and $\sigma^2_{i\,t} = Var(R_{i\,t+1}|\boldsymbol{X}_{i\,t})$. If in addition $\lambda_W \geq 1$ and $\lambda_L \geq 1$, then *the optimal portfolio selection rule simplifies to*

$$\delta_{it}^* = \begin{cases} 1 & \textit{if } \mu_{it} - F_Z^{-1}(\frac{\lambda_W}{1+\lambda_W})\sqrt{\sigma_{it}^2} > c_{K-1} \\ -1 & \textit{if } \mu_{it} + F_Z^{-1}(\frac{\lambda_L}{1+\lambda_L})\sqrt{\sigma_{it}^2} \leq c_1 \\ 0 & \textit{otherwise} \end{cases} . \tag{6}$$

Proposition 2 shows that, in the additive regression model, the inclusion of a stock in the portfolio depends on the stock's expected return and volatility. Figure 2 depicts the selection region obtained for $\lambda_W = \lambda_L = 1$ (left panel) and $\lambda_W = \lambda_L = 1.1$ (right panel). The area of the selection region depends on λ_W and λ_L. As in the general case, increasing λ_W reduces the area of the "buy" region, whereas increasing λ_L reduces the area of the "sell" region. Setting $\lambda_W = \lambda_L = 1$ recovers the classic expected return sorting procedure, which can thus be seen as a special case of the optimal selection rule for regression. As pointed out in Ledoit *et al.* (2019), deciding which stocks to hold solely on the basis of expected return without any regard for volatility may be undesirable. Setting $\lambda_W > 1$ and $\lambda_L > 1$ is consistent with a risk-averse investor that, when comparing two stocks with the same expected returns, prefers the one with the lower volatility. In particular, the optimal rule recommends a long (short) position in a stock if its expected return is sufficiently high (low) so as to compensate for its volatility. Finally, we remark that although the assumptions on the standardized error Z_{it} are somewhat restrictive, they allow us to derive an insightful expression for the optimal portfolio selection rule for regression.

[FIGURE 2 ABOUT HERE]

A few remarks are in order. The first remark concerns the choice of the return classes. When casting the detection of predictive signals as a classification problem, a natural approach is to set the winner and loser class thresholds c_{K-1} and c_1 as the top and bottom quantiles of the cross-sectional distribution of *realized* returns, as in Rapach *et al.* (2024). By contrast, when the predictive task is framed as a return prediction problem, a natural approach is to set the winner and loser class thresholds c_{K-1} and c_1 as the top and bottom quantiles of the cross-section of *expected* returns. In fact, in this case our optimal portfolio selection rule for regression subsumes (when $\lambda_L = \lambda_W = 1$) the classic long-short portfolio selection rule based on expected return sorts, in which the investor buys stocks whose expected returns are above the top quantile of the cross-section of expected returns and sells those whose expected returns are below the bottom quantile of the cross-section of expected returns. More generally, our framework allows return classes to be defined based on either the distribution of realized or expected returns. Each definition may be of value to investors, and determining which is more suitable for practical applications is an empirical question. The second remark concerns the return classification problem. Within our statistical decision theory framework, return classification—and the associated focus on return class probabilities—arises from the investor's objective of solving a specific problem: the construction of the optimal long-short portfolio implied by an appropriate loss function associated with portfolio selection. This parallels the return prediction case, where the focus on expected returns arises from the investor's goal of accurately predicting returns under the square loss.

2.2 Implementation of the Optimal Selection Rules

In this section, we discuss how to implement the optimal portfolio selection rules for both classification and regression. While there are multiple ways to operationalize these rules, we focus on implementations that align with standard practices in the empirical asset pricing literature (Kelly *et al.*, 2020; Rapach *et al.*, 2024).

To implement the optimal portfolio selection rule for classification in practice, we need to set the return class variable $Y_{i\,t}$ and estimate the conditional class probabilities $\mathbb{P}(Y_{i\,t+1} = k|\boldsymbol{X}_{i\,t})$ for each $k = 1, \ldots, K$. We resort to return decile classes (i.e. $K = 10$), that is, we define $Y_{i\,t}$ as

$$Y_{i\,t} = k \Leftrightarrow \{R_{(\lceil[(k-1)/10]n\rceil)\,t} < R_{i\,t} \leq R_{(\lceil(k/10)n\rceil)\,t}\}\ ,$$

where $R_{(q)\,t}$ denotes the q-th order statistic of returns at time t for $q = 1, \ldots, n$ and $R_{(0)\,t}$ is defined as $-\infty$. We assume that the conditional class probabilities are a function of the characteristics alone, that is, there are functions $g_k : \mathbb{R}^p \to [0, 1]$ for $k = 1, \ldots, K$ such that

$$g_k(\boldsymbol{X}_{i\,t}) = \mathbb{P}(Y_{i\,t+1} = k|\boldsymbol{X}_{i\,t}) \text{ for } k \in 1, \ldots, K,$$

and $\sum_{k=1}^{K} g_k(\boldsymbol{X}_{i\,t}) = 1$. The conditional class probabilities are then estimated by pooling information over the cross-section and time dimensions. In the empirical application, we rely on a pooled multinomial logistic regression and pooled extreme gradient boosting for classification. The investor can then construct the portfolios by applying the optimal rule in Proposition 1.

To implement the optimal portfolio selection rule for regression in practice, we need to estimate the conditional expected returns $\mathbb{E}(R_{i\,t+1}|\boldsymbol{X}_{i\,t})$, the conditional variances $\text{Var}(R_{i\,t+1}|\boldsymbol{X}_{i\,t})$ and the distribution of standardized residuals F_Z. We assume that the conditional expected returns are a function of the characteristics alone, that is there is function $m : \mathbb{R}^p \to \mathbb{R}$ such that

$$m(\boldsymbol{X}_{i\,t}) = \mathbb{E}(R_{i\,t+1}|\boldsymbol{X}_{i\,t}) \ ,$$

and that the conditional variances solely depend on i, that is $\sigma_i^2 = \text{Var}(R_{i\,t+1}|\boldsymbol{X}_{i\,t})$. The conditional expected returns are then estimated by pooling information over the cross-section and time dimensions. In the empirical application we rely on a pooled linear regression and pooled extreme gradient boosting for regression. The conditional variance of the i-th stock is estimated as the variance of the residuals $\hat{\epsilon}_{i\,t} = R_{i\,t+1} - \hat{m}(\boldsymbol{X}_{i\,t})$ of that stock. Finally, we assume that the standardized residual distribution is a standard Gaussian[1], set the class thresholds c_1 and c_{K-1} to $\hat{m}_{(\lceil 0.1n \rceil)\,t}$ and $\hat{m}_{(\lceil 0.9n \rceil),t}$, and construct the corresponding probabilities as

$$\begin{aligned} g_K(\boldsymbol{X}_{i\,t}) &= \Phi\Big(\frac{\hat{m}(\boldsymbol{X}_{i\,t}) - c_{K-1}}{\sqrt{\hat{\sigma}_i^2}}\Big), \ \text{and} \\ g_1(\boldsymbol{X}_{i\,t}) &= \Phi\Big(\frac{c_1 - \hat{m}(\boldsymbol{X}_{i\,t})}{\sqrt{\hat{\sigma}_i^2}}\Big), \end{aligned}$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of a standard Gaussian. The investor can then construct the optimal portfolio by applying the optimal rule in Proposition 2.

In both the classification and regression approaches, the investor must choose the costs associated with misclassification errors, λ_W and λ_L. Since the rules we consider are optimal given a choice of λ_W and λ_L, no statistical criteria is available to guide the selection of the loss tuning parameters. We assume that the investor ranks portfolios based on their Sharpe ratios and wishes to maximize this metric. Therefore, we propose selecting the pair λ_W and λ_L corresponding to the portfolio that achieves the highest Sharpe ratio on the basis of the data. Alternative criteria—such as expected returns or portfolio variance—could also be considered. This type of criterion-based tuning is standard in the selection and ranking literature, where

[1] We remark that any symmetric distribution that admits an inverse would be a viable alternative. We use the standard Gaussian for simplicity.

data-driven choices of misclassification costs are typically carried out to control the type-I error of the selection procedure (Gu and Koenker, 2023).

It is important to note that in our empirical implementation the class thresholds are order statistics and thus are functions of the data that are not fixed, as assumed in the theoretical framework. Nonetheless, given the large cross-sectional dimension of the data, we argue that these estimated thresholds ought to converge to their population counterparts, ensuring consistency with the theoretical setup.

3 Empirical Application

We construct long-short portfolios on the basis of stocks characteristics using a large cross-section of U.S. stocks. Our data consist of monthly stock prices for all firms listed on the New York Stock Exchange, American Stock Exchange, or Nasdaq. We consider ordinary equities (share codes 10 and 11) from the Center for Research in Security Prices (CRSP) spanning the period from January 1957 to December 2021. The data forms an unbalanced panel with an average of 5000 stocks per month. We use the 94 firm characteristics employed in Kelly *et al.* (2020), which are based on those considered in Green *et al.* (2017).[2] We map characteristics into the $[0, 1]$ interval according to cross-sectional rankings as in Kelly *et al.* (2020) and Freyberger *et al.* (2020), among others. We replace missing characteristics with the cross-sectional median for each time period, and we append delisting returns when available.

We consider the optimal portfolio selection rules for classification and regression, which are implemented as described in Section 2.2. We benchmark our optimal selection rules against standard portfolio selection rules used in the literature. In the classification setup, the standard approach employed in the literature (Rapach *et al.*, 2024) is to label a stock according to $\arg\max_k \mathbb{P}(Y_{i\,t+1} = k|\boldsymbol{X}_{i\,t})$. That is, buy (sell) a stock if its most likely class is the winner (loser) class. In the regression setup, the standard approach employed in the literature (Kelly *et al.*, 2020) is to label a stock as a winner (loser) if it has expected returns higher (lower) than 90% of the cross-section of expected returns. The implementation of the rules requires estimates of the conditional probability of being a winner or a loser (in the classification case) or the expected return and conditional volatility of each stock (in the return prediction case). In the classification case we consider predictions based on a (multinomial) logistic regression as well as eXtreme

[2] We download the characteristics data from Dacheng Xiu's website.

Gradient Boosting with a (multi-class) classification loss function (XGBC). Portfolios based on these methods are labelled as LIN-C and ML-C, respectively. In the regression case we consider predictions based on linear regression as well as eXtreme Gradient Boosting with the square loss function (XGBR). Portfolios based on these methods are labelled as LIN-R and ML-R, respectively. We consider extreme gradient boosting models with 100 rounds of boosting and we treat the depth of the tree and the learning rate of each boosting round as tuning parameters. In particular, we consider tree-depths $d \in \{2, 3\}$ and learning rates $\eta \in \{0.01, 0.1, 0.25, 0.5, 0.75, 1\}$.

We split the sample into training, validation, and testing sets. The first training period starts in January 1957 and ends in January 1974. The first validation period starts in January 1974 and ends in December 1986. The first testing period starts in January 1987. We fix the validation set length to 13 years and reestimate all models annually using an expanding training window. Tuning parameters for the XGBC and XGBR models are selected based on their performance over the rolling validation sample, following Kelly *et al.* (2020). We use the selected model to construct pseudo-out-of-sample forecasts for the current validation and testing periods. Idiosyncratic volatilities are estimated as the volatility of the entire sequence of pseudo-out-of-sample forecast errors. That is, we use an expanding window that starts in January 1974 to estimate idiosyncratic volatilities. We construct portfolios based on a 150×150 grid of equally spaced values from 0 to 1.50 for λ_W and λ_L. Prior to each month in the out-of-sample period, we choose the λ_W and λ_L that yield the portfolio with the highest Sharpe ratio over the whole sequence of pseudo-out-of-sample observations, which starts in January 1974.

3.1 Performance of the Optimal Portfolios

In this section, we investigate the properties of the portfolios obtained from all forecasting models and selection rules. We find that the optimal portfolios based on classification with machine learning deliver higher expected returns and Sharpe ratios than the benchmark portfolios.

[TABLE 1 ABOUT HERE]

Table 1 reports the average monthly returns, the volatility, the downside volatility constructed as the volatility of negative returns, the skewness, the annualized Sharpe ratio, the annualized Sortino ratio, the minimum, maximum and the quartiles of the return distribution of each of the portfolios considered, as well as the the average percentage of stocks selected. Panel

A of Table 1 reports results for the equally-weighted portfolios. Panel B reports results for the value-weighted portfolios.

A number of remarks are in order. First, most equally-weighted optimal portfolios have higher expected returns and Sharpe ratios than their corresponding benchmarks, and the same is true for most value-weighted optimal portfolios. In addition, and despite not being a direct target of our framework, most optimal portfolios have higher Sortino ratios and skewness than their corresponding benchmarks. Second, optimal portfolios select substantially fewer stocks than benchmark portfolios. The optimal ML-C portfolio trades, on average, on 12.52% of the stocks available in the cross-section. In contrast, the benchmark ML-C portfolio trades on about 34.67% of the stocks available in the cross-section. Optimal LIN-R portfolios are the exception, with the optimal portfolio selecting, on average, 32.31% of the stocks available in the cross-section, in contrast to the 20% of stocks selected by the benchmark portfolio. Third, optimal portfolios based on return classification have higher Sharpe ratios than their regression based counterparts. In line with the findings in Rapach *et al.* (2024), the optimal ML-C portfolio is the portfolio with the highest expected returns and Sharpe ratio for both equally and value-weighted portfolios.

[FIGURE 3 ABOUT HERE]

Figure 3 reports the natural logarithm of the cumulative returns obtained by the equal and value-weighted optimal ML-C portfolios along with their corresponding benchmark. Cumulative returns for the optimal ML-C portfolios stochastically dominate those from the benchmark ML-C portfolios.

3.1.1 Accounting for Transaction costs

We carefully account for transaction costs following Ledoit and Wolf (2025) to provide a realistic assessment of competing portfolio selection rules. After accounting for transaction costs, only the optimal portfolios based on classification with machine learning continue to outperform the market.

Due to data availability, we carry out the analysis of transaction costs on a reduced sample on both the cross-section and time-series dimensions. On the time-series dimension, we focus our analysis to data from January 2000 to December 2021, the end of our sample. On the cross-section, we restrict the sample to stocks with daily bid-ask spreads available on CRSP. We follow Ledoit and Wolf (2025) and set transaction costs to one half of the stock-specific bid-ask spread. Because we include a larger universe of stocks than that considered in Ledoit and Wolf (2025), and in particular our universe includes small stocks, we estimate the bid-ask spread using the quoted spreads, defined as the average of daily bid-ask spreads divided by the daily quote midpoints within a month. We report results obtained by estimating bid-ask spreads as in Ledoit and Wolf (2025) in Table OA.1 of the Online Appendix.

[TABLE 2 ABOUT HERE]

Table 2 reports our results. Once transactions costs are taken into account, benchmark portfolios based on machine learning forecasts marginally—if at all—outperform the market. Optimal portfolios based on machine learning still generate sizable gains to investors. The value-weighted optimal ML-C portfolio yields a Sharpe ratio of 0.84, the highest of all portfolios considered. In contrast, the value-weighted benchmark ML-C portfolio yields a Sharpe ratio of 0.48. Moreover, the equally-weighted optimal ML-C portfolio yields a Sharpe ratio of 0.61, and the equally-weighted optimal ML-R portfolio yields a Sharpe ratio of 0.5. In contrast, portfolios based on linear methods (LIN-C and LIN-R) deliver Sharpe ratios that fall short of the market, which yields a Sharpe ratio of 0.46 over the same period.

[FIGURE 4 ABOUT HERE]

Figure 4 reports the natural logarithm of cumulative returns from optimal and benchmark ML-C portfolios accounting for transaction costs. We report both equally-weighted portfolios (solid lines) and value-weighted portfolios (dashed lines). As before, optimal portfolios outperform their corresponding benchmarks.

3.1.2 Risk-adjusted Returns

We report risk-adjusted returns for value-weighted portfolios based on optimal and benchmark selection rules and across all forecasting models. Risk-adjusted returns from optimal portfolios are generally higher than their corresponding benchmarks. Once transaction costs are accounted for, only the optimal portfolio based on classification with machine learning delivers positive and significant risk-adjusted returns. Results for equally-weighted portfolios can be found in Table OA.2 of the Online Appendix.

[TABLE 3 ABOUT HERE]

Table 3 reports the regression coefficients of the portfolio returns (in columns) on the FF5+ model. Alphas are multiplied by 100. The FF5+ model successfully explains the excess returns from the benchmark LIN-C and ML-C portfolios. In contrast, optimal LIN-C and ML-C portfolios yield significant risk-adjusted returns. All portfolios constructed on the basis of return predictions (LIN-R and ML-R) lead to positive and significant risk-adjusted returns. Risk-adjusted returns from the optimal portfolios are larger than those from their corresponding benchmark portfolios, with the exception of the LIN-R portfolio. Broadly speaking, loadings of optimal and benchmark portfolios on each of the risk factors considered are similar within model, but vary across models. Finally, the share of variation in portfolio returns explained by the FF5+ model is substantially smaller for the optimal than for the benchmark portfolios.

[TABLE 4 ABOUT HERE]

Table 4 reports the regression coefficients of the net-of-transaction-costs value-weighted portfolio returns (in columns) on the FF5+ model. Results for equally-weighted portfolios can be found in Table OA.3 of the Online Appendix. Alphas are multiplied by 100. Once transaction costs are accounted for, only the optimal ML-C portfolio yields positive and significant risk-adjusted returns. Overall, differences in risk loadings across portfolio selection rules are unlikely explain the spread in excess returns, and once transaction costs are accounted for, only the optimal ML-C portfolio leads to positive and significant risk-adjusted returns.

3.2 Dissecting Optimal Portfolios

3.2.1 Stock Characteristics

We study how the characteristics of the constituent stocks vary across the constructed portfolios. The most marked differences between long and short positions emerge along characteristics such as short-term reversal, 12-2 months momentum, change in 6-month momentum, sales-to-price ratio, and industry momentum.

Table 5 reports the percentiles of each characteristic (rows) for the average stock bought and sold by each of the machine learning portfolios (columns), as well as their differences. Characteristics are ordered by their relevance to the optimal ML-C portfolio, defined as the absolute difference in characteristic percentiles between the long and short legs. For readability, the table is truncated to the top 40 characteristics. Table 6 reports analogous results for the linear portfolios, sorted by the characteristic relevance to the optimal LIN-C portfolio. The full results for all characteristics are provided in Tables OA.4 and OA.5 in the Online Appendix.

[TABLES 5 and 6 ABOUT HERE]

Characteristics based on past returns—such as short-term reversal and 12-2 months momentum—play a relevant role across all portfolios. In fact, they are the characteristics whose values differ the most between the long and the short leg of all portfolios considered. Lower past month returns are associated with a higher probability of becoming a winner in the following period, and higher past month returns are associated with a higher probability of becoming a loser in the following period. Conversely, having a low 12-2 months momentum ranking is associated with a higher probability of becoming a loser in the following period, but stocks in the long leg of the constructed portfolios typically hover around the median of the 12-2 months momentum. Moreover, optimal portfolios tend to select stocks with more extreme characteristic values than benchmark portfolios. This pattern also holds when comparing classification-based portfolios with their regression-based counterparts, and machine learning portfolios with portfolios constructed using linear methods.

3.2.2 Predicting Winners and Losers

We study the classification properties of each of the forecasts. We find that identifying losers is easier than identifying winners, that machine learning methods have higher discriminatory power than linear methods, that classification methods outperform their regression counterparts, and that both classification and regression with linear or machine learning methods outperform random guessing.

To visualize the classification properties of each of the forecasts, we plot two independent Receiver Operating Characteristic (ROC) curves for the classification of winners and losers. Although we cast portfolio construction as a multi-class classification problem, we explore the properties of the forecasting models in identifying losers and winners independently for simplicity. Figure 5 reports the ROC curves for the classification of winners (top) and losers (bottom) across several models. Solid lines are constructed from probabilities estimated from multi-class classification models. Dashed lines are constructed from probabilities obtained using the implementation of the optimal selection rules for regression as described in Section 2.2. We remind the reader that the probabilities obtained from the classification implementation represent the likelihood that a stock's realized return exceeds (falls below) the top (bottom) decile of realized returns. In the case of regression, the probabilities represent the likelihood that a stock's realized return exceeds (falls below) the top (bottom) decile of expected returns. Blue lines represent the XGB model for regression and classification, and red lines represent linear models for regression and classification. The dots mark the performance of selecting stocks in the top (bottom) decile of conditional mean forecasts in terms of true positive rate and false positive rate.

[FIGURE 5 ABOUT HERE]

Classification models achieve better performance than regression models in identifying winners and losers, yielding a higher true positive rate for the same false positive rate. Labelling stocks in the top decile of XGBR predicted returns as winners incurs in a true positive rate of about 15.4% and a false positive rate of about 9.4%. For linear regression forecasts, the true positive rate is of 12.5%, and the false positive rate is of about 9.7%. In contrast, for a false positive rate of 9.4%, the XGBC model has a true positive rate of about 22.5%, whereas the logistic model has a true positive rate of about 21.2%. Labelling stocks in the bottom decile of XGBR predicted returns as losers incurs in a true positive rate of about 19.3% with a false positive rate of about 9%. For linear regression forecasts, the true positive rate is of 14.3%, and the false positive rate is of about 9.5%. In contrast, for a false positive rate of 9%, the XGBC model has a true positive rate of about 32.3%, whereas the logistic model has a true positive rate of about 30.4%. Among both the classification and regression models, machine learning methods outperform linear methods, and both outperform random guessing.

[TABLE 7 ABOUT HERE]

Table 7 reports the confusion matrices of all portfolios. Each row in the table reports the percentage of the predicted class (in rows) that correspond with the realized class (in columns), for both benchmark and optimal portfolios. All optimal portfolios achieve higher precision than their corresponding benchmarks, where precision is defined for each class as the proportion of correct predictions among all predictions made for that class. Similarly, portfolios based on classification have higher precision than their regression counterparts, and those based on machine learning have higher precision than their linear counterparts.

3.2.3 Tuning Parameter Selection

We next investigate the role of tuning parameter selection in the performance of the optimal portfolios. Sharpe ratios vary smoothly with respect to λ_W and λ_L, and each portfolio exhibits a broad region of (λ_W, λ_L) combinations that produce high Sharpe ratios. The results in this section support our earlier finding that the optimal ML-C portfolio consistently achieves the highest Sharpe ratios among all portfolios, and that these Sharpe ratios are attainable across a wide region of λ_W and λ_L.

[FIGURE 6 ABOUT HERE]

Figure 6 contains the out-of-sample annualized Sharpe ratios achieved across different models and (λ_W, λ_L) pairs. The highest Sharpe ratio achieved across all portfolios is of 3.63, which is obtained by ML-C with $\lambda_W = 0.25$ and $\lambda_L = 0.45$. This choice of parameters implies that the investor should buy stocks that have a probability of being a winner of at least 20% and sell stocks that have a probability of being loser of at least 31%. Note that the cost of buying a loser stock is smaller than the cost of selling a winner. ML-R portfolios have a maximum Sharpe ratio of 2.84, achieved by setting $\lambda_W = 1.03$ and $\lambda_L = 1.05$, for comparison, we note that setting $\lambda_W = \lambda_L = 1$, the standard sorting strategy, would produce a Sharpe ratio of 2.56. LIN-C portfolios achieve a maximum Sharpe ratio of 2.43 by setting $\lambda_W = 0.28$ and $\lambda_L = 0.41$. Finally, LIN-R portfolios achieve a maximum Sharpe ratio of about 2 by setting $\lambda_W = 1$ and $\lambda_L = 1.04$. Standard portfolio sorts would provide a Sharpe ratio of about 1.95.

We remark that the maximum Sharpe ratios described above are not feasible: investors would not have known to choose the optimal values of λ_W and λ_L for the whole out-of-sample period at the beginning of the out-of-sample period. Their feasible counterparts, however, display strong performance, as documented in the previous sections.

4 Conclusion

We develop a novel methodology for constructing long-short portfolios based on stock characteristics. Our approach is grounded in statistical decision theory. We consider an investor who aims to buy the top-performing stocks and sell the worst-performing ones in the cross-section of stocks. We derive the optimal portfolio selection rule implied by a loss function that explicitly accounts for different types of misclassification errors in portfolio construction. The resulting optimal rule prescribes buying or selling stocks based on their estimated probabilities of being winners or losers, conditional on stock characteristics. When returns are generated by an additive regression model, this simplifies to a decision rule that selects stocks based on their expected returns and volatilities, again conditional on stock characteristics. We apply our framework to U.S. equity data and evaluate the performance of our optimal portfolio selection rules. Portfolios constructed using our classification-based approach implemented with machine learning techniques consistently outperform benchmark strategies in terms of Sharpe ratios and risk-adjusted returns. These performance gains persist once transaction costs are accounted for. Overall, our findings highlight that portfolios constructed with optimal selection rules on the basis of stock characteristics deliver substantial economic gains to investors, which indicates that predictive signals for the cross-section of stock returns may extend beyond expected returns.

References

Andrews, I., Kitagawa, T., and McCloskey, A. (2024). Inference on winners. *The Quarterly Journal of Economics*, **139**(1), 305–358.

Bali, T., Beckmeyer, H., Morke, M., and Weigert, F. (2023). Option return predictability with machine learning and big data. *The Review of Financial Studies*, **36**(9), 3548–3602.

Brandt, M. W., Santa-Clara, P., and Valkanov, R. (2009). Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *The Review of Financial Studies*, **22**(9), 3411–3447.

Cattaneo, M. D., Crump, R. K., Farrell, M., and Schaumburg, E. (2020). Characteristic-sorted portfolios: Estimation and inference. *Review of Economics and Statistics*, **102**(3), 531–551.

Cattaneo, M. D., Crump, R. K., and Weining, W. (2023). Beta-sorted portfolios. *Working Paper*.

Chen, L., Pelger, M., and Zhu, J. (2023). Deep learning in asset pricing. *Management Science*, **70**(2), 714–750.

Daniel, K., Mota, L., Rottke, S., and Santos, T. (2020). The cross-section of risk and returns. *The Review of Financial Studies*, **33**(5), 1927–1979.

Del Viva, L., Sala, C., and Souza, A. B. (2024). Directional information in equity returns. *SSRN*.

Fama, E. and French, K. (1995). Size and book-to-market factors in earnings and returns. *Journal of Finance*, **50**, 131–155.

Fama, E. F. and French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance*, **47**(2), 427–465.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, **3**, 3–56.

Freyberger, J., Neuhierl, A., and Weber, M. (2020). Dissecting characteristics nonparametrically. *The Review of Financial Studies*, **33**(5), 2326–2377.

Gelman, A. and Price, P. N. (1999). All maps of parameter estimates are misleading. *Statistics in Medicine*, (18), 3221–3234.

Green, J., Hand, J. R., and Zhang, F. X. (2017). The characteristics that provide independent information about average u.s. monthly stock returns. *The Review of Financial Studies*, **30**(12), 4389–4436.

Gu, J. and Koenker, R. (2023). Invidious comparisons: Ranking and selection as compound decisions. *Econometrica*, **91**(1), 1–41.

Han, C. (2022). Bimodal characteristic returns and predictability enhancement via machine learning. *Management Science*, **68**(10), 7065–7791.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. 2nd edition.

He, S., Lv, L., and Zhou, G. (2024). Empirical asset pricing with probability forecasts. *SSRN*.

Hirano, K. and Porter, J. R. (2009). Asymptotics for statistical treatment rules. *Econometrica*, **77**(5), 1683–1701.

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, **48**(1), 65–91.

Kelly, B., Xiu, D., and Gu, S. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies*, **33**(5), 2223–2273.

Ledoit, O. and Wolf, M. (2025). Markowitz portfolios under transaction costs. *The quarterly review of economics and finance*, **100**.

Ledoit, O., Wolf, M., and Zhao, Z. (2019). Efficient sorting: A more powerful test for cross-sectional anomalies. *Journal of Financial Econometrics*, **17**(4), 645–686.

Leippold, M., Wang, Q., and Zhou, W. (2022). Machine learning in the chinese stock market. *Journal of Financial Economics*, **145**(2), 64–82.

Lewellen, J. (2015). The cross-section of expected stock returns. *Critical Finance Review*, **4**, 1–44.

Makridakis, S., Spiliotis, E., Hollyman, R., Petropoulos, F., Swanson, N., and Gaba, A. (2024). The m6 forecasting competition: Bridging the gap between forecasting and investment decisions. *International Journal of Forecasting*.

Olmo, J. and McGee, R. (2022). Optimal characteristic portfolios. *Quantitative Finance*, **22**(10), 1853–1870.

Rapach, D. E., Coulombe, P. G., Montes Schutte, E. C., and Schwenk-Nebbe, S. (2024). The anatomy of portfolio performance: A shapley-based approach. *Working Paper*.

Zaffaroni, P. and Zhou, G. (2024). Asset pricing: Cross-section predictability. In *Oxford Research Encyclopedia of Economics and Finance*. Oxford University Press.

Table 1: Performance of Long - Short portfolios

Panel A. Equally-weighted Portfolios

	LIN-R		LIN-C		ML-R		ML-C	
	BM	Opt.	BM	Opt.	BM	Opt.	BM	Opt.
Avg. Exc. Returns	2.66	2.36	1.58	3.55	3.61	5.49	2.78	6.12
Vol.	4.29	3.73	3.00	5.18	4.56	6.50	3.06	6.98
Downside Vol.	1.99	1.74	2.42	3.76	2.47	3.12	1.73	3.83
Skew.	1.23	1.26	-0.37	-0.04	1.18	1.28	0.29	2.87
Ann. Sharpe Ratio	1.95	1.97	1.54	2.21	2.56	2.80	2.88	2.92
Ann. Sortino Ratio	4.22	4.21	1.92	3.06	4.74	5.82	5.09	5.31
Min.	-10.76	-9.71	-13.13	-19.02	-11.73	-17.97	-7.10	-22.51
25%	0.31	0.09	0.13	0.75	1.20	1.72	1.01	2.55
Median	2.26	2.12	1.59	3.40	3.34	4.89	2.79	5.64
75%	4.27	3.82	3.01	6.33	5.28	8.12	4.35	8.81
Max.	24.28	22.76	12.14	25.59	31.60	42.92	17.07	76.61
% of Selected Stocks	20.02	32.31	38.34	9.60	20.05	8.68	34.67	12.52

Panel B. Value-weighted Portfolios

	LIN-R		LIN-C		ML-R		ML-C	
	BM	Opt.	BM	Opt.	BM	Opt.	BM	Opt.
Avg. Exc. Returns	1.59	1.32	0.73	1.93	1.68	2.40	1.24	4.00
Vol.	4.96	4.88	5.53	6.55	6.04	6.67	6.38	10.48
Downside Vol.	3.62	3.78	4.35	4.73	5.21	5.31	6.01	7.00
Skew.	0.09	-0.06	-0.12	0.28	-0.74	-0.52	-1.40	2.82
Ann. Sharpe Ratio	0.94	0.77	0.31	0.90	0.83	1.12	0.54	1.24
Ann. Sortino Ratio	1.29	0.99	0.39	1.24	0.96	1.41	0.58	1.86
Min.	-20.56	-22.20	-28.17	-33.97	-36.02	-33.94	-52.38	-48.46
25%	-0.68	-0.78	-1.49	-1.30	-0.93	-0.90	-1.24	-0.44
Median	1.62	1.42	0.91	1.78	1.91	2.60	1.52	3.59
75%	3.84	3.56	3.17	4.90	4.75	5.82	4.42	7.17
Max.	25.39	25.49	27.23	43.79	25.85	25.88	27.07	114.25
% of Selected Stocks	20.02	32.31	38.34	9.60	20.05	8.68	34.67	12.52

This table reports the average monthly returns, the volatility, the downside volatility constructed as the volatility of negative returns, the skewness, the annualized Sharpe ratio, the annualized Sortino ratio, the minimum, maximum and the quartiles of the return distribution of each of the portfolios considered, as well as the average percentage of stocks selected. Panel A reports results for equally-weighted portfolios, and Panel B for value-weighted portfolios.

Table 2: Performance of Long - Short portfolios net of transaction costs

Panel A. Equally-weighted Portfolios

	LIN-R		LIN-C		ML-R		ML-C	
	BM	Opt.	BM	Opt.	BM	Opt.	BM	Opt.
Avg. Exc. Returns	0.32	0.21	-0.38	-0.77	0.49	1.09	0.35	1.35
Vol.	4.47	3.89	3.45	4.99	4.62	6.63	3.25	6.93
Downside Vol.	2.62	2.23	3.19	3.93	2.94	3.76	2.51	3.97
Skew.	0.69	0.85	-1.52	-0.62	0.62	0.85	-0.55	4.16
Ann. Sharpe Ratio	0.15	0.07	-0.50	-0.62	0.28	0.50	0.24	0.61
Ann. Sortino Ratio	0.26	0.13	-0.55	-0.79	0.44	0.89	0.31	1.07
Min.	-15.33	-13.70	-16.46	-20.54	-17.02	-19.86	-12.26	-24.11
25%	-2.17	-1.66	-1.86	-3.35	-1.88	-2.56	-1.34	-1.63
Median	-0.19	-0.25	-0.00	-0.31	0.38	0.54	0.41	1.06
75%	2.19	1.88	1.47	2.11	2.42	4.13	2.30	3.96
Max.	19.68	18.52	7.52	14.05	23.08	31.24	10.94	73.83
% of Selected Stocks	20.03	32.55	37.93	8.84	20.05	7.91	34.61	11.47

Panel B. Value-weighted Portfolios

	LIN-R		LIN-C		ML-R		ML-C	
	BM	Opt.	BM	Opt.	BM	Opt.	BM	Opt.
Avg. Exc. Returns	0.50	0.50	0.52	0.80	0.61	0.75	0.86	2.97
Vol.	4.70	4.33	4.45	9.17	5.75	7.99	5.33	11.73
Downside Vol.	3.67	3.28	3.48	6.65	4.60	5.97	3.96	6.16
Skew.	-0.36	-0.04	-0.38	0.13	-0.34	-0.17	-0.00	4.13
Ann. Sharpe Ratio	0.28	0.30	0.31	0.26	0.29	0.27	0.48	0.84
Ann. Sortino Ratio	0.35	0.39	0.39	0.35	0.36	0.36	0.64	1.60
Min.	-21.67	-18.02	-19.54	-48.11	-30.11	-30.80	-20.56	-29.57
25%	-1.65	-1.48	-1.63	-3.17	-2.11	-3.07	-1.60	-1.90
Median	0.77	0.59	0.54	0.50	0.55	0.54	0.63	2.44
75%	2.70	2.09	2.47	4.59	3.50	4.91	3.63	6.66
Max.	17.44	18.53	15.30	50.37	27.22	36.20	23.13	116.47
% of Selected Stocks	20.03	32.55	37.93	8.84	20.05	7.91	34.61	11.47

This table reports the net of transaction costs average monthly returns, the volatility, the downside volatility constructed as the volatility of negative returns, the skewness, the annualized Sharpe ratio, the annualized Sortino ratio, the minimum, maximum and the quartiles of the return distribution of each of the portfolios considered, as well as the average percentage of stocks selected. Panel A reports results for equally-weighted portfolios, and Panel B for value-weighted portfolios. Transaction costs are estimated as one-half of the average daily quoted spread within a month.

Table 3: Risk-adjusted returns of value-weighted portfolios

	LIN-R		LIN-C		ML-R		ML-C	
	BM	Opt.	BM	Opt.	BM	Opt.	BM	Opt.
Alpha	0.96***	0.74***	0.23	1.40***	1.04***	1.70***	0.43	2.96***
	(4.72)	(3.82)	(0.85)	(4.13)	(4.27)	(5.76)	(1.41)	(4.87)
MKT	0.02	0.01	0.05	0.24***	0.09	0.18**	0.27***	0.63***
	(0.31)	(0.13)	(0.76)	(2.88)	(1.37)	(2.24)	(2.63)	(2.69)
SMB	−0.49***	−0.55***	−0.07	0.04	−0.71***	−0.62***	−0.08	0.11
	(−5.43)	(−5.80)	(−0.52)	(0.31)	(−5.14)	(−3.77)	(−0.67)	(0.54)
HML	0.02	0.03	0.18	0.04	0.06	0.08	0.08	0.10
	(0.17)	(0.36)	(1.26)	(0.27)	(0.51)	(0.57)	(0.49)	(0.30)
RMW	0.27**	0.32***	0.61***	0.24	0.43**	0.27	0.72***	0.67*
	(2.32)	(3.05)	(5.63)	(1.33)	(2.38)	(1.18)	(4.53)	(1.71)
CMA	0.48***	0.50***	0.15	0.12	0.51***	0.48**	0.16	0.15
	(3.25)	(4.06)	(0.66)	(0.66)	(2.74)	(2.10)	(0.63)	(0.45)
MOM	0.58***	0.56***	0.40***	0.38**	0.74***	0.79***	0.71***	0.50**
	(8.97)	(7.84)	(4.40)	(2.35)	(8.71)	(7.34)	(4.31)	(2.18)
STR	0.60***	0.43***	0.11	0.25	0.10	0.21*	0.04	0.18
	(7.42)	(5.61)	(1.00)	(1.35)	(0.91)	(1.77)	(0.45)	(0.48)
LTR	0.22*	0.15	−0.24	−0.12	−0.05	−0.10	−0.32*	−0.12
	(1.83)	(1.40)	(−1.60)	(−0.76)	(−0.32)	(−0.68)	(−1.94)	(−0.42)
T	420	420	420	420	420	420	420	420
R^2	0.53	0.53	0.22	0.14	0.58	0.45	0.35	0.19

This table reports the risk-adjusted returns of the value-weighted portfolios considered. We consider a model that includes the Market, Size, Book-to-Market, Profitability, Investment, Long and short term reversals, as well as the Momentum factor. Newey west t-statistics are reported in parenthesis. Alphas are expressed in percentage points.

Table 4: Net of transaction costs risk-adjusted returns

	LIN-R		LIN-C		ML-R		ML-C	
	BM	Opt.	BM	Opt.	BM	Opt.	BM	Opt.
Alpha	−0.02	0.08	0.01	0.69	0.29	0.20	0.09	1.79***
	(−0.09)	(0.55)	(0.06)	(1.11)	(1.26)	(0.46)	(0.39)	(2.87)
MKT	0.16***	0.12**	−0.15*	−0.38*	0.02	0.12	0.12	0.35
	(2.73)	(2.52)	(−1.95)	(−1.86)	(0.28)	(1.22)	(1.25)	(1.29)
SMB	0.21	0.23***	0.42***	0.13	0.02	0.09	0.26***	0.40
	(1.62)	(2.78)	(3.78)	(0.47)	(0.19)	(0.57)	(2.85)	(1.53)
HML	−0.15*	−0.07	0.23	0.13	0.01	−0.23	0.13	−0.51
	(−1.66)	(−0.81)	(1.65)	(0.41)	(0.08)	(−1.14)	(0.94)	(−1.02)
RMW	−0.02	−0.10	0.27*	0.13	−0.10	0.16	0.76***	1.23**
	(−0.20)	(−1.25)	(1.91)	(0.29)	(−0.65)	(0.73)	(4.12)	(2.12)
CMA	0.53***	0.33**	0.55***	0.05	0.51**	0.57	0.54***	0.66*
	(3.02)	(2.31)	(3.19)	(0.14)	(2.27)	(1.62)	(3.09)	(1.83)
MOM	0.44***	0.44***	0.35***	0.29	0.80***	0.90***	0.58***	0.29*
	(6.21)	(8.15)	(4.59)	(1.31)	(12.35)	(8.77)	(8.67)	(1.66)
STR	0.64***	0.58***	0.39***	0.54**	0.25***	0.38***	0.07	0.54
	(8.89)	(11.72)	(5.11)	(2.03)	(3.01)	(3.06)	(0.77)	(1.45)
LTR	0.20	0.22**	−0.29*	−0.03	0.02	0.21	−0.19	0.20
	(1.57)	(2.22)	(−1.71)	(−0.08)	(0.12)	(0.84)	(−0.93)	(0.47)
T	264	264	264	264	264	264	264	264
R^2	0.54	0.61	0.41	0.07	0.52	0.35	0.53	0.13

This table reports the net of transaction costs risk-adjusted returns of the value-weighted portfolios considered. We consider a model that includes the Market, Size, Book-to-Market, Profitability, Investment, Long and short term reversals, as well as the Momentum factor. Newey west t-statistics are reported in parenthesis. Alphas are expressed in percentage points. Transaction costs are estimated as one-half of the average daily quoted spread within a month.

Table 5: Stock Characteristics: Machine Learning portfolios

	ML-R						ML-C					
	BM			Optimal			BM			Optimal		
	Long	Short	Diff.	Long	Short	Diff.	Long	Short	Diff.	Long	Short	Diff.
mom1m	30.82	70.09	-39.27	26.50	77.79	-51.29	30.68	53.37	-22.69	19.68	58.93	-39.25
mom12m	49.79	31.06	18.72	43.83	24.87	18.95	54.75	34.32	20.43	42.21	26.70	15.51
chmom	41.84	54.30	-12.46	40.51	52.61	-12.10	43.09	52.27	-9.18	38.99	54.16	-15.16
bm	55.27	42.73	12.54	56.68	43.41	13.27	54.37	40.53	13.84	50.48	36.99	13.49
sp	55.66	44.02	11.65	57.63	44.46	13.17	56.36	42.21	14.15	51.73	39.51	12.23
maxret	57.03	76.61	-19.58	61.34	80.62	-19.28	63.42	78.27	-14.84	75.59	86.24	-10.65
cashpr	45.78	55.60	-9.82	44.08	54.78	-10.70	47.31	55.98	-8.67	48.46	57.90	-9.43
indmom	54.96	40.94	14.02	52.59	39.21	13.38	55.30	44.80	10.50	53.28	43.92	9.36
agr	60.84	46.91	13.93	62.19	47.06	15.13	58.74	52.42	6.32	63.54	54.74	8.80
mom6m	47.35	34.48	12.87	42.96	25.76	17.20	51.72	37.80	13.92	39.90	31.57	8.33
lgr	44.29	53.01	-8.72	43.80	53.43	-9.63	44.53	51.65	-7.11	44.08	51.53	-7.45
mvel1	33.06	40.37	-7.32	25.65	38.95	-13.30	29.40	32.64	-3.24	19.41	26.49	-7.08
mom36m	39.75	47.13	-7.38	37.28	46.14	-8.87	38.04	41.57	-3.52	31.71	38.70	-6.99
invest	44.06	53.18	-9.12	42.92	53.29	-10.38	44.76	50.33	-5.57	42.56	49.20	-6.64
rd_mve	56.31	50.39	5.92	56.32	49.36	6.95	56.29	53.96	2.33	60.60	54.08	6.52
hire	43.81	52.78	-8.97	42.41	52.51	-10.10	45.31	50.19	-4.88	42.37	48.53	-6.16
turn	47.70	54.74	-7.04	45.05	53.49	-8.44	51.83	58.78	-6.95	54.56	60.62	-6.06
bm_ia	53.81	49.94	3.87	54.44	49.19	5.26	55.21	50.43	4.78	54.87	48.82	6.05
sgr	44.28	52.74	-8.46	42.93	52.43	-9.50	46.09	50.51	-4.41	43.78	49.65	-5.86
ill	64.05	59.31	4.74	71.05	61.93	9.12	66.00	63.44	2.56	74.12	68.27	5.85
lev	50.42	43.48	6.93	52.02	45.16	6.87	47.49	41.76	5.73	46.32	40.55	5.78
zerotrade	54.53	44.78	9.75	58.38	45.87	12.50	50.18	43.78	6.40	48.12	42.47	5.65
chempia	44.23	51.39	-7.16	42.93	51.40	-8.47	44.58	48.77	-4.19	41.59	47.21	-5.63
dolvol	37.20	42.65	-5.45	30.47	40.17	-9.70	36.36	40.01	-3.65	30.15	35.64	-5.49
cfp	45.62	41.15	4.47	44.32	41.84	2.49	46.07	35.08	10.99	36.02	30.56	5.46
chcsho	48.71	55.88	-7.17	48.16	54.99	-6.83	50.12	57.79	-7.67	53.25	58.50	-5.26
cfp_ia	48.24	43.75	4.49	46.55	43.29	3.25	49.45	39.94	9.51	40.52	35.38	5.14
age	49.29	41.30	7.99	48.53	41.21	7.31	43.83	37.03	6.80	39.84	34.70	5.14
grltnoa	44.48	51.47	-6.99	43.71	51.80	-8.09	45.03	48.66	-3.63	43.07	48.17	-5.10
egr	41.07	50.22	-9.15	39.21	49.53	-10.31	43.26	44.83	-1.57	37.32	41.83	-4.51
roavol	58.82	57.42	1.41	59.63	55.91	3.71	62.98	62.85	0.13	67.55	63.16	4.39
orgcap	57.36	51.79	5.57	58.57	50.32	8.25	58.67	55.31	3.35	60.92	56.58	4.35
cashdebt	41.68	42.89	-1.21	39.24	42.10	-2.85	43.19	33.54	9.65	31.53	27.23	4.30
std_dolvol	58.74	57.99	0.75	63.71	60.19	3.53	60.00	57.76	2.24	64.09	60.03	4.07
grcapx	45.49	51.47	-5.98	44.46	51.46	-6.99	46.18	49.05	-2.88	44.14	48.16	-4.02
chinv	45.50	50.91	-5.40	44.70	50.90	-6.21	45.38	48.72	-3.34	43.73	47.62	-3.89
rd	58.50	48.98	9.52	58.06	48.62	9.44	53.32	51.66	1.67	56.40	52.62	3.78
currat	51.74	52.83	-1.09	50.71	51.39	-0.68	55.36	52.94	2.42	53.72	50.16	3.57
retvol	62.76	74.75	-11.99	68.32	78.44	-10.11	70.00	80.63	-10.63	84.78	88.11	-3.33
tang	52.82	53.85	-1.04	52.75	52.43	0.32	56.70	56.62	0.08	58.63	55.31	3.32

This table reports the average percentile of each characteristic in the stocks bought and the stocks sold by the portfolio in the column, as well as their differences. Characteristics are sorted according to their relevance for the optimal ML-C portfolio, defined as the spread between the average percentile of the stocks bought and stocks sold. We truncate the table at 40 characteristics for readability.

Table 6: Stock Characteristics: Linear portfolios

	LIN-R						LIN-C					
	BM			Optimal			BM			Optimal		
	Long	Short	Diff.	Long	Short	Diff.	Long	Short	Diff.	Long	Short	Diff.
mom1m	25.82	73.45	-47.63	27.53	70.53	-43.00	29.17	55.71	-26.54	16.42	59.97	-43.55
mom12m	59.93	36.60	23.33	58.02	38.40	19.62	53.58	32.92	20.66	46.16	20.54	25.63
chmom	36.17	61.29	-25.11	38.04	59.86	-21.82	42.11	53.41	-11.30	36.97	55.77	-18.80
sp	59.06	38.04	21.02	57.28	39.66	17.61	57.28	42.60	14.68	53.17	36.12	17.05
indmom	62.63	36.37	26.25	60.67	38.81	21.86	55.88	43.69	12.19	57.86	41.69	16.17
mom6m	53.51	42.39	11.12	52.72	43.51	9.20	50.72	36.59	14.13	43.36	27.31	16.04
bm	58.69	38.58	20.11	56.69	39.55	17.14	54.74	41.11	13.63	48.69	33.60	15.09
agr	62.03	38.48	23.54	60.73	40.79	19.94	59.02	51.85	7.16	69.24	54.93	14.31
cashpr	42.83	62.14	-19.31	44.50	60.74	-16.23	46.94	56.19	-9.25	47.96	60.40	-12.44
lgr	42.73	57.65	-14.92	43.64	56.20	-12.55	44.25	51.98	-7.73	41.32	52.74	-11.42
invest	42.03	59.72	-17.69	43.06	57.81	-14.75	44.94	50.93	-5.99	38.99	49.75	-10.76
mom36m	40.09	56.65	-16.56	40.63	54.79	-14.16	38.50	41.64	-3.15	26.07	36.46	-10.38
maxret	48.17	61.61	-13.44	50.17	62.49	-12.32	62.75	78.30	-15.55	77.21	87.54	-10.34
rd_mve	58.31	45.31	13.00	57.46	46.49	10.97	56.59	53.07	3.52	64.21	53.92	10.29
chcsho	43.73	60.39	-16.66	45.21	59.23	-14.02	48.36	58.35	-10.00	51.86	61.91	-10.05
lev	53.14	40.85	12.29	51.55	41.10	10.45	47.91	41.60	6.31	47.89	38.68	9.22
cfp	53.84	40.65	13.19	51.99	40.72	11.26	46.92	35.01	11.91	33.77	25.47	8.31
hire	43.35	57.66	-14.31	44.37	56.36	-12.00	45.62	50.42	-4.80	40.04	48.31	-8.27
grltnoa	43.70	56.88	-13.18	44.30	55.28	-10.98	45.23	49.07	-3.84	40.79	48.45	-7.65
egr	41.76	57.11	-15.35	42.62	55.54	-12.91	43.00	45.12	-2.12	32.45	39.91	-7.46
sgr	43.41	57.00	-13.58	44.47	55.98	-11.51	46.28	50.69	-4.42	42.44	49.86	-7.42
chinv	42.86	55.65	-12.80	43.82	54.57	-10.75	45.17	49.48	-4.30	40.77	48.05	-7.28
grcapx	43.88	56.70	-12.82	44.67	55.41	-10.74	46.19	49.75	-3.56	41.24	48.40	-7.15
cfp_ia	54.38	42.93	11.45	53.04	43.42	9.63	49.94	40.00	9.94	38.36	31.35	7.01
rd	56.19	46.07	10.12	55.31	46.75	8.56	53.38	51.24	2.14	59.20	52.57	6.64
cashdebt	48.27	46.42	1.86	47.66	45.95	1.71	44.77	33.46	11.31	27.02	20.43	6.60
chempia	44.73	55.23	-10.50	45.21	54.07	-8.86	45.31	48.82	-3.51	40.06	46.64	-6.59
turn	47.18	58.24	-11.06	48.45	58.01	-9.55	48.59	57.97	-9.39	54.63	60.83	-6.20
ear	54.94	44.67	10.28	54.02	45.50	8.52	51.86	46.22	5.63	50.21	44.30	5.90
depr	56.18	46.62	9.56	56.60	48.89	7.71	59.36	57.37	1.99	66.22	60.58	5.64
mvel1	38.03	55.70	-17.67	37.90	52.84	-14.94	28.12	31.84	-3.72	15.80	21.30	-5.50
nincr	57.26	43.01	14.24	56.30	44.57	11.73	52.96	48.25	4.71	53.34	47.97	5.37
age	52.28	43.24	9.04	50.73	43.08	7.65	43.91	37.10	6.81	37.28	32.13	5.14
std_turn	59.39	49.78	9.61	58.63	51.12	7.51	58.01	58.19	-0.18	66.69	61.58	5.11
orgcap	56.79	46.43	10.36	56.70	48.25	8.45	58.51	55.19	3.32	62.90	57.83	5.07
roavol	54.67	54.87	-0.20	55.42	55.61	-0.19	62.69	63.08	-0.39	71.09	66.06	5.03
betasq	51.09	56.03	-4.94	52.59	56.90	-4.32	54.08	62.66	-8.58	61.02	65.96	-4.94
roaq	50.50	43.09	7.41	49.79	43.73	6.06	45.37	35.55	9.81	32.99	28.10	4.90
roic	48.13	44.91	3.22	47.36	44.66	2.71	42.05	33.38	8.67	26.72	21.87	4.85
chatoia	51.85	48.62	3.23	51.72	49.02	2.70	53.25	50.29	2.96	55.53	50.72	4.82

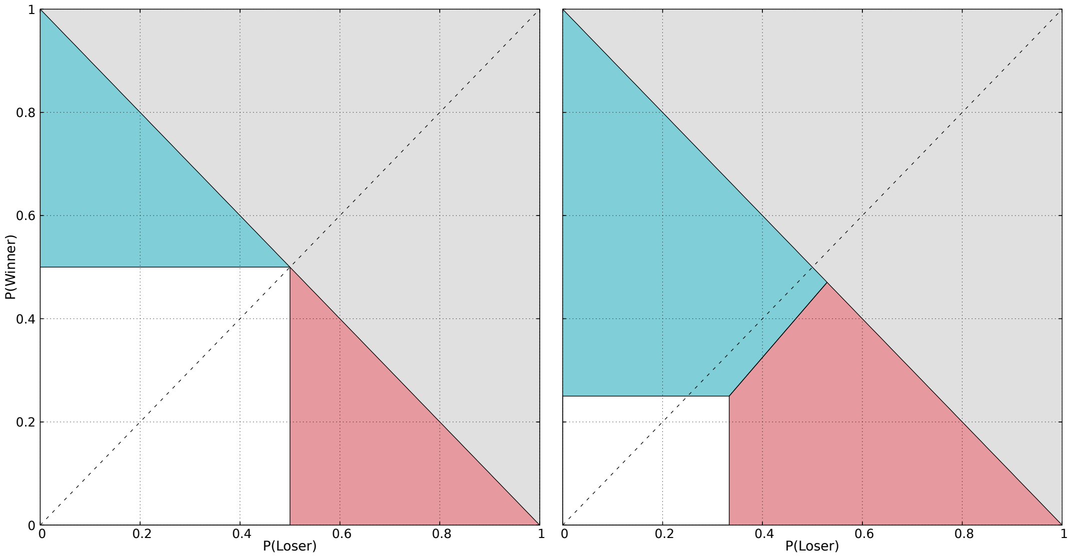
This table reports the average percentile of each characteristic in the stocks bought and the stocks sold by the portfolio in the column, as well as their differences. Characteristics are sorted according to their relevance for the optimal LIN-C portfolio, defined as the spread between the average percentile of the stocks bought and stocks sold. We truncate the table at 40 characteristics for readability.

Table 7: Classification Properties: Confusion Matrix

		Benchmark			Optimal		
Portfolio	Predicted Class	Loser	Neutral	Winner	Loser	Neutral	Winner
	Loser	14.26	75.52	10.21	14.84	74.15	11.00
LIN-R	Neutral	9.42	80.89	9.68	8.52	82.39	9.10
	Winner	10.14	77.30	12.56	11.07	75.88	13.05
	Loser	21.93	61.61	16.46	29.99	51.34	18.67
LIN-C	Neutral	4.50	89.30	6.20	8.13	82.92	8.95
	Winner	13.99	70.36	15.65	22.31	56.22	21.48
	Loser	19.25	67.69	13.07	20.71	66.91	12.38
ML-R	Neutral	8.32	82.72	8.96	9.25	81.28	9.46
	Winner	13.94	70.61	15.46	15.68	66.88	17.45
	Loser	23.08	60.42	16.50	28.92	53.38	17.70
ML-C	Neutral	4.78	88.81	6.41	7.55	83.89	8.57
	Winner	14.57	68.14	17.30	21.66	56.00	22.34

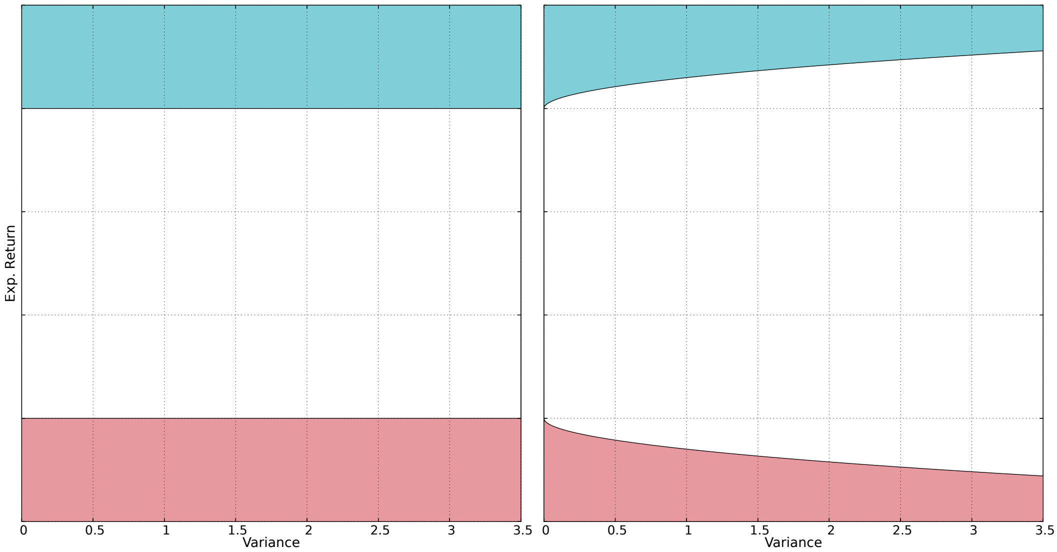
This table reports the classification performance of each of the considered portfolios, as assessed in a confusion matrix. Each row reports the percentage of the predicted class (in rows) that correspond with the realized class (in columns), for both benchmark and optimal portfolios.

Figure 1: Optimal Selection Regions



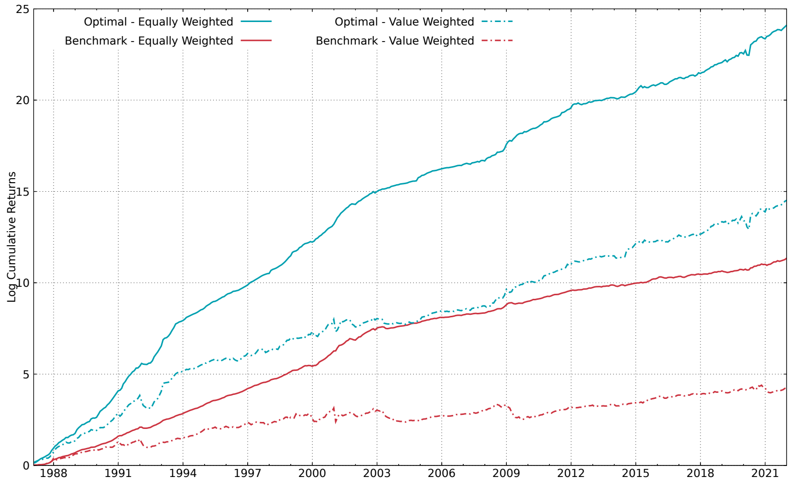
This figure reports the selection regions obtained by setting $\lambda_W = \lambda_L = 1$ (left panel) and $\lambda_W = \frac{1}{3}, \lambda_L = \frac{1}{2}$ (right panel). Blue colors demark the buy region, red colors are the sell regions, and white colors are the no-trade zone. The greyed out area is not achievable since probabilities must add up to 1.

Figure 2: Selection Regions in Location Scale Models



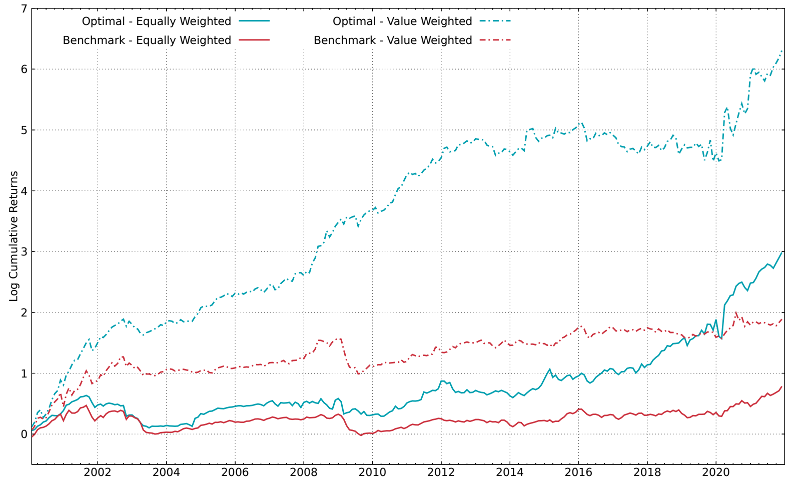
This figure reports the selection regions obtained by setting $\lambda_W = \lambda_L = 1$ (left panel) and $\lambda_W = \lambda_L = 1.1$ (right panel). Blue colors demark the buy region, red colors are the sell regions, and white colors are the no-trade zone.

Figure 3: Cumulative Returns: ML-C



This figure reports the log cumulative returns of portfolios built using XGBC. We report both portfolios created using the optimal selection regions (blue lines) and portfolios obtained by classifying stocks according to the most likely class (Benchmark). We also report equally-weighted returns (solid lines), and value-weighted returns (dashed lines).

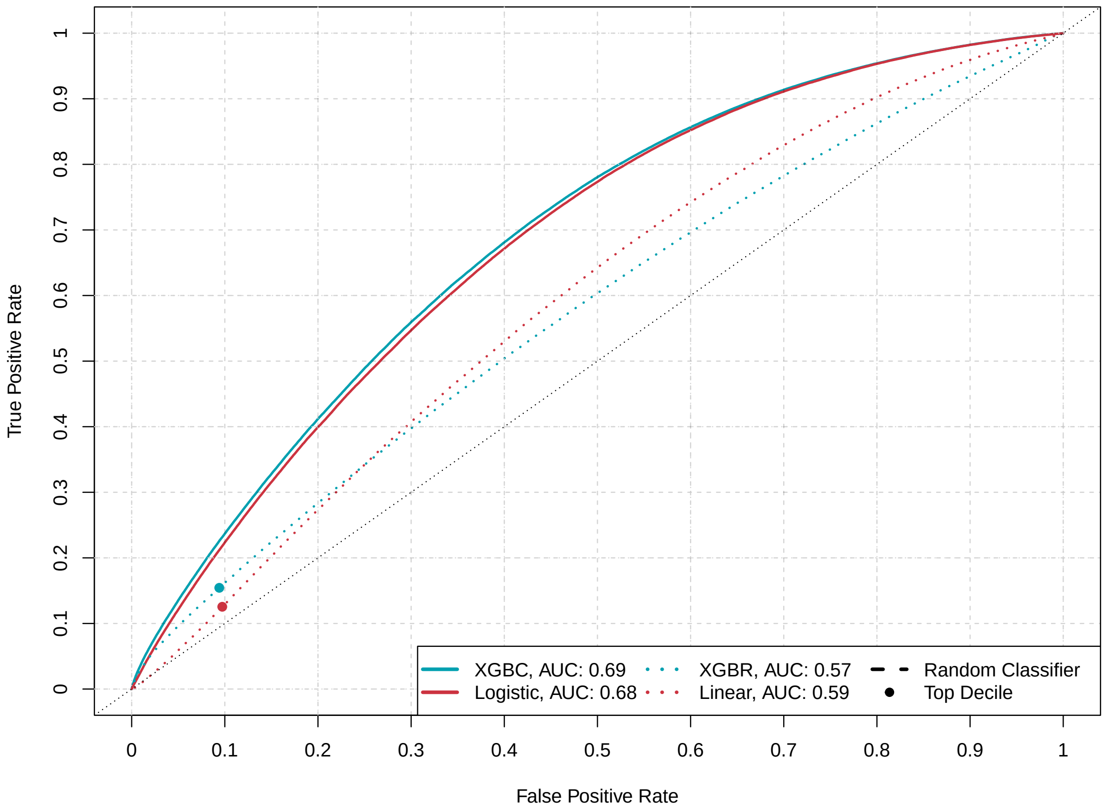
Figure 4: Net of Transaction Costs cumulative returns: ML-C



This figure reports the log cumulative returns of portfolios built using XGBC and accounting for transaction costs. We report both portfolios created using the optimal selection regions (blue lines) and portfolios obtained by classifying stocks according to the most likely class (Benchmark). We also report equally-weighted returns (solid lines), and value-weighted returns (dashed lines).

Figure 5: Receiver Operating Characteristic (ROC) Curves

Panel A. ROC for Winners



Panel B. ROC for Losers

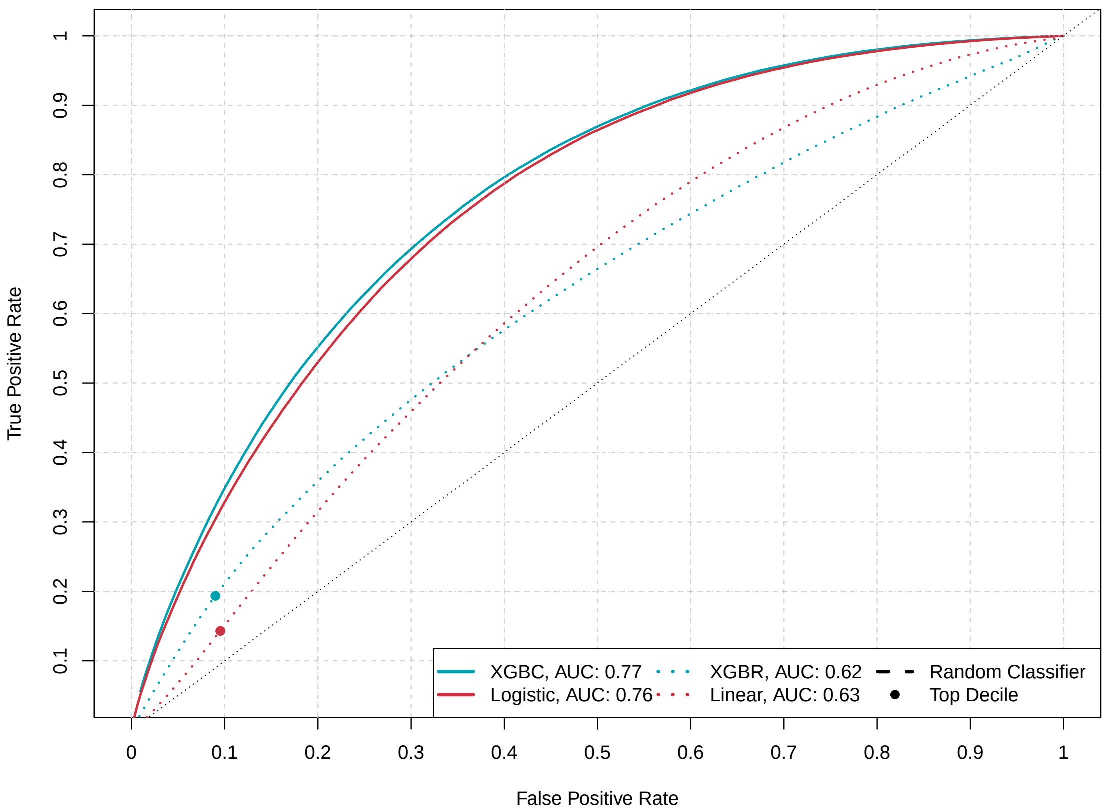
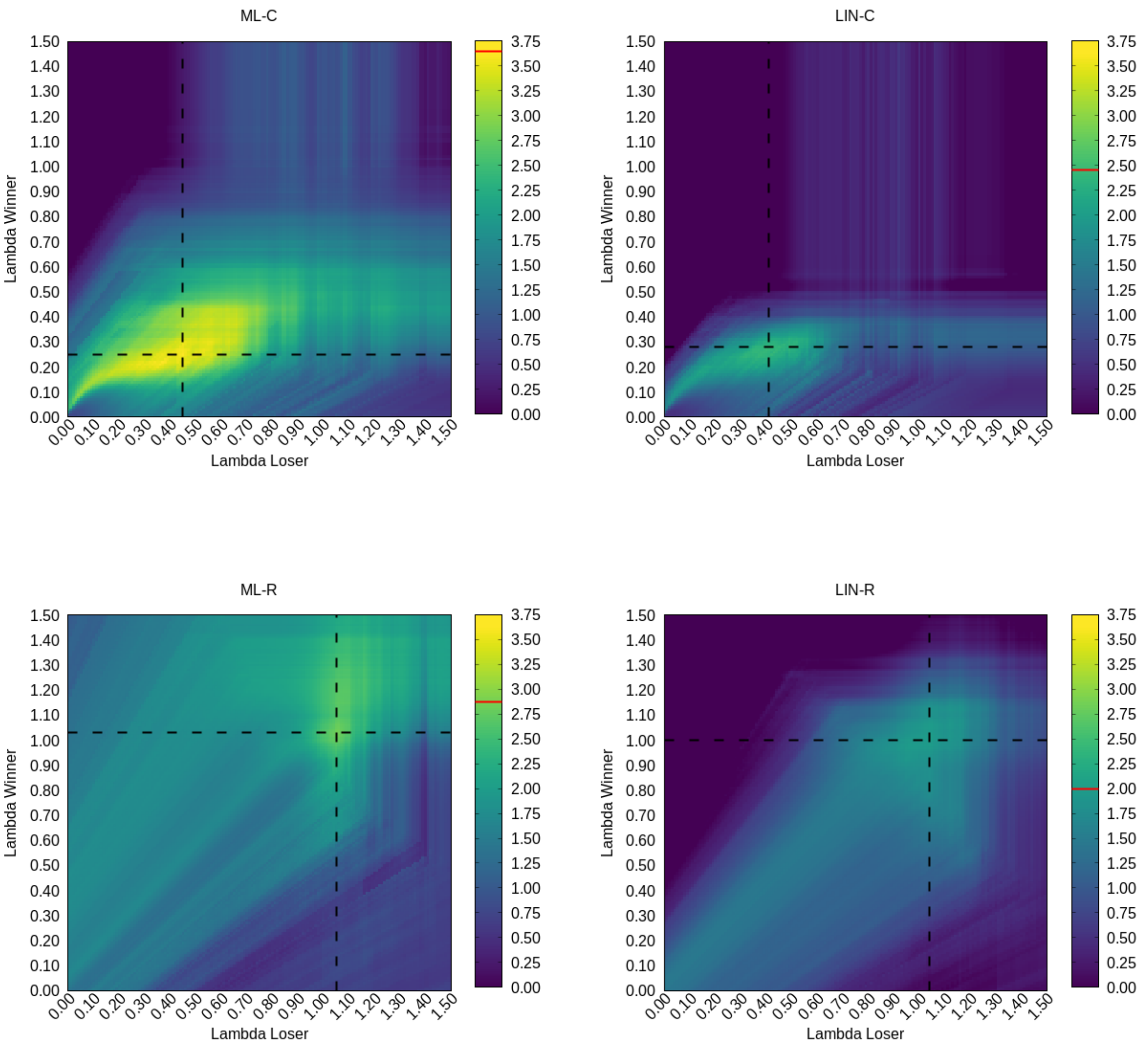


Figure 6: Lambda and Portfolio Properties



This figure reports the relationship between λ_W, λ_L, and portfolio Sharpe ratios. The top panels reflect the performance of classification models. The bottom panels reflect the performance of regression models. The red lines in the legend boxes represent the maximum Sharpe ratio achieved by the corresponding portfolio. The dashed lines represent the (unfeasible) optimal parameters.

Online Appendix

How to Bet on Winners and Losers

A Proofs

Proof of Proposition 1. We need to show that $\delta^* \in \arg\min_\delta \mathbb{E}(\mathcal{L}_{\lambda_W,\lambda_L}(\delta))$ is given by

$$\delta^*_{it} = \begin{cases} 1 & \text{if } p^W_{it} \geq \max\left(\frac{\lambda_W}{1+\lambda_W}, \frac{\lambda_W-\lambda_L}{1+\lambda_W} + \frac{1+\lambda_L}{1+\lambda_W} p^L_{it}\right) \\ -1 & \text{if } p^L_{it} \geq \max\left(\frac{\lambda_L}{1+\lambda_L}, \frac{\lambda_L-\lambda_W}{1+\lambda_L} + \frac{1+\lambda_W}{1+\lambda_L} p^W_{it}\right) \\ 0 & \text{otherwise} \end{cases} , \tag{A.1}$$

where $p^L_{it} = \mathbb{P}(Y_{it+1} = 1|\boldsymbol{X}_{it})$ and $p^W_{it} = \mathbb{P}(Y_{it+1} = K|\boldsymbol{X}_{it})$. The contribution of stock i to the overall loss is given by

$$\begin{array}{cc|ccccc} & & \text{Loser} & & \text{Neutral} & & \text{Winner} \\ & & Y_{it+1}=1 & Y_{it+1}=2 & \dots & Y_{it+1}=K-1 & Y_{it+1}=K \\ \hline \text{Sell} & \delta_{it}=-1 & 0 & \lambda_L & \dots & \lambda_L & 1+\lambda_L \\ \text{No-trade} & \delta_{it}=0 & 1 & 0 & \dots & 0 & 1 \\ \text{Buy} & \delta_{it}=1 & 1+\lambda_W & \lambda_W & \dots & \lambda_W & 0 \end{array} .$$

Note that the value of the loss is known for each region $\{\mathcal{R}_k\}_{k=1}^K$, so the only randomness stems from the probability that the return of stock i falls in region $\mathcal{R}_k$. We define

$$\mathbb{P}(R_{it} \in \mathcal{R}_k|\boldsymbol{X}_{it}) = p_k(\boldsymbol{X}_{it})$$

and we note that

$$\mathbb{E}\mathcal{L}_{\lambda_W,\lambda_L}(\delta) = \sum_{i=1}^n \mathbb{E}\mathcal{L}^{(i)}_{\lambda_W,\lambda_L}(\delta_{it}) ,$$

so it suffices to look at a single stock i. For a single stock, the optimal decision leads to the smallest expected loss. In fact, the investor only needs to compare $\mathbb{E}\mathcal{L}^{(i)}_{\lambda_W,\lambda_L}(1)$ with $\mathbb{E}\mathcal{L}^{(i)}_{\lambda_W,\lambda_L}(0)$ and $\mathbb{E}\mathcal{L}^{(i)}_{\lambda_W,\lambda_L}(-1)$, where

$$\begin{aligned} \mathbb{E}\mathcal{L}^{(i)}_{\lambda_W,\lambda_L}(1) &= (1+\lambda_W)p_1(\boldsymbol{X}_{it}) + \lambda_W(1 - p_K(\boldsymbol{X}_{it}) - p_1(\boldsymbol{X}_{it})) \\ \mathbb{E}\mathcal{L}^{(i)}_{\lambda_W,\lambda_L}(0) &= p_K(\boldsymbol{X}_{it}) + p_1(\boldsymbol{X}_{it}) \\ \mathbb{E}\mathcal{L}^{(i)}_{\lambda_W,\lambda_L}(-1) &= (1+\lambda_L)p_K(\boldsymbol{X}_{it}) + \lambda_L(1 - p_K(\boldsymbol{X}_{it}) - p_1(\boldsymbol{X}_{it})) . \end{aligned}$$

The investor chooses $\delta_{i\,t} = 1$ if $\mathbb{E}\mathcal{L}^{(i)}_{\lambda_W,\lambda_L}(1) < \min\{\mathbb{E}\mathcal{L}^{(i)}_{\lambda_W,\lambda_L}(0), \mathbb{E}\mathcal{L}^{(i)}_{\lambda_W,\lambda_L}(-1)\}$, which happens when

$$p_K(\boldsymbol{X}_{i\,t}) \geq \max\Big(\frac{\lambda_W}{1+\lambda_W}, \frac{\lambda_W - \lambda_L}{1+\lambda_W} + \frac{1+\lambda_L}{1+\lambda_W} p_1(\boldsymbol{X}_{i\,t})\Big).$$

Conversely, the investor chooses $\delta_{i\,t} = -1$ if $\mathbb{E}\mathcal{L}^{(i)}_{\lambda_W,\lambda_L}(-1) < \min\{\mathbb{E}\mathcal{L}^{(i)}_{\lambda_W,\lambda_L}(1), \mathbb{E}\mathcal{L}^{(i)}_{\lambda_W,\lambda_L}(0)\}$, which happens when

$$p_1(\boldsymbol{X}_{i\,t}) \geq \max\Big(\frac{\lambda_L}{1+\lambda_L}, \frac{\lambda_L - \lambda_W}{1+\lambda_L} + \frac{1+\lambda_W}{1+\lambda_L} p_K(\boldsymbol{X}_{i\,t})\Big) .$$

If neither condition is satisfied then $\mathbb{E}\mathcal{L}^{(i)}_{\lambda_W,\lambda_L}(0) < \min\{\mathbb{E}\mathcal{L}^{(i)}_{\lambda_W,\lambda_L}(1), \mathbb{E}\mathcal{L}^{(i)}_{\lambda_W,\lambda_L}(-1)\}$, and the investor does not trade on stock i at time t.

□

Proof of Proposition 2. This follows from noting that, in the additive regression model given in (4), we have

$$\begin{aligned}
p_K(\boldsymbol{X}_{i\,t}) &= \mathbb{P}(R_{i\,t+1} > c_{K-1}) \\
&= \mathbb{P}\Big(\frac{R_{i\,t+1} - \mu_{i\,t}}{\sqrt{\sigma^2_{i\,t}}} > \frac{c_{K-1} - \mu_{i\,t}}{\sqrt{\sigma^2_{i\,t}}}\Big) \\
&= 1 - \mathbb{P}\Big(\frac{R_{i\,t+1} - \mu_{i\,t}}{\sqrt{\sigma^2_{i\,t}}} \leq \frac{c_{K-1} - \mu_{i\,t}}{\sqrt{\sigma^2_{i\,t}}}\Big) \\
&= 1 - F_Z\Big(\frac{c_{K-1} - \mu_{i\,t}}{\sqrt{\sigma^2_{i\,t}}}\Big) \\
&= F_Z\Big(\frac{\mu_{i\,t} - c_{K-1}}{\sqrt{\sigma^2_{i\,t}}}\Big)
\end{aligned}$$

where the last line follows from the symmetry of F_Z. Analogously, we have

$$\begin{aligned}
p_1(\boldsymbol{X}_{i\,t}) &= \mathbb{P}(R_{i\,t+1} \leq c_1) \\
&= \mathbb{P}\Big(\frac{R_{i\,t+1} - \mu_{i\,t}}{\sqrt{\sigma^2_{i\,t}}} \leq \frac{c_1 - \mu_{i\,t}}{\sqrt{\sigma^2_{i\,t}}}\Big) \\
&= F_Z\Big(\frac{c_1 - \mu_{i\,t}}{\sqrt{\sigma^2_{i\,t}}}\Big) ,
\end{aligned}$$

and the first part of the proposition follows. Next, we need to show that if $\lambda_W \geq 1$ and $\lambda_L \geq 1$, then the optimal rule

$$\delta^*_{it} = \begin{cases} 1 & \text{if } F_Z\Big(\frac{\mu_{it}-c_{K-1}}{\sqrt{\sigma^2_{it}}}\Big) \geq \max\Big(\frac{\lambda_W}{1+\lambda_W}, \frac{\lambda_W-\lambda_L}{1+\lambda_W} + \frac{1+\lambda_L}{1+\lambda_W}F_Z\Big(\frac{c_1-\mu_{it}}{\sqrt{\sigma^2_{it}}}\Big)\Big) \\ -1 & \text{if } F_Z\Big(\frac{c_1-\mu_{it}}{\sqrt{\sigma^2_{it}}}\Big) \geq \max\Big(\frac{\lambda_L}{1+\lambda_L}, \frac{\lambda_L-\lambda_W}{1+\lambda_L} + \frac{1+\lambda_W}{1+\lambda_L}F_Z\Big(\frac{\mu_{it}-c_{K-1}}{\sqrt{\sigma^2_{it}}}\Big)\Big) \\ 0 & \text{otherwise} \end{cases}, \tag{A.2}$$

is equivalent to the rule given by:

$$\delta^s_{it} = \begin{cases} 1 & \text{if } \mu_{it} - F_Z^{-1}(\frac{\lambda_W}{1+\lambda_W})\sqrt{\sigma^2_{it}} > c_{K-1} \\ -1 & \text{if } \mu_{it} + F_Z^{-1}(\frac{\lambda_L}{1+\lambda_L})\sqrt{\sigma^2_{it}} \leq c_1 \\ 0 & \text{otherwise} \end{cases}. \tag{A.3}$$

We have the following 6 cases to analyse. Case 1: $\delta^s_{it} = 1 \Rightarrow \delta^*_{it} = 1$; case 2: $\delta^s_{it} \neq 1 \Rightarrow \delta^*_{it} \neq 1$; case 3: $\delta^s_{it} = -1 \Rightarrow \delta^*_{it} = -1$; case 4: $\delta^s_{it} \neq -1 \Rightarrow \delta^*_{it} \neq -1$; case 5: $\delta^s_{it} = 0 \Rightarrow \delta^*_{it} = 0$, and case 6: $\delta^s_{it} \neq 0 \Rightarrow \delta^*_{it} \neq 0$. We prove cases 1 and 2. Cases 3 and 4 follow analogously, and cases 5 and 6 are a consequence of cases 1 through 4.

Case 1. $\delta^s_{it} = 1 \Rightarrow \delta^*_{it} = 1$. We need to show that

$$F_Z\Big(\frac{\mu_{it} - c_{K-1}}{\sqrt{\sigma^2_{it}}}\Big) > \frac{\lambda_W}{1+\lambda_W}$$

implies

$$F_Z\Big(\frac{\mu_{it} - c_{K-1}}{\sqrt{\sigma^2_{it}}}\Big) \geq \max\Big(\frac{\lambda_W}{1+\lambda_W}, \frac{\lambda_W-\lambda_L}{1+\lambda_W} + \frac{1+\lambda_L}{1+\lambda_W}F_Z\Big(\frac{c_1-\mu_{it}}{\sqrt{\sigma^2_{it}}}\Big)\Big).$$

If $\max\Big(\frac{\lambda_W}{1+\lambda_W}, \frac{\lambda_W-\lambda_L}{1+\lambda_W} + \frac{1+\lambda_L}{1+\lambda_W}F_Z\Big(\frac{c_1-\mu_{it}}{\sqrt{\sigma^2_{it}}}\Big)\Big) = \frac{\lambda_W}{1+\lambda_W}$ then the result immediately follows. We remain to show that $\mu_{it} - F_Z^{-1}(\frac{\lambda_W}{1+\lambda_W})\sqrt{\sigma^2_{it}} > c_{K-1}$ implies

$$\frac{\lambda_W}{1+\lambda_W} \geq \frac{\lambda_W-\lambda_L}{1+\lambda_W} + \frac{1+\lambda_L}{1+\lambda_W}F_Z\Big(\frac{c_1-\mu_{it}}{\sqrt{\sigma^2_{it}}}\Big).$$

To see this, note that $\mu_{it} - F_Z^{-1}(\frac{\lambda_W}{1+\lambda_W})\sqrt{\sigma^2_{it}} > c_{K-1}$ implies $\mu_{it} > c_{K-1}$ as long as $\lambda_W \geq 1$ and F_Z has median 0. In addition, since $c_{K-1} > c_1$ by construction, $F_Z\Big(\frac{\mu_{it}-c_{K-1}}{\sqrt{\sigma^2_{it}}}\Big) \geq \frac{1}{2}$ implies

$F_Z\left(\frac{c_1-\mu_{i\,t}}{\sqrt{\sigma^2_{i\,t}}}\right) \leq \frac{1}{2}$. Next, note that

$$\frac{\lambda_W}{1+\lambda_W} \geq \frac{\lambda_W-\lambda_L}{1+\lambda_W} + \frac{1+\lambda_L}{1+\lambda_W}\left(\frac{1}{2}\right)$$

if $\frac{\lambda_L}{1+\lambda_L} \geq \frac{1}{2}$, that is, $\lambda_L \geq 1$.

Case 2. $\boldsymbol{\delta^s_{i\,t} \neq 1 \Rightarrow \delta^*_{i\,t} \neq 1}$**.** We next need to show that $\mu_{i\,t} - F_Z^{-1}(\frac{\lambda_W}{1+\lambda_W})\sqrt{\sigma^2_{i\,t}} < c_{K-1}$ implies $F_Z\left(\frac{\mu_{i\,t}-c_{K-1}}{\sqrt{\sigma^2_{i\,t}}}\right) < \max\left(\frac{\lambda_W}{1+\lambda_W}, \frac{\lambda_W-\lambda_L}{1+\lambda_W} + \frac{1+\lambda_L}{1+\lambda_W}F_Z\left(\frac{c_1-\mu_{i\,t}}{\sqrt{\sigma^2_{i\,t}}}\right)\right)$. As before, if $\max\left(\frac{\lambda_W}{1+\lambda_W}, \frac{\lambda_W-\lambda_L}{1+\lambda_W} + \frac{1+\lambda_L}{1+\lambda_W}F_Z\left(\frac{c_1-\mu_{i\,t}}{\sqrt{\sigma^2_{i\,t}}}\right)\right) = \frac{\lambda_W}{1+\lambda_W}$ the result immediately follows. Conversely, by assumption we have

$$F_Z^{-1}\left(\frac{\mu_{i\,t}-c_{K-1}}{\sqrt{\sigma^2_{i\,t}}}\right) \leq \frac{\lambda_W}{1+\lambda_W}$$

which implies

$$F_Z^{-1}\left(\frac{\mu_{i\,t}-c_{K-1}}{\sqrt{\sigma^2_{i\,t}}}\right) \leq \frac{\lambda_W}{1+\lambda_W} \leq \max\left(\frac{\lambda_W}{1+\lambda_W}, \frac{\lambda_W-\lambda_L}{1+\lambda_W} + \frac{1+\lambda_L}{1+\lambda_W}F_Z\left(\frac{c_1-\mu_{i\,t}}{\sqrt{\sigma^2_{i\,t}}}\right)\right).$$

$\square$

Table OA.1: Performance of Long - Short portfolios net of transaction costs

Panel A. Equally-weighted Portfolios

	LIN-R		LIN-C		ML-R		ML-C	
	BM	Opt.	BM	Opt.	BM	Opt.	BM	Opt.
Avg. Exc. Returns	2.04	1.77	1.29	2.48	2.78	4.35	2.18	4.57
Vol.	4.82	4.18	3.50	5.52	5.18	7.17	3.39	7.15
Downside Vol.	2.00	1.71	2.53	3.69	2.50	3.17	1.72	3.98
Skew.	1.27	1.40	-0.23	0.13	1.57	1.31	0.60	3.76
Ann. Sharpe Ratio	1.38	1.36	1.16	1.48	1.77	2.04	2.10	2.16
Ann. Sortino Ratio	3.31	3.33	1.60	2.21	3.68	4.62	4.13	3.87
Min.	-11.30	-10.08	-13.39	-19.53	-12.04	-18.18	-7.36	-22.81
25%	-0.99	-0.76	-0.53	-0.76	-0.03	0.05	0.20	1.32
Median	1.22	1.21	1.29	2.21	2.04	3.38	2.05	4.12
75%	3.65	3.29	2.89	5.54	4.83	7.40	3.73	7.33
Max.	24.23	22.66	11.96	25.08	34.81	45.39	16.68	75.97
Wrong Direction Rate	20.03	32.55	37.93	8.84	20.05	7.91	34.61	11.47

Panel B. Value-weighted Portfolios

	LIN-R		LIN-C		ML-R		ML-C	
	BM	Opt.	BM	Opt.	BM	Opt.	BM	Opt.
Avg. Exc. Returns	0.71	0.68	0.86	1.78	0.82	1.12	1.11	3.77
Vol.	4.73	4.38	4.54	9.44	5.76	8.00	5.33	11.76
Downside Vol.	3.51	3.20	3.37	6.81	4.48	5.81	3.81	6.13
Skew.	-0.13	0.21	-0.15	0.22	-0.17	-0.02	0.14	4.06
Ann. Sharpe Ratio	0.43	0.44	0.56	0.61	0.42	0.43	0.64	1.07
Ann. Sortino Ratio	0.58	0.60	0.75	0.84	0.54	0.59	0.89	2.06
Min.	-20.49	-16.83	-19.44	-46.87	-29.91	-30.59	-19.40	-29.26
25%	-1.52	-1.38	-1.29	-2.41	-1.93	-2.67	-1.38	-1.54
Median	0.88	0.65	0.84	1.38	0.74	0.85	0.88	3.17
75%	2.81	2.35	2.84	5.70	3.61	5.21	3.80	7.83
Max.	19.50	20.70	16.37	52.67	28.32	37.50	23.93	117.25
% of Selected Stocks	20.03	32.55	37.93	8.84	20.05	7.91	34.61	11.47

This table reports the net of transaction costs average monthly returns, the volatility, the downside volatility constructed as the volatility of negative returns, the skewness, the annualized Sharpe ratio, the annualized Sortino ratio, the minimum, maximum and the quartiles of the return distribution of each of the portfolios considered, as well as the average percentage of stocks selected. Panel A reports results for equally-weighted portfolios, and Panel B for value-weighted portfolios. Transaction costs are estimated as in Ledoit and Wolf (2025).

Table OA.2: Risk-adjusted Returns: Equally-weighted

	LIN-R		LIN-C		ML-R		ML-C	
	BM	Opt.	BM	Opt.	BM	Opt.	BM	Opt.
Alpha	2.28***	1.99***	1.27***	3.26***	3.33***	5.26***	2.43***	5.63***
	(9.04)	(9.59)	(10.16)	(10.11)	(9.60)	(10.47)	(12.30)	(12.31)
MKT	−0.10**	−0.08**	−0.07**	−0.09	−0.10	−0.12	−0.05*	0.03
	(−2.16)	(−1.99)	(−2.40)	(−1.50)	(−1.45)	(−1.32)	(−1.81)	(0.23)
SMB	0.08	0.07	−0.00	0.09	−0.03	−0.03	0.03	0.20
	(1.24)	(1.30)	(−0.03)	(0.78)	(−0.22)	(−0.20)	(0.66)	(1.32)
HML	−0.07	−0.04	0.15**	−0.06	−0.21*	−0.49***	0.04	−0.33
	(−0.65)	(−0.43)	(2.57)	(−0.51)	(−1.67)	(−2.74)	(0.56)	(−1.45)
RMW	0.00	0.04	0.46***	0.42***	0.02	−0.06	0.54***	0.58**
	(0.03)	(0.40)	(7.34)	(3.50)	(0.15)	(−0.32)	(7.81)	(2.17)
CMA	0.34*	0.33**	0.11	0.15	0.32	0.27	0.13	0.21
	(1.96)	(2.15)	(1.05)	(0.66)	(1.36)	(0.91)	(1.07)	(0.81)
MOM	0.31***	0.28***	0.26***	0.19*	0.24***	0.15	0.23***	0.10
	(2.99)	(3.29)	(7.01)	(1.82)	(2.80)	(1.22)	(7.34)	(0.93)
STR	0.71***	0.62***	0.17***	0.32***	0.55***	0.68***	0.21***	0.58***
	(8.25)	(8.45)	(3.26)	(2.84)	(6.54)	(6.48)	(5.32)	(2.86)
LTR	0.43***	0.35***	−0.11	0.02	0.34***	0.67***	0.02	0.19
	(4.01)	(3.70)	(−1.61)	(0.15)	(2.84)	(3.76)	(0.32)	(1.11)
T	420	420	420	420	420	420	420	420
R^2	0.60	0.61	0.54	0.37	0.52	0.52	0.65	0.48

This table reports the risk-adjusted returns of the portfolios considered. We consider a model that includes the Market, Size, Book-to-Market, Profitability, Investment, Long and short term reversals, as well as the Momentum factor. Newey west t-statistics are reported in parenthesis. Alphas are expressed in percentage points.

Table OA.3: Risk-adjusted Returns: Equally-weighted with Transaction Costs

	LIN-R		LIN-C		ML-R		ML-C	
	BM	Opt.	BM	Opt.	BM	Opt.	BM	Opt.
Alpha	−0.09	−0.18	−0.69***	−1.14***	0.29	0.89**	−0.03	0.51
	(−0.44)	(−1.06)	(−4.64)	(−3.42)	(1.08)	(2.02)	(−0.16)	(1.46)
MKT	0.02	0.03	0.02	0.11	−0.06	0.02	0.03	0.32*
	(0.28)	(0.57)	(0.55)	(1.64)	(−0.47)	(0.14)	(0.63)	(1.83)
SMB	0.14*	0.10	−0.09	−0.02	−0.01	0.02	−0.04	0.26
	(1.90)	(1.25)	(−1.24)	(−0.14)	(−0.07)	(0.09)	(−0.57)	(1.38)
HML	−0.12	−0.08	0.24***	0.09	−0.18	−0.50**	0.16**	−0.39
	(−0.97)	(−0.74)	(3.39)	(0.67)	(−1.25)	(−2.52)	(2.03)	(−1.28)
RMW	0.04	0.06	0.39***	0.38***	−0.12	−0.12	0.46***	0.82**
	(0.34)	(0.66)	(4.80)	(2.81)	(−0.78)	(−0.60)	(5.21)	(2.12)
CMA	0.43**	0.40***	0.07	0.08	0.46**	0.37	0.19	0.33
	(2.52)	(2.92)	(0.51)	(0.42)	(2.40)	(1.33)	(1.63)	(1.34)
MOM	0.34***	0.32***	0.35***	0.29***	0.26***	0.16	0.29***	0.17**
	(3.46)	(3.95)	(7.70)	(3.47)	(2.70)	(1.18)	(8.30)	(2.04)
STR	0.67***	0.57***	0.11*	0.19**	0.56***	0.67***	0.15***	0.51**
	(7.55)	(8.21)	(1.95)	(2.10)	(5.22)	(4.39)	(2.96)	(2.25)
LTR	0.29**	0.25**	−0.22**	−0.27	0.08	0.34	−0.16*	−0.12
	(2.47)	(2.27)	(−2.54)	(−1.63)	(0.59)	(1.60)	(−1.81)	(−0.46)
T	264	264	264	264	264	264	264	264
R^2	0.50	0.50	0.50	0.15	0.26	0.21	0.49	0.17

This table reports the risk-adjusted returns of the portfolios considered. We consider a model that includes the Market, Size, Book-to-Market, Profitability, Investment, Long and short term reversals, as well as the Momentum factor. Newey west t-statistics are reported in parenthesis. Alphas are expressed in percentage points.

Table OA.4: Stock Characteristics: Machine Learning Portfolios

	ML-R						ML-C					
	BM			Optimal			BM			Optimal		
	Long	Short	Diff.	Long	Short	Diff.	Long	Short	Diff.	Long	Short	Diff.
tang	52.82	53.85	-1.04	52.75	52.43	0.32	56.70	56.62	0.08	58.63	55.31	3.32
gma	49.40	49.41	-0.01	48.92	48.39	0.53	52.99	46.43	6.56	47.04	43.88	3.16
salecash	49.88	45.53	4.35	50.76	46.09	4.67	48.76	42.48	6.28	44.30	41.19	3.11
depr	55.63	52.88	2.75	55.81	50.87	4.94	59.37	57.69	1.68	62.82	59.75	3.07
ms	44.39	46.85	-2.46	42.49	46.64	-4.15	42.70	42.63	0.06	38.67	41.73	-3.06
salerec	51.71	50.47	1.24	51.85	49.98	1.86	52.76	49.75	3.00	51.85	48.88	2.97
ep	39.68	39.93	-0.25	36.84	39.67	-2.84	39.34	30.05	9.29	27.47	24.54	2.94
betasq	51.78	57.70	-5.92	49.63	55.83	-6.20	56.49	62.18	-5.69	61.05	63.97	-2.92
quick	50.75	53.04	-2.29	49.35	51.44	-2.09	54.70	53.33	1.37	53.85	51.00	2.85
ear	51.96	45.48	6.48	50.70	45.05	5.65	51.51	46.84	4.67	48.82	45.98	2.84
beta	51.47	57.46	-5.99	49.22	55.60	-6.38	56.24	61.67	-5.43	60.49	63.07	-2.58
pchcapx_ia	48.42	50.61	-2.19	48.11	50.47	-2.37	48.35	50.04	-1.69	47.15	49.58	-2.43
roaq	43.85	41.58	2.27	41.25	40.67	0.58	44.63	36.12	8.51	35.06	32.65	2.41
roic	40.18	42.72	-2.53	37.69	42.44	-4.75	40.71	33.73	6.98	30.27	27.93	2.35
tb	46.04	45.45	0.59	44.48	45.41	-0.93	44.45	41.30	3.15	39.61	37.33	2.28
nincr	52.45	47.43	5.02	50.98	47.05	3.93	52.59	49.07	3.52	50.98	49.00	1.98
mve_ia	45.18	46.34	-1.17	43.98	46.51	-2.52	42.23	43.23	-1.00	40.02	41.99	-1.97
std_turn	51.19	58.72	-7.54	50.68	59.09	-8.41	57.97	60.78	-2.81	61.70	63.66	-1.97
cash	52.97	54.84	-1.87	51.77	52.65	-0.88	56.49	58.07	-1.58	59.41	57.58	1.83
roeq	43.70	41.70	2.00	41.26	40.96	0.30	43.58	37.19	6.39	35.92	34.19	1.73
ps	45.06	44.27	0.78	43.25	44.17	-0.92	45.29	38.97	6.32	37.80	36.10	1.70
cinvest	50.04	49.67	0.37	49.93	50.20	-0.27	50.92	49.22	1.70	50.29	48.59	1.70
secured	52.79	53.25	-0.46	53.87	52.81	1.06	56.56	54.85	1.71	57.22	55.56	1.67
baspread	63.64	71.25	-7.61	68.73	72.86	-4.13	71.82	82.17	-10.35	86.54	88.17	-1.63
chatoia	52.65	49.97	2.67	52.26	49.56	2.70	52.23	50.95	1.27	52.52	50.97	1.55
stdcf	56.29	54.62	1.66	56.69	53.80	2.89	56.21	59.44	-3.23	61.82	60.31	1.52
stdacc	56.27	54.43	1.84	56.85	53.65	3.20	56.01	59.05	-3.04	61.55	60.13	1.42
herf	50.42	51.68	-1.27	51.42	52.75	-1.33	52.95	50.81	2.14	51.14	52.51	-1.37
chtx	52.68	44.88	7.80	51.13	44.09	7.04	50.07	47.68	2.39	47.84	46.49	1.35
pchdepr	52.46	50.58	1.88	53.03	50.63	2.40	52.97	52.41	0.56	55.01	53.73	1.29
convind	50.18	49.74	0.44	49.99	49.90	0.09	48.68	50.15	-1.47	49.65	50.86	-1.21
pchquick	47.45	48.85	-1.40	46.92	48.41	-1.49	49.25	46.44	2.81	45.64	44.46	1.18
pctacc	49.19	52.91	-3.73	49.49	52.19	-2.70	49.07	53.53	-4.46	52.91	54.04	-1.13
saleinv	46.63	48.53	-1.90	45.75	48.49	-2.75	47.70	46.61	1.09	46.63	45.52	1.11
pchgm_pchsale	48.64	48.19	0.45	47.71	47.47	0.24	48.86	46.89	1.96	46.43	45.32	1.11
pchsaleinv	50.72	49.42	1.30	50.51	49.09	1.42	51.44	49.91	1.53	51.32	50.22	1.09
rsup	47.81	48.11	-0.31	46.33	47.32	-0.99	48.84	47.01	1.83	44.58	45.52	-0.94
pchcurrat	47.34	49.11	-1.77	46.83	48.70	-1.87	49.00	46.57	2.43	45.50	44.56	0.93
acc	45.17	48.58	-3.42	44.91	49.10	-4.19	45.60	44.89	0.71	42.18	43.11	-0.93
pchsale_pchinvt	50.77	49.64	1.12	50.59	49.30	1.29	51.49	50.24	1.25	51.60	50.70	0.90
aeavol	49.73	49.04	0.70	48.77	48.35	0.42	51.53	48.66	2.87	49.06	48.18	0.88
idiovol	63.99	65.60	-1.61	67.30	64.07	3.23	73.20	77.66	-4.46	82.96	83.73	-0.77
rd_sale	52.68	52.60	0.08	51.88	51.53	0.35	53.75	56.43	-2.68	58.19	57.67	0.52
chpmia	48.64	50.14	-1.50	48.50	49.89	-1.39	49.22	49.50	-0.29	48.83	49.34	-0.51
pchsale_pchrect	51.24	50.02	1.21	51.29	49.88	1.41	51.43	50.73	0.70	51.99	51.50	0.50
pchsale_pchxsga	48.65	50.35	-1.70	48.16	49.74	-1.58	49.97	50.05	-0.08	49.21	49.67	-0.46
operprof	39.97	43.49	-3.51	37.78	43.91	-6.12	39.78	35.74	4.04	31.72	32.06	-0.34
absacc	55.67	54.43	1.24	56.53	53.36	3.17	57.02	59.21	-2.19	61.48	61.17	0.32
pricedelay	52.64	50.77	1.86	54.87	51.31	3.56	51.46	52.14	-0.68	53.59	53.87	-0.28
dy	43.20	40.16	3.04	42.05	41.51	0.54	35.62	35.63	-0.01	35.86	36.13	-0.27
divi	50.04	51.73	-1.69	50.34	52.87	-2.53	49.67	49.88	-0.21	49.86	50.12	-0.26
realestate	48.40	48.98	-0.58	48.02	49.91	-1.89	47.38	46.96	0.43	45.40	45.61	-0.21
securedind	50.56	53.01	-2.45	51.28	53.64	-2.37	54.24	52.74	1.51	53.30	53.13	0.16
divo	52.83	49.48	3.34	53.77	49.45	4.32	50.20	50.17	0.04	50.24	50.08	0.16
sin	49.92	50.44	-0.53	49.97	50.89	-0.92	49.87	49.82	0.05	49.76	49.85	-0.09

This table reports the average percentile of each characteristic in the stocks bought and the stocks sold by the portfolio in the column, as well as their differences. Characteristics are sorted according to their relevance for the optimal ML-C portfolio, defined as the spread between the average percentile of the stocks bought and stocks sold. This table reports the values for characteristics ranked 41st through 94th.

Table OA.5: Stock Characteristics: Linear Portfolios

	LIN-R						LIN-C					
	BM			Optimal			BM			Optimal		
	Long	Short	Diff.	Long	Short	Diff.	Long	Short	Diff.	Long	Short	Diff.
beta	50.78	56.04	-5.27	52.27	56.86	-4.59	53.84	62.19	-8.35	60.03	64.77	-4.74
gma	51.16	49.44	1.72	51.15	49.92	1.22	53.58	46.85	6.73	45.66	40.93	4.73
mve_ia	44.19	52.22	-8.03	44.44	50.99	-6.55	41.01	44.01	-3.00	37.40	41.97	-4.57
salerec	52.19	50.27	1.92	52.05	50.39	1.66	53.23	49.76	3.47	52.73	48.25	4.48
bm_ia	53.68	48.86	4.82	53.37	49.26	4.11	54.50	51.20	3.30	53.76	49.29	4.48
ep	48.14	42.08	6.05	46.93	41.88	5.05	40.73	29.71	11.02	23.01	18.65	4.35
pchsaleinv	53.31	46.87	6.44	52.84	47.58	5.26	52.19	49.33	2.86	54.03	49.76	4.28
pchcapx_ia	48.25	52.29	-4.04	48.34	51.83	-3.49	48.15	50.45	-2.30	46.14	50.36	-4.22
quick	52.69	50.98	1.72	53.01	51.64	1.37	55.25	52.96	2.30	53.16	49.01	4.15
currat	51.94	51.78	0.16	52.38	52.35	0.03	55.73	52.91	2.82	52.13	48.02	4.11
pchsale_pchinvt	53.27	46.99	6.28	52.85	47.73	5.12	52.21	49.67	2.54	54.35	50.45	3.89
retvol	53.82	58.50	-4.68	55.47	59.90	-4.43	69.18	80.64	-11.46	86.06	89.94	-3.89
ps	48.49	47.00	1.49	47.94	46.62	1.32	45.96	38.38	7.57	35.90	32.06	3.84
saleinv	51.30	47.42	3.88	50.68	47.38	3.31	48.51	45.89	2.62	47.55	43.78	3.77
tang	53.99	49.84	4.15	54.14	50.78	3.36	56.52	56.21	0.31	58.80	55.05	3.75
cash	54.94	49.90	5.03	55.34	51.33	4.01	56.81	56.64	0.17	60.53	56.94	3.59
tb	48.89	47.82	1.07	48.26	47.25	1.01	45.12	41.09	4.03	37.58	34.00	3.58
dolvol	40.66	56.28	-15.61	41.10	54.19	-13.09	34.33	39.01	-4.68	27.94	31.43	-3.49
roeq	51.23	41.57	9.67	50.29	42.43	7.85	43.98	36.53	7.44	34.31	30.83	3.49
salecash	48.50	48.80	-0.30	48.08	48.25	-0.16	48.69	43.61	5.07	42.99	40.03	2.96
ms	50.11	46.98	3.13	49.67	46.99	2.68	43.10	42.04	1.06	36.55	39.28	-2.73
chtx	52.00	47.41	4.59	51.63	47.96	3.66	50.37	47.35	3.02	48.99	46.28	2.70
acc	42.87	52.82	-9.95	43.57	51.93	-8.36	45.75	45.41	0.34	38.29	40.79	-2.51
std_dolvol	58.72	45.16	13.56	57.89	46.60	11.29	61.31	58.26	3.05	66.02	63.53	2.49
pctacc	46.58	53.28	-6.71	47.51	53.32	-5.81	48.88	53.21	-4.33	52.75	55.22	-2.47
baspread	54.35	57.40	-3.05	56.44	59.22	-2.77	70.94	81.32	-10.38	87.32	89.70	-2.38
herf	49.05	52.07	-3.02	49.52	52.02	-2.50	52.84	51.71	1.13	50.65	52.92	-2.26
operprof	48.28	44.31	3.98	47.27	44.12	3.15	41.10	35.31	5.79	29.74	27.50	2.25
dy	45.46	44.51	0.95	44.06	43.17	0.89	35.14	36.34	-1.20	34.37	36.53	-2.15
ill	57.02	46.96	10.06	57.11	48.77	8.34	66.47	65.12	1.34	75.56	73.53	2.03
stdcf	51.91	54.80	-2.90	52.40	54.65	-2.25	55.87	59.22	-3.34	64.79	62.77	2.03
pchgm_pchsale	50.73	48.58	2.15	50.48	48.64	1.84	48.78	46.64	2.14	45.97	43.94	2.02
pchdepr	51.54	48.54	3.00	51.67	49.28	2.39	52.71	52.44	0.27	56.54	54.57	1.98
pricedelay	50.82	48.95	1.87	50.76	49.25	1.51	51.40	52.37	-0.97	54.40	56.30	-1.90
convind	49.33	51.06	-1.73	49.35	50.78	-1.43	48.25	50.19	-1.94	49.42	51.24	-1.82
aeavol	52.04	49.09	2.95	51.68	49.38	2.30	51.46	48.89	2.58	49.15	47.37	1.79
rd_sale	52.90	50.47	2.43	53.07	51.09	1.99	53.62	55.77	-2.14	60.39	58.70	1.69
idiovol	56.71	56.79	-0.08	58.87	59.20	-0.33	70.64	78.19	-7.55	88.12	89.77	-1.64
securedind	47.20	55.65	-8.45	47.99	55.10	-7.11	53.75	54.35	-0.61	53.74	55.36	-1.62
pchsale_pchxsga	49.38	50.61	-1.23	49.64	50.71	-1.07	50.49	49.61	0.88	50.56	49.14	1.42
chpmia	48.47	51.16	-2.68	48.65	50.93	-2.28	48.97	49.77	-0.80	48.46	49.78	-1.32
realestate	49.07	49.56	-0.49	48.86	49.25	-0.39	47.91	46.86	1.05	45.44	44.16	1.28
stdacc	51.30	54.86	-3.56	51.90	54.69	-2.78	55.29	59.04	-3.75	63.78	62.70	1.09
pchsale_pchrect	51.58	48.76	2.82	51.51	49.17	2.34	51.39	50.96	0.43	53.33	52.35	0.98
divi	49.46	50.39	-0.93	49.53	50.29	-0.76	49.51	49.99	-0.48	49.75	50.56	-0.81
pchquick	48.57	50.92	-2.35	48.54	50.55	-2.01	49.09	46.50	2.59	43.66	42.89	0.77
rsup	49.50	49.52	-0.02	49.50	49.65	-0.16	49.65	46.54	3.12	44.77	44.03	0.74
zerotrade	48.74	45.69	3.05	48.36	45.48	2.88	51.04	46.14	4.90	44.70	45.36	-0.66
secured	50.91	51.77	-0.86	51.31	52.16	-0.85	56.07	55.47	0.60	57.07	56.54	0.53
absacc	55.45	51.05	4.40	55.50	51.96	3.54	56.65	59.12	-2.47	64.85	64.52	0.32
pchcurrat	47.91	51.47	-3.56	47.99	51.02	-3.03	48.83	46.68	2.15	43.20	43.03	0.17
cinvest	37.71	66.71	-29.00	39.34	62.85	-23.50	49.84	49.89	-0.05	49.72	49.61	0.12
divo	50.32	49.75	0.57	50.25	49.79	0.46	50.14	50.08	0.06	50.21	50.13	0.08
sin	49.76	50.36	-0.60	49.76	50.23	-0.47	49.90	49.88	0.03	49.82	49.90	-0.07

This table reports the average percentile of each characteristic in the stocks bought and the stocks sold by the portfolio in the column, as well as their differences. Characteristics are sorted according to their relevance for the optimal ML-C portfolio, defined as the spread between the average percentile of the stocks bought and stocks sold. This table reports the values for characteristics ranked 41st through 94th.