How To Bet On Winners and Losers

PRELIMINARY AND INCOMPLE, PLEASE DO NOT CIRCULATE

Christian Brownlees* André B.M. Souza*

April 5, 2025

Abstract

We study the construction of long-short portfolios on the basis of cross-sectional return predictions. We derive an optimal portfolio construction procedure that takes the form of a return classification rule. Selecting stocks on the basis of expected return predictions, the standard practice in the literature, is also optimal in special cases of the general framework. An empirical application to US stocks highlights that the portfolios constructed using the proposed procedure outperform portfolios constructed using the standard tools in the literature, and the outperformance persists when transaction costs are duly accounted for.

*Universitat Pompeu Fabra and Barcelona School of Economics. e-mail: christian.brownlees@upf.edu ESADE Business School. e-mail: andre.souza@esade.edu.

1 Introduction

Portfolio sorts are extensively used in empirical finance to assess the economic value of predictive signals for the cross-section of stock returns. The standard procedure consists in constructing an investment strategy by sorting stocks into deciles of a signal, buying stocks in the top decile and selling those in the bottom decile. The economic value of the predictive signal is then assessed by the properties of returns from the *top-minus-bottom* portfolio.

In the standard framework, predictive signals provide economic value to investors insofar as they help discriminating stocks associated with high returns from those with low returns, *i.e winners* from *losers*. Portfolio sorts achieve discrimination by estimating the mapping from the predictive signal to expected returns, and grouping stocks into deciles of expected returns (Jegadeesh and Titman, 1993; Kelly *et al.*, 2020). This procedure, however, amounts to solving the intermediate and more general problem of conditional mean estimation[1]. A direct approach is to treat the problem of telling apart *winners* and *losers* as a classification task, *i.e* estimate a discriminant function based on predictive signals, a procedure that has recently attracted attention from the empirical literature (Rapach *et al.*, 2024; He *et al.*, 2024; Han, 2022).

In this work, we introduce a framework to formalize the construction of long-short portfolios based on predictive signals for the cross section of returns. We cast portfolio construction as a stock selection problem in which the investor is concerned with identifying which stocks will feature in the top and bottom of the cross-sectional return distribution. We introduce a loss function based on the discrimination properties of competing selection rules and derive the optimal portfolio construction rule. The optimal portfolio construction rule is a function of the probabilities that a stock will feature in the top and bottom of the cross-sectional return distribution. In an empirical application, we document that the optimal portfolio construction rule leads to portfolios with higher Sharpe ratios and expected returns when compared to portfolios constructed from the standard sorting procedure.

We consider an investor who wishes to buy (sell) stocks expected to out(under)perform their peers. In our framework, the investor's problem is to identify which stocks will underperform, market perform, or outperform, and take positions accordingly. Since the investor's decision problem is categorical, it is natural to evaluate candidate selection rules using loss functions suitable for (multi-class) classification problems. The class of loss functions we consider allows for false discovery penalties to be incorporated in the problem. In particular, buying a stock that underperforms its peers incurs in higher costs than buying a stock that

[1]Vapnik (1999) argues that the main principle for solving problems using a restricted amount of information is to avoid solving a more general problem as an intermediate step.

market performs.

We derive optimal portfolio construction rules for the class of loss functions introduced. The optimal rules do not depend on parametric assumptions about the data-generating process, and selects stocks according to the probabilities of the stock outperforming and underperforming the market. The shape of the selection region for both *winners* and *losers* depends on the costs associated with false discoveries, which we treat as tuning parameters to be chosen by the investor. Importantly, the optimal construction rule accomodates predictive signals that impact the entire distribution of returns, and not just expected returns. In the special case in which the conditional distribution of returns belong to a location-scale family, the optimal construction rules are a function of a stock's expected returns and volatility. Investors take long (short) positions in a stock if the expected return is higher (lower) than a threshold that depends on the tuning parameter and increases (decreases) with the volatility of the stock. A specific choice of tuning parameters leads to the standard expected return sorts widely employed in the literature. We remark that, in line with the empirical finance literature, our procedure is intended to aid the investor in finding *which* stocks to invest rather than *how much* to invest in each stock. For the latter, we follow standard practice in the literature and consider both equally-weighted portfolios and value-weighted portfolios.

We put our framework to test in a sample of U.S. stock data from 1957 to 2021. Our dataset consists of more than 7500 unique assets over 756 months, and we assume investors have access to the 94 characteristics used in Green *et al.* (2017); Kelly *et al.* (2020). We combine characteristics into predictive signals using both regression and multi-class classification methods, as well as generalized linear models (GLM), Ordinary Least Squares(OLS) and tree-based models (eXtreme Gradient Boosting, XGB). We compare portfolios constructed using the optimal construction rules — which we label *Optimal* — with their counterparts constructed using the standard tools in the literature, which depends on the type of forecast employed. For regression methods, we partition conditonal mean forecasts into deciles, and create the high-minus-low *Decile* portfolios by going long in stocks in the top decile, and short in stocks in the bottom decile (Kelly *et al.*, 2020). Conversely, for classification methods, we follow standard practice and select stocks into the most likely class (MLC). Stocks for which the most likely class is the top decile are bought, and stocks for which the most likely class is the bottom decile are sold.

Optimally constructed portfolios achieve substantially higher expected returns than their standard counterparts. The highest average return achieved by the equal-weighted (value-weighted) standard portfolios is of 3.51% (1.68%) per month, achieved by decile portfolios based on XGB expected return forecasts. The highest average returns achieved by the

equal-weighted (value-weighted) optimal portfolios is of 6.12% (4.00%) per month, achieved by optimally constructed portfolios based on XGB classification forecasts, more than twice the highest average returns obtained by traditional sorts. In addition, optimally constructed portfolios also achieve higher Sharpe ratios than their standard counterparts. The highest Sharpe ratios achieved by standard portfolio sorts is of 2.88 and 0.94 for equal-weighted and value-weighted portfolios, respectively. In contrast, the highest Sharpe ratios achieved by the optimal portfolios is of 2.92 and 1.24 for equal-weighted and value-weighted portfolios, respectively.

We account for transaction costs by estimating stock specific bid-ask spreads, as in Ledoit and Wolf (2025). Due to data availability, we consider the impact of transaction costs in the optimal portfolios starting from January, 2000, until the end of the sample. When transaction costs are carefully accounted for, we find that out of all portfolios constructed using standard sorting procedure with either value or equal weights, only the most-likely-class portfolio constructed with XGB(C) and value-weighting has a Sharpe ratio of 0.48, marginally higher than the market Sharpe ratio of 0.46 in the same period. In contrast, the optimally constructed portfolios based on XGB(C) has Sharpe ratios that are about 30% higher than the market Sharpe ratio, using equal weights, and about 80%, using value weights.

Overall, we find that predictive signals obtained with classification models combined with optimal portfolio construction rules outperform portfolios based on conditional mean forecasts, be it constructed using standard sorting tools or with optimal construction rules. This suggests that considering the impact of characteristics on the conditional distribution of returns, beyond their effect on the conditional mean, may be relevant. Moreover, we find that tree-based models outperform linear (and generalized linear) models across the board, but the performance gains are substantially larger when optimal portfolio construction rules are employed.

We then explore which characteristics are associated with the constructed portfolios. Characteristics based on past returns, such as momentum and short-term-reversal, seem to be the most relevant characteristics for inclusion of a stock in the portfolios. The average stock in the optimal XGB(C) long portfoio is on the lowest quintile of short-term-reversal and on the second lowest quintile of 12-2 months momentum, whereas the average stock in the short portfolio is on the top 60th percentile of short-term-reversal and on the bottom 25th percentile of 12-2 months momentum. Moreover, the optimal portfolio typically selects small stocks, with the average stock in the long leg being at the 20th percentile of market value, and the short leg at the 26th percentile. For generalized linear models, we find that

characteristics based on past returns play a more pronounced role, whereas size, measured as market capitalization, is less relevant, particularly for OLS. This suggests that interactions of characteristics play a role in the conditional distribution of returns, a point also raised in Kelly *et al.* (2020) and Freyberger *et al.* (2020).

We examine whether the Fama-French 5-factor model (Fama and French, 1995, FF5) augmented with momentum, short-term-reversal and long-term-reversal is able to account for the excess returns generated by our portfolios. The FF5 model is able to account for excess returns from the GLM and XGB(C) models constructed with the standard sorting procedure, but not when optimal portfolio construction rules are employed. In fact, alphas from all optimally constructed portfolios are larger than those obtained through standard sorting procedure, and all are statistically significant at any reasonable threshold for t-statistics. Moreover, the share of variation of portfolio returns explained by the FF5 model is substantially smaller for optimally constructed portfolios than for the portfolios constructed using the standard procedure. Once transaction costs are accounted for, only the XGB(C) optimal portfolio generates positive and significant alphas.

We attribute the outperformance of optimally selected portfolios based on classification methods to the strong discriminatory power achieved by these methods. Classification methods are able to achieve about twice as much true positive rates as regression methods, for the same level of false positive rates.

Our paper relates to several strands of the literature. The empirical finance literature has long employed characteristic sorted portfolios to construct investment strategies and identify pricing anomalies. (Jegadeesh and Titman, 1993; Fama and French, 1992, 1993). More recently, portfolios constructed from expected returns (Lewellen, 2015; Kelly *et al.*, 2020, to name a few) and conditional probability (Rapach *et al.*, 2024; He *et al.*, 2024) sorts have been employed to construct investment strategies and to ascertain the economic value of predictability. Theoretical properties of portfolio sorts have recently attracted attention from the literature. Cattaneo *et al.* (2020, 2023) develop the theoretical properties of characteristic-sorted portfolios as nonparametric estimators of expected returns. Daniel *et al.* (2020) show that characteristic-sorted portfolios may capture not only priced risk associated with the characteristic but also unpriced risk. Patton and Timmermann (2010) develop tests to assess null of monotonicity of the expected return of characteristics-sorted portfolios. Ledoit *et al.* (2019); Olmo and McGee (2022) construct mean-variance optimal characteristic-sorted portfolios. The properties of ranking and selection of top performing entities on the basis of estimated sample means has been studied in the econometrics (Gu and Koenker, 2023; Hirano and Porter, 2009; Andrews *et al.*, 2024) and statistics (Gelman

and Price, 1999) literature.

The remainder of the paper is organized as follows. Section 2 introduces the framework and derives optimal portfolio construction rules. Section 3 discusses the implementation of the optimal portfolio construction rules. Section 4 contains the empirical application, and Section 5 concludes.

2 How to Bet on Winners and Losers

We consider an investor that may trade in n stocks, and we denote by R_i the return of stock i, for $i = 1, \ldots, n$. We assume that the objective of the investor is to construct a zero-net-investment portfolio by buying stocks that will outperform their peers (i.e *winners*) and selling their underperforming counterparts (i.e *losers*). Formally, we assume the investor has a labelling function, $c(R_i)$, that maps observed returns into investment targets:

$$c(R_i) = \begin{cases} 1 & \text{if } R_i \geq \mu_W \quad (\textit{winners}) \\ 0 & \text{if } \mu_L < R_i < \mu_W \quad (\textit{neutral}) \\ -1 & \text{if } R_i \leq \mu_L \quad (\textit{losers}) \end{cases},$$

where $\mu_L \leq \mu_W \in \mathbb{R}$ are return thresholds set by the investor. Clearly, $c(R_i)$ is unknown prior to the realization of returns. Therefore, the investor's objective is to predict $c(R_i)$ on the basis of the information available in the portfolio formation period.

To aid in their task, we assume the investor has access to a vector of $\boldsymbol{X} \in \mathbb{R}^p$ of characteristics for each stock. The investor must choose a portfolio selection rule $\boldsymbol{w} : \mathbb{R}^{n \times p} \to \{-1, 0, 1\}^n$ that maps characteristics into a set of decisions to buy, sell, or not take any positions in stock i, for $i = 1, \ldots, n$. In our framework, the investor problem is a multi-class classification problem with ordered outcomes. An appropriate loss function in this setting is the nonparametric ordinal loss function, which can be obtained as the sum of the misclassification losses for the *winners* and *losers*. In other words, define the misclassification loss for the *winner* and *loser* portfolios to be:

$$\mathcal{L}^W_{\lambda_W}(\boldsymbol{w}) = \sum_{i=1}^{n} \underbrace{\mathbb{1}(\{c(R_i) = 1\} \cap \{w_i \neq 1\})}_{\text{False Negative Error}} + \lambda_W \underbrace{\mathbb{1}(\{c(R_i) \neq 1\} \cap \{w_i = 1\})}_{\text{False Positive Error}}, \text{ and}$$

$$\mathcal{L}^L_{\lambda_L}(\boldsymbol{w}) = \sum_{i=1}^{n} \underbrace{\mathbb{1}(\{c(R_i) = -1\} \cap \{w_i \neq -1\})}_{\text{False Negative Error}} + \lambda_L \underbrace{\mathbb{1}(\{c(R_i) \neq -1\} \cap \{w_i = -1\})}_{\text{False Positive Error}},$$

where $\lambda_W, \lambda_L \in [0, \infty)$ is the relative cost of a false discovery. Combining the two losses, we

define the nonparametric ordinal loss as:

$$\mathcal{L}_{\boldsymbol{\lambda}}(\boldsymbol{w}) = \mathcal{L}^{W}_{\lambda_W}(\boldsymbol{w}) + \mathcal{L}^{L}_{\lambda_L}(\boldsymbol{w}) \ , \tag{1}$$

where $\boldsymbol{\lambda} = (\lambda_W, \lambda_L)'$. Naturally, the value of loss is unknown prior to the realization of returns, so the investor's objective is to find

$$\boldsymbol{w}^* \in \arg\min_{\boldsymbol{w}} \mathbb{E}(\mathcal{L}_{\boldsymbol{\lambda}}(\boldsymbol{w})) \ , \tag{2}$$

where $\boldsymbol{\lambda} = (\lambda_W, \lambda_L)'$. The following proposition characterizes the optimal portfolio selection rule.

Proposition 1 (Optimal Portfolio Selection Rule)**.** *Let* $p_W = \mathbb{P}(R_i > \mu_W|\boldsymbol{X})$ *and* $p_L = \mathbb{P}(R_i < \mu_L|\boldsymbol{X})$*, where* $\boldsymbol{X} = (\boldsymbol{X}_1', \ldots, \boldsymbol{X}_n')'$ *is the vector of stacked characteristics. The selection rule* $\boldsymbol{w}^*$ *is such that*

$$w_i^* = \begin{cases} 1 & \text{if } p_W \geq \max\left(\frac{\lambda_W}{1+\lambda_W}, \frac{\lambda_W - \lambda_L}{1+\lambda_W} + \frac{1+\lambda_L}{1+\lambda_W} p_L\right) \\ -1 & \text{if } p_L \geq \max\left(\frac{\lambda_L}{1+\lambda_L}, \frac{\lambda_L - \lambda_W}{1+\lambda_L} + \frac{1+\lambda_W}{1+\lambda_L} p_W\right) \\ 0 & \text{otherwise} \end{cases} \ . \tag{3}$$

The optimal portfolio selection rule $\boldsymbol{w}^*$ buys (sells) stocks that have returns higher (lower) than μ_W (μ_L) with high enough probability, and does not trade on stocks that are not likely to feature on the desired buy and sell regions. In other words, the optimal rule selects stocks that are likely to be future *winners* or *losers*. Figure 1 plots the selection region implied by equation (3) obtained by setting $\lambda_W = \lambda_L = 1$, on the left panel, and $\lambda_W = \frac{1}{3}$ and $\lambda_L = \frac{1}{2}$, on the right panel. We color the buy region in blue, the sell region in red, and the no-trade region in white.

[FIGURE 1 ABOUT HERE]

As can be seen in Figure 1, the parameters (λ_W, λ_L) control the size and shape of the selection region, and are therefore closely related to portfolio characteristics such as the expected return, variance, Sharpe ratio, and the size (the number of stocks included) of the portfolio. 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 also note that the optimal portfolio selection rule in (3) does not make any assumptions about the

distribution of returns.

It is useful to illustrate the optimal portfolio selection rule when $R_i \sim \mathcal{F}_{\theta_i}$, where $\mathcal{F}_{\theta_i}$ is a zero-median location-scale distribution parametrized by $\theta_i = \{\mu_i, \sigma_i\}$, with μ_i a location parameter and σ_i a scale parameter. The proposition below characterizes the optimal selection rule in location-scale models:

Proposition 2 (Optimal Portfolio Selection Rule in Location-Scale Models)**.** *For $\lambda_W, \lambda_L \geq 1$, the optimal portfolio selection rule in location-scale models is given by*

$$w_i^* = \begin{cases} 1 \text{ if } \mu_i - F^{-1}(\frac{\lambda_W}{1+\lambda_W})\sigma_i \geq \mu_W \\ -1 \text{ if } \mu_i + F^{-1}(\frac{\lambda_L}{1+\lambda_L})\sigma_i \leq \mu_L \\ 0 \text{ otherwise} \end{cases}, \tag{4}$$

where F is the cumulative distribution function of $\mathcal{F}_{\{0,1\}}$, and F^{-1} its inverse.

Proposition 2 shows that in location-scale models the inclusion of a stock in the portfolio depends on the stock's mean and variance. In particular, if the distribution has median equal to zero, then setting $\lambda_W = \lambda_L = 1$ implies a long position in a stock if $\mu_i \geq \mu_W$, a short position if $\mu_i \leq \mu_L$, and no position otherwise. Hence, in this particular case, our procedure is equivalent to the standard procedure used in the literature of sorting stocks according to expected returns. An investor that sets $\boldsymbol{\lambda} = \mathbf{1}$ wishes to obtain high expected returns regardless of risk considerations. This may be undesirable, as pointed out in Ledoit *et al.* (2019). Choosing $\lambda_W > 1$, implies the inclusion of a stock in the portfolio depends on the stock's expected return and volatility. In particular, the investor should take a long position in a high risk stock only if this stock has a high enough expected return, and λ_W controls the risk-return tradeoff for the investor, with analogous results for λ_L. Figure 2 depicts the selection region obtained for $\lambda_W = \lambda_L = 1$ (left panel) and $\lambda_W = \lambda_L = 1.1$ (right panel).

[FIGURE 2 ABOUT HERE]

The area of the selection region — and hence the number of stocks included in the portfolio — depends on (λ_W, λ_L) and the marginal distribution of returns $\{\mathcal{F}_{\theta_i}\}_{i=1}^n$. 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. In particular, increasing λ_W implies that stocks with expected returns higher than μ_W are included in the portfolio as long as their volatility is "small enough", and so λ_W can be thought of as a penalty for volatility. Notice that the standard portfolio sorting procedure can be seen as a special case of our optimal selection rules for location scale models and by setting $\lambda_W = \lambda_L = 1$. In general, however, whether

portfolios constructed with $\lambda_W = \lambda_L = 1$ have better properties than those obtained with other choices of (λ_W, λ_L) depends on the (unknown) joint distribution of returns. We remark that if the investors are interested in controlling the size of the portfolio, they may vary $(\mu_W, \mu_L, \lambda_W, \lambda_L)$ so as to include $q\%$ of stocks in the selection region. Finally, we note that setting $\lambda_W < 1$ or $\lambda_L < 1$ would imply "risk-loving" investors that, when comparing two stocks with the same expected returns, would prefer the one with the higher risk.

2.1 Discusson

A number of remarks are in order. First, the loss function in Equation (1) incorporates the ordering of the classes in the sense that buying a stock that has returns greater than μ_L incurs a lower loss than buying stocks with returns lower than μ_L. An appealing feature of the proposed loss is that it does not rely on parametric assumptions on the distribution of returns, in contrast to the loss used in standard ordinal regression, for example.

Second, it is standard practice to denote a stock as a *winner* if it is on the top $q\%$ of *expected returns*. Since expected returns are unobservable, the standard procedure typically sorts stocks according to characteristics thought to predict the cross-section of expected returns (Jegadeesh and Titman, 1993; Fama and French, 1995). In contrast, we denote a stock as a *winner* if it has *realized returns* higher than μ_W, or, alternatively, if it is on the top $q\%$ of realized returns. Both definitions may be of interest to investors. Whereas the standard definition relies on proxies for unobservable expected returns, our definition relies on observable realized returns. As a remark, we note that returns for the portfolio consisting of buying the top $q\%$ of realized returns are lower bounded by returns for the portfolio consisting of buying the top $q\%$ of expected returns [2].

Third, the loss in (1) is a variant of standard loss functions used in binary decision problems, and is akin to the one used in Gu and Koenker (2023). The performance thresholds (μ_W, μ_L) may be set to a particular benchmark threshold, such as 0, the risk-free rate, or returns on the market. Conversely, one may set the thresholds to the q-th percentile of the cross-section of realized returns. The cost of trading in the wrong direction (λ_W, λ_L) may be chosen by the investor ex-ante. Alternatively, the investor may choose (λ_W, λ_L) to match some investing objective. We describe the selection of loss function parameters in detail in Section 3.

Finally, the portfolio selection rules that we consider determine inclusion or exclusion of a stock in the portfolio. Hence, we aim to answer *which* stocks the investor should buy, rather than *how much* of each stock they should buy. As a consequence, portfolios constructed

[2] This follows from Jensen's inequality.

from such rules are equally weighted portfolios by default. An often-used alternative is to assign weights that are proportional to the market capitalization of selected stocks, *i.e* value-weighting. In the empirical application, we report portfolios construct by assigning equal and value weights to its components.

3 Implementation of the Optimal Selection Rules

The implementation of the optimal selection rules described above requires: (i) a choice of $c(R_i)$, a function to label which stocks are winners and which stocks are losers, (ii) forecasts of $\mathbb{P}(R_i > \mu_W|\boldsymbol{X})$ and $\mathbb{P}(R_i < \mu_L|\boldsymbol{X})$, and (iii) a choice of λ_W and λ_L for the construction of the selection rule. In what follows, we describe each of the above ingredients.

Choice of $c(R_i)$. The first ingredient in the construction of portfolios is the function $c(R_i)$, which categorizes stocks as *winners*, *losers*, or *neutral* as a function of realized returns. In this work, we label *losers* as the stocks with returns below the 10% quantile of $\{R_i\}_{i=1}^n$, and *winners* the stocks with returns above the 90% quantile of $\{R_i\}_{i=1}^n$. We choose this definition to follow, as closely as possible, the standard practice in the literature. Although we do not pursue this path, other choices may be entertained. For example, one may wish to define as *winners* the stocks on the top quintile of the cross-sectional return distribution, and *losers* those on the bottom quintile. Alternatively, one could wish to buy stocks returns higher than a *fixed* threshold of, say, 2%, per month and sell those with returns below, say, 0%.

Forecasts of $\mathbb{P}(R_i > \mu_W|\boldsymbol{X})$ and $\mathbb{P}(R_i < \mu_L|\boldsymbol{X})$. The second ingredient required for portfolio construction are forecasts of $\mathbb{P}(c(R_i) = 1|\boldsymbol{X})$ and $\mathbb{P}(c(R_i) = -1|\boldsymbol{X})$, where, as before, $\boldsymbol{X}$ is the vector of stacked stock characteristics. There are several alternatives to estimate these probabilities. One may employ use multi-class classification models to estimate the probabilities directly. Note that doing this requires labeling the data according to $c(R_i)$ and estimating probabilities with an appropriate loss function, for example, the multinomial loss function. We refer to this procedure as the classification framework. Conversely, one may estimate conditional means and volatilities and plug-in these quantities in Equation 2, replacing F with some assumed zero-median location scaled marginal distribution for returns, for example, a Gaussian distribution. Conditional means may be estimated using the standard mean squared error loss function, for example. Conditional volatilities may be estimated the standard deviations of model residuals, or with GARCH type models. We refer to this procedure as the regression framework.

Choosing λ_W and λ_L: Standard practice. Once the investor has defined the *winners* and *loser* stocks and obtained stock-level probabilities, they must decide which stocks should feature in which leg of their portfolio. The standard approach to portfolio construction based on conditional mean forecasts is to plug-in forecasts of R_i into $c(\cdot)$, that is, buy a stock if $c(\mathbb{E}[R_i]) = 1$ and sell it if $c(\mathbb{E}[R_i]) = -1$. This corresponds to using the regression framework described above with $\lambda_W = \lambda_L = 1$. We label this approach *Decile* based classification throughout. In the classification setting, the standard approach employed in the literature (Rapach *et al.*, 2024) is to label a stock according to $\arg\max_k \mathbb{P}(c(R_i) = k|\boldsymbol{X})$. That is, buy a stock if the stock is more likely to be a *winner* than a *loser* or *neutral*. We label this approach the *Most Likely Class* classification (MLC).

Choosing λ_W and λ_L: Data-based. In our framework, the assignment of a stock into a portfolio depends on the forecasted class probabilities and the parameters (λ_L, λ_W) which define the costs associated with misclassifying stocks. In practice, we treat (λ_W, λ_L) as tuning parameters, which are recursively chosen based on past performance according to some metric. Unfortunately, the choice of (λ_W, λ_L) is problem-specific and there are no statistical loss functions to aid in this regard. Fortunately, however, the portfolio construction problem has a clear target: to obtain high Sharpe ratios. We therefore select (λ_W, λ_L) recursively to maximize portfolio Sharpe Ratio on a hold-out sample, and we label these portfolios as *Optimal*. Clearly, other targets (based on expected returns or variances, for example) may be entertained.

4 Empirical Application

We consider the construction of portfolios from characteristics-based forecasts for US 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 on average 5000 stocks per time period. We use the 94 firm characteristics employed in Kelly *et al.* (2020), which are based on those considered in Green *et al.* (2017).[3]. 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.

Similarly to Kelly *et al.* (2020); Rapach *et al.* (2024), all the models we consider pool

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

cross-section information to estimate parameters. In the regression framework, we model the conditional mean of returns as $\mathbb{E}(R_{i\,t}|\boldsymbol{X}_t) = f(\boldsymbol{X}_{i\,t})$, with $f(\cdot)$ depending on i and t only through $X_{i\,t}$. We estimate the idiosyncratic volatility as the residual volatility of each stock.[4] In the classification framework, we model the conditional distribution of returns as follows. To ensure comparability with standard practice, we follow Rapach *et al.* (2024) and estimate a 10-class classification model where the class labels are the corresponding return decile memberships. That is, we estimate $\mathbb{P}(R_{\lceil qn \rceil\,t} \leq R_{i\,t} < R_{\lceil (q+0.1)n \rceil\,t}|\boldsymbol{X}_t) = g_q(\boldsymbol{X}_{i\,t})$ for $q = \{0, 0.1, \ldots, 0.9\}$ where $g_q(\cdot)$ depends on i and t only through $X_{i\,t}$. This is necessary to ensure that the of most likely class selection leads to (somewhat) "balanced" outcomes. With 3 classes which account for respectively 10, 80 and 10 % of the data (*winners*, *neutrals*, *losers*), most likely class selection will likely select *neutrals* for the overwhelming majority of stocks. The relevant probabilities for our selection rules are the ones obtained for $q = 0$ and for $q = 0.9$.

As in Lewellen (2015); Kelly *et al.* (2020); He *et al.* (2024), we consider conditional mean forecasts from Ordinary Least Squares (OLS) and probability forecasts from a Generalized Linear Model (GLM). In particular, we use the logit model. As in Rapach *et al.* (2024), we consider eXtreme Gradient Boosting with a regression loss function (XGB(R)), and a multi-class classification loss function (XGB(C)). We consider tree-based XGB models. We consider 100 rounds of boosting, where we treat the depth of the tree and the learning rate of each round as tuning parameters. We consider tree-depths, $d \in 2, 3, 4$, 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 1957-01-01 and ends in 1974-12-01. The first validation period starts in 1974-12-01 and ends in 1986-01-01. The first testing year is 1986-01-01. We fix the validation set size to 12 years. Each year, we recursively estimate the models with an expanding training window, as in Kelly *et al.* (2020). Tuning parameters for the XGB(R) and XGB(C) models are chosen on the validation set, as in Kelly *et al.* (2020). Prior to each rebalancing period, we re-estimate idiosyncratic volatilities and choose (λ_W, λ_L) based on data from all the previous months that has not been used in the training sample.

4.1 Performance of Optimal Portfolios

In this section, we investigate the properties of the portfolios obtained from all forecasting models and selection rules.

[4]We require at least 2 months of prior data to estimate idiosyncratic volatilities.

[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. In addition, we report the average percentage of stocks selected, the average correct classification rate, as well as the wrong direction rate, defined by the probability of buying a *loser* or selling a *winner*. Panel A of Table 1 reports the results for the equally weighted portfolio. Panel B reports results for the value weighted portfolio. Decile portfolios are constructed by buying stocks in the top decile and selling stocks in the bottom decile of the relevant forecast. MLC is classification according to the Most Likely Class.

A number of remarks are in order. First, and in line with the findings in Rapach *et al.* (2024), we find that XGB(C) forecasts lead to portfolios that perform better than XGB(R), GLM and OLS models. Second, optimally selected portfolios tend to select substantially fewer stocks to invest in than decile sorting or most likely class classification (MLC). In particular, optimally selected XGB(C) portfolios trade, on average, on 12.52% of the stocks available at each point in time. In contrast, portfolios constructed by selecting the most likely class trade on about 34.67% of the stocks at each point in time. OLS based portfolios are the exception, with the optimal portfolio trading on about 29.16% of the stocks available at each point in time. Third, the *optimally* selected portfolios have higher expected returns, Sharpe and Sortino ratios than the standard selection rule across all forecasting models As a remark, we note that since all portfolios are self-financing, in theory one could obtain higher returns by using leverage, for example. In practice, however, leveraging incurs in costs, and self-financing portfolios require margins to be held, so "scaling-up" is not a trivial matter. For this reason, given two similar Sharpe ratio portfolios, investors would likely prefer the one with the higher expected returns. In addition, and despite not being a direct target of the proposed framework, *Optimal* portfolios exhibit higher skewness than their "standard" counterparts. Finally, equally-weighted portfolios display substantially better performance than value-weighted portfolios. This is in line with the literature, and likely due to the influence of small stocks. Most of the previous remarks remain valid in the case of value-weighted portfolios. The one exception is that OLS based optimal portfolios display a Sharpe ratio that is marginally lower than their Decile counterpart, but still with higher average returns, skewness and Sortino Ratio. Moreover, for classification models, the gains from optimally choosing portfolios increase substantially, with the Sharpe ratio of optimally

selected XGB(C) portfolios more than doubling that of the standard procedures employed in the literature.

[FIGURE 3 ABOUT HERE]

Figure 3 reports the natural logarithm of the cumulative returns obtained by XGB(C) using either equal or value weights, and optimal and most likely class classification. Cumulative returns for the optimally chosen portfolio stochastically dominate those from portfolios formed according to the standard tools.

Accounting for Transaction costs Our results showcase substantial return premiums for XGB(C) investors with optimally chosen stocks. To provide a realistic assessment of competing portfolio selection rules based on past performance, we must account for transaction costs.

We therefore carefully account for transaction costs, following Ledoit and Wolf (2025). 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 consider two measures of transaction costs: the (i) Quoted Spreads (QS) and (ii) the transaction cost measure used in Ledoit and Wolf (2025) (LW). Due to data availability, we consider a reduced sample on both the cross-section and time-series dimensions. On the time-series dimension, we focus our analysis of transaction costs to data from January, 2000. On the cross-section, we require CRSP spreads to be available for the stocks.

We construct portfolios following the procedures outlined in Section 3, and portfolios are rebalanced monthly.

[TABLE 2 ABOUT HERE]

Table 2 reports our results using Quoted Spreads. Clearly, accounting for transaction costs substantially reduces all Sharpe Ratios. In particular, OLS and GLM no longer provide profitable portfolios to be constructed. However, XGB(R) and XGB(C) still provide profitable alternatives. In particular, equally weighted optimally chosen XGB(C) portfolios provide excess returns of about 1.35% net of transaction costs, leading to a Sharpe Ratio of 0.61, about 50% higher than the Market sharpe ratio of about 0.46 for the same time period. In addition, when transaction costs are accounted for, value-weighted portfolios perform better than equally weighted portolios, with Sharpe ratios reaching 0.84, nearly double the market Sharpe Ratio.

[FIGURE 4 ABOUT HERE]

Figure 4 reports the natural logarithm of cumulative returns from XGB(C) portfolios accounting for transaction costs. We report both equal (solid lines) and value (dashed lines) weights. As before, optimally chosen portfolios outperform their standard counterparts.

Stock Characteristics We next study whether the portfolios constructed may be seen as variations of the standard long-short portfolios constructed on the basis of characteristics, or whether the good performance of the models considered is achieved by "mixing" several characteristics.

[TABLE 3 ABOUT HERE]

Table 3 reports the average percentile of each characteristic in the stocks bought and the stocks sold by the portfolio in the column. Characteristics are sorted according to their relevance for XGB(C) - Optimal, defined as the spread between the average percentile of the characteristic in the long minus the short leg, taken in absolute value. We truncate the table at 30 characteristics for readability. *Decile* portfolios are obtained by buying the top 10% of stocks according to the relevant forecast. *MLC* portfolios are obtained by buying the stocks for which the most likely class according to the relevant forecast is *winner* and selling their *loser* counterparts. *Optimal* portfolios are obtained by the optimal selection regions with parameters selected through cross-validation. The full table can be found in the Internet Appendix.

Characteristics based on past returns, such as momentum and short-term-reversal, seem to be the most relevant characteristics for inclusion of a stock in the portfolios. Short-term reversal and 12-month momentum seem to play a relevant role across the selected stocks. The average stock in the XGB(C) optimal buy region have previous month's return (str) in the bottom quintile of the cross-section, as well as featured in the bottom 42% of 12-month momentum. In contrast, stocks in the XGB(C) optimal sell region are in the top 58% of previous' month return, and in the bottom 27% of 12-month momentum. Moreover, the optimal portfolio typically selects small stocks, with the average stock in the long leg being at the 20th percentile of market value, and the short leg at the 26th percentile. For generalized linear models, we find that characteristics based on past returns play a more pronounced role, whereas size, measured as market capitalization, is less relevant, particularly for OLS. This suggests that interactions of characteristics play a role in the conditional distribution of returns, a point also raised in Kelly *et al.* (2020) and Freyberger *et al.* (2020).

Risk Adjusted Returns We report risk-adjusted alphas for all models. We focus exposition on the value-weighted alphas, but equally weighted returns can be found in the appendix.

[TABLE 5 ABOUT HERE]

Table 5 reports the coefficients obtained by regressing the portfolio on the column on the risk factor on the rows. Alphas are multiplied by 100.

[TABLE 6 ABOUT HERE]

Table 6 reports the coefficients obtained by regressing the portfolio on the column on the risk factor on the rows. Alphas are multiplied by 100. he FF5 model is able to account for excess returns from the GLM and XGB(C) models constructed with the standard sorting procedure, but not when optimal portfolio construction rules are employed. In fact, alphas from all optimally constructed portfolios are larger than those obtained through standard sorting procedure, and all are statistically significant at any reasonable threshold for t-statistics. Moreover, the share of variation of portfolio returns explained by the FF5 model is substantially smaller for optimally constructed portfolios than for the portfolios constructed using the standard procedure. Once transaction costs are accounted for, only the XGB(C) optimal portfolio generates positive and significant alphas.

4.2 Dissecting the Performance of Optimal Portfolios

Predicting Winners and Losers To visualize the classification properties of the constructed forecasts, we start by plotting two independent Receiver Operating Characteristic (ROC) curve for the classification of *winners* and *losers*. Although the portfolio construction problem is a multiclass classification problem, we start by investigating the properties of the classifiers 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. Dashed lines are constructed from probabilities obtained using the algorithm described above, and assuming a Gaussian distribution for returns. Solid lines are constructed from probabilities estimated from multi-class classification models. Blue lines represent the XGB model for regression and classification, and red lines represent generalized linear models. 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*, displaying a higher true positive rate for the same false positive rate. Labelling stocks with the top decile of XGB(R) predicted returns as *winners* incurs in a true positive rate of about 15% with a false positive rate of about 9.5%. For OLS, 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 9.5%, the XGB(C) model has a true positive rate of about 22.5%, whereas the GLM has a true positive rate of about 21.2%. Labelling stocks with the bottom decile of XGB(R) predicted returns as *losers* incurs in a true positive rate of about 19.3% with a false positive rate of about 8.9%. For OLS, 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 8.9%, the XGB(C) model has a true positive rate of about 32.2%, whereas the GLM has a true positive rate of about 30.2%. Among the classification models, XGB(C) performs slightly better than GLM. Among the regression models, OLS performs slightly better than XGB(R), which in turn performs better than random classification.

Identifying *losers* is an easier task than identifying winners, as can be seen by comparing the area under the ROC curve (i.e, the AUC) of models across the two plots, and both regression and classification methods perform better than random guessing.

Tuning Parameter Selection We next investigate the role of the tuning parameter selection in the optimal portfolios constructed.

[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 models is of 3.62, which is obtained by the XGB(C) model 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 *winner* greater than 20% and 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.

XGB(R) has a maximum Sharpe ratio of 2.70, achieved by setting $\lambda_W = 1.04$ 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 2.54. The GLM model achieves a maximum Sharpe ratio of 2.42 by setting $\lambda_W = 0.28$ and $\lambda_L = 0.41$. Finally, the OLS model achieves a maximum Sharpe Ratio of 1.98 by setting $\lambda_W = 1.06$ and $\lambda_L = 1.04$. Standard portfolio sorts would

provide a Sharpe ratio of about 1.95.

The portfolios obtained above include, on average, 7-10% of stocks, in contrast to at least 20% of stocks, as is the case for the top minus bottom decile strategy. Clearly, 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.

5 Conclusion

We study the construction of long-short portfolios on the basis of cross-sectional return predictions. We derive an optimal portfolio construction procedure that takes the form of a return classification rule. Selecting stocks on the basis of expected return predictions, the standard practice in the literature, is also optimal in special cases of the general framework. An empirical application to US stocks highlights that the portfolios constructed using the proposed procedure outperform portfolios constructed using the standard tools in the literature. This outperformance persists when transaction costs are duly accounted for.

References

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

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*.

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.

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.

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.

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

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

Patton, A. J. and Timmermann, A. (2010). Monotonicity in asset returns: New tests with applications to the term structure, the capm, and portfolio sorts. *Journal of Financial Economics*, **98**(3), 605–625.

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

Vapnik, V. (1999). *The Nature of Statistical Learning Theory*. Springer, New York., second edition.

Table 1: Portfolio Performance: Long − Short portfolio

Panel A. Equally Weighted Portfolios

	OLS		GLM		XGB(R)		XGB(C)	
	Decile	Opt.	MLC	Opt.	Decile	Opt.	MLC	Opt.
Avg. Exc. Returns	2.66	4.92	1.58	3.55	3.61	5.49	2.78	6.12
Vol.	4.29	8.21	3.00	5.18	4.56	6.50	3.06	6.98
Downside Vol.	1.99	2.98	2.42	3.76	2.47	3.12	1.73	3.83
Skew.	1.23	2.16	-0.37	-0.04	1.18	1.28	0.29	2.87
Ann. Sharpe Ratio	1.95	1.98	1.54	2.21	2.56	2.80	2.88	2.92
Ann. Sortino Ratio	4.22	5.44	1.92	3.06	4.74	5.82	5.09	5.31
Min.	-10.76	-15.79	-13.13	-19.02	-11.73	-17.97	-7.10	-22.51
25%	0.31	0.61	0.13	0.75	1.20	1.72	1.01	2.55
Median	2.26	3.34	1.59	3.40	3.34	4.89	2.79	5.64
75%	4.27	7.21	3.01	6.33	5.28	8.12	4.35	8.81
Max.	24.28	65.71	12.14	25.59	31.60	42.92	17.07	76.61
% of Selected Stocks	20.02	29.16	38.34	9.60	20.05	8.68	34.67	12.52
Correct Classification Rate	13.34	16.76	19.44	26.78	17.34	18.48	20.82	27.89
Wrong Direction Rate	10.46	14.39	15.64	20.50	13.86	15.13	12.60	19.97

Panel B. Value Weighted Portfolios

	OLS		GLM		XGB(R)		XGB(C)	
	Decile	Opt.	MLC	Opt.	Decile	Opt.	MLC	Opt.
Avg. Exc. Returns	1.59	2.30	0.73	1.93	1.68	2.40	1.24	4.00
Vol.	4.96	8.32	5.53	6.55	6.04	6.67	6.38	10.48
Downside Vol.	3.62	3.26	4.35	4.73	5.21	5.31	6.01	7.00
Skew.	0.09	5.43	-0.12	0.28	-0.74	-0.52	-1.40	2.82
Ann. Sharpe Ratio	0.94	0.86	0.31	0.90	0.83	1.12	0.54	1.24
Ann. Sortino Ratio	1.29	2.19	0.39	1.24	0.96	1.41	0.58	1.86
Min.	-20.56	-18.68	-28.17	-33.97	-36.02	-33.94	-52.38	-48.46
25%	-0.68	-1.57	-1.49	-1.30	-0.93	-0.90	-1.24	-0.44
Median	1.62	1.50	0.91	1.78	1.91	2.60	1.52	3.59
75%	3.84	4.54	3.17	4.90	4.75	5.82	4.42	7.17
Max.	25.39	106.89	27.23	43.79	25.85	25.88	27.07	114.25
% of Selected Stocks	20.02	29.16	38.34	9.60	20.05	8.68	34.67	12.52
Correct Classification Rate	13.34	16.76	19.44	26.78	17.34	18.48	20.82	27.89
Wrong Direction Rate	10.46	14.39	15.64	20.50	13.86	15.13	12.60	19.97

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. In addition, we report the average percentage of stocks selected, the average correct classification rate, and the average wrong direction rate, where wrong direction means buying a *loser* or selling a *winner*. Panel A reports results for equally weighted portfolios, and Panel B for value weighted portfolios. *Decile* portfolios are obtained by buying the top 10% of stocks according to the relevant forecast. *MLC* portfolios are obtained by buying the stocks for which the most likely class according to the relevant forecast is *winner* and selling their *loser* counterparts. *Opt.* portfolios are obtained by the optimal selection regions with parameters selected through cross-validation.

Table 2: Long − Short portfolio with Transaction Costs

Panel A. Equally Weighted Portfolios

	OLS		GLM		XGB(R)		XGB(C)	
	Decile	Opt.	MLC	Opt.	Decile	Opt.	MLC	Opt.
Avg. Exc. Returns	0.32	0.82	-0.38	-0.77	0.49	1.09	0.35	1.35
Vol.	4.47	8.08	3.45	4.99	4.62	6.63	3.25	6.93
Downside Vol.	2.62	3.54	3.19	3.93	2.94	3.76	2.51	3.97
Skew.	0.69	2.07	-1.52	-0.62	0.62	0.85	-0.55	4.16
Ann. Sharpe Ratio	0.15	0.30	-0.50	-0.62	0.28	0.50	0.24	0.61
Ann. Sortino Ratio	0.26	0.68	-0.55	-0.79	0.44	0.89	0.31	1.07
Min.	-15.33	-17.63	-16.46	-20.54	-17.02	-19.86	-12.26	-24.11
25%	-2.17	-3.49	-1.86	-3.35	-1.88	-2.56	-1.34	-1.63
Median	-0.19	-0.28	-0.00	-0.31	0.38	0.54	0.41	1.06
75%	2.19	3.12	1.47	2.11	2.42	4.13	2.30	3.96
Max.	19.68	51.21	7.52	14.05	23.08	31.24	10.94	73.83
% of Selected Stocks	20.03	29.78	37.93	8.84	20.05	7.91	34.61	11.47
Correct Classification Rate	13.24	16.40	19.62	26.35	17.38	18.36	20.80	27.08
Wrong Direction Rate	11.16	15.21	16.02	21.34	14.83	16.85	12.79	21.15

Panel B. Value Weighted Portfolios

	OLS		GLM		XGB(R)		XGB(C)	
	Decile	Opt.	MLC	Opt.	Decile	Opt.	MLC	Opt.
Avg. Exc. Returns	0.50	1.02	0.52	0.80	0.61	0.75	0.86	2.97
Vol.	4.70	10.40	4.45	9.17	5.75	7.99	5.33	11.73
Downside Vol.	3.67	4.21	3.48	6.65	4.60	5.97	3.96	6.16
Skew.	-0.36	4.47	-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	0.74	0.39	0.35	0.36	0.36	0.64	1.60
Min.	-21.67	-20.08	-19.54	-48.11	-30.11	-30.80	-20.56	-29.57
25%	-1.65	-3.78	-1.63	-3.17	-2.11	-3.07	-1.60	-1.90
Median	0.77	-0.09	0.54	0.50	0.55	0.54	0.63	2.44
75%	2.70	4.11	2.47	4.59	3.50	4.91	3.63	6.66
Max.	17.44	105.69	15.30	50.37	27.22	36.20	23.13	116.47
% of Selected Stocks	20.03	29.78	37.93	8.84	20.05	7.91	34.61	11.47
Correct Classification Rate	13.23	16.40	19.61	26.35	17.37	18.36	20.79	27.07
Wrong Direction Rate	11.16	15.20	16.02	21.34	14.83	16.85	12.79	21.15

This table reports the average monthly returns in excess of the risk-free rate, 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. In addition, we report the average λ, the average percentage of stocks selected, the average misclassification loss and false discovery rates implied by each of the selection rules used to construct portfolios. *Decile* portfolios are obtained by buying the top 10% of stocks according to the relevant forecast. *MLC* portfolios are obtained by buying the stocks for which the most likely class according to the relevant forecast is *winner* and selling their *loser* counterparts. *Opt.* portfolios are obtained by the optimal selection regions with parameters selected through cross-validation. Panel A reports results for the Optimally selected portfolios, and Panel B for the standard decile sorts. All returns are net of transaction costs.

Table 3: Stock Characteristics: Tree-Based Models

	XGB(R)				XGB(C)			
	Decile		Optimal		MLC		Optimal	
	Long	Short	Long	Short	Long	Short	Long	Short
mom1m	30.82	70.09	26.50	77.79	30.68	53.37	19.68	58.93
mom12m	49.79	31.06	43.83	24.87	54.75	34.32	42.21	26.70
chmom	41.84	54.30	40.51	52.61	43.09	52.27	38.99	54.16
bm	55.27	42.73	56.68	43.41	54.37	40.53	50.48	36.99
sp	55.66	44.02	57.63	44.46	56.36	42.21	51.73	39.51
maxret	57.03	76.61	61.34	80.62	63.42	78.27	75.59	86.24
cashpr	45.78	55.60	44.08	54.78	47.31	55.98	48.46	57.90
indmom	54.96	40.94	52.59	39.21	55.30	44.80	53.28	43.92
agr	60.84	46.91	62.19	47.06	58.74	52.42	63.54	54.74
mom6m	47.35	34.48	42.96	25.76	51.72	37.80	39.90	31.57
lgr	44.29	53.01	43.80	53.43	44.53	51.65	44.08	51.53
mvel1	33.06	40.37	25.65	38.95	29.40	32.64	19.41	26.49
mom36m	39.75	47.13	37.28	46.14	38.04	41.57	31.71	38.70
invest	44.06	53.18	42.92	53.29	44.76	50.33	42.56	49.20
rd_mve	56.31	50.39	56.32	49.36	56.29	53.96	60.60	54.08
hire	43.81	52.78	42.41	52.51	45.31	50.19	42.37	48.53
turn	47.70	54.74	45.05	53.49	51.83	58.78	54.56	60.62
bm_ia	53.81	49.94	54.44	49.19	55.21	50.43	54.87	48.82
sgr	44.28	52.74	42.93	52.43	46.09	50.51	43.78	49.65
ill	64.05	59.31	71.05	61.93	66.00	63.44	74.12	68.27
lev	50.42	43.48	52.02	45.16	47.49	41.76	46.32	40.55
zerotrade	54.53	44.78	58.38	45.87	50.18	43.78	48.12	42.47
chempia	44.23	51.39	42.93	51.40	44.58	48.77	41.59	47.21
dolvol	37.20	42.65	30.47	40.17	36.36	40.01	30.15	35.64
cfp	45.62	41.15	44.32	41.84	46.07	35.08	36.02	30.56
chcsho	48.71	55.88	48.16	54.99	50.12	57.79	53.25	58.50
cfp_ia	48.24	43.75	46.55	43.29	49.45	39.94	40.52	35.38
age	49.29	41.30	48.53	41.21	43.83	37.03	39.84	34.70
grltnoa	44.48	51.47	43.71	51.80	45.03	48.66	43.07	48.17
egr	41.07	50.22	39.21	49.53	43.26	44.83	37.32	41.83

This table reports the average percentile of each characteristic in the stocks bought and the stocks sold by the portfolio in the column. Characteristics are sorted according to their relevance for XGB(C) - Optimal, defined as the spread between the average percentile of the stocks bought and stocks sold. We truncate the table at 30 characteristics. *Decile* portfolios are obtained by buying the top 10% of stocks according to the relevant forecast. *MLC* portfolios are obtained by buying the stocks for which the most likely class according to the relevant forecast is *winner* and selling their *loser* counterparts. *Opt.* portfolios are obtained by the optimal selection regions with parameters selected through cross-validation.

Table 4: Stock Characteristics: Linear Models

	OLS				GLM			
	Decile		Optimal		MLC		Optimal	
	Long	Short	Long	Short	Long	Short	Long	Short
mom1m	25.82	73.45	38.00	71.10	29.17	55.71	16.42	59.97
mom12m	59.93	36.60	50.57	38.26	53.58	32.92	46.16	20.54
chmom	36.17	61.29	45.19	60.00	42.11	53.41	36.97	55.77
sp	59.06	38.04	53.25	39.44	57.28	42.60	53.17	36.12
indmom	62.63	36.37	53.25	38.59	55.88	43.69	57.86	41.69
mom6m	53.51	42.39	49.06	43.40	50.72	36.59	43.36	27.31
bm	58.69	38.58	52.86	39.35	54.74	41.11	48.69	33.60
agr	62.03	38.48	54.84	40.70	59.02	51.85	69.24	54.93
cashpr	42.83	62.14	47.30	60.95	46.94	56.19	47.96	60.40
lgr	42.73	57.65	47.46	56.30	44.25	51.98	41.32	52.74
invest	42.03	59.72	47.00	57.91	44.94	50.93	38.99	49.75
mom36m	40.09	56.65	44.77	54.97	38.50	41.64	26.07	36.46
maxret	48.17	61.61	53.36	62.65	62.75	78.30	77.21	87.54
rd_mve	58.31	45.31	52.87	46.51	56.59	53.07	64.21	53.92
chcsho	43.73	60.39	49.11	59.31	48.36	58.35	51.86	61.91
lev	53.14	40.85	50.79	40.94	47.91	41.60	47.89	38.68
cfp	53.84	40.65	48.05	40.71	46.92	35.01	33.77	25.47
hire	43.35	57.66	47.01	56.50	45.62	50.42	40.04	48.31
grltnoa	43.70	56.88	47.31	55.36	45.23	49.07	40.79	48.45
egr	41.76	57.11	46.05	55.62	43.00	45.12	32.45	39.91
sgr	43.41	57.00	47.21	56.08	46.28	50.69	42.44	49.86
chinv	42.86	55.65	47.61	54.56	45.17	49.48	40.77	48.05
grcapx	43.88	56.70	47.72	55.50	46.19	49.75	41.24	48.40
cfp_ia	54.38	42.93	48.79	43.46	49.94	40.00	38.36	31.35
rd	56.19	46.07	51.12	46.84	53.38	51.24	59.20	52.57
cashdebt	48.27	46.42	46.00	46.00	44.77	33.46	27.02	20.43
chempia	44.73	55.23	47.06	54.15	45.31	48.82	40.06	46.64
turn	47.18	58.24	48.96	58.09	48.59	57.97	54.63	60.83
ear	54.94	44.67	51.25	45.46	51.86	46.22	50.21	44.30
depr	56.18	46.62	53.04	48.86	59.36	57.37	66.22	60.58

This table reports the average percentile of each characteristic in the stocks bought and the stocks sold by the portfolio in the column. Characteristics are sorted according to their relevance for GLM - Optimal, defined as the spread between the average percentile of the stocks bought and stocks sold. We truncate the table at 30 characteristics. *Decile* portfolios are obtained by buying the top 10% of stocks according to the relevant forecast. *MLC* portfolios are obtained by buying the stocks for which the most likely class according to the relevant forecast is *winner* and selling their *loser* counterparts. *Opt.* portfolios are obtained by the optimal selection regions with parameters selected through cross-validation.

Table 5: Risk Adjusted Returns: Value-weighted

	OLS		GLM		XGB(R)		XGB(C)	
	Decile	Opt.	MLC	Opt.	Decile	Opt.	MLC	Opt.
Alpha	0.96***	1.85***	0.23	1.40***	1.04***	1.70***	0.43	2.96***
	(4.72)	(4.78)	(0.85)	(4.13)	(4.27)	(5.76)	(1.41)	(4.87)
MKT	0.02	0.10	0.05	0.24***	0.09	0.18**	0.27***	0.63***
	(0.31)	(0.73)	(0.76)	(2.88)	(1.37)	(2.24)	(2.63)	(2.69)
SMB	−0.49***	−0.11	−0.07	0.04	−0.71***	−0.62***	−0.08	0.11
	(−5.43)	(−0.73)	(−0.52)	(0.31)	(−5.14)	(−3.77)	(−0.67)	(0.54)
HML	0.02	−0.63*	0.18	0.04	0.06	0.08	0.08	0.10
	(0.17)	(−1.83)	(1.26)	(0.27)	(0.51)	(0.57)	(0.49)	(0.30)
RMW	0.27**	0.37	0.61***	0.24	0.43**	0.27	0.72***	0.67*
	(2.32)	(1.03)	(5.63)	(1.33)	(2.38)	(1.18)	(4.53)	(1.71)
CMA	0.48***	0.43	0.15	0.12	0.51***	0.48**	0.16	0.15
	(3.25)	(1.19)	(0.66)	(0.66)	(2.74)	(2.10)	(0.63)	(0.45)
MOM	0.58***	0.00	0.40***	0.38**	0.74***	0.79***	0.71***	0.50**
	(8.97)	(0.03)	(4.40)	(2.35)	(8.71)	(7.34)	(4.31)	(2.18)
STR	0.60***	0.63**	0.11	0.25	0.10	0.21*	0.04	0.18
	(7.42)	(1.98)	(1.00)	(1.35)	(0.91)	(1.77)	(0.45)	(0.48)
LTR	0.22*	0.51**	−0.24	−0.12	−0.05	−0.10	−0.32*	−0.12
	(1.83)	(2.40)	(−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	0.14	0.22	0.14	0.58	0.45	0.35	0.19

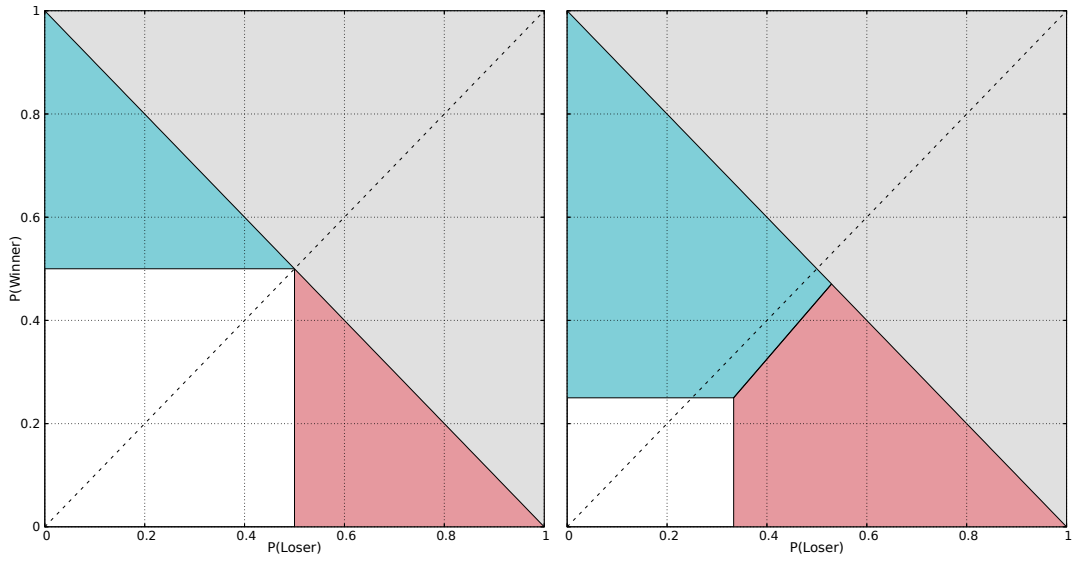
This table reports the risk-adjusted returns of the portfolios considered. We consider model that includes the Market, Size, Book-to-Market, Profitability, Investment, Long and short term reversals, as well as the Momentum factor. Newey west t-statistics are reported in parenthesis. Alphas are expressed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

Table 6: Risk Adjusted Returns: Value-weighted, with Transaction Costs

	OLS		GLM		XGB(R)		XGB(C)	
	Decile	Opt.	MLC	Opt.	Decile	Opt.	MLC	Opt.
Alpha	−0.02	0.65	0.01	0.69	0.29	0.20	0.09	1.79***
	(−0.09)	(1.38)	(0.06)	(1.11)	(1.26)	(0.46)	(0.39)	(2.87)
MKT	0.16***	0.08	−0.15*	−0.38*	0.02	0.12	0.12	0.35
	(2.73)	(0.37)	(−1.95)	(−1.86)	(0.28)	(1.22)	(1.25)	(1.29)
SMB	0.21	0.77***	0.42***	0.13	0.02	0.09	0.26***	0.40
	(1.62)	(3.94)	(3.78)	(0.47)	(0.19)	(0.57)	(2.85)	(1.53)
HML	−0.15*	−0.85*	0.23	0.13	0.01	−0.23	0.13	−0.51
	(−1.66)	(−1.82)	(1.65)	(0.41)	(0.08)	(−1.14)	(0.94)	(−1.02)
RMW	−0.02	0.02	0.27*	0.13	−0.10	0.16	0.76***	1.23**
	(−0.20)	(0.04)	(1.91)	(0.29)	(−0.65)	(0.73)	(4.12)	(2.12)
CMA	0.53***	0.16	0.55***	0.05	0.51**	0.57	0.54***	0.66*
	(3.02)	(0.43)	(3.19)	(0.14)	(2.27)	(1.62)	(3.09)	(1.83)
MOM	0.44***	−0.21	0.35***	0.29	0.80***	0.90***	0.58***	0.29*
	(6.21)	(−1.63)	(4.59)	(1.31)	(12.35)	(8.77)	(8.67)	(1.66)
STR	0.64***	0.79**	0.39***	0.54**	0.25***	0.38***	0.07	0.54
	(8.89)	(2.11)	(5.11)	(2.03)	(3.01)	(3.06)	(0.77)	(1.45)
LTR	0.20	0.67***	−0.29*	−0.03	0.02	0.21	−0.19	0.20
	(1.57)	(2.65)	(−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	0.26	0.41	0.07	0.52	0.35	0.53	0.13

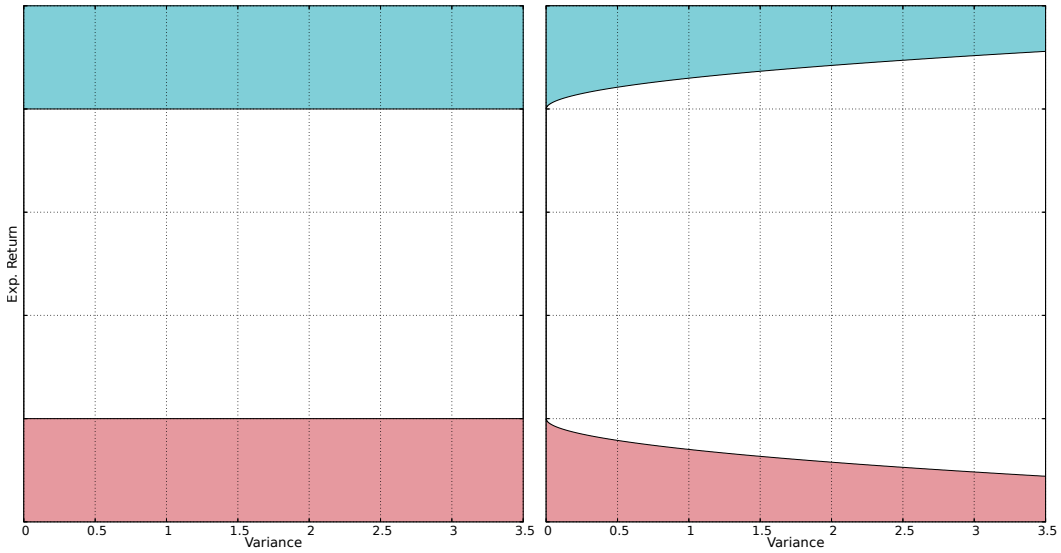
This table reports the risk-adjusted returns of the portfolios considered. We consider model that includes the Market, Size, Book-to-Market, Profitability, Investment, Long and short term reversals, as well as the Momentum factor. Newey west t-statistics are reported in parenthesis. Alphas are expressed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

Figure 1: Optimal Selection Regions



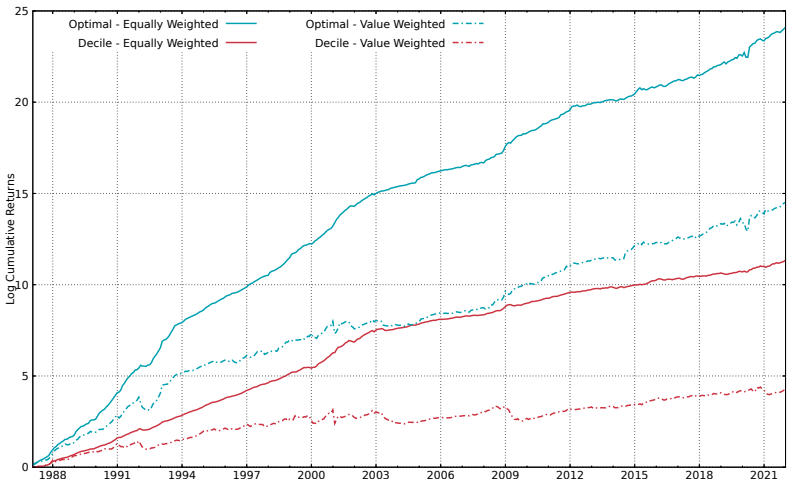
This figure reports the selection regions obtained by setting $\lambda_W = \lambda_L = 1$ (left panel) and $\lambda_W = \frac{1}{3}, \lambda_L = \frac{1}{2}$ (right panel). Blue colors demark the buy region, red colors are the sell regions, and white colors are the no-trade zone. The greyed out area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

Figure 2: Selection Regions in Location Scale Models



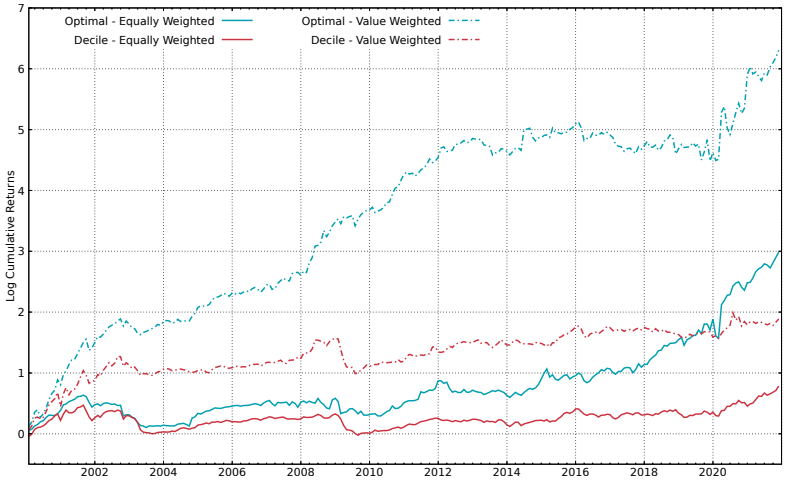
This figure reports the selection regions obtained by setting $\lambda_W = \lambda_L = 1$ (left panel) and $\lambda_W = \lambda_L = 1.1$ (right panel). Blue colors demark the buy region, red colors are the sell regions, and white colors are the no-trade zone.

Figure 3: Cumulative Returns XGB(C): Top Decile v Optimal



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

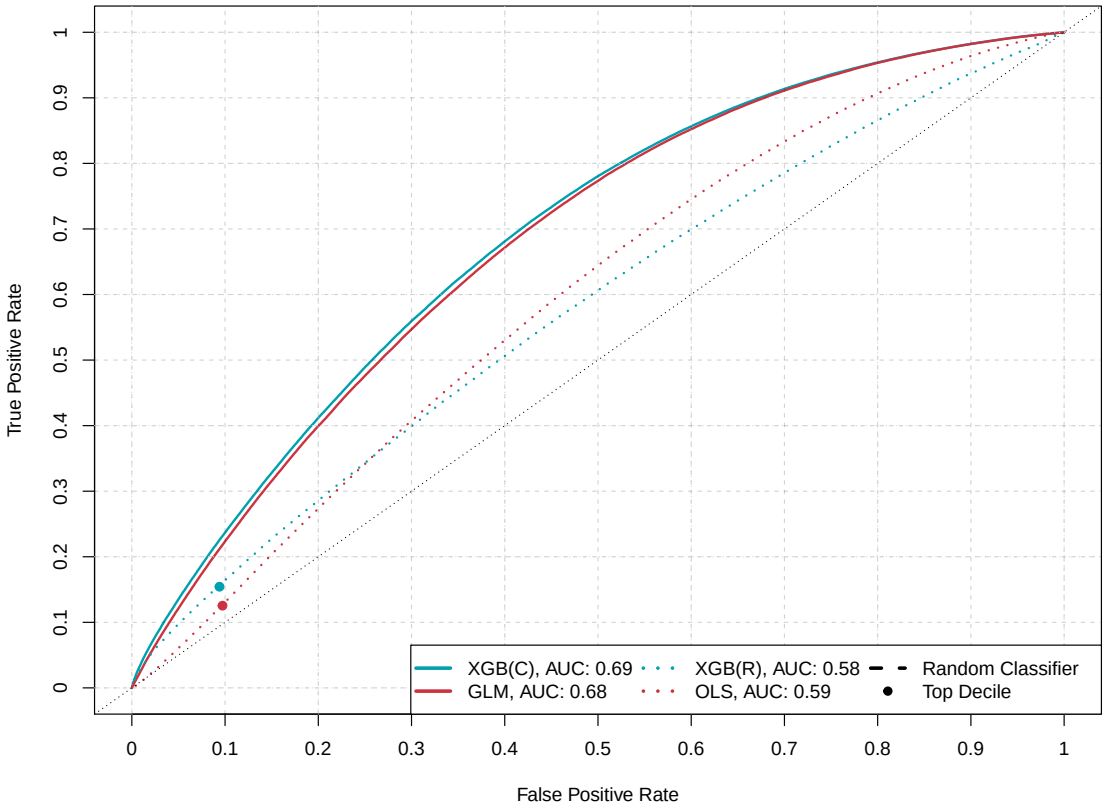
Figure 4: Cumulative Returns XGB(C) with Transaction Costs



This figure reports the log cumulative returns of portfolios built using XGB(C) 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 (MLC). We also report equally weighted returns (solid lines), and value weighted returns (dashed lines).

Figure 5: Receiver Operating Characteristic (ROC) Curves

Panel A: Long



Panel B: Short

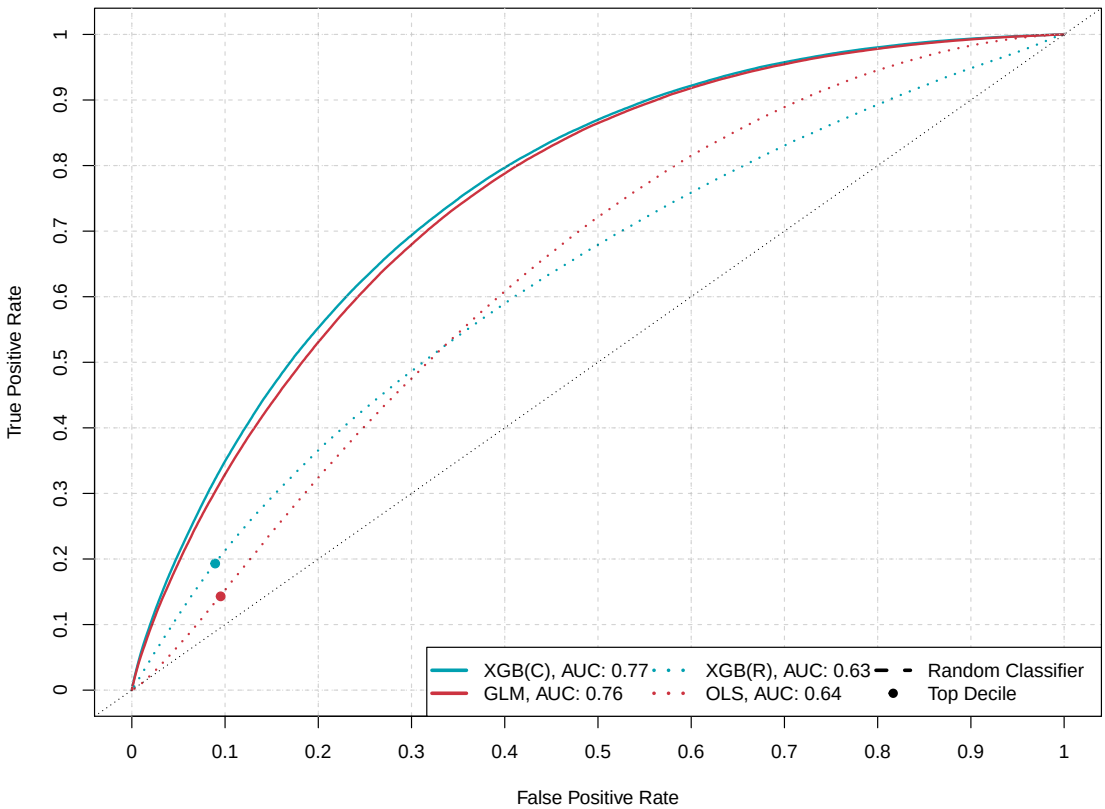
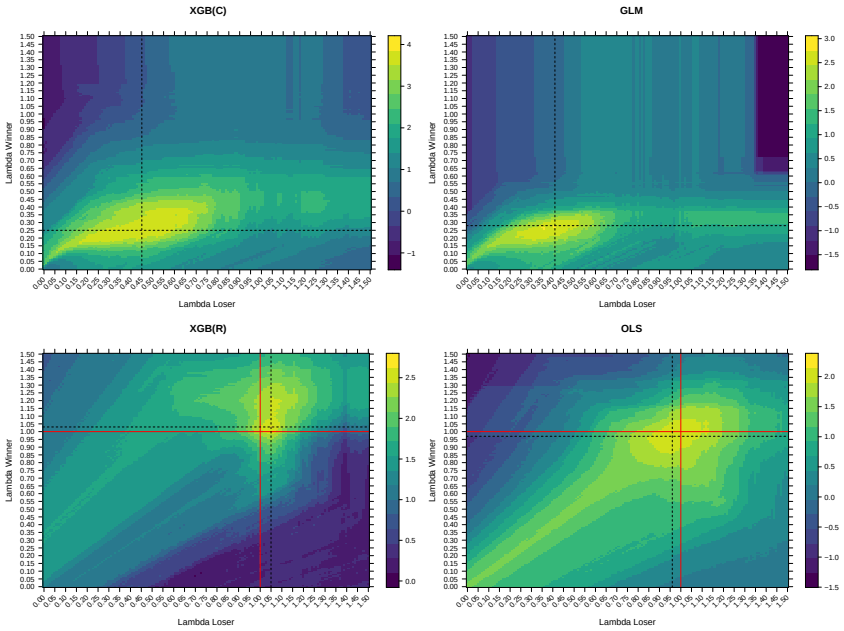


Figure 6: Lambda and Portfolio Properties



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 represent the standard decile sorting procedure. The dashed lines represent the location of the (unfeasible) optimal parameters.